On Determining Smartphone Microphone Directivity with Application to Beamforming

R. Brinkman
T. de Rooij

Delft University of Technology

TUDelft
Delft University of Technology

Challenge the future

Abstract

An equiangular sampling scheme and accessory measurement set-up for determining the directivity gain pattern of a commercially available smartphone is presented. Determining this directivity gain pattern based on these measurements was not possible, because next to the impulse response of the microphone also the impulse response of the acoustic system. Different smartphone set-ups are tested and the directivity of a smartphone on a surface differs from a smartphone in mid-air. The determined directivity gain pattern is suitable for use by a beamforming algorithm, but it resulted in no improvement of the performance of the algorithm.

Interpolation of the measured results was not considered, as the system for which the measurements are intended cannot measure accurate smartphone orientations. Other work has shown that the orientation of a modern smartphone is not yet known within $10°$ margins, therefore interpolation of the measured results is not considered. This could be included in future work, along with to the determination of indoor position of a smartphone, a sampling scheme with fewer samples and a means to reverse the contribution of the acoustic system to the measurements.

Acknowledgements

Contents

List of Figures

Chapter 1
Introduction

A conference call is a telephone call in which two or more parties, possibly consisting of more than one person, can simultaneously communicate. Conference calls are frequently used in large companies because they allow parties to communicate efficiently without travelling to the same location. They facilitate team meetings or other occasions in which employees at different locations want to communicate by telephone.

Many large companies currently make use of expensive equipment in specialized rooms. These systems are often difficult to configure and even then sound quality can not be guaranteed. Because the audio signals are recorded with one microphone, most of the time positioned at the centre of the table, there is a lot of noise and options for noise cancellation are limited. A more accessible and convenient solution would be to only use the participants smartphones for such a system: a clear conference call could be held anywhere using a device nearly everyone already possesses.

Another application of such a system could be to help people with hearing aids to communicate. Someone with a hearing aid has the same kind of problems as a modern conference call system: they cannot filter sounds from different directions, the sound volume is the same from every direction. For example, when someone with a hearing aid is in a pub, it is very difficult for that person to concentrate on the voices of the people around him. This is due to the voices of a person at for instance a table to his left being at the same volume as the voice of the person sitting in front of him. This person could be helped if just a few people placed their smartphones on table and started an application, which connects to the hearing aid and dampens the distracting noises from the other tables at the pub.

An example of a possible conference calling system configuration is shown in Figure 1.1. Several smartphones are located on a table in a regular office environment. A number of desired audio sources (e.g. people speaking) and undesired noise sources (e.g. an air conditioning unit or background conversation) are located around the table. The smartphones are placed in an arbitrary but known positions and the locations of the desired and undesired sources are also assumed to be known.

The smartphones are running a mobile application from which the sound data is acquired. This data is then transmitted via a WiFi network and the signals are synchronised in order to make beam-
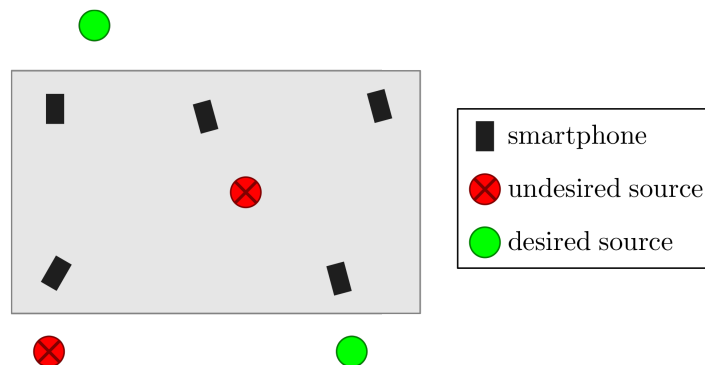

Figure 1.1: An example of a conference calling system configuration

1

forming possible [3]. Beamforming is a technique in the field of signal processing with the purpose to transmit or receive signals directionally [4]. The described system applies acoustic beamforming. Hennecke and Fink [5] concluded that modern smartphones are suitable for acquiring audio signals as input for a beamforming algorithm. The received signals from the smartphone microphones can be processed into one resulting audio stream.

These microphone signals can be used to amplify the desired signal and suppress background noise. This is done by spatial filtering, using the given location information of the smartphone microphones and desired and noise sources. There exist several acoustic beamforming algorithms, each having their own advantages and disadvantages [6]. Beamforming will be further explained in this thesis in Section 2.1.1, as this thesis focuses on a three-dimensional directivity gain pattern to improve the result of a beamforming algorithm. A directivity gain pattern, from now on referred to as *directivity*, gives information about the directional gain of, for instance, a microphone. The directional gain indicates how a signal of a certain frequency originating from a certain place is attenuated and delayed.

Research on the application of smartphone microphone directivity in a beamforming algorithm has been performed before [2], but this research only describes the directivity in one plane and not for all directions. Still, this research points out that including the two-dimensional microphone directivity in a beamforming algorithm considerably improves the results of the beamforming algorithm. It is thus plausible that a three-dimensional microphone directivity will enhance the performance of a beamforming algorithm even more. The aim of this thesis is to determine the three-dimensional directivity of a smartphone microphone. This directivity will be implemented in a larger system to apply beamforming on audio signals recorded by smartphones. The directivity behaviour will be determined by examining the impulse response of the microphone for a variety of sample directions.
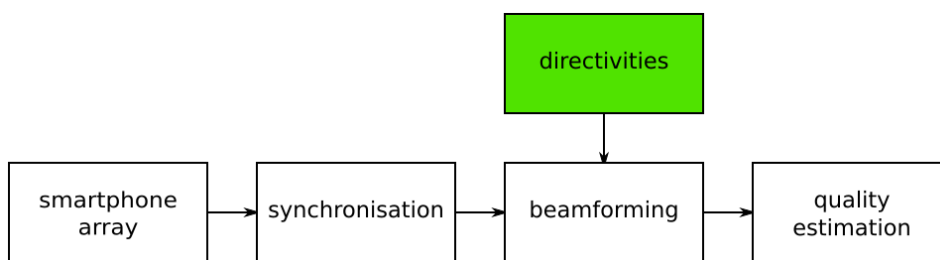
Figure 1.2: The contribution of the directivity measurements in the whole system.

1.1. Objective

An overview of the whole system is shown in Figure 1.2. This thesis describes one of the three parts of the design to make a conference calling system as describe above possible. The part concerning the smartphone array, including communication between smartphones and the computer, and the synchronisation of the audio signals is left to Bosma and Smeding [3]. The implementation of different beamforming algorithms and the quality comparison for these different algorithms was performed by Van Wijngaarden and Wouters [6].

The green box marks the study of this thesis: determining the directivity gain pattern of a smartphone microphone and apply this directivity to a beamforming algorithm. This thesis will describe the process of measurement, processing and displaying the data. An important question that needs to be answered is how representative one measurement is for different settings of the smartphone and for different phones of the same type. For this project Google's Nexus 5™ mobile phone is used. Not much information is available about the microphones used in this smartphone or their direc-

tivities. Therefore the directivities of the microphones will have to be determined by measurement, with as much accuracy as possible in the given time. The objectives for the function, as a part of the beamforming system, are specified more precisely in the schedule of requirements, in Appendix A.

1.2. Thesis structure

This thesis starts with the theoretical background of measuring microphone directivities in Chapter 2. Subsequently the available resources, measurement set-up and decisions made will be discussed in Chapter 3. The results of these measurements were not directly usable, but needed to be equalized to support the beamforming algorithm. This process is described in Chapter 4. After this, the final results and a discussion of these results will be presented in Chapter 5 in Section 5.1 and Section 5.2 respectively. Finally, a conclusion and our recommendations for future work are presented in 7.
Ethical considerations regarding the project are added in Appendix B.

