Master of Science Thesis in Electrical Engineering
Department of Electrical Engineering, Linköping University, 2021

Investigation of Light and Ultrasound Injected Signals in Microphones

Robin Djerv

Master of Science Thesis in Electrical Engineering

Investigation of Light and Ultrasound Injected Signals in Microphones

Robin Djerv

LiTH-ISY-EX–21/5440–SE

Supervisor: **Ted Johansson**
ISY, Linköpings universitet
Niccolò de Milleri
Infineon Technologies AG
Ruben Pecas
Infineon Technologies AG

Examiner: **Jacob Wikner**
ISY, Linköpings universitet

Division of Integrated Circuits and Systems
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

Abstract

Voice commanded systems (VCS) have been proved to be vulnerable to signal injections mimicking voice commands and explored security flaws in market available products for the time of each respective study that originally discovered those methods. Signal injection caused with the help of amplitude modulated ultrasonic waves (being known as DolphinAttacks - DA) were proved to work on several such devices in 2017. In 2019, another study were also successful in achieving signal injections using modulated laser also known as LightCommands (LC). This thesis has investigated the occurring circumstances which enables such injections. Simulations and laboratory trials have shown a thermoacoustic origin enabling LC to be injected and the response differs with respect to microphones physical size. DA utilizes the non-linearity of microphones and more linear microphones have indeed been shown to withstand DA better and physical parameters have been shown to indicate how DA may be optimized for successful injections. The results have been used to provide ideas on how a VCS system can be designed to be more resilient.

Sammanfattning

Röststyrda System har visat sig vara sårbara mot signalinjektioner som härmar röstkommandon och utnyttjar kryphål hos produkter som fanns på marknaden i samtid när studierna som först tog upp kryphålen publicerades. Signalinjektoner inducerade med hjälp av amplitudmodulerat ultraljud (känt som DolphinAttacks - DA) bevisades fungera på flertalet enheter år 2017. 2019 visade en annan studie framgång med signalinjektion genom modulerad laser, även känt som LightCommands (LC). Detta examensarbete har utrett de bakomliggande faktorer som möjliggör sådana injektioner. Simuleringar och laboratorieexperiment har visat att termoakustiska effekter möjliggör LC med resultat som beror på mikrofoners fysiska storlek. DA nyttjar ickelinjäritet hos mikrofoner och linjärare mikrofoner har visat sig stå emot DA bättre och det har visat sig att DA kan optimeras för bättre lyckade injektioner. Resultaten har används för att bidra till idéer och resonemang från föregående studier på hur lösningar mot LC och DA skulle kunna implementeras och göra mikrofoner och dess tillhörande system tåligare mot sådana angrepp.

Acknowledgments

I would first like to thank my supervisors Niccolò and Ruben at Infineon for having supported me during the period of the thesis. They made me learn alot more within the field of electronics that I did not know so much about before this thesis.

I would also like to thank my examiner Prof. J Jacob Wikner and Prof. Ted Johansson who has been my academical supervisor. They have both always been guiding, supporting and inspiring during my time as a master student in the area of electronics.

Last but not least. I would like to thank my family and close friends for helping me and supporting me to endure the lockdown due to a pandemic we all had to go through and in particular it became tougher when national lockdown occurred after moving to a new country as an expat.

Linköping, January 2021

Contents

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**
1.1 Motivation 1
1.2 Aim 2
1.3 Research Objectives 2
1.4 Report Structure 2

2 Theory **3**
2.1 MEMS Microphones - Overview 3
2.1.1 Dual Backplate Microphones 5
2.1.2 Constant Voltage 5
2.1.3 Constant Charge 6
2.1.4 MEMS Key Terms 7
2.2 LightCommands 8
2.2.1 Thermoelectrical Analogy 9
2.2.2 Relation between Thermal and Acoustic Models in SPICE . 9
2.3 DolphinAttack 11
2.3.1 Concept 11
2.3.2 Mathematical Description of an AM Signal 11
2.3.3 Non-linearity in MEMS Microphones 12

3 Methodology **13**
3.1 Pre-Study 13
3.2 Delimitations 14
3.2.1 Microphone Models 14
3.3 LightCommands 15
3.3.1 Simulations 15
3.3.2 Laboratory 19
3.4 DolphinAttack 19
3.4.1 Simulations 19

3.4.2 Laboratory 20

4 Results **25**
4.1 LightCommands 25
4.1.1 Thermal Model 25
4.1.2 Full LC-MEMS Model Frequency Response 27
4.1.3 Transient Comparisons 28
4.1.4 Voice Injections using LightCommands 31
4.2 DolphinAttack 34
4.2.1 Simulations 34
4.2.2 Laboratory 35
4.2.3 Observed Multiple Order Non-Linearity 41
4.2.4 Voice Injections using DolphinAttack 42

5 Discussion **47**
5.1 Potential solutions 47
5.1.1 LightCommands 47
5.1.2 DolphinAttack 48
5.2 Limitations 50

6 Conclusion **51**
6.1 Conclusions on LightCommands 51
6.2 Conclusions on DolphinAttack 52
6.3 Future Work 52
6.3.1 LightCommands 52
6.3.2 DolphinAttack 53

Bibliography **55**

List of Figures

2.1 Cross sectional depiction of a MEMS microphone. 3
2.2 Graphical depiction of a Helmholtz resonator. 7

3.1 Schematic of the 2D thermal block. 17
3.2 Hierarchical block 17
3.3 Front Cavity as RC-link, R11=Conductive thermal resistance, C10=Thermal capacitance, and R15=Convective resistance with sound inlet. . . . 18
3.4 AC-response with small ultrasonic transducer and main microphone. 21
3.5 Frequency response of the chamber. 22
3.6 Equalizer filter 22
3.7 Comparison measured frequency (red) response and simulated (grey). 23

4.1 Frequency response for the thermal model. 26
4.2 Frequency response in main microphone when lid is incremented in height. 26
4.3 Frequency response comparison when thermal model and acoustic models are combined to have an LC input. 27
4.4 Magnitude comparison between acoustic pressure signal and laser injected signal with absorption of 0.1-1 mW. 28
4.5 Comparison using simulated and laser generated 35 Hz square signal. 28
4.6 Comparison using simulated and laser generated 50 Hz square signal. 29
4.7 Comparison using simulated and laser generated 100 Hz square signal. 29
4.8 Comparison using simulated and laser generated 250 Hz square signal. 30
4.9 Comparison using simulated and laser generated 500 Hz square signal. 30
4.10 Frequency response comparison for main microphone with linear scale x-axis. 31
4.11 "Hey Siri" comparison original audio an LC injected. 31
4.12 "Alexa" comparison original audio an LC injected. 32
4.13 "English Vowels" comparison original audio an LC injected. 32
4.14 Simulation on main mic. 34

4.15 Zoomed in at (supposed) injections. 34
4.16 Simulated injection at single backplate microphone. 35
4.17 Comparison between injections in main microphone at 450 Hz when the carrier frequency is swept. 36
4.18 Comparison between strongest and weakest injection when targeting the main microphone. 36
4.19 Comparison between injections in LabSec microphone at 450 Hz when the carrier frequency is swept. 37
4.20 Comparison between strongest and weakest injection when targeting the LabSec microphone. 37
4.21 Injection using baseband frequency sweep with 94 dBSPL ultrasound signal hitting the Helmholtz resonance on main mic. 38
4.22 Injection using baseband frequency sweep with 94 dBSPL ultrasound signal on LabSec. 38
4.23 Resonance slope comparison. 40
4.24 Multiple order injections comparison between main and LabSec (dotted). 41
4.25 Comparison original voice vs injection when DA is 94 dBSPL. . . . 42
4.26 Comparison original voice vs injection when DA is 100 dBSPL. . . 43
4.27 Comparison original voice vs injection when DA is 106 dBSPL. . . 43
4.28 Comparison original voice vs injection when DA is 112 dBSPL. . . 44
4.29 Comparison original voice vs injection when DA is 107 dBSPL. . . 44
4.30 Comparison original voice vs injection when DA is 109 dBSPL. . . 45
4.31 Comparison original voice vs injection when DA is 112 dBSPL. . . 45

List of Tables

2.1 Analogy between thermal and electrical units. 9
2.2 Comparison of force proportion between single and differential topology with constant voltage. 12

3.1 Comparison of analogies. 16

4.1 Comparison of the power density ratio TTS generated voices in the range of 100-1000 and 1000-7000 Hz. 33
4.2 Properties of audio signal used for sweeping carrier frequency trials. 36
4.3 Magnitude relation between injection signals when sweeping baseband frequency on LabSec microphone. 39
4.4 Magnitude relation between injection signals when sweeping baseband frequency on main microphone. 39

1 Introduction

1.1 Motivation

It has been shown in the recent years that devices with voice commanded systems - VCS (i.e Alexa, Siri etc) are vulnerable to signal injections from non-audio sources which cause a security risk. Zheng et al. proved in 2017 that they could be successfully hacked using amplitude modulated ultrasonic signals. Such signals are inaudible to humans but non-linearity of the microphones caused internal demodulation making audio signals getting injected and thus, make the the VCS interpret it as a voice command and execute commands such as turning on airplane mode, calling someone et cetera. This was proven to work on many devices that were available at the market by the time of the study. This method became known as "DolphinAttack" (DA) [1].

Two years later (2019), a new method was discovered by Sugawara et al. were they achieved similar goals and hacked several VCS available on the market at that time. This method utilized amplitude modulated laser instead of ultrasound and was given the name "LightCommands" (LC) [2]. Those two studies together indicated major security flaws in the microphones and products using them as they were proved to be prone to injections.

1.2 Aim

This thesis serves as a continuation of the previous studies. While they mainly focused on hacking several VCS with different parameters such as distance, background noise, injection messages et cetera. They proved that most of the tested voice commanded devices could be hacked with LC and DA in various circumstances at the time of the writing for each respective study. The continuation focuses on the physical and electrical parts of the microphones to give a deeper insight in what is causing LC and DA to work. The aims are to verify LC and DA in a simulation environment and provide data that can be used to understand correlation and causation. Furthermore, trials in a laboratory environment will also be conducted to validate the simulation models and provide further information that can be used for post-hack processing.

1.3 Research Objectives

The investigation of laser and -ultrasound induced signals are described by these following objectives;

- Verify DolphinAttacks and LightCommands in simulations.
- Perform DolphinAttacks and LightCommands in laboratory to confirm and validate the results of the simulations.
- Propose new insights and ideas for possible solution based on older and newer knowledge.

1.4 Report Structure

The structure of the thesis is as follows: Chapter 2 reviews the background of how a MEMS microphone functions. Furthermore, the idea behind DA and LC is also explained and shown on how it is supposed to act on a MEMS microphone given literature describing the nature of MEMS and previous research studies investigating the methods. Lastly, theory behind the methods is described to show how the simulation and laboratory experiment were executed. The methods themselves are described in chapter 3 explaining how the LC and DA in the simulations and laboratory experiments will be executed. Chapter 4 presents the results of the simulations and experiments. Chapter 5 and 6 bring up the discussion including analysis of the result, limitations of the thesis, ideas for future work, and conclusions of the thesis.

2
Theory

This chapters covers the concept of MEMS microphones, how they work and also covers the theory behind DolphinAttack and LightCommands and as well as other relevant formulas, equations and expressions that are used or encountered in this thesis.

2.1 MEMS Microphones - Overview

A microphone is a transducer that converts an acoustical signal to an electrical signal. There are different principles of microphones such as piezoelectric, piezoresistive, optical and capacitive [3]). This study focuses on capactivie microphones since they dominate the market [4] and the market value for MEMS microphone was 1207.7 MUSD in 2018 [5]. A cross sectional picture of how a MEMS microphone looks like is shown in figure 3.7 .

Figure 2.1: *Cross sectional depiction of a MEMS microphone.*

A capacitive microphone can be described as a parallel plate capacitor consisting of a membrane and a backplate with an air gap between acting as a dielectric material [3]. The capacitance between them can be expressed as

$$C = \frac{\epsilon_0 \cdot A}{d} \tag{2.1}$$

where

- ϵ_0 is the dielectric constant [m].
- A is area [m^2].
- d is the distance between the membrane and backplate and can be expressed as $d = d_0 - d(t)$ where d_0 is the static distance when the membrane is at rest and d(t) is the change in distance when the membrane vibrates [m].

The charge when there is a voltage across the plates is applied is expressed as

$$Q = V \cdot C. \tag{2.2}$$

The energy between the plates in static case is

$$E = \frac{1}{2} \cdot C \cdot V^2. \tag{2.3}$$

By inserting the capacitance equation (2.1) into the expression for energy 2.3 and derive it with the respect to the distance d, the electrostatic force becomes

$$F = \frac{\partial E}{\partial d} = \frac{1}{2} \cdot \frac{\epsilon_0 \cdot A}{d^2} \cdot V^2 = \frac{1}{2} \cdot \frac{Q^2}{\epsilon_0 \cdot A}. \tag{2.4}$$

In a physical implementation is a microphone biased to have either the charge or voltage constant. This is done by either by an external bias voltage source or a permanent stored charge [3]. For a constant voltage approach (V=V_0) where the charge and force is time varying, The expressions for the electrostatic force and charge can be written as

$$F(t) = \frac{1}{2} \cdot \frac{\epsilon_0 \cdot A}{d^2} \cdot V_0^2 \tag{2.5}$$

and

$$Q(t)_{Vconst} = \frac{\epsilon_0 \cdot A \cdot V_0}{d}. \tag{2.6}$$

The equivalent expressions for a constant charge approach (Q=Q_0) are

$$F(t)\frac{1}{2}\cdot\frac{Q_0^2}{\epsilon_0\cdot A} \tag{2.7}$$

and

$$V(t)_{Qconst}=\frac{d\cdot Q_0}{\epsilon_0\cdot A\cdot}. \tag{2.8}$$

Both approaches have connections to the output of the backplates to process the charge/voltage in order to generate an analog/digital output depending on choice of full microphone design.

2.1.1 Dual Backplate Microphones

Dual backplate topology (also known as differential microphones [6]) differs such as they consist of two backplates instead of one and creates a differential output into the signal processing schematic.

For a dual backplate microphone. The capacitances between each backplate and the membrane are

$$C_1=\frac{\epsilon_0\cdot A}{d_0-d(t)} \tag{2.9}$$

and

$$C_2=\frac{\epsilon_0\cdot A}{d_0+d(t)}, \tag{2.10}$$

since the perturbing of the membrane with distance $d(t)$ makes the air gap increase on one side while deceasing on the other side thus changing the capacitances differential. The capacitances can further be expressed as

$$C_1=C_{10}+\Delta C_1 \tag{2.11}$$

and

$$C_2=C_{20}+\Delta C_2 \tag{2.12}$$

where C_{10} is the mean capacitance and ΔC_1 is the change in capacitance.

2.1.2 Constant Voltage

The charge for each backplate using constant voltage approach is

$$Q_1=V_0\dot{(}C_{10}+\Delta C_1) \tag{2.13}$$

and

$$Q_2=-V_0\dot{(}C_{20}+\Delta C_2) \tag{2.14}$$

respectively and acting as output to a processing circuit.

The electrostatic forces for each backplate on a dual backplate microphone with constant voltages are

$$F(t)_1 = \frac{1}{2} \cdot \frac{\epsilon_0 \cdot A}{(d_0 - d(t))^2} \cdot V_0^2 \tag{2.15}$$

and

$$F(t)_2 = -\frac{1}{2} \cdot \frac{\epsilon_0 \cdot A}{(d_0 + d(t))^2} \cdot V_0^2 \tag{2.16}$$

thus acting in opposite directions. The total electrostatical force is equal to

$$F(t) = \frac{1}{2} \cdot \frac{\epsilon_0 \cdot A}{(d_0 - d(t))^2} \cdot V_0^2 - \frac{1}{2} \cdot \frac{\epsilon_0 \cdot A}{(d_0 + d(t))^2} \cdot V_0^2 \tag{2.17}$$

where the expression can be further simplified to

$$F(t) = 2 \cdot V_0^2 \frac{\epsilon_0 \cdot A}{(d_0^2 + d(t)^2)^2}. \tag{2.18}$$

2.1.3 Constant Charge

For a constant charge microphone. The Voltage across each capacitor is

$$V_1 = \frac{Q}{C_{10} + \Delta C_1} \tag{2.19}$$

and

$$V_2 = \frac{Q}{C_{10} + \Delta C_2}. \tag{2.20}$$

The electrostatic forces on each backplate are as well differential and can be written as

$$F(t) = \frac{1}{2} \cdot \frac{Q_0^2}{\epsilon_0 \cdot A} \tag{2.21}$$

and

$$F(t) = -\frac{1}{2} \cdot \frac{Q_0^2}{\epsilon_0 \cdot A}, \tag{2.22}$$

and the resulting total force acting on the membrane is

$$F(t) = \frac{1}{2} \cdot \frac{Q_0^2}{\epsilon_0 \cdot A} - \frac{1}{2} \cdot \frac{Q_0^2}{\epsilon_0 \cdot A} = 0. \tag{2.23}$$

As long as the charge remains constant, the electrostatic force will be zero [3].

2.1.4 MEMS Key Terms

Some key terms relevant in the thesis are;

Helmholtz Resonance is an inherent resonance frequency in a MEMS microphone. The inheritance is due to the geometry of the sound inlet and the front cavity. Excitation near the neck forces the air particles in the neck to oscillate which creates a pressure difference in the chamber. The frequency for when it occurs is [4][7] [8]

$$f = \frac{c}{2 \cdot \pi} \cdot \sqrt{\frac{S}{V \cdot L}} \tag{2.24}$$

where c is the speed of sound, S is the cross sectional area of the neck, V is the volume of the chamber, and L_{eff} is the effective length of the neck see picture below;

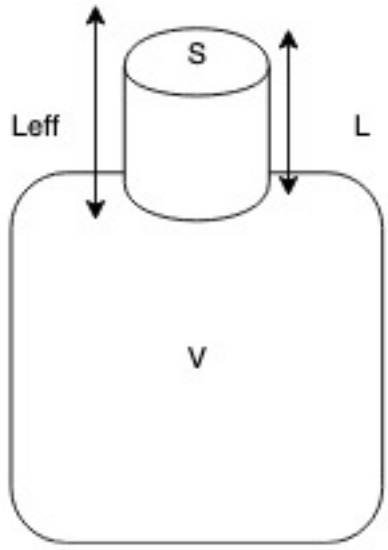


Figure 2.2: *Graphical depiction of a Helmholtz resonator.*

Acoustic Overload Point, abbreviated AOP is in general defined as the sound pressure value in dBSPL where the total harmonic distortion (THD) of the microphone exceeds 10% [6].

Sound Pressure - SPL is the pressure ratio between the pressure of the sound and the lowest detectable pressure that human ears can detect. It is given by the following equation

$$dB_{SPL} = 20 \cdot \log_{10}(p/p_0) \tag{2.25}$$

where p is the sound pressure p_0 is the sound pressure and $p_0 = 20$ [μPa] (0 dB) is the reference pressure and the auditory threshold for human ears [4].

dBFS, stands for Decibel Full Scale which is the expression for the sensitivity of a digital microphone. The reference lies at 1 kHz 94 dBSPL sine signal which corresponds to a negative number in dBFS and 0 dBFS is full scale output in digital microphones. A lower (more negative number) of dBFS at 94 dBSPL is equal to a more sensitive microphone that can capture higher sound pressure without distortions [9].

Parametric Array is a phenomenon where ultrasonic signal gets demodulated and creates an audible sound [10]. While this is not the purpose of this project, it is of interest since it could be a source of errors in DA trials.

2.2 LightCommands

The original paper by Sugawara et al. describing LightCommands mentions photoacoustics as the main effect behind the phenomenon which occurs when the laser's photons hit the membrane. Laser induced photoacoustic is known to be of thermoacoustic nature [11] and an earlier student project in-house have shown that thermoacoustic effect plays a role in inducing unwanted signals in MEMS microphones[12]. This indicates that membrane heating is the leading cause of laser induced signals.

Lumped Capacitance Model

Before making a thermal model, it is important to know the value of Biot number since it determines whether the system can be approximated to a lumped capacitance system or not [13]. The condition required to make lumped capacitor model accurate is when Biot number is smaller than 0.1 and the expression for the Biot number is

$$B_i = \frac{h}{k} \cdot L \tag{2.26}$$

where

- L is the characteristic length [m].
- h is the convective heat transfer coefficient $[\frac{W}{m^2 \cdot K}]$.
- k is the thermal conductivity $[\frac{W}{mK}]$.

and L is expressed as

$$L = \frac{V}{A_s} \tag{2.27}$$

where for a component

- V is the volume
- A_s is the surface area [14].

2.2.1 Thermoelectrical Analogy

The analogy between normal electrical units and thermal analogs are shown in the table below;

Table 2.1: *Analogy between thermal and electrical units.*

Electric parameter	Thermal analogy	Symbol
Voltage	Temperature [K]	T
Current	Heat Transfer rate [W]	Q
Electrical Resistance	Thermal Resistance [K/W]	R
Capacity	Thermal Capacity [W/K]	C

The expressions for the analogs and the relations between them are

$$\Delta T = Q \cdot R \quad \text{(analog to Ohm's law)} \tag{2.28}$$

$$R = \frac{L}{\lambda \cdot A} \quad \text{(thermal resistance)} \tag{2.29}$$

$$C = c \cdot \rho \cdot V = c \cdot m \quad \text{(thermal capacity)} \tag{2.30}$$

where

- L is length [m].
- A is cross sectional area [m^2]
- λ is thermal conductivity [W/Km]
- c is specific heat capacity [m]
- ρ is density [kg/m^3]
- V is volume [m^3]
- and m is mass [kg].

Given these relations and analogs, resistors, capacitors, voltage sources, voltage controlled voltage sources and current sources can be used to build a thermal model in SPICE [12].

2.2.2 Relation between Thermal and Acoustic Models in SPICE

The ideal gas law can be used to acquire the volume change from a temperature change by making an approximation assuming that the back cavity is an isobaric environment (constant pressure) and no gas exchange with ambient world occurs.

The equation for the ideal gas law is

$$P \cdot V = n \cdot R \cdot T \tag{2.31}$$

where;

- P is pressure [N/m^2],
- V is volume [m^3],
- T is temperature [K],
- n is the amount of gas substance [mol],
- and R is ideal gas constant [J/Kmol] [15].

By using two different states of the back cavity,

$$P_{const} \cdot \Delta V = n \cdot R \cdot \Delta T \tag{2.32}$$

$$P_{const} \cdot V_{idle} = n \cdot R \cdot T_{room} \tag{2.33}$$

By substituting P_{const}, the following expression is acquired

$$\Delta V = \frac{V_{idle} \cdot \Delta T}{T_{room}} \tag{2.34}$$

where the temperature is room temperature and the volume is constant with resting membrane and one state where a change in volume occurs due to change in temperature.

Temperature change ΔT is represented as a voltage in a thermal system and since T_{room} and V_{idle} are to be considered constant values, ΔV is represented as a voltage in its node in the same SPICE schematic. By getting volume flow [m^3/s] from the volume, the relation between thermal and acoustic models can be acquired by using the following relation

$$I = C \cdot \frac{\partial V}{\partial t} \tag{2.35}$$

which is the relation between voltage, capacitance and current [16]. By setting C=1, The capacitor acts as a derivator and produces the volume flow as current $\frac{V}{S}$. Volume flow is analog to current in an electroacoustical analogy [7].

2.3 DolphinAttack

2.3.1 Concept

The concept of a DolphinAttack is to use an ultrasonic AM signal carrying a voice command for instance a wake-up word for a VCA and utilize the non-linearity of MEMS microphones to inject the voice command and activate the device [1].

2.3.2 Mathematical Description of an AM Signal

An amplitude modulated signal can mathematically be described as

$$S_{in} = m(t) \cdot c(t) + c(t) \tag{2.36}$$

where m(t) is the baseband signal containing the transmitted information i.e a voice. A simple case is when the baseband is a single sine (single tone) and for that case be written as $m(t) = A_m \cdot sin(\omega_m \cdot t + \phi_m)$ and c(t) is the carrier frequency and can be written as $c(t) = A_c \cdot sin(\omega_c \cdot t + \phi_c)$ with $\omega_m = 2 \cdot \pi \cdot f_m$ and $\omega_c = 2 \cdot \pi \cdot f_c$ being the angular frequency for each signal [17]. For simplicity, both ϕ_m ϕ_c are set to zero in further derivation.

Expanding the equation 2.36 by inserting m(t) and c(t), the expressions turns into

$$S_{in} = A_m \cdot sin(\omega_m \cdot t) \cdot A_c \cdot sin(\omega_c \cdot t) + A_c \cdot sin(\omega_c \cdot t) \tag{2.37}$$

which can be further rewritten as

$$S_{in} = \frac{A_m}{2} sin((\omega_c - \omega_m) \cdot t) + A_c \cdot sin(\omega_c \cdot t) + \frac{A_m}{2} sin((\omega_c + \omega_m) \cdot t). \tag{2.38}$$

The ratio between A_m and A_c is called the modulation index m and can be written as

$$m = \frac{A_m}{A_c}. \tag{2.39}$$

The ratio determines the amplitude characteristic of the signal where the absolute amplitude goes between $A_c \cdot (1 + m)$ and $A_c \cdot (1 - m)$ [17].