1.3. Definitions and notation

Throughout this thesis, there will be definitions and (mathematical) notation, which will be presented in advance. If $x[n]$ is a discrete time signal, with N entries and $n = 1, 2, \ldots, N$, $\mathbf{x}$ is the vector notation of the signal and $X[\omega]$ ($\mathbf{X}$ in vector) is its discrete Fourier domain counterpart. $\mathbb{N}_0$ denotes the set of natural numbers, with 0 includes: 0, 1, 2, 3,… and $\mathbb{N}$ denotes the set of natural numbers larger than 0: 1, 2, 3, 4,…. For matrices non-bold capitals will be used. The convolution $c[n]$ of two signals $a[n]$ and $b[n]$ is written using an asterisk:

$$c[n] = (a * b)[n].$$

For vector multiplication, assume two vectors $\mathbf{a}$ and $\mathbf{b}$, then its dot product or inner product $\mathbf{c}$ is written as:

$$\mathbf{c} = \langle \mathbf{a}, \mathbf{b} \rangle$$

and its cross product or vector product $\mathbf{d}$ written as:

$$\mathbf{d} = \mathbf{a} \times \mathbf{b}.$$

For rounding the following signs are used: $\lceil \bullet \rceil$ for rounding up and $\lfloor \bullet \rfloor$ for rounding down.
Quantities in this thesis are often expressed on a logarithmic scale using decibel (dB) units. Since the units in these thesis are field units (unless specified otherwise), the field unit form of the decibel scale is used:

$$20 \log_{10} | \bullet |.$$

From this follows the inverse of the dB-scale:

$$10^{\frac{\bullet}{20}}.$$

The definitions of the spherical coordinate-system can be found in Figure 1.3a. The azimuth, θ, is the angle between the projection of the vector pointing from the origin to the point on the x, y-plane and the x-axis. The elevation, ϕ, is the angle between the projection of the vector pointing from the origin to the point on the x, z-plane and the z-axis. In Figure 1.3b, the definitions of the different sides of a smartphone are given. The top of the smartphone, we will be called the listeningside, the bottom the speechside, the left side of the screen is the left side of the phone, the other side is the right side. For specifying the orientation of the phone, the terms 'screen up' and 'screen down' or 'face up' and 'face down' are also used.

If the phone is placed in the spherical coordinate system, the listeningside of the phone is pointing towards $(\phi, \theta) = (90°, 0°)$. Laying in the (x, y)-plane, with the screen face up towards the $+z$-direction.

The last thing to note is that all code for this thesis was written for MATLAB Student R2014a. For some some of the hardware used in this thesis special drivers are needed, which can be found via the reference given when the hardware is mentioned.

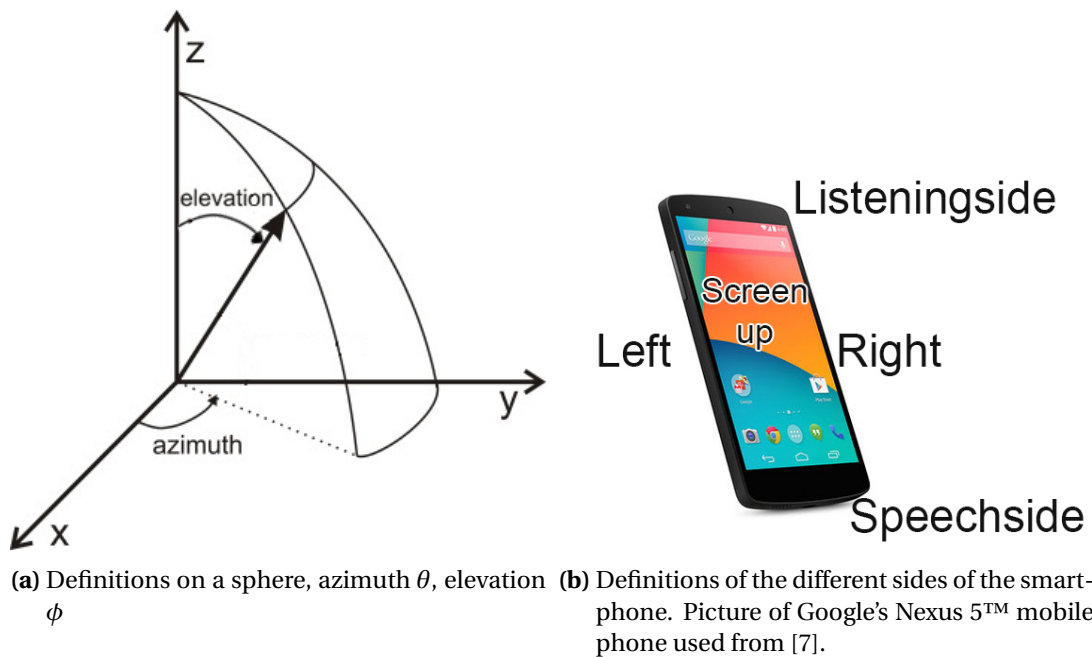**(a)** Definitions on a sphere, azimuth θ, elevation ϕ **(b)** Definitions of the different sides of the smartphone. Picture of Google's Nexus 5™ mobile phone used from [7].

Figure 1.3: Definitions used throughout this thesis.

Background

In this chapter most of the theory about measuring the directivity of a smartphone microphone is addressed. The related research will be discussed first, after which the operations of a beamforming algorithm is briefly explained, followed by a comparison between different sampling schemes for spherical sampling and methods to measure the impulse response. Of these impulse response determination methods, two techniques will be explained in more detail, since they will be used in the measurements (further explanation of this choice is given in Chapter 3). In the next chapter, the final choices concerning the directivity measurements will be made on base of this presented theory.

2.1. Related Research

This research is a continuation of the work of Gaubitch et al. [2], who concluded the directivity of a smartphone microphone could be used to improve beamforming algorithms. While some smartphone directivity measurements were already conducted for this paper, these were only performed in a two-dimensional plane around the smartphone, instead of a full three-dimensional sphere [2]. The smartphone available for the measurements is Google's Nexus 5™ mobile phone [8]. No information about the microphone used in this mobile phone is made available by Google [8] or LG [9], the two parties that produced it. Information about microphone control and automatic gain control from the smartphone is also not given [3].

Determining the directivity is, in short, determining the impulse response of the microphone with respect to signals from different directions. This comes down to two parts: impulse response determination and sampling in space. To take the measured directivities into account in the beamforming algorithm, the measured results must be converted to a more suitable form for the algorithm. The directivities are measured for a certain number of data points. The beamforming algorithm also needs the directivities for an intermediate value, thus interpolation is necessary. These interpolation calculations will take time and computing power, a possible source of trouble in the beamforming algorithm [6]. The algorithm will use the information from the measurements multiple times per second [6], so performing this interpolation during the actual beamforming is undesirable. Instead, all the interpolation is done as a preprocessing step as the algorithm only works with non-moving and known places in space [6]. As a result, when directivity data for a given orientation is requested, only a look-up is performed

In the next section, the operations of a beamforming algorithm is discussed, as well as how to account for the directivity of (for instance) a smartphone microphone in these algorithms.

2.1.1. Beamforming

This thesis is about the directivity of a smartphone microphone, which is to be used in a beamforming algorithm. It is therefore important to know more about beamforming in general, to specify the type of beamforming used in this research and to recognize the role of the microphone directivity in such an algorithm.

Beamforming is a widely used technique in the field of signal processing with the purpose to transmit or receive signals directionally. Several algorithms for achieving this goal are described in

the literature [10]. Beamforming makes use of an array of sensors to achieve spatial filtering [4]. A beamformer processes the spatial samples of propagating wave fields, collected by the sensor array. Signals with overlapping frequency content but different spatial origin are separated. Desired signals from a certain direction can then be estimated despite the presence of noise and interfering signals. In this way, the beamformer can select the desired signal and eliminate the undesired one.