In practical case when creating a DA; m(t) is a voice that covers a frequency spectrum and c(t) is an ultrasonic sine signal. The condition $\omega_c - \omega_m > 40000 \cdot \pi$ (20 kHz) must be satisfied to prevent and signal become audible to humans [18].

2.3.3 Non-linearity in MEMS Microphones

MEMS microphones are known to have several non-linear properties [3]

- Membrane deflection
- Non-uniform gap
- Parasitic capacitances
- Electrostatic forces and capacitance relations
- Capacitor mismatch (Differential microphones).

The electrostatic forces and capacitances 2.3.3 are causing non-linear behaviour depending on microphone model and approach [3]. Using constant voltage approach, the electrostatic forces have non-linear relation to the varying distance between membrane and backplates.

Table 2.2: *Comparison of force proportion between single and differential topology with constant voltage.*

Single/Differential	Proportion	Equation
Single	$F \propto \frac{1}{d^2}$	2.5
Differential	$F \propto \frac{1}{(d_0^2+d(t)^2)^2}$	2.18

However, with constant charge approach, the equations expressed for electrostatic force in a single backplate (2.7) and dual backplate (2.23) have no non-linear proportionality between force and distance. All microphones used in this thesis are based on constant charge approach and are used on conceptual levels (no parasitics) meaning with respect to the theory, it is expected that they will be spared of some non-linear effects.

3
Methodology

This chapter describes how the thesis was carried out. It describes LC and DA each and how they were reproduced in simulations as well as in laboratory. Some failed approaches are also mentioned to motivate why the final methods were chosen and to show the importance of good conditions for proper experiments.

3.1 Pre-Study

Both LC and DA are fairly new concepts and there is not much literature regarding both methods. Other literature includes studies about MEMS microphones to uncover their functionality and how it may relate to LC and DA. The nonlinearity of a MEMS microphone have been studied in earlier papers and while they are not emphasizing the impact of DA, it gives an insight in what could be the underlying mechanisms.

During the time span where the work for this thesis was done, no other paper regarding LC than the original one written by Sugawara et al. [2] had been found in the search for those, thus indicating that LC has not been well studied. Previous in-house studies and experiments on thermal(acoustic) effects on MEMS Microphones were studied to get a better understanding of how the interdisciplinary connection between the physics and electronics of a MEMS microphone system.

3.2 Delimitations

Previous studies on LC and DA have focused primarily on benchmark and on trying out the versatility and feasibility of both hacking methods on different consumer products with different circumstances such as distance, power, background noise et cetera. To get additional knowledge and understandings of DA and LC, a few available in-house MEMS Microphones were chosen for the investigations, the trials, and simulations focused rather on understanding the nature of LC and DA to provide new understanding and knowledge.

3.2.1 Microphone Models

The selected microphones are all capacitive MEMS microphone using constant charge topology. They were chosen due to availability and relevance. To not reveal exactly which microphones were used, they were given nicknames.

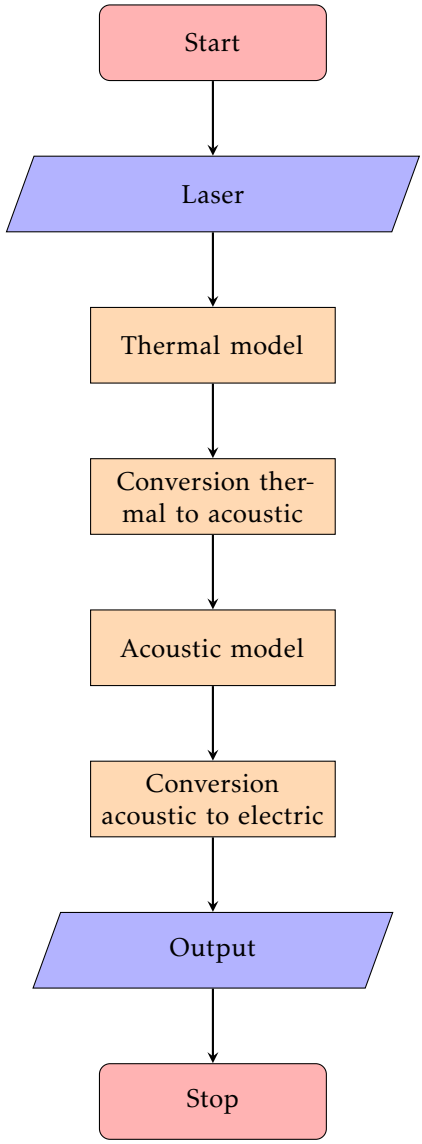
- Main: Used for simulations and laboratory for both LC and DA and serves as the main microphone in this thesis. This microphone is of a differential (Dual backplate) model.
- Sec: An older model similar to the Main model being a differential. This one was used for LC related simulations since previous thermal experiments had been conducted and had data available to make a proper simulation model. It differs from the Main one mainly in being physically smaller.
- DaSec: Available single backplate of older model. Since it is analog, the output is in dbV instead of dbFS.
- LabSec: Single backplate used for laboratory experiments since it was the physically available single backplate microphone for the thesis.

3.3 LightCommands

3.3.1 Simulations

The simulation model were made by connecting the laser input to an existing in-house acoustic model of the microphone with an electrical output.

A flow chart on the complete simulation model is shown below;



Electroacoustical Analogy

The acoustical model of the microphone is an existing in-house model. It was built using electrical components as acoustic analogs using electroacoustical analogy and to some extent electromechanical analogy see table right below [7].

Table 3.1: *Comparison of analogies.*

Mechanical	**Acoustic**	**Electrical**	**SPICE**
Force	Instantaneous pressure	Voltage	Voltage source
Velocity	Volume Flow	Current	Current source
Mass	Acoustic Mass	Inductance	Inductor
Mechanical Compliance	Acoustic Compliance	Capacity	Capacitor
Mechanical Impedance	Acoustic Impedance	Electrical Resistance	Resistor

Thermal Model

The Biot number is acquired to determine the needed complexity of the model. The convective coefficient h is around 5-15 since no forced convection occurs in the microphone, k for air is 0.0262 [12] . Each part were measured and had the characteristic length spanning from order of 10^{-4} to 10^{-7}. which makes the Biot number $0.1 \gg B$ and allows this model to be build as a lumped capacitance system where the thermal conduction dominates the thermal flow.

Model Components: The model consists of several parts of the microphone making up the whole system;

- ASIC
- Lid
- Membrane
- Microphone pillars
- Front cavity
- Back cavity
- PCB

Most parts were created as hierarchical 2D+ blocks or simple RC-links if the heat flow was assumed to flow in one direction. The blocks consist of four thermal resistances that represent the thermal flow in each 2-dimensional direction and one capacitance that represents the heat capacity. To avoid netlist errors, voltage sources without any potential differences (V=0) were used to allow direct connections with the thermal body of each component.

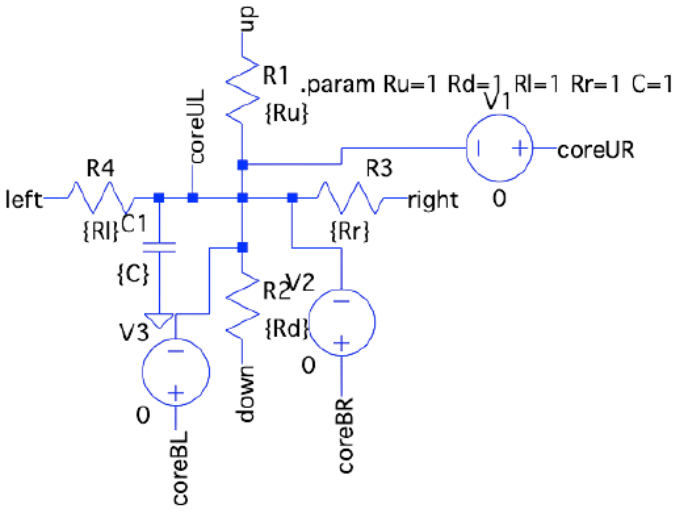


Figure 3.1: *Schematic of the 2D thermal block.*

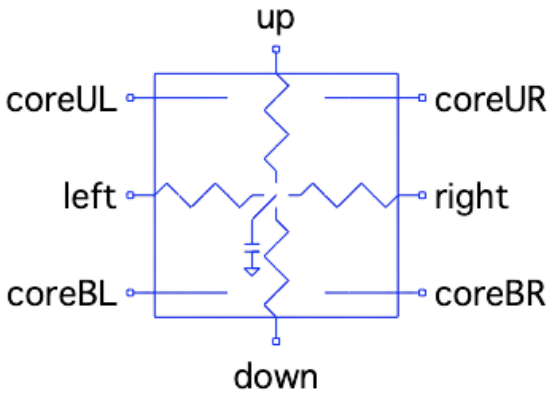


Figure 3.2: *Hierarchical block*

Simplification Certain approximations and simplifications were done to make the model easier to work with. The backplates were not included due to their perforation holes that take up big parts of their surface and volume and allow the heat to flow from the membrane to both cavities. The air gaps between the membrane and backplates are omitted since their volumes are much smaller than the one of the front and back cavity and were expected to make minor to no contributions to the results.

The front cavity, lid, and PCB with its layers were simplified into simple RC-links as the heat flow flow one dimensional in this model (see 3.3 below for an example). The heat dissipated into the back cavity were modelled to flow one dimensional through the PCB and lid while the front cavity receives heat from the membrane and flows from the membrane through the front cavity and into the outside world through convection. The lid and front cavity are connected through a resistor resembling free convection flow with the outside world. All blocks are also approximated to consist of one type of material i.e the ASIC block is assumed to be pure silicon to make model construction easier. The system is also assumed to be an isobaric environment since older in-house investigations have done the same approximations for the same systems.

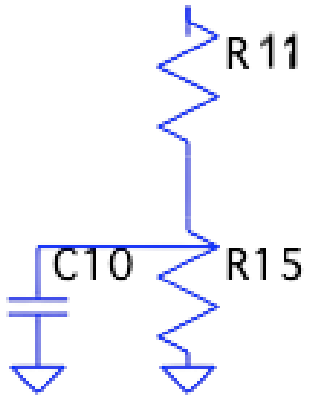


Figure 3.3: *Front Cavity as RC-link, R11=Conductive thermal resistance, C10=Thermal capacitance, and R15=Convective resistance with sound inlet.*

Input: The input is a current source connected to the membrane to resemble thermal power absorbed by the membrane from the laser.

Output: The output of the model is the average temperature change ΔT of the back cavity. The acquired value were used in an acoustic model. By using the ideal gas law for an isobaric case, the volume change was acquired. Since the temperature is analog to voltage, a voltage controlled voltage source were used to acquire the volume change. The output of the voltage controlled voltage source is voltage in SPICE but the value resembled the volume change.

Conversion Thermal to Acoustic: A capacitor was used as a derivator (setting C=1) to acquire volume flow using the relation

$$I = C \cdot \frac{\partial V}{\partial t} \tag{3.1}$$

where the volume flow was represented as the current through the capacitor 3.1. The volume flow was used as an input through a current source into a resembling back cavity node in the in-house acoustic model.

Value Acquisition: All model components were measured in width, length and height to get their surface areas and volumes to calculate thermal resistances and capacitances. All necessary values were put into a text file as parameters and upload as a text file into SPICE.

Signals: Experiments were done using square waves at 35 Hz, 50 Hz, 100 Hz, 250 Hz, 500 Hz and 1000 Hz respectively. Those frequencies were used in earlier similar in-house projects and are thus chosen for this one as well. No voice samples were simulated due to time and performance limitations.

3.3.2 Laboratory

To achieve LC in a laboratory environment, the following equipments were used;

- Signal source: AFG3000 signal generator for square waves and Audio Precision APX525 for audio signals.
- Laser driver circuit: A pre-designed current driver circuit used to modulate the laser amplitude with a signal.
- Isolation box: To reduce the ambient noise.
- Laser: The chosen laser a small laser pointer providing 1 mW with an wavelength of 740 nm (red). The distance from laser to microphone was 5 cm and the laser point is focused to have a width equal to the width of the sound port to maximize the the efficiency of the experiment.

3.4 DolphinAttack

3.4.1 Simulations

An existing in-house acoustic conceptual model schematic was used for simulating DolphinAttacks. This saved time as no model had to be created and due to the acoustic input signal and no conversion had to be done since the input is of electroacoustical analogy. The conceptual models do not include parasitic capacitors, AOP and symmetry in differential cases which caused some non-linearity's to not be included.

3.4.2 Laboratory

Equipment needed to conduct DA experiments were

- Audio Precision APX525 (generator and analyzer)
- Transmitter
- Receiver (MEMS Microphone)

An available AFG31000 Signal generator was at first considered but whenever an AM-signal was generated, leaks of single tones on the frequencies $f_c - X * f_m$ where $X \in 0, 1, 2....n$ were discovered meaning that it would interfere with the potential injections and lead to possible false positives for validation tests with single injections. It was at first considered to build a high pass filter in between to filter it out but it caused more issues as a flat passband was not achieved and it distorted the signal itself by minor filtering of lower sideband. It was also turned out to not be suitable for transmitting external signal i.e a voice from an audio file since it could only run files with tfx-formats and audio files converted into tfx had a poor reproduced quality and often required more samples than allowed by the machine.

Instead, an Audio Precision APX525 was chosen as a signal generator and the AM-signals were formatted as ".wav" created in matlab with generated sine signals and recorded audio files. To decide which method to use - several transmitters were carried out. They were chosen due to availability.

Kemo L002 Ultrasonic Speaker: Used to keep vermins away from gardens one was purchased for trials because it was cheap and available to purchase from an online retailer.

Small Ultrasonic Transducer: Small discrete transducer similar to the transducer used by in the first study of DolphinAttack [1] to create a budget version. The issue with this speaker was mainly two things; despite using the highest possible output from the Audio precision, the received measured sound pressure were not high and it also showed that the AC response for the speaker itself had more influence rather than the microphone which is seen in 3.4. While this show interest aspects, using the speaker to investigate injections in microphones with respect to the microphones properties would be hard and not feasible for the purpose of the thesis.

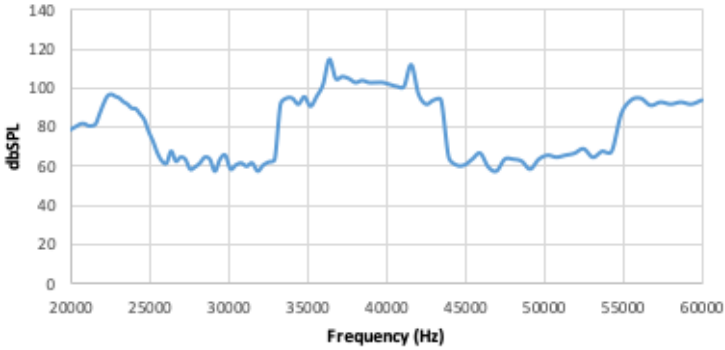


Figure 3.4: *AC-response with small ultrasonic transducer and main microphone.*

Pressure Chamber/Fixture: By using an in-house fixture with built in speaker, pressure chamber and a receiving microphone. They did however have a very big attenuation for frequencies above 2 kHz and non-linearity above 10 kHz making those unfeasible for carrying out real tests since the environment had too big influence. For these reasons, neither of the available transmitters mentioned above could be used to extract results.

Fixed Speaker in Anechoic Chamber: Being the most advanced, expensive and complex one, linear up to 60 kHz but not being able to create high outputs in the ultrasonic domain continuously without risk of breaking it; This speaker set-up was utilized to be able to carry out controlled DAs and get the needed results that could be used for further analysis given the limitations. Given the pre-fixed setup. The microphones were put 20 cm from the speaker to receive maximum sound pressure. A reference microphone was used to create an AC-response of the anechoic chamber and the inverse was used as an equalizing filter.

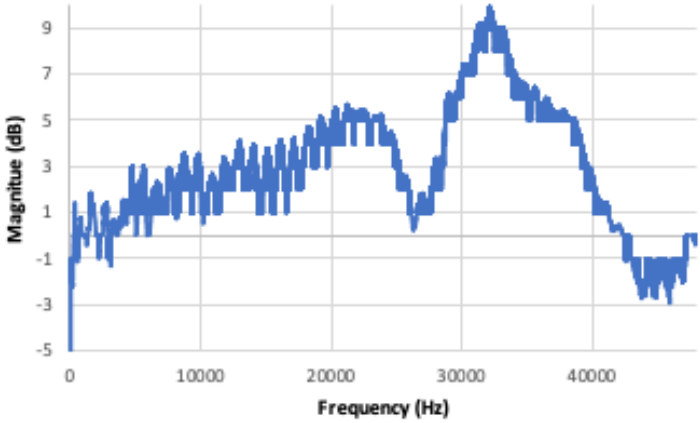


Figure 3.5: *Frequency response of the chamber.*

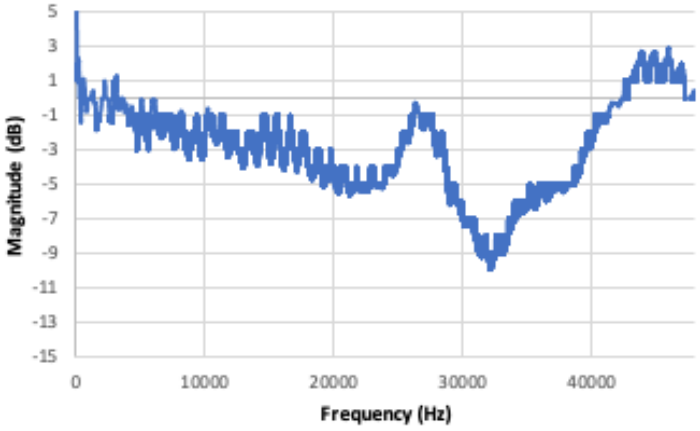


Figure 3.6: *Equalizer filter*

After using an equalizer filter, the main microphone was tested to see how their frequency response were related to the conceptual model shown in 3.7.

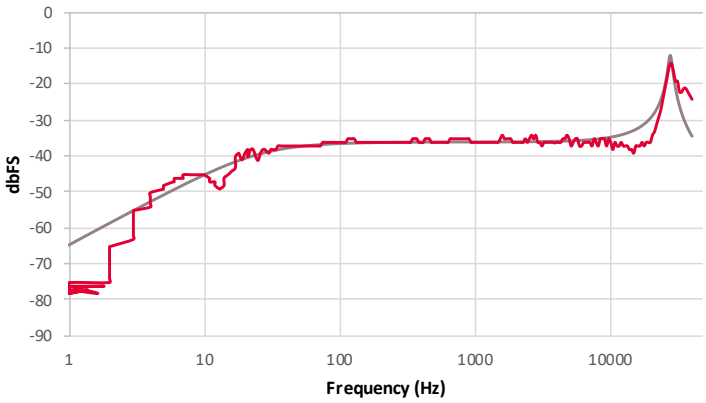


Figure 3.7: *Comparison measured frequency (red) response and simulated (grey).*

Three main experiments were made to validate DA in the laboratory.

- f_c: Sweeping the carrier frequency is conducted to investigate how it affects the quality of the injections when the frequency varies and give hopeful insights to follow up from original study where the success rate was depending on the carrier frequency among some parameters.
- f_m: sweeping the f_m given a fixed f_c is important to see whether the injections will exhibit some kind of frequency response which could be crucial in understanding and finding properties that differs from injected signals and regular audio signals.
- Voice Injection: Injections occurs where the carrier frequency lies at the Helmholtz resonance frequency and different output power are used to see how the injection correlates with the pressure of the ultrasonic signal. This tested the range and feasibility of succeeding and also investigated whether the injection had a clear difference from an original audio signal.

The sound pressure for carrier and baseband sweep was 94 dBSPL to make fair comparisons and avoid results getting biased from the signals themselves.

4
Results

This chapter provides the results for LightCommand and DolphinAttacks each respectively.

4.1 LightCommands

The results from LightCommands are shown here, thermal responses from simulations and laboratory are displayed as well are comparison of transient laser pulses in simulation and laboratory and finally voice injections from laboratory. No voice injections were simulated due to limitations in SPICE and computer power.

4.1.1 Thermal Model

The frequency response of The thermal model of the microphones show low-pass characteristics and can be approximated as a two-pole system.

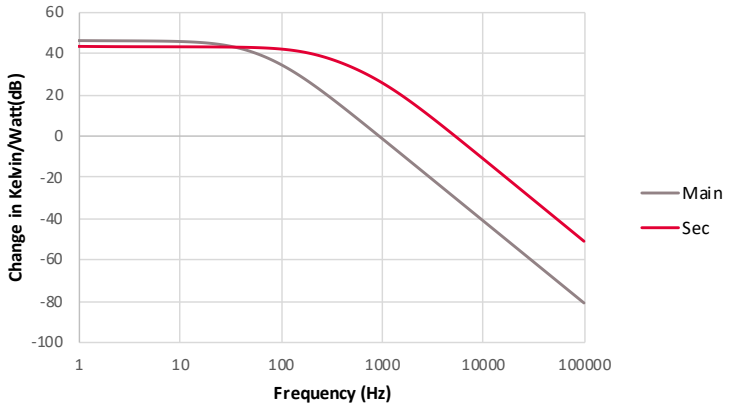


Figure 4.1: *Frequency response for the thermal model.*

Comparing the two microphones, they both show the clear same pattern but the sec microphone has a higher cut-off frequency. This verifies that the thermal response to a laser signal varies depending on the microphone which is true for these two models.

Influence of the Physical Dimension

In regard of the graph of 4.1 indicating how the physical dimensions affect the thermal response. Simulations with an increased lid height (increased back cavity volume) in main microphone showed a proportion between bigger volume and smaller cut-off frequency shown in 4.2.

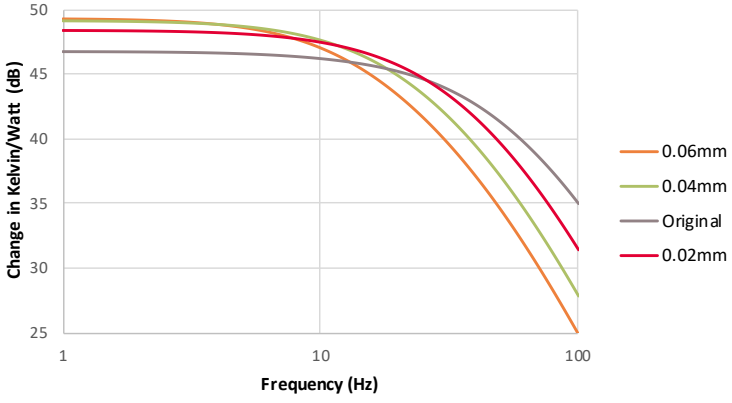


Figure 4.2: *Frequency response in main microphone when lid is incremented in height.*

4.1.2 Full LC-MEMS Model Frequency Response

The thermal model and the acoustic model merged, using thermal power in the membrane as a system input, new AC characteristics for the microphone has a bandpass characteristic in contrast to the acoustic response with a constant response in the audible band where small effects of the Helmholtz resonance is also noticeable in figure 4.3.

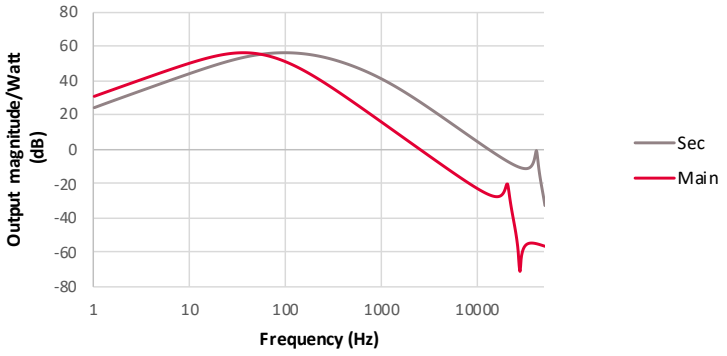


Figure 4.3: *Frequency response comparison when thermal model and acoustic models are combined to have an LC input.*

The results show in simulations that the frequency response differs when the signal origin differs and the physical size plays a big role in the thermal response.

Magnitude Comparison of Frequency Response from Acoustic and LC-origin

Comparing the frequency response when the microphone receive LC and when it receives an acoustic signal in figure 4.4 below show that 0.01-0.1 mW heat absorption generates a signal with similar magnitude of a signal generated by acoustic pressure of 1 Pa (94 dBSPL) at least for lower frequencies (below 1 kHz). Furthermore, it does also show that the microphone could reach AOP if the absorbed power is not much higher than 1 mW. This indicates a possible upper limit on desired power for successful injection although it may be much higher than 1 mW when the injected signal is not a single tone but rather a spectrum.