Because of the variety of beamforming applications, several distinctions have to be made to determine the type of beamforming in this project. The beamforming discussed here is about voice signal reception with smartphone microphones. Beamforming can be used with different types of sensor arrays. Sensor arrays may vary in geometry, number and size [4]. The array geometry could vary from linear, circular, spherical to arbitrary geometries. This research focuses on so-called ad-hoc microphone arrays, which are microphone sensor arrays with arbitrary geometries [6].

Beamforming can be classified as either near-field or far-field, with different algorithms required for each. The situation depends on the dimensions of the array [11]. If the wave field originates from a distance far greater than the dimensions of the array, the situation can be considered far-field. There will be no difference in Direction of Arrival (DOA) for every sensor in the array, because the wave field can then be assumed to have a flat wave front. However, in the case of near-field beamforming the distance between the source and the array is comparable to the dimensions of the array. This will lead to a noticeable difference in DOA and amplitude of the signal at every sensor. During a conference call, the distance between the participant (source) and the smartphone microphone will be small. Therefore, this is considered a near-field scenario [6].

The last distinction to be made concerns the choice between broadband and narrowband beamforming techniques [12]. This choice depends on the signals incident on the array. The size of an array in terms of operating wavelength is important in measuring array performance. Consider a linear array with a fixed number of elements and fixed inter-element distance. For high frequency signals (with small wavelength) incident on the array, this fixed array will appear large and the main beam will be narrow. In case of low frequency signals (with large wavelength), this same array will appear small and the main beam appears wide. For application of the processed signals in speech communication the beamforming algorithm has to operate well in a frequency band from 300 Hz to 3500 Hz [13]. A speech signal is a very wideband signal, which covers some four octaves. If a speech signal is used in a narrowband array, this will give a disturbing speech output. This is the result of a varying beam width caused by the wide frequency range. The interfering signal will not be uniformly attenuated but instead low-pass filtered over its entire band [12]. In speech applications the narrowband assumption is therefore never valid and a beamformer designed specifically for broadband applications need to be used.

Beamforming with smartphone microphone arrays is a form of acoustic beamforming. It can be applied in a room with multiple speakers, which is the case during a conference call. The desired signal originates from a speaker's mouth and is hindered by the signals caused by other speakers and room reverberation [12]. In this case, spatial filtering can be applied because the interfering sources usually originate from points in space apart from the unwanted signal origins. When the locations of the speakers and smartphones are known, the beamformer algorithm can be used to select the voices and sounds you want to hear and reject noise.

Each ith microphone in the array observes a signal $Y_i(\omega)$ of the form of equation (2.1), with $S(\omega)$ is the desired voice signal. The presence of noise sources is indicated by $V_i(\omega)$. The signal travels through the room and through the microphone, each affecting the desired signal. The acoustic transfer function of the room is given by $H_i(\omega)$ [2],

$$Y_i(\omega) = S(\omega)H_i(\omega)G_i(\omega,\theta,\phi) + V_i(\omega). \tag{2.1}$$

The frequency response of the smartphone microphone can be defined by $G_i(\omega,\theta,\phi)$. As can be seen, this frequency response depends on both the frequency and orientation of the smartphone

microphone relative to the source (the orientation is the direction from which the signal arrives with respect to the speechside of the smartphone, given in ϕ and θ).

This frequency *and* orientation dependency of the microphone is called the microphone directivity. The directivity is related to how a microphone receives sounds from a source dependent on their relative orientation. An approaching signal will thus be recorded with different intensities, according to the side it comes from. These differences need to be considered in the calculation of the weights by the beamforming algorithm [2, 14] . This thesis is about the microphone directivity, more details about the beamforming algorithms are given in [6].

2.1.2. Sampling schemes

For determining the directivity of a microphone, the impulse response of the microphone will need to be determined for different points in space. A logical sampling scheme (the placement of sampling points in space) is a sphere, with the smartphone in its center, for it has equal distances all sampling points to the smartphone. The advantage of this is that there is no correction needed with respect to the attenuation because all distances are the same.

Sampling and interpolation methods on a sphere are rooted in non trivial mathematics [15]. Zhang et al. [1] show that there exists a minimum number of measurement points M on a sphere

$$M \equiv \left\lceil \left(\frac{e\pi s f}{c} + 1 \right)^2 \right\rceil. \tag{2.2}$$

With $e = \exp(1)$, s the radius of the sphere, f the highest frequency of the signal to be measured and c speed of sound.

Zhang et al. [1] also describe and compare four different sampling schemes for a sphere: three existing sampling schemes and one developed by Zhang et al. [1], IGLOO, an example of the distribution of sampling points on a sphere using the IGLOO method can be found in Figure 2.1, for which they took in consideration that it is desirable to keep the rotations to a minimum number of steps [1]. Their results are given in Table 2.1.

Equiangular grid is a grid equally divided in latitudes and longitudes. The biggest drawback is the overly densely sampled pole-region, which is reflected in the high number of samples needed compared with the ideal number.

Gauss-Legendre sampling takes the points as roots of a Legendre polynomial and corresponding weights are determined by the Gauss-Legendre method. The downside of this method is that there is no regularity in the sampling region and that the sampling grids for different resolutions are completely dissimilar. As such, a lower-resolution sampling grid cannot simply be taken as a subset of a higher-resolution grid.

HEALPix requires a lot of sample points with respect to the other three sampling methods and the stated minimum (equation (2.2)). Also, the azimuthal positions change with the elevation.

IGLOO divides the sphere into a number of base regions subject to a minimum distortion criterion [1]. The data posseses an exact discrete azimuthal symmetry which allows fast and precise spherical harmonic transform computation. The sample locations are more suitable for automatic measurement.

Very recently new research on grid sampling on spheres has been done[1] [16, 17], which might provide an improvement in following work.

---

[1]These papers will be published in the proceedings of the International Conference on Acoustics, Speech and Signal Processing 2015 (ICASSP'2015), which took place after the start of the Bachelor Graduation Project.
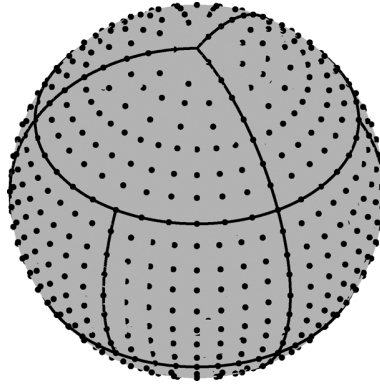
Figure 2.1: An example of a IGLOO sampling scheme, Fig. 1 from [1]. *Picture of the 3:6:3 equal area division, which divides the sphere into 12 base regions, three at either cap and six* 60° × 60° *equatorial regions. Here, each base region is sampled with 64 points.*

	Equiangular	Gauss-Legendre	HEALPix	IGLOO
Number of samples (for 20 kHz bandwidth)	8836	4371	12288	3072
Equal area division	No	No	Yes	Almost
Hierarchical	Yes	No	Yes	Yes
Iso-longitude	Yes	Yes	No	No

Table 2.1: Comparison between four different methods for sampling on a sphere, Table I from [1]. The ideal number of samples is 2209.

2.1.3. Interpolation

The resulting impulse responses can be labelled with two dimensions: an elevation ϕ and an azimuth θ. The complex gain of a signal however is also dependent on frequency. This complex gain can be determined by using the Fourier transform of the impulse response. Interpolation is needed to find the impulse response on a full sphere. However [3, 18] the azimuth and elevation determination of a modern smartphone is not so precise. It often differs more than 10°. Therefore interpolation in the spatial domain is not a great concern. The ϕ and θ error due to imprecise orientation measurement outweighs the error caused by nearest-neighbour interpolation. It would be a waste of time to compute the complex gain for a given spatial coordinates, which probably differ about 10° from the real place.