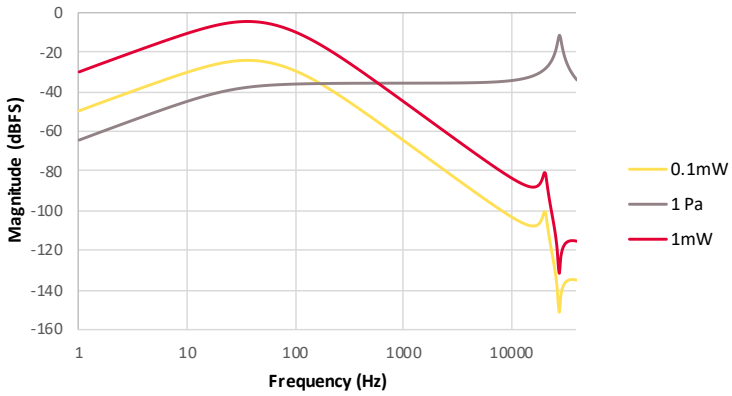


Figure 4.4: *Magnitude comparison between acoustic pressure signal and laser injected signal with absorption of 0.1-1 mW.*

4.1.3 Transient Comparisons

Using the setup up in the laboratory. Trials by measuring the frequency response and transient square waves were done and compared to simulations. Y-axis is digital magnitude spanning from -1 to 1 where 1 resembles 0 dbFS (130 dBSPL) and X-axis represents the time span that was extracted from the whole signal. Their scales were individualised to provide good visual looks on the signals with different frequencies.

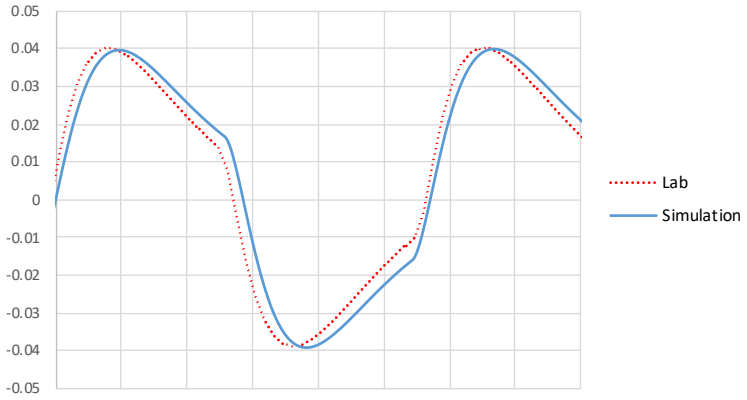


Figure 4.5: *Comparison using simulated and laser generated 35 Hz square signal.*

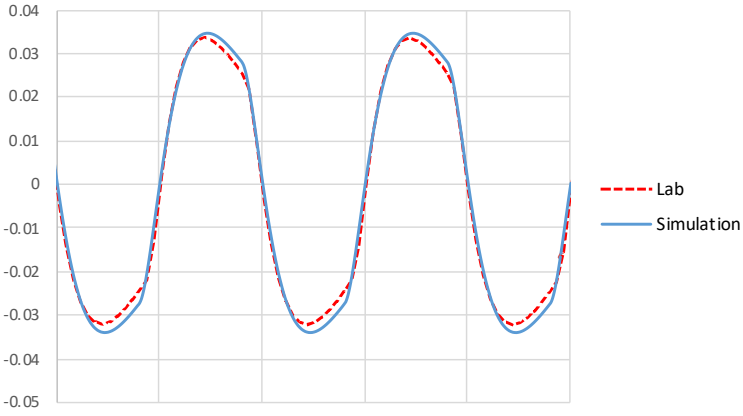


Figure 4.6: *Comparison using simulated and laser generated 50 Hz square signal.*

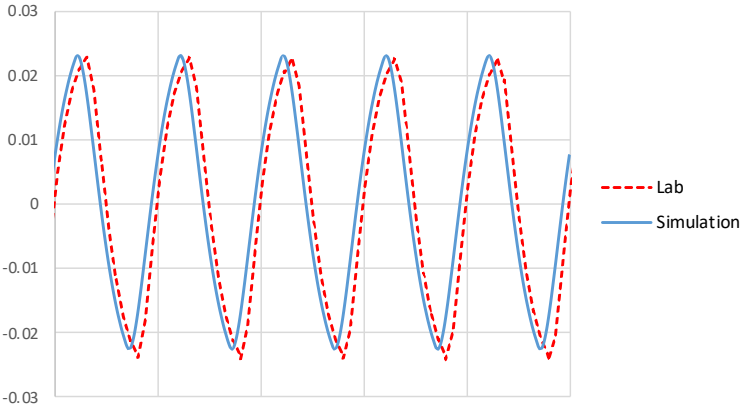


Figure 4.7: *Comparison using simulated and laser generated 100 Hz square signal.*

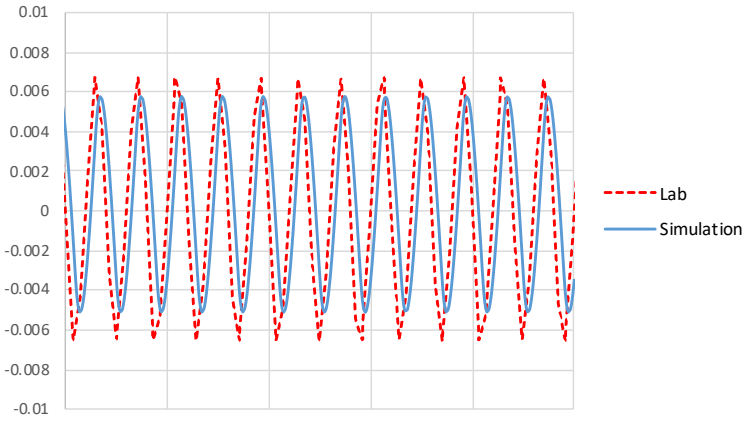


Figure 4.8: *Comparison using simulated and laser generated 250 Hz square signal.*

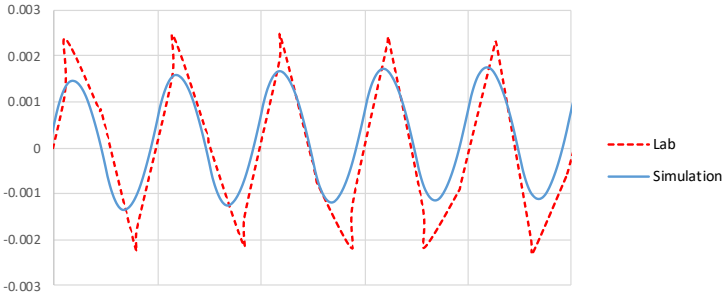


Figure 4.9: *Comparison using simulated and laser generated 500 Hz square signal.*

Transient simulations are all showing a good overlapping between simulated signals and measured signals. The simulation model predicts LightCommands with good precision. The frequency response for the main microphone itself also has a good overlapping as seen in 4.10.

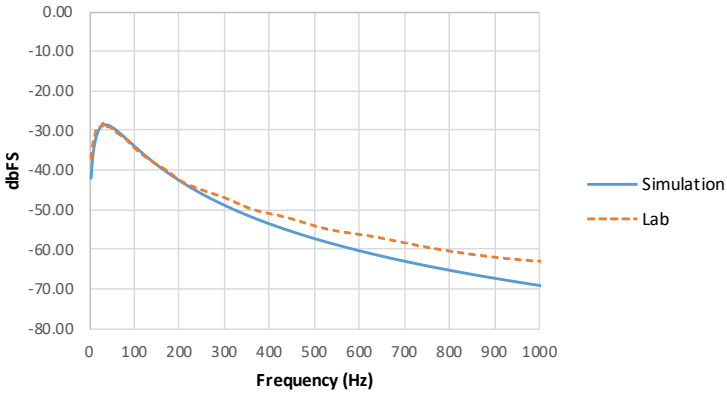


Figure 4.10: *Frequency response comparison for main microphone with linear scale x-axis.*

4.1.4 Voice Injections using LightCommands

Comparing the frequency response of a laser induces sine and an acoustical sine signal shows a clear difference between acoustical induced signals and laser induced signals. Below are two examples of the author using the command words *Hey Siri* and *Alexa* with normalized y-axises where frequencies above 600 Hz are suppressed and lower frequencies are dominant in the spectrum between 100-2000 Hz with red being from original audio source and dotted blue LC.

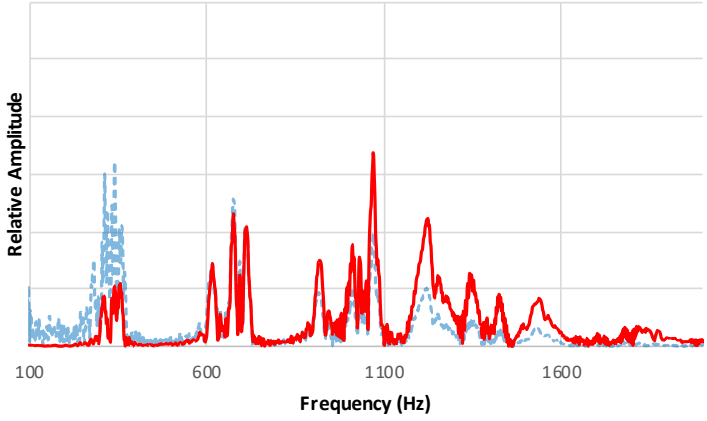


Figure 4.11: *"Hey Siri" comparison original audio an LC injected.*

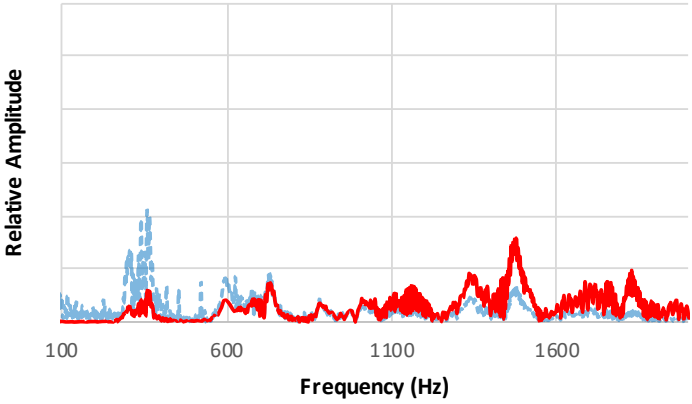


Figure 4.12: *"Alexa" comparison original audio an LC injected.*

Although there is a small data set, it is a clear pattern indicated. Laser injected voices suffer a filtering which makes the injected signal differ from the original voice. An additional recording with vowels shown in 4.13 amplifies the indication of low frequency dominating tones and higher tones being filtrated.

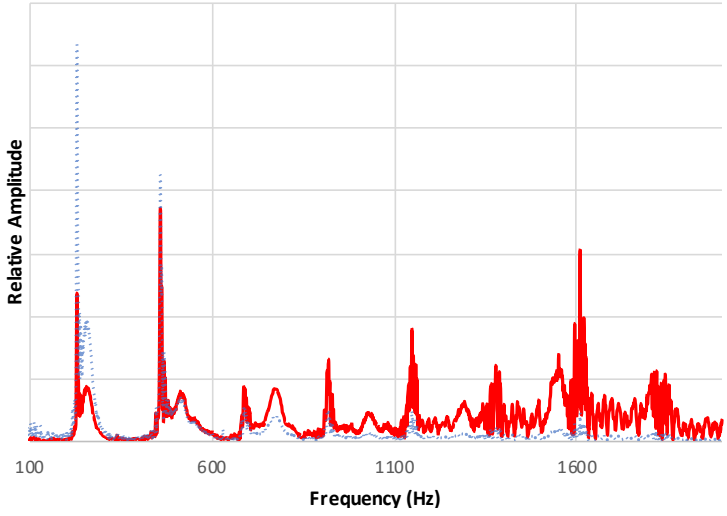


Figure 4.13: *"English Vowels" comparison original audio an LC injected.*

Power Density Ratio

The signal processing showed clearly that the thermoacoustic low-pass characteristics make original audio signals and laser signals clearly distinguishable. Down below are results from several Text-To-Speech/TTS (from http://www.fromtexttospeech.com/)

voices with several different native languages saying the same command word and their power density ratio between their voice in the frequency band between 100-1000 Hz and 1000-7000. While these are simple calculations, it indicates strong clues on how a software algorithm could distinguish the difference between a voice and an LC injected signal.

Table 4.1: *Comparison of the power density ratio TTS generated voices in the range of 100-1000 and 1000-7000 Hz.*

Name	**Nationality**	**Ratio Voice**	**Ratio LC**
Alessandra	Italian	5.6928	2.3267e+04
Giovanni	Italian	4.0670	2.3683e+04
Harry	British	3.3955	5.7710e+03
Emma	British	18.4203	3.5447e+04
Mateo	Spanish	4.7357	1.3135e+04
Isabella	Spanish	6.4433	1.9568e+04
Nadine	German	75.1609	9.7930e+03
Michael	German	21.3994	2.4986e+04
Rodrigo	Portuguese	3.4899	7.0945e+03
Valentina	Russian	37	2.4585e+04
Gabriel	French	6.4046	1.3924e+04
Alice	American	6.5125	2.3267e+04
Jenna	American	7.6884	7.6744e+04
Daisy	American	36.1019	8.3009e+04
George	American	4.1942	5.0010e+03
John	American	8.3550	2.7325e+03

While there are individual differences within the same category, there a two distinct populations where the calculated ratio differs by at least one magnitude and often even more which provides good clues on how to distinguish real sounds from LC.

As for the amplitude of the injected signal: The LC injected signals were about 1 per thousand of the original audio for all cases (30 dBSPL difference) when the main microphone received 1 mW red laser. From an audibility perspective, the recorded LC-injections sounded like a whisper indicating that a 1 mW red laser may not sufficient in waking up a device in a noisy environment.

4.2 DolphinAttack

4.2.1 Simulations

Attempts of injecting a signal in a conceptual differential model turned out fruitless. Figure 4.14 and figure 4.16 show an attempt where a Hann window was used, the carrier frequency was chosen to be Helmholtz resonance and the baseband signal was a single 4000 Hz sine signal. No other trials brought anything either.

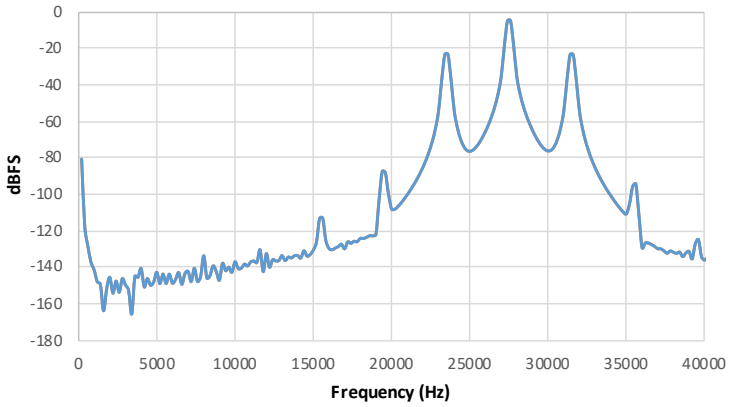


Figure 4.14: *Simulation on main mic.*

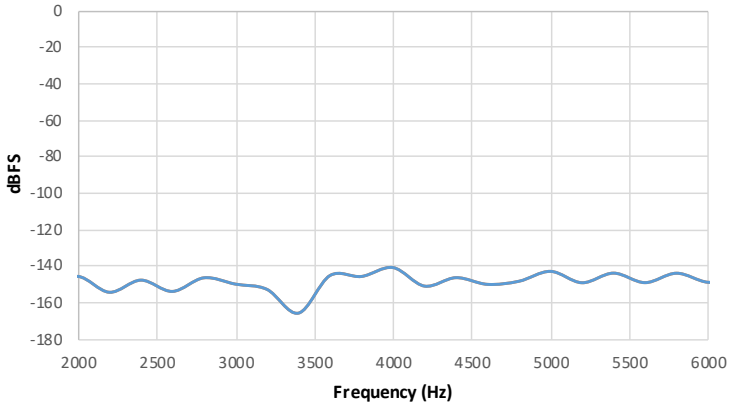


Figure 4.15: *Zoomed in at (supposed) injections.*

Voice commanded systems (VCS) have been proved to be vulnerable to signal injections mimicking voice commands and explored security flaws in market available products for the time of each respective study that originally discovered

those methods. Signal injection caused with the help of amplitude modulated ultrasonic waves (being known as DolphinAttacks - DA) were proved to work on several such devices in 2017. In 2019, another study were also successful in achieving signal injections using modulated laser also known as LightCommands (LC). This thesis has investigated the occurring circumstances which enables such injections. Simulations and laboratory trials have shown a thermoacoustic origin enabling LC to be injected and the response differs with respect to microphones physical size. DA utilizes the non-linearity of microphones and more linear microphones have indeed been shown to withstand DAs better and physical parameters have been shown to indicate how DA may be optimized for successful injections. The results have been used to provide ideas on how a VCS system can be designed to be more resilient.

This shows that differential microphones are not prone to DolphinAttacks by design but rather by non-ideality. The single backplate microphone (DaSec) did however show a successful injection in conceptual model seen in 4.16 which indicates that the weakness to DolphinAttacks is inherited in the hardware design.

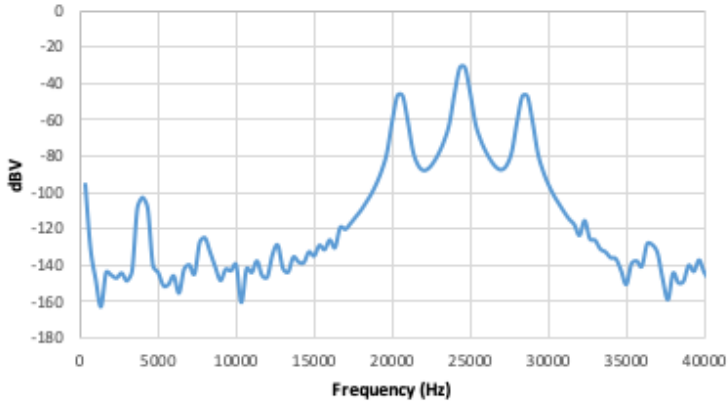


Figure 4.16: *Simulated injection at single backplate microphone.*

4.2.2 Laboratory

Carrier Sweep

The results when targeting the main microphone in an anechoic chamber with these following parameters shown in 4.2.

led to the injections shown in figure 4.17 indicating rather weak injections ranging from 10-30 dBSPL depending on carrier frequency.

Table 4.2: Properties of audio signal used for sweeping carrier frequency trials.

Signal type	AM modulated ultrasonic with sine baseband signal
Carrier frequency	23-33 kHz
Baseband frequency	450 Hz
Sound pressure received	94 dBSPL
Modulation index	100%

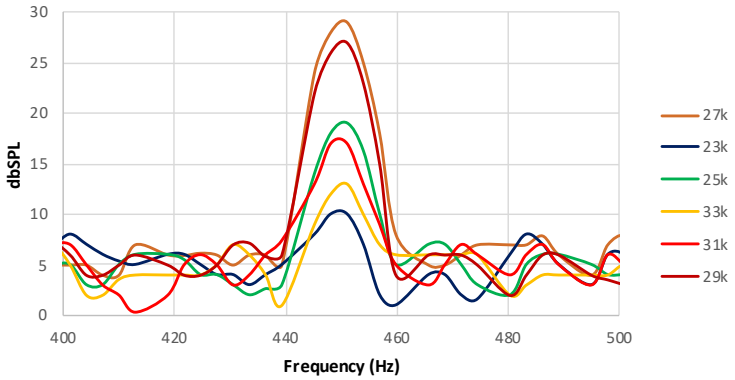


Figure 4.17: Comparison between injections in main microphone at 450 Hz when the carrier frequency is swept.

A comparison between the strongest (where $f_c = 27k$) and weakest signal (where $f_c = 23k$) is shown in figure 4.18.

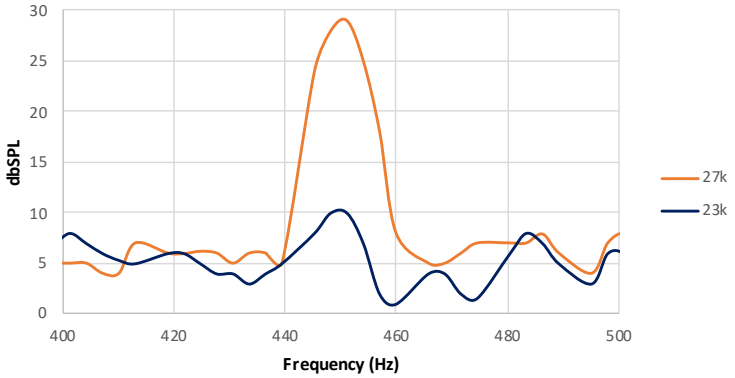


Figure 4.18: Comparison between strongest and weakest injection when targeting the main microphone.

The strongest and the weakest injection differ around 20 dBSPL/factor 100 in sound pressure. This validates the results from original study by Zhang et al. [1] that emphasis the importance of right carrier frequency for successful attacks and also indicates that a differential microphone receives rather small injection if the received ultrasonic signal is around 94 dBSPL. Using same comparison but targeting the LabSec microphone with same conditions, the injections done are shown in figure 4.19.

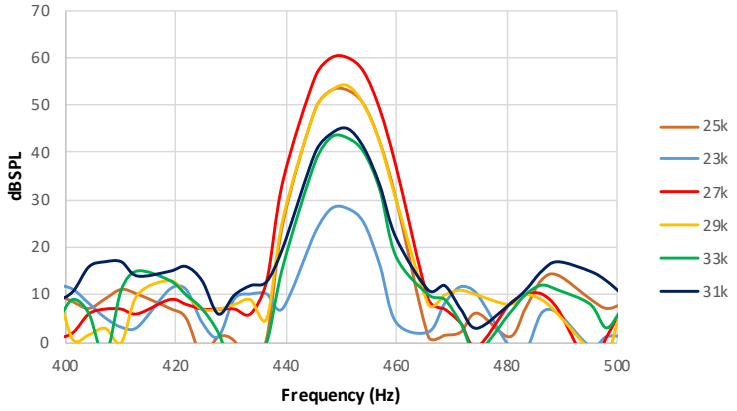


Figure 4.19: *Comparison between injections in LabSec microphone at 450 Hz when the carrier frequency is swept.*

Comparing the strongest and weakest injection shows in 4.19 a difference of 30 dBSPL showing a major difference in injection strength.

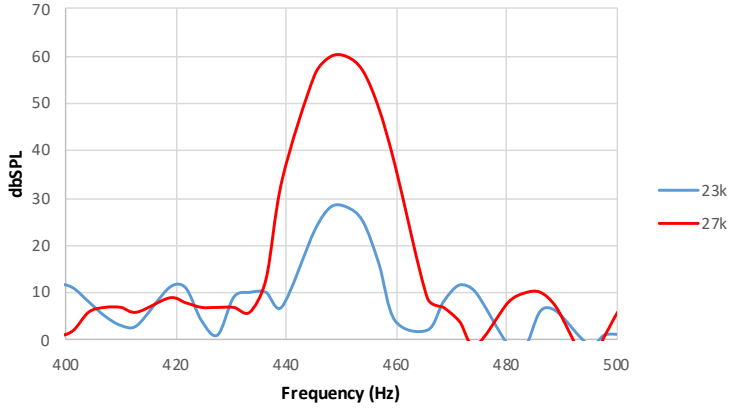


Figure 4.20: *Comparison between strongest and weakest injection when targeting the LabSec microphone.*

The choice of carrier frequency has a huge impact on the injected signal strength where the factors of injected signal strength could differ by 20-30 dBSPL.

Baseband Sweep

Fixing the carrier frequency to the microphones respective Helmholtz variants (main f_c = 27.8 kHz and LabSec f_c = 27.2 kHz), the frequencies were swept (with modulation index=1) using sine signals from 250 Hz and going up one octave except using 7000 instead of 8000 as highest frequency to prevent the side band signal with frequency $f_c - f_m$ from leaking into the audible spectrum.

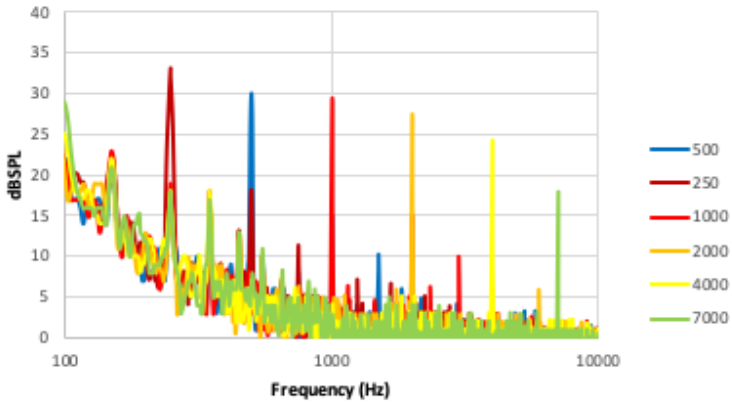


Figure 4.21: *Injection using baseband frequency sweep with 94 dBSPL ultrasound signal hitting the Helmholtz resonance on main mic.*

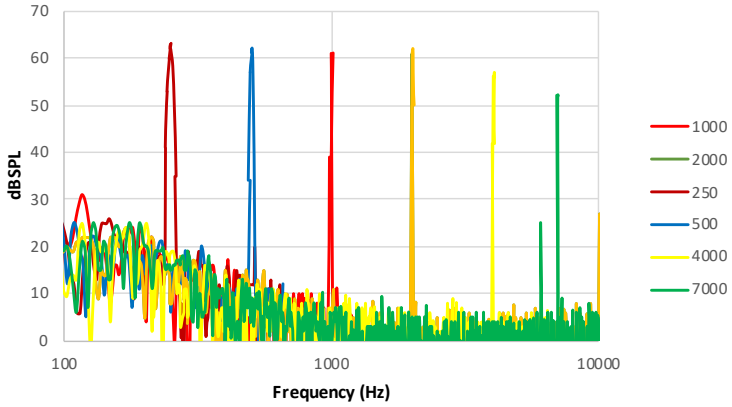


Figure 4.22: *Injection using baseband frequency sweep with 94 dBSPL ultrasound signal on LabSec.*

Minor Low Pass Behaviour: Comparing the magnitudes of each injection in both microphones show a relation of attenuation of the injections when the baseband frequency goes up. The relations for each microphone are shown in 4.3 and 4.4 indicate a slightly frequency depending attenuation pattern.