Interpolation in the frequency domain can be done using the reconstruction formula [19, p. 420] (example with the spectrum $X(\omega)$ of the sequence $x[n]$ in (2.3)), which provides a properly weighted linear combination of the original spectral samples. This can be performed more easily using a technique called zero-padding [19, p. 420-425], where zeros are added to the end of the time-domain signal before transformation to the Fourier domain. Zero padding does not introduce any new information about the frequency content of the signal, it merely serves to interpolate the spectrum to the desired size. The zero padding will be done in the beamforming algorithm, explained in Chapter 6,

$$X[\omega] = \sum_{k=0}^{N-1} X\left[\frac{2\pi}{N}k\right] P\left[\omega - \frac{2\pi}{N}k\right]$$

$$P\left[\frac{2\pi}{N}k\right] = \begin{cases} 1, & k = 0 \\ 0, & k = 1, 2, \ldots, N-1. \end{cases}$$

(2.3)

2.1.4. Impulse response

The response of a system to a single impulse (Delta pulse, discussed later) is called the impulse response of a system. The impulse response of a smartphone microphone is assumed to be *linear* and time-invariant, so it can be described by a linear and time-invariant system (LTI system) [20, p. 17] [21, 22]. This assumption is made because the source (loudspeaker) and receiver (microphone) are in a fixed place during each measurement and the smartphone is not exposed to significant temperature changes.

A property of an LTI system is that its behaviour can be characterized by its impulse response [19, p. 105]. This means that for a given input signal, the output signal can be calculated using the impulse response of the system. This also implies that the impulse response can be found by a convolution of a known input signal and a measured output signal. The frequency response of a system is the Fourier transform of the impulse response and the other way around. The input signal and corresponding deconvolution technique should give the deconvolved impulse response a maximized signal-to-noise ratio [22]. It is desirable to know the behaviour, and thus to find the impulse response, of the smartphone microphone for a wide range of frequencies. It is therefore necessary to use an input test signal that contains many frequencies equally.

The theoretically most simple signal to use for impulse response measurements in continuous time is the Dirac delta pulse $\delta(t-\tau)$ [20, p. 158-159] (2.4), in digital signal processing the unit sample sequence $\delta[n]$ is most often used [19, p. 42] (2.5). This is a hypothetical distribution representing an infinitely high impulse of infinitesimal duration, giving it a flat power spectrum - it contains all frequencies equally.

Convolution of the Dirac delta with a function yields the function that was convolved with it, and as such an impulse response is directly usable for modelling the response of a system [20, p. 169-171]. Unfortunately, the Dirac delta is only a hypothetical distribution. Real approximations to it suffer from a very large dynamic range (the ratio between the smallest and largest amplitude of the signal) - the signal power is very localised in time, which makes it difficult to put enough power into the pulse for adequate signal-to-noise ratio (S/N) [23]

$$\delta(t-\tau) \quad = \quad \begin{cases} \infty & \text{for } t = \tau \\ 0 & \text{elsewhere,} \end{cases} \qquad \text{with:} \int_{-\infty}^{\infty} \delta(t)\,\mathrm{d}t = 1, \tag{2.4}$$

$$\delta[n] \quad = \quad \begin{cases} 1 & \text{for } n = 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.5}$$

The following three discussed impulse response methods aim to keep the flat power spectrum of the Dirac delta, while trying to improve on its dynamic range behaviour. They are commonly used methods for measuring impulse responses. The three impulse response measurement methods taken into account in this work are maximum length sequences (MLS) [23], time stretched pulses (TSP) [24] and the sine sweep technique [22].

2.2. Impulse response measurement methods

These three techniques have different advantages and disadvantages, as compared by Stan et al. [22] and Thomas [21]. The MLS technique is strongly immune to all kinds of noise and it has a relatively low optimal sound level [22]. The TSP technique on the other hand is much less sensitive to distortion by non-linearities than the MLS technique [22], but is much less suited for noisy environments.

The Sine Sweep technique is concluded to be the best method for impulse response measurements in quiet rooms by Stan et al. [22]. It is insensitive to harmonic distortion and has an excellent

signal-to-noise ratio. Its calibration is a lot less tedious to obtain good results, but is not recommended for measurements in noisy environments. However, Thomas [21] notes that the Sine Sweep technique concentrates excitation energy to a very narrow spectrum. Because speech does not tend to contain single tones, the sine sweep technique is less useful for applications in speech. In addition, the sine sweep technique does not yield phase information [21]. The phase information can be used by the beamforming algorithm, so the sine sweep technique does not seem appropriate for the intended purpose, therefore this technique will not be discussed any further.

2.2.1. Time Stretched Pulse (TSP)

The time-stretched pulse (TSP) technique is an impulse response measurement method which was introduced by Aoshima [24]. It is based on specifying a wideband, spectrally flat signal, then taking the inverse Fourier transform to yield a suitable time-domain signal. If the phase of the signal is taken to be zero, an impulse-like signal results which, as discussed above, is unsuitable for practical applications. The time-stretched pulse can be considered to be the output of a phase-shifting filter applied to an impulse signal, with a transfer function given by

$$H[\omega] = \exp[j(12\omega^2/10000)]. \tag{2.6}$$

The numbers in this equation are chosen such that this filter has a magnitude of 1, so it will conserve the wideband frequency content of the signal and only shift its phase. This added phase shift results in a stretched signal in time. The impulse response can be recovered by using an inverse filter with transfer function (2.7) on the Fourier transform of the received signal

$$H^{-1}[\omega] = \exp[-j(12\omega^2/10000)]. \tag{2.7}$$

The technique of Aoshima is suitable for sound signals with small time duration. For long impulse responses, Aoshima's technique exhibits a discontinuity at frequencies over $f_s/2$ (with f_s the sampling frequency) and below zero, which is caused by aliasing. As the transfer functions of acoustic arrangements often display long impulse responses, Suzuki et al. [25] designed the Optimized Aoshima Time Stretched Pulse technique (OATSP). They first generalized Aoshima's TSP to the function

$$H[k] = \begin{cases} \exp(jpk^2) & 0 \le k < N/2 \\ 1 & k = N/2 \\ H^*[N-k] & N/2 < k < N. \end{cases} \tag{2.8}$$

Where $N = 2^i$, with $i \in \mathbb{N}$, and the variable p determines the stretch of the signal. Discontinuities in phase can arise because $H(N/2)$ is always set to the real value 1. To remove this discontinuity, they introduced an integer m that determines the stretch of the pulse, given by equation (2.9), which is then substituted in equation (2.8). As can be seen from equation (2.10), this integer m prevents the occurrence of $H(N/2) = 1$ and thus removes the discontinuities. The OATSP gives an almost ideal characteristic to measure impulse responses shorter and even longer than its specific length N. The code to generate and analyse the TSP sequences, by Thomas [21], uses this OATSP method:

$$p[N/2]^2 = m\pi \tag{2.9}$$

$$H[k] = \begin{cases} \exp(j4m\pi k^2/N^2) & 0 \le k \le N/2 \\ H^*[N-k] & N/2 < k < N. \end{cases} \tag{2.10}$$

2.2.2. Maximum Length Sequence (MLS)

A maximum length sequence is a pseudorandom two-level signal consisting of values +1 and -1, commonly generated by a linear feedback shift-register (LFSR). The binary value 0 and 1 are mapped to -1 and +1 respectively [23]. The signal is periodic with a period $P = 2^N - 1$ with N the number of bits: $N \in \mathbb{N}$. The frequency behaviour of an MLS signal is flat, except for a small DC offset [21].