Table 4.3: *Magnitude relation between injection signals when sweeping baseband frequency on LabSec microphone.*

Baseband sine frequency	dBSPL	Relative to 250 Hz injected signal
250	63	0
500	62	-1
1000	61	-2
2000	62	-1
4000	57	-6
7000	52	-11

Table 4.4: *Magnitude relation between injection signals when sweeping baseband frequency on main microphone.*

Baseband sine frequency	dBSPL	Relative to 250 Hz injected signal
250	33	0
500	30	-3
1000	29.4	-3.3
2000	27.6	-5.4
4000	24	-9
7000	18	-15

These results show signs to be aligned with the shape of the Helmholtz resonance peak between those microphones in 4.23 where the main microphone has a bigger relative slope between resonance frequency and 20 kHz compared to the sec microphone.

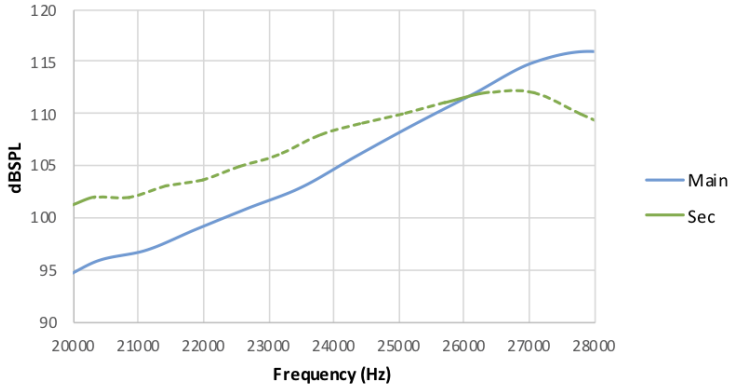


Figure 4.23: *Resonance slope comparison.*

The baseband sweep concludes three important results;

- Baseband frequency matters in terms of how strong the injected signal becomes and thus indicate a minor low-pass behaviour.
- The shape of Helmholtz resonance curve may play a role of how said low-pass behaviour looks like.
- Single backplate microphone receives stronger injections than a differential microphone.

4.2.3 Observed Multiple Order Non-Linearity

While doing the sweeps, indications of higher order non-linearity were pronounced. 4.24 shows the case when the baseband signal is a 250 Hz sine and several tones are located on octaves of the fundamental frequency.

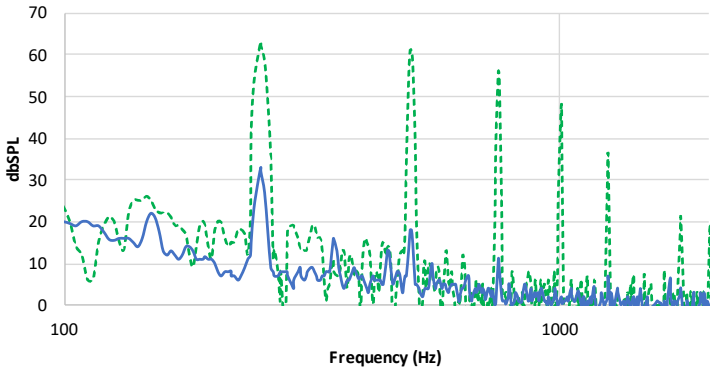


Figure 4.24: *Multiple order injections comparison between main and LabSec (dotted).*

It was more prominent in the LabSec microphone. This indicates that the full mechanism which enables DolphinAttack is not strictly of second non-linear nature but rather higher orders.

4.2.4 Voice Injections using DolphinAttack

Several injections were done in both microphones where the voice is a command word. The carrier was set to Helmholtz frequency to ensure maximal injection.

4.25 shows a comparison between the original audio sample and injection to estimate differences and similarities. All following comparisons are:

- Done on single backplate microphone.
- Logarithmic x-axis.
- DA injection is transparent blue (dotted).
- Frequency span ranges from 100 to 7000 Hz.
- Normalized amplitude to be between 0 and 1.
- Helmholtz resonance causes the peak of the carrier frequency to be amplified by ca 15 dB.

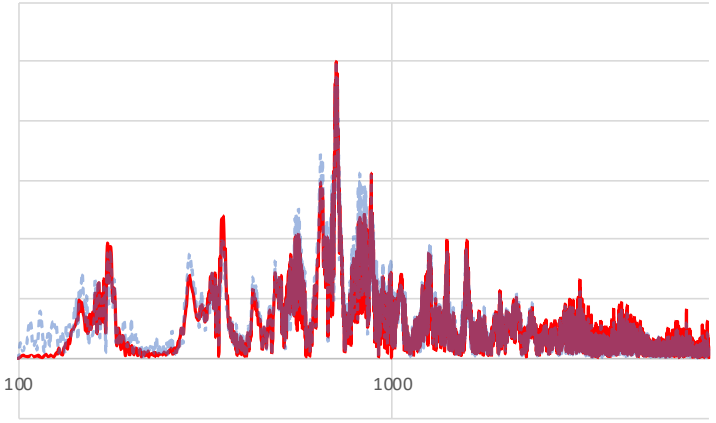


Figure 4.25: *Comparison original voice vs injection when DA is 94 dBSPL.*

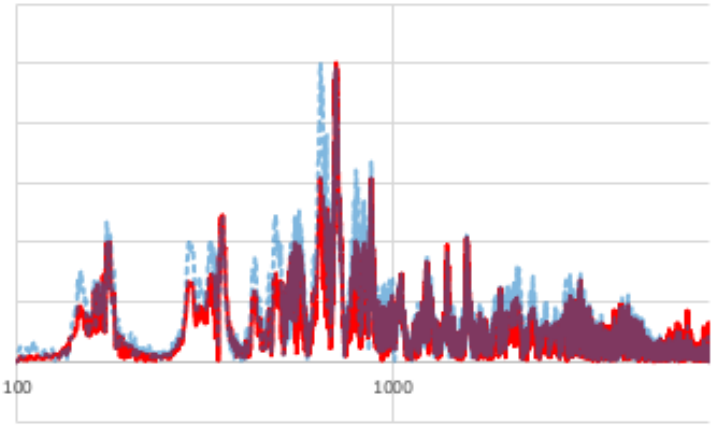


Figure 4.26: *Comparison original voice vs injection when DA is 100 dBSPL.*

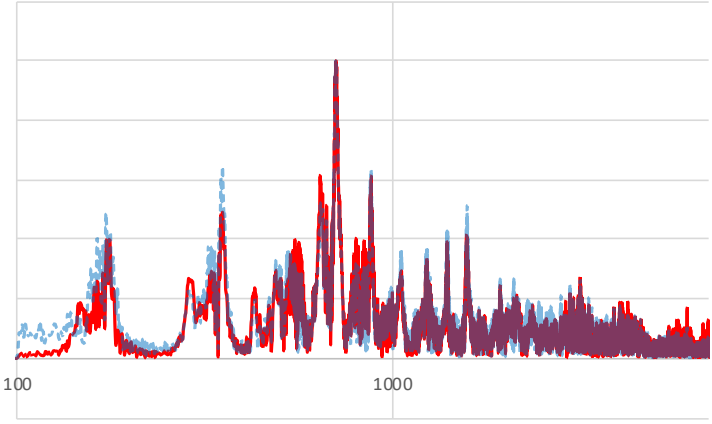


Figure 4.27: *Comparison original voice vs injection when DA is 106 dBSPL.*

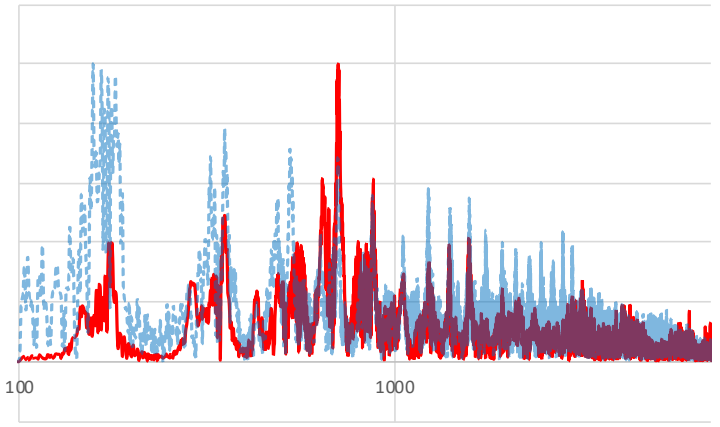


Figure 4.28: *Comparison original voice vs injection when DA is 112 dBSPL.*

For lower sound pressure, the signals are similar and could be hard to distinguish from real voices but increased sound pressure increases the difference which is noticeable in 4.28. This indicates that "less is more" could possibly apply to successful DolphinAttacks when attempting to make the injection appear nearly indistinguishable from an original voice. When attempting the same on the dual backplate microphone with the same conditions in 4.2.4 but Helmholtz resonance amplifies 20 dB instead of 15 dB: No audible injection occurred below a reception of received 107 dBSPL seen in figure 4.31 where the injection is to some extent detectable while being close to the noise floor. It was audible yet a very weak injection and required careful listening to catch it This proves the higher threshold and resilience of a dual backplate microphone.

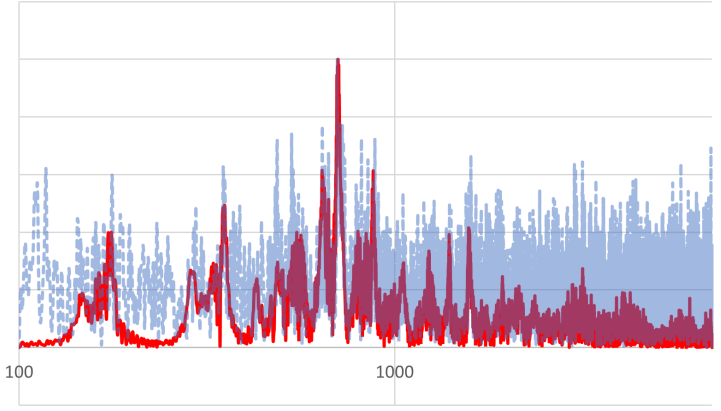


Figure 4.29: *Comparison original voice vs injection when DA is 107 dBSPL.*

The only rather "good" injection occurred as seen in figure 4.30 when it received 109 dBSPL which amplified the signal to 129 dBSPL due to the resonance which was very close to the AOP at 130 dBSPL.

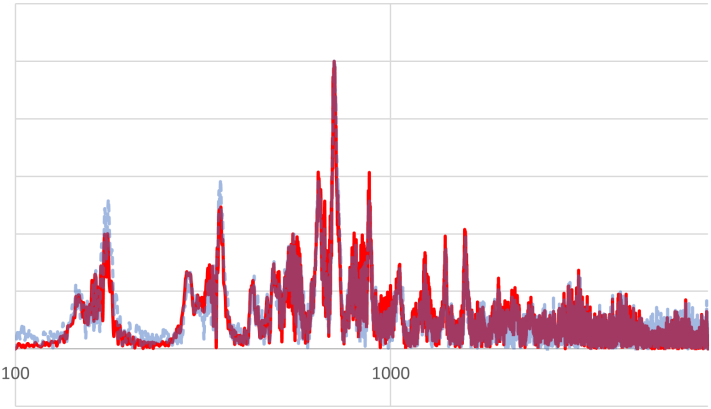


Figure 4.30: *Comparison original voice vs injection when DA is 109 dBSPL.*

Upon receiving strong enough to surpass AOP due to Helmholtz, injections were slightly audible although heavily distorted.