Figure 2.2 illustrates an implementation of a LSFR as a block diagram. The LFSR repeatedly applies an exclusive or (XOR) operation on two coefficients. The chosen coefficients can be calculated from a mathematical derived recursive formula. After each XOR operation, the LFSR shifts a bit and performs the operation again. The LFSR initializes with N ones or N random numbers. This method has a good S/N ratio, which makes it suited for measurements in noisy environments [23].

Assume $s[n]$ to be the MLS input signal of the system and $x[n]$ ato bes the output signal of the system, with $h[n]$ the impulse response of the system. When circular convolution (2.11) [19, p. 439] is applied on the output of this system, the impulse response of the system can be obtained. This is because the cross-correlation of $s[n]$ and $x[n]$ (defined as (2.12) [20, p. 228-229]) has the same result as the circular convolution. If the input and output signals are the same ($x[n] = s[n]$), the cross correlation of these signals approaches the Delta-function with a scaling factor c (2.13). Because $(s * h)[n] = x[n]$ and from this follows (2.14): in this way the circular convolution can be used to find the impulse response, differing from the real one by a scaling factor;

$$h[m] = \sum_{n=0}^{N-1} s[n]x[m-n] \qquad m = 0, 1, \ldots, N-1 \tag{2.11}$$

$$\text{crosscor}(s, x)[n] = s[-n] * x[n] \tag{2.12}$$

$$= \sum_{m=-\infty}^{\infty} s[m]x[m-n]$$

$$\text{crosscor}(s, s)[n] \approx c \cdot \delta[n] \tag{2.13}$$

$$\text{crosscor}(s, (s * h))[n] = s[-n] * (s * h)[n] = s[-n] * s * h \tag{2.14}$$

$$\approx c \cdot \delta * h = c \cdot h.$$
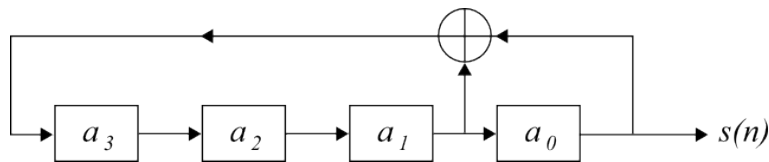
Figure 2.2: A block diagram of a fourth order MLS, with a_i the coefficients. The addition is modulo 2 (equivalent to the exclusive-or operation).

Background - Conclusion

In this chapter first some related research was addressed. This research is a continuation of Gaubitch et al. [2]; they concluded the directivity of a smartphone microphone could improve beamforming algorithms. The inner working of these beamforming algorithms was explained, and the choice for an ad-hoc near-field broadband beamformer was elucidated. After that, different sampling schemes on a sphere were presented with their advantages and disadvantages. Finally three different impulse response determination methods were compared, and it was concluded that the Maximum Length Sequence and Time-Stretched Pulse techniques are more appropriate for the intended purpose than the Sine Sweep technique and therefore those two were explained in more detail.

<div align="right">

Chapter 3

Measurements

</div>

For determining the directivity of the smartphone microphone, measurements must be conducted. This process of determining the measurement set-up is discussed in this chapter. The hardware which was at our disposal will be adressed first, because it influenced some choices made. After that, the application of two different sampling schemes will be discussed, from which one is chosen, and the choice of impulse response technique is discussed. This chapter is concluded with the measurement set-up, some preliminary results and a recap of this chapter. A summary of this chapter and an overview of our final measurement set-up (for replication) can be found in Appendix C.

During the measurements there were some drawbacks, which will be explained further in the last section of this chapter (Section 3.5). A solution will be presented in Chapter 4. These drawbacks were caused by the response of the acoustic system of the measurement set-up, which had more influence on the results than expected.

3.1. Measurement hardware

In order to determine the directivity of the smartphone microphone, the impulse response of the microphone is measured in the anechoic room at the Delft University of Technology. This room gives as little disturbance as possible from the surroundings. Figure 3.1 shows the final measurement set-up, this set-up is already presented here to clarify the terms of the hardware. As mentioned
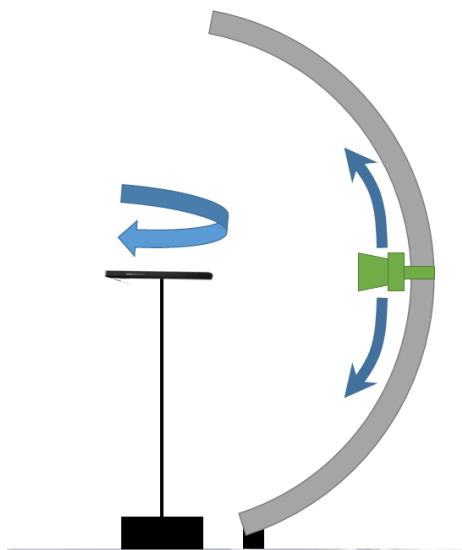
(a) A sketch of the measurement set-up

(b) A picture of the measurement set-up in the anechoic room at the Faculty of TNW at the Delft University of Technology

Figure 3.1: Measurement set-up

earlier, Google's Nexus 5™ mobile phone[8] is used and from now on the term *smartphone* will refer to this specific phone.

An aluminium arc with a radius of 1.21 meter will be positioned in the middle of the anechoic room. This arc has a range from $-75°$ tot $75°$ (as a reference: $0°$ is the horizon). A loudspeaker (with a length of 0.21 meter) is attached to this arc, which can be moved along the arc. The loudspeaker and accompanying hanging system have a length of 0.21 meter, so the radius of the arc with loudspeaker is 1 meter.

The phone is positioned on a turntable[1] in such a way that the distance between the speaker cone and the smartphone microphone is 1 meter for all measurements. This turntable can only be turned in full degrees.

In order to record the sound on the smartphone, an application specifically made for this project was used, which allowed the smartphone to send the recordings via WiFi directly to MATLAB. The implementation of this smartphone application is detailed in [3]. The audio source consisted of a loudspeaker connected to a high-fidelity audio interface, an RME FireFace 800 [27], connected to an audio amplifier[2]. The loudspeaker used, is a Tymphany 4" Midrange loudspeaker, type number M10MD-39-08 [28], which specifications can be found in Figure D.1. This speaker is contained by a custom enclosure, for placement on the arc.

A Brüel & Kjær Free-field microphone [29] was available to use to determine the impulse response of the acoustic equipment without the disturbance by the smartphone microphone. Specifications of this microphone can be found in Figure D.2.

3.2. Sampling scheme

There are many different ways to sample a sphere [1], as also discussed in Section 2.1.2. The turntable can be controlled with MATLAB, the elevation of the loudspeaker on the arc however has to be adjusted by hand. Therefore, sampling schemes with as few manual operations as possible will be considered. Replacing the loudspeaker is very time consuming and therefore undesirable.

To minimize the error margins, a maximum distance between two sampling points of half the minimum wavelength of the desired signal is taken into account. A speed of sound of $v_{air} = 344$ m/s is considered and the maximum frequency in a human voice is around $f_{voice_{max}} = 8$ kHz [30], although for most applications in communication and speech enhancement a maximum of just 3.5 kHz is taken into account [31, p. 58]. Because the objective of this project concerns application in speech, frequencies above 8 kHz are of less value. This gives a maximum distance between two sampling points of 2.15 cm (3.1).

The sphere which will be sampled, is the sphere with centre the microphone of the smartphone and radius the length of the smartphone. Within this sphere the behaviour of the sound is influenced by the smartphone and outside this sphere not, so the choice for the smallest sphere is made. The smartphone has a length of $\ell_{phone} = 13.87$ cm. The length of the equator of the sphere is $\ell_{eq} = 2\pi\ell_{phone} \approx 87.1$ cm.