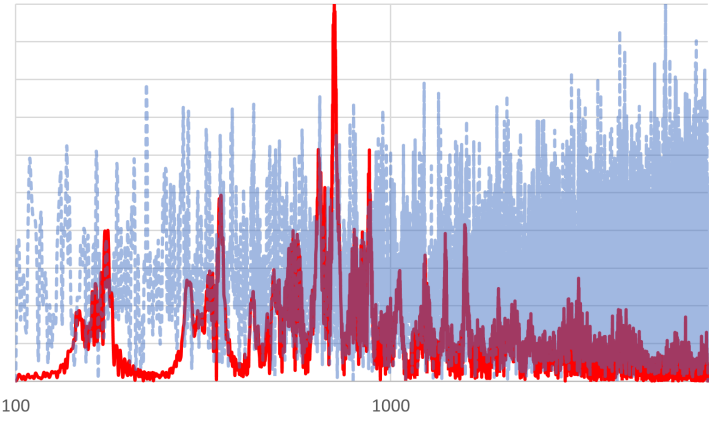


Figure 4.31: *Comparison original voice vs injection when DA is 112 dBSPL.*

The results show a small window lying close to AOP were good injections can be done in the dual backplate microphone.

The lab results all-together prove the differences of injecting a signal in different types of microphones. For a single backplate, it was relatively easy but if the

sound pressure was high enough, it introduced new frequencies making the injected signal differ from the original voice which potentially could be a giveaway for an hypothetical software algorithm designed to find an DA. A dual backplate required a strong injection that had to be within the magnitude of some single decibels to not get heavily distorted and self revealing.

Hacking Trials

Although not being the main purpose, some devices were subject to some trials tested since they were available. The small ultrasonic transducer managed to trigger the Siri-software of an Iphone 5s using close range and a carrier frequency of 29 kHz with a range of ca 5 cm. This proved that as of spring 2020, an iPhone 5S is susceptible to DolphinAttacks.

An Amazon Echo device was successfully hacked by 27 kHz DA from a distance of ca 1.5 meters and could as well be triggered by a LightCommand if the laser hit the the mid microphone of the device. This shows that there are products that as of spring 2020 were still susceptible to DA and LC.

5
Discussion

In this section the methods and the results of the thesis will be discussed as well will possible solutions be discussed.

5.1 Potential solutions

5.1.1 LightCommands

Software Solutions

The power ratio shown in table 4.1.4 indicated a strong filtration of higher tones and made clear separation between real voices and LC injected analogs. While this may not be fine tuned enough, it proves a great potential on what software engineers could possibly look after when designing voice commanded systems that are precise enough to never mistake a voice from an injection.

It is important to take into consideration that each microphone may have a unique thermal filter pattern which will give difference injection patterns from microphone to microphone. A software engineer developing speech recognition algorithms should have some knowledge on the specific microphone and its application in the specific system to create an optimized algorithm to match the pattern of the thermal response for the microphone.

If the attacker is aware of the thermal response then it could maybe be possible to use an equalizer to match and inject a signal similar to a real voice. Such action may however require knowledge on which microphones are used by the products on the market but should be taken into consideration if LC turns out to be a true security risk and for engineers developing protection against LC.

Laser Blockage

This study has not investigated blockage or reflection of the laser. However, the AC-response in 4.10 indicated a peak around -30 dbFS which corresponds to 100 dbSPL with 1 mW red laser and 10% i.e 0.1 mW absorption. These results could use as an indicator on how much attenuation would be needed in an light blocking application. It was proposed that blockage could be a possible solution in the original LC paper [2] and this thesis result may serve as a reference to how much thermal absorption attenuation can be implemented in the acoustic pathway without degrading the acoustic performance to a level where it affects user experience.

5.1.2 DolphinAttack

Dual Backplate/Differential Microphone

The results have shown that signals can be injected in dual backplate microphone. However, the results also show that it is very unfeasible and performing DolphinAttack on one in reality may be very hard and in many practical cases, impossible. The received sound pressure from the microphone had to be high and preferably use a carrier frequency which is the same as or close to the Helmholtz resonance which would make it next to impossible to hack with small discrete speakers. By using dual backplate microphone in voice commands assistance-devices, the threshold level of successfully be able to perform successful attack is very likely to increase by a big factor and may decrease the risk to a very high extent. In case of an attacker that would use very strong ultrasonic speakers at a close distance and one clear injection is done, harmonic distortion caused by the high sound pressure could possibly be used as a clue and should possibly be investigated further by software engineers developing speech recognition algorithms.

Software Solutions

Vibration Domain: A study made by Wang et al. [19] showed a possible solution against DA with a high precision by measuring the signal in the vibration domain using an accelerometer. Their method had a detection rate of 99.9%. This thesis does not provide any new content on that subject but the confirmation of injection occurring inside the microphone further proves their solution proposal could make a device resistant against DA, at least airborne AM-modulated waves.

Multi-Microphone Crosscheck: A device using several microphones and in particular at least one differential microphone and one single backplate would likely be able to detect a DA due to the difference in the perceived voice signal where the differential may detect a very weak or no injection. An enhanced speech recognition algorithm could utilize this to repel DA.

Possible Target Traces The studies have shown that several factors from hardware design affect the injected signal which releases clues about a potential hacking attempt.

- Shape of Helmholtz peak may affect the pattern of a resembling low pass filter but with a very small attenuation
- Non-linear behaviour of the microphone of second and higher order.
- Each microphone (and device) may create a unique pattern combining the factors which makes DolphinAttacks differ slightly on each unique microphone.

Investigations on how to create a predicting non-linear distortion model has been done by Maziewski et al. [20] where models taking up to third order could predict most of their measured non-linear distortions. Their models and reasoning is supported by the results pointing on multiple order non-linearity shown in 4.24.

Both this study and their indicate which clues and behaviours a potential speech recognition algorithm possibly could aim to detect in order to discover an attempted DolphinAttack. Those clues include the shape of frequency response in the ultrasonic range in combination with a non-linear prediction model to find deviating patterns on unique microphone models to properly adjust the detection precision.

Possible AOP distortions could possibly be included in the pattern recognition because the distortions from the AOP would occur together when a voice command appears specially if the DolphinAttack has a sound pressure close to AOP either by a strong source or by amplification from Helmholtz resonance. If the distortion also appears once again when the execution occurs, it increases the probability that there is an ongoing DA.

5.2 Limitations

This thesis met some limitations along the way as methods were chosen that may have had an impact on the final results. The main limitations along the way were

- Investigating LC and DA limited the quantitative data and trials due to parallel work.
- Conducting satisfying DA were harder than expected since most of the transmitters did not satisfy experiment conditions.
- The simulation models were done where at conceptual level which may have limited the DA simulations.
- No DA voice injections were done on carrier frequencies outside of Helmholtz Frequency due to the limitations in output power from the speaker which mean that potentially interesting data got left out.
- The covid-19 pandemic was on the rise at the beginning of the thesis which led to home office work and limited time in the laboratory.

It was in particular DA that got affected by the limitations. It could be verified in conceptual models that the weakness against DA is inherited due non-linearity that comes from the construction of the MEMS, however, it could not be determined whether which sources were dominant. For differential microphones, there was correlation between the AOP and success of injection but it can't be determined from this thesis alone. This topic should be further studied.

6
Conclusion

The thesis has given a greater understanding of how LC and DA work. It has elaborated previous studies and also showed how to simulate them in SPICE and has given new clues on precise targets for protection.

6.1 Conclusions on LightCommands

The major conclusions that the thesis has shown on LightCommands are;

- Signal injections occur due to thermoacustic effects when the membrane absorbs the laser.
- The thermal response is (most likely) unique for each microphone leading to different signal injections from the same source.
- LC can be simulated and verified with high precision in SPICE using a thermal model that links laser as input to an acoustic model of the microphone.
- LC Injected voice signals will be different from acoustical signals because of a different frequency response in the microphone making them differ from real voices.

.

6.2 Conclusions on DolphinAttack

The major conclusions that the thesis has shown on DolphinAttacks are

- The non-linearity is of multiple orders.
- The injection gets stronger if the attack hits the Helmholtz resonance of the microphone.
- Dual backplate microphones can withstand a DA very well and make it an unfeasible method in many practical cases or even be impossible to perform successfully.
- For single backplate, higher sound pressure tends to make the signal differ more in the frequency domain compared to signals from real voices.

.

6.3 Future Work

Based on the results of this thesis, proposal for future works, includes;

6.3.1 LightCommands

- Always developing a thermal model when developing new MEMS microphones to enable prediction of LC injected signals in simulations and provide data that can be used to help SR algorithms recognize such attacks.
- Investigate further into unique thermal patterns of microphones to determine general LC injection appearance.
- Investigate how the impact of color (wavelength) of the laser has on successfully injecting LC. This could in particularly be of interest for MEMS developers or system designers considering increasing the reflection of the microphone/system to reduce absorption.

.

6.3.2 DolphinAttack

- Investigating whether the the non-linear distortions from AOP can contribute to the injection from a DA and if it could be utilized when attempting DA detection.
- Make further investigation on how to predict non-linear behaviour in microphones and apply it in future SR-algorithms.
- Examine whether the Helmholtz resonance amplification can be damped without compromising the audio quality since a bigger resonance opens up to perform a successful DA using lower sound pressure.

A similar method to DA known as SurfingAttack (SA) has been proven to work as well for same purposes by using ultrasound waves propagate through other material than air [21]. While SA has not been investigated in this thesis and nothing can be presented about SA, the similarities make it an interesting case to also follow up on for those considering implementing any protection against signal injections caused by ultrasonic waves.

Bibliography

[1] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack," *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, Oct 2017.

[2] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu, "Light commands: Laser-based audio injection attacks on voice-controllable systems," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020.

[3] M. David Thomas, *Design, Fabrication, and Characterization of a MEMS Dual-Backplate Capacative Microphone*. PhD thesis, University of Florida, 2007.

[4] S. Zawawi, A. Hamzah, B. Majlis, and F. Mohd-Yasin, "A review of mems capacitive microphones," *Micromachines*, vol. 11, p. 484, 05 2020.

[5] "Mems microphones market size, share & trends analysis report by type (digital, analog), by snr, by technology (capacitive, piezoelectric), by application, by region, and segment forecasts, 2019 - 2025," 2019.

[6] M. Fueldner and A. Déhe, "Dual back plate silicon mems microphone : Balancing high performance !," (Nürnberg), 2015.

[7] D. Cattin, *Design, Modelling and Control of IRST Capacitive MEMS Microphone*. PhD thesis, University of Trento, 2009.

[8] K. Poskus, "Acoustic analysis of wave-guide and mems microphone in camera including thermoviscous losses," Master's thesis, Blekinge Tekniska Högskola, 06 2018.

[9] "Microphone specifications explained," June 2016. `https://invensense.tdk.com/wp-content/uploads/2015/02/AN-1112-v1.1.pdf`.

[10] P. Joseph, "Sound from ultrasound : the parametric array as an audible sound source," 08 2005.

[11] F. Gao, R. Kishor, X. Feng, S. Liu, R. Ding, R. Zhang, and Y. Zheng, "An analytical study of photoacoustic and thermoacoustic generation efficiency towards contrast agent and film design optimization," *Photoacoustics*, vol. 7, pp. 1 – 11, 2017.

[12] J. A. Hasenbichler, "Thermoacoustical simulations," (TU Wien), 2016.

[13] M. Parti, "Mass transfer blot numbers," *Periodica Polytechnica Mechanical Engineering*, vol. 38, no. 2-3, pp. 109–122, 1994.

[14] F. Incropera, D. DeWitt, T. Bergman, and A. Lavine, *Fundamentals of Heat and Mass Transfer*. Wiley, 2007. https://books.google.se/books?id=_P9QAAAAMAAJ.

[15] K. M. Tenny and J. S. Cooper, "Ideal gas behavior," Jan 2021. https://www.ncbi.nlm.nih.gov/books/NBK441936/.

[16] S. Söderkvist, *Kretsteori och Elektronik*, p. 103. Tryckeriet Erik Larsson AB, 4 ed., 2005.

[17] N. I. Corp, "Amplitude modulation." Accessed: 2020-06-01 https://www.ni.com/sv-se/innovations/white-papers/06/amplitude-modulation.html.

[18] Purves D, Augustine GJ, Fitzpatrick D, et al., editors. Neuroscience. 2nd edition. Sunderland (MA): Sinauer Associates; 2001. The Audible Spectrum. Available from: https://www.ncbi.nlm.nih.gov/books/NBK10924/.

[19] C. Wang, S. A. Anand, J. Liu, P. Walker, Y. Chen, and N. Saxena, "Defeating hidden audio channel attacks on voice assistants via audio-induced surface vibrations," in *Proceedings of the 35th Annual Computer Security Applications Conference*, ACSAC '19, (New York, NY, USA), p. 42–56, Association for Computing Machinery, 2019.

[20] P. Maziewski, J. Banas, D. Koszewski, D. Stanczak, and P. Trella, "analysis of nonlinear distortions in a digital mems microphone," *journal of the audio engineering society*, may 2020.

[21] Q. Yan, K. Liu, Q. Zhou, H. Guo, and N. Zhang, "Surfingattack: Interactive hidden attack on voice assistants using ultrasonic guided wave," in *Network and Distributed Systems Security (NDSS) Symposium*, 2020.