There are two sampling schemes taken in account, which both feature a low number of movements about the ϕ-direction, which shortens the execution time of the experiment,

$$\Delta\ell_{max} = \frac{\lambda_{min}}{2} = \frac{1}{2}\frac{v_{air}}{f_{voice_{max}}} = \frac{344}{16000} = 2.15 \text{ cm.} \tag{3.1}$$

---

[1]This turntable is a part of the remote controllable Brüel & Kjær 9640 turntable system. [26]

[2]The amplifier that was used is a custom-built 8-channel amplifier built around the Philips TDA8560Q integrated circuit, capable of delivering 25W per channel into a 4Ω load.

3.2.1. Equiangular

An equiangular sampling scheme has equal angles between two neighbouring samples, so the sphere is divided into longitudes and latitudes. In the direction of θ samples will be taken over 360° and in the ϕ-direction over 180°. It is therefore important that the difference in angle between two sampling points is a common divisor of 180 and 360. The distance around the equator is the circle with the largest distance between two sampling points on the sphere. On the equator the maximum distance between two sampling points thus is[1]

$$\Delta \ell_{\max} \geq \ell_{\mathrm{eq}} \cdot \frac{360°}{\Delta \theta}.$$

The largest common divisor smaller than 8.9 is 7.5. This yields 24 longitudes and 48 latitudes, for a total of $23 \cdot 48 + 2 = 1106$ sampling points (when taken into account that on the poles there is only one sampling point).

In practice this cannot be done, because the turntable only turns in full degrees. The first common divisor smaller than 8.9 is 6, this gives 30 longitudes and 60 latitudes, so a total of $29 \cdot 60 + 2 = 1742$ sampling points. It would be more comprehensive to choose angles of 9°, which lowers the largest frequency, but it is more practical because it results in $19 \cdot 40 + 2 = 762$ sampling points.

3.2.2. IGLOO

The other sampling scheme taken into account is the IGLOO sampling scheme [1]. The IGLOO sampling scheme divides the surface of the sphere in 12 faces, from now on called base faces, of approximately the same surface area. The following base faces are used:

North pole

- $\phi \in [0°, 60°], \theta \in [0°, 120°]$
- $\phi \in [0°, 60°], \theta \in [120°, 240°]$
- $\phi \in [0°, 60°], \theta \in [240°, 360°]$

Middle

- $\phi \in [60°, 120°], \theta \in [0°, 60°]$
- $\phi \in [60°, 120°], \theta \in [60°, 120°]$
- $\phi \in [60°, 120°], \theta \in [120°, 180°]$

- $\phi \in [60°, 120°], \theta \in [180°, 240°]$
- $\phi \in [60°, 120°], \theta \in [240°, 300°]$
- $\phi \in [60°, 120°], \theta \in [300°, 360°]$

South pole

- $\phi \in [120°, 180°], \theta \in [0°, 120°]$
- $\phi \in [120°, 180°], \theta \in [120°, 240°]$
- $\phi \in [120°, 180°], \theta \in [240°, 360°]$.

A sphere, partitioned in this way, with 64 sampling points per base face is shown in Figure 2.1.

Assuming the same minimal distance between two sampling points on the equator of $\Delta \ell$, the sphere will be divided with $\Delta \phi = 7.5°$, because $\phi \in [0°, 180°]$ can be divided in three parts of 60°. $\Delta \theta_{\mathrm{middle}} = 7.5°$ in the middle, because this can also be divided in pieces of 60°. This gives a total number of sampling points in the middle part of $s_{\mathrm{middle}} = 7 \cdot 48 = 336$.

For the pole faces the sampling is not equiangular like in the middle part. The following is applicable to both the north as the south pole of the sphere. The $\Delta \phi = 7.5$ is the same as in the middle part. Assume $\Delta \theta_{\mathrm{pole}}$ (3.2) and $\Delta \theta$ have to be a common divisor of 120. $\ell_{60°} = 2 \cdot (\cos(30°) \cdot \ell_{\mathrm{phone}}) \cdot \pi = 74.98$ cm. It then follows that $\Delta \theta_{\mathrm{pole}} \leq 10.32°$, so a $\Delta \theta_{\mathrm{pole}} = 10°$ will be used,

$$\Delta \theta_{\mathrm{pole}} \leq \frac{360° \cdot \Delta \ell_{\max}}{\ell_{60°}}. \tag{3.2}$$

---

[1] For small angles the distance over the sphere is approximately the same as the direct distance, therefore this equation will be used.

The poles are divided in three faces of 120°. The sampling points have to be placed so they still meet the requirement of $\Delta\ell_{max} = 2.15$ cm. Per face this gives a distribution like can be found in Figure 3.2. This ideal case is by coincidence also the sphere given in Figure 2.1 (page 8).

This sampling scheme gives $s_{pole} = 3 \cdot (5 \cdot 12 + 2 \cdot 6 + 3) + 1 = 226$ sampling points at the poles. The full sampling scheme gives a total of 788 sampling points. This theoretical case is not usable, because the turntable cannot turn half degrees. Because the middle of sphere has to be sampled equiangular, this gives $\Delta\theta_{middle} = \Delta\phi = 6°$ (9 is no divisor of 60), from which oversampling follows and this gives much more datapoints than the 762 of the equiangular method.
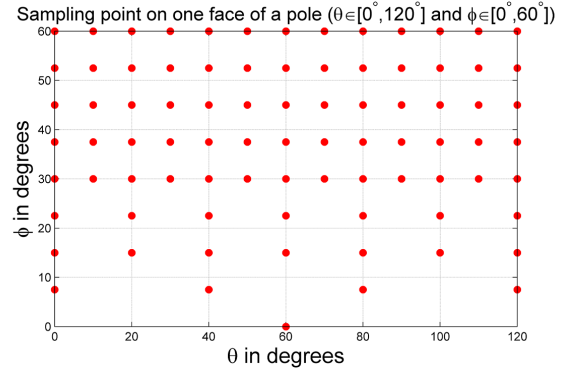
Figure 3.2: Distribution of sampling points on one face of the pole

3.2.3. Sampling scheme of choice

The IGLOO sampling scheme seems appropriate for smaller angles (or larger spheres), because the requirements at $\Delta\theta$ of being a common divisor of 60 and 120 are easy applicable for those cases. The sampling scheme used for the measurements is the equiangular scheme, because it has less sampling points than the IGLOO method and it is accurate for frequencies up to[1]

$$f_{measure} = \frac{360 \cdot v_{air}}{2 \cdot \Delta\theta \cdot \ell_{eq}} = \frac{360 \cdot 344}{2 \cdot 9 \cdot 0.871} \approx 7898 \text{ Hz.}$$

According to the method of Ajdler et al. [32, eq. (5)], the density of this sampling scheme should be accurate enough for frequencies up to 21.9 kHz (3.3), which is higher than the maximum tone a human can hear (about 20 kHz).

$$\omega_t = \frac{cl_\theta}{0.137} = \frac{v_{sound} \cdot \Delta\theta}{0.137} = \frac{334 \cdot 9}{0.137} \approx 21.9 \text{ kHz} \tag{3.3}$$

3.3. The measurement set-up

The decisions concerning the measurement set-up are based on the hardware, theory and techniques mentioned above. The choice of measurement scheme is already explained in the above section.

In case of the impulse response measurement methods, both the TSP and MLS methods are used. These measurement techniques should give the same impulse response result, which makes them suitable to verify the measurements while measuring. If both techniques gave the same result, the measurement can be assumed to be succeeded.

3.3.1. Settings

As much equipment was used for these measurements, values for many parameters had to be chosen. First of all, the sound volume settings of the loudspeaker need to be in the desired range: loud enough for the phone to register, but not so loud it could saturate or force the speaker. The volume settings are verified by looking at the signal amplitude measured by the smartphone. Many smartphones operate an automatic gain control: a controlled signal output despite variation of the input

---

[1]If the assumption from the beginning of section 3.2 is followed.

signal. This automatic gain control could give distortions in the measurement results. Keeping the signal amplitude measured by the smartphone between 10% to 15% of its maximum amplitude scale will prevent this.

For the setting of the volume of the computer, it was important to not let the audio interface clip. The software belonging to the Fireface clearly shows whether there is clipping or not. The clipping could be prevented by lowering the volume of the operating system (Windows in this case) or multiplying the MATLAB-generated signal with a gain factor. The final used settings were a volume of 5% in Windows and a gain factor of -7 dB (power-dB) in MATLAB.

Generating and analysing the TSP and MLS signals is done with MATLAB code by Thomas [21]. In order to use these codes, the order of the TSP and MLS, and the number of repetitions of each signal had to be chosen. At the MATLAB side, we chose for both the MLS and the TSP a sequence of order 15, with ten repetitions. These sequences were played and recorded with a sampling frequency of 48 kHz. This sampling frequency gives a Nyquist frequency of 24 kHz, which is more than enough for our application (remember $f_{\text{voice}_{\max}} = 8\,\text{kHz}$). The combination of the length of the signal and the sampling frequency made it possible to do approximately two measurements per minute, if multiplied by 762, a full sphere takes about 6 hours to measure. When also the time to check and the manual movements of the speaker are taken into account, a full sphere takes about 8 hours (one work day) to measure. So these choices give a achievable time schedule, with a good resolution and frequency range.

Measurements were conducted on different days, so the results of different days may differ due to analogue settings on the amplifier. Therefore some showed preliminary results will not be alike. In the next chapter, our choice of equalizing the different measurements will be explained.

3.3.2. Hardware placement

The measurement set-up is shown in Figure 3.1a. In this set-up the following settings will be tested:

- The smartphone in mid-air
  To measure the smartphone in mid-air, it will be positioned on a small supporting stick, covered with foam. The foam is sound absorbing and smaller than the phone. It supports the phone but does not change the path of the sound. For measurements $\phi \geq 90°$ the phone will be placed upside down on the foam and will be turned the other way around. This way, the foam and the stick will disturb the signal as little as possible and the measurement will be the same as when the speaker is set-up larger than $\phi = 90°$.

- The smartphone on a surface
  For the measurements on a surface there is another set-up available: a plywood plate with dimensions of $290 \times 240 \times 8$ mm placed on a stick. The difference between the impulse response in mid-air and the impulse response on a surface is interesting because it will show whether the directivity of the phone in mid-air is representative of the behaviour on a surface. The overall goal is to use the directivity in a system in which the phones will be positioned on a surface. If the results on a table are too different from those in mid air, the conclusion may be that there is too much difference between mid-air and table to use the mid-air measurements in the beamforming algorithm.

  The smartphone has two configurations on a surface:

  - Face up
    With the back of the phone on the table.
  - Face down
    With the screen of the phone on the table.

The last part of the measurement set-up concerns angles on the arc larger than 75° and smaller than −75°, because the arc does not reach that far. To reach angles in [75, 90] and [−90, −75] the turntable can be placed at an angle, to cover the last 15°. For instance: to measure at an angle of 82°, the speaker will be set at 75° and the table will be tilted 7°. This of course causes the smartphone to slide, so it will be secured with double-sided tape (1 mm thick).

For the smartphone in mid-air at angles smaller than −75° the described set-up with the double-sided tape will be used, but with the smartphone turned upside down. For the set-up with the smartphone on a surface, this cannot be applied, so there will be no measurements done at angles lower than −45°, because an audio signal coming from angles smaller than 0° will probably be attenuated very fast.

For the measurements with the table, the same sampling sphere will be used.

Pictures of the measurement set-up can be found in Figure C.1 in Appendix C.

3.4. Expectations

There were already some expected results before the start of the measurements, mostly based on the measurements done previously by Gaubitch et al. [2] (Figure 3.3) and the physical characteristics of Google's Nexus 5™ mobile phone. A non-symmetric directivity pattern in the θ-axis is expected as the microphone of Google's Nexus 5™ mobile phone is not positioned in the middle of the smartphone but about two centimeters to the right. When turning the smartphone with the turntable, the microphone will thus not be completely turned away at $\theta = 180°$. Furthermore, a turned away microphone should give less gain than a microphone directed towards the loudspeaker. Another expectation is that the measurements on a surface and in mid-air will give different results. The surface is expected to cause reflections of the signal, so there will be multipath interference patterns to observe. For $\phi > 90°$ a high gain loss is expected because the surface will hamper the signal for such angles.

Figure 3.3: Device A (Google's Nexus 5™ mobile phone) from Figure 1 from [2]: *Average responses for three smartphone models in one-third octave bands. The results are presented as the average responses per smartphone model.*

3.5. Preliminary results

The preliminary results of our measurements, plotted on a linear frequency axis, are shown in Figure 3.4a and 3.4b at the end of this chapter. The first things to notice are the very low gains at the lowest and highest frequencies. For the used decibel-scale 0 dB stands for no gain, which in this case means that the sent audiosignal is received with a gain of 1 at the microphone speaker. The given decibel scale comes from the settings of the computer, equipment and the attenuation through the air, which all cause the signal to have very high weakening. These gains are the result of the loudspeaker, which frequency response for frequencies larger than 20 kHz is unknown [28] and therefore considered low. This can be seen in Figure D.1 in Appendix D. For frequencies smaller than approximately 125 Hz, the gain of the speaker is also decreasing very quickly.

To let not cloud our results by these limitations of the loudspeaker these plots are replotted with a limited frequency axis. These are shown in Figure 3.4c and 3.4d.

One very important thing to notice is the presence of the line of lower gain at about 13.5 kHz. This line was not expected, because these measurements were performed with the same smart-
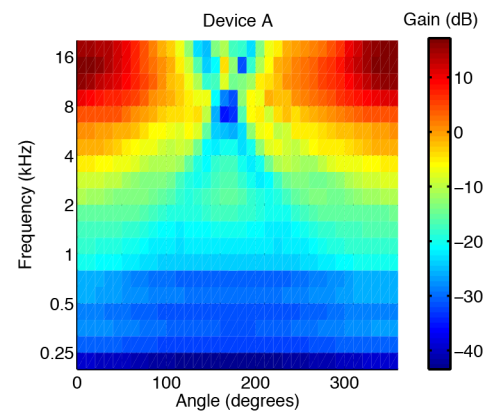
phone as used by [2] (Figure 3.3). Although these results are plotted in octave bands, there is no line of reduced gain visible around 13.5 kHz. In the search for the cause, different smartphones, microphones and another loudspeaker of the same type [28] were tested. It is concluded that this reduced gain was probably a result of the use of the available banana connector cable of which 10 meter was used to transport the signal from the audio amplifier to the loudspeaker. Because this cable is not as good as a normal transmission line (like a coaxial cable), there could be some frequency dependencies in the cable.

To determine the smartphone directivity, the frequency dependencies of the acoustic system need to be eliminated from the result. Therefore the microphone [29] is used: the impulse response of the acoustic system is measured using both MLS and TSP and these measured responses are used for equalization. How this equalization is done will be explained in Section 4.2. The impulse responses given by the TSP and MLS signals are very much alike. From now on all the plotted impulse responses will be those using the TSP-method as they are the same as the MLS results.

Measurements - Conclusion

In this chapter an equiangular sampling scheme (with latitudes and longitudes 9° apart) for determining the smartphone directivity was presented. Measurements using this scheme were conducted using the TSP and MLS impulse response determination techniques. A summary of this chapter can be found in Appendix C, for replication of our measurements. The preliminary results of these measurements were given in Figure 3.4. These results point out some effects, which did not match with the expectations. Therefore a way of equalizing the acquired dataset will be presented in the next chapter.

(a) TSP method

(b) MLS method

(c) TSP method, limited frequency axis
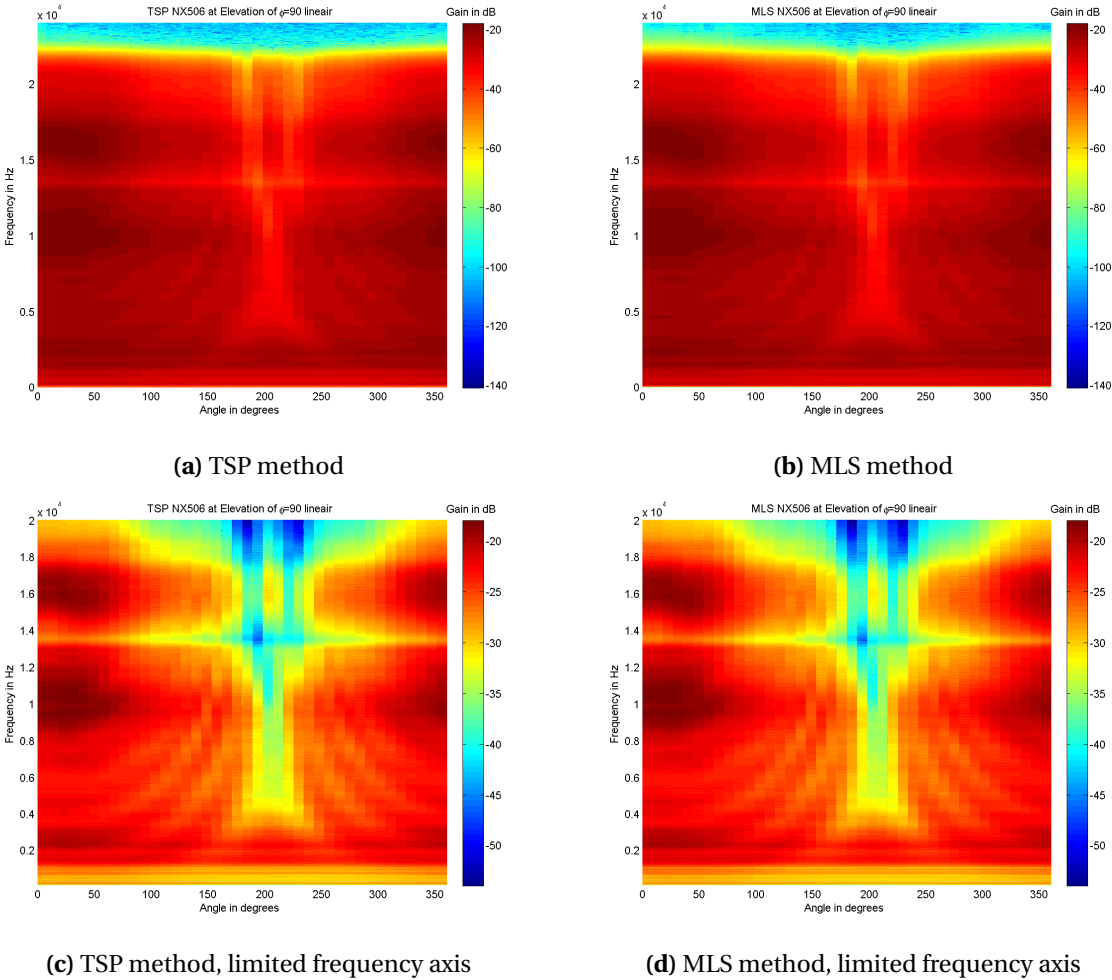

(d) MLS method, limited frequency axis

Figure 3.4: Preliminary measurement results of Google's Nexus 5™ mobile phone, labelled with number 6, at $\phi = 90$ degrees, with full frequency axis (from 0 Hz to 24 kHz) for the given methods.

Equalization

In this chapter two things will be discussed. At first a few adjustments are discussed to treat the recorded signals in such a way that they give the right results and to make sure they are displayed in the same dB-scales so they are comparable to each other. After that some equalizing-theory is discussed to solve the problem of the influence of the acoustic system on the measurements, presented in the last chapter. Which we will conclude with the solution for this problem and again some preliminary results and a short recap.

4.1. Processing the recordings

There are a few differences between different recordings. For a start, the phones do not start with recording at the same time as the signal is played through the loudspeaker. To compensate for these different delays, the impulse responses are altered so the delay is 141 samples (equation (4.1)) for all measurements.

The audio signal from the loudspeaker travels through space before it is recorded by the smartphone microphone. Every measured impulse response thus needs to only consist of a part of the signal travelling through space and a part affected by the microphone. The number of samples it would take the sound from the loudspeaker to the smartphone microphone is computable from the facts that the distance between the loudspeaker and the smartphone microphone is 1 meter, and the speed of sound and the sampling frequency are known too,

$$\frac{\text{distance}}{v_{\text{air}}} \cdot f_s = \frac{1}{344} \cdot 48000 \Rightarrow \left\lceil \frac{1}{344} \cdot 48000 \right\rceil = 141 \text{ samples.} \tag{4.1}$$

Next to this, there needs to be a 'point of reference' to compare the measurements with different settings with each other and to choose a setting for the dB-scale. In the presented results (Section 5.1), the equalized gain of the smartphone microphone (labelled with number 6) at a frequency of $f = 200$ Hz at $\theta = 0°$ and $\phi = 90°$ is chosen as a reference point of 0 dB. Measurements with other settings will be referenced to this reference point via the measurements with the microphone. The preliminary results in this chapter (Section 4.4) are calibrated to this point, the results in the rest of this thesis are.

4.2. Equalizer design

When measuring the impulse response of the microphone, also the impulse response of the acoustic system is taken into account (the acoustic system consists of: the loudspeaker, the audio interface (in this case the Fireface 800), the audio-amplifier and the cables). To determine the impulse response of the microphone in the smartphone as good as possible, the impulse response of the acoustic system should be inverted. Therefore the same measurements as with the smartphone are conducted with a microphone [29] with an almost perfect frequency response (Figure D.2).

The response of a system can be written as (4.2) in the z-domain [19, eq. (5.4.6)]. The zeros of the system are the zeros of $A(z)$ and the poles of the system are the zeros of $B(z)$,

$$H(z) = \frac{A(z)}{B(z)}. \tag{4.2}$$

If a system is minimum phase [19, p. 331-337], meaning it is causal and invertible, the inverse of the system is relatively easy to determine. A minimum phase system is easy to recognise by its pole-zero plot. If all poles and zeros are within the unit circle, the system is minimum phase, so the system can be inverted. In this case (Figure 4.1), all the poles of the system are within the unity circle, like the majority of the zeros, but the handful of exceptions of zeros outside the unity circle makes the system in theory non-invertible.

This can be seen from equation (4.2). For the inverse of the system the following equation determines the inverse of the system $H(z)$: $H^{-1}(z)$

$$H(z)H^{-1}(z) = 1.$$

With $H(z)$ from equation (4.2), H^{-1} must be equal to

$$\frac{B(z)}{A(z)}.$$

Here one sees that the zeros of $H(z)$ become the poles of $H^{-1}(z)$ and the poles of $H(z)$ become the zeros of $H^{-1}(z)$. If there are poles outside the unit circle, the system is unstable, which is the case, because the original system has zeros outside the unit circle (Figure 4.1).

The systems seems non-invertible, this most possibly caused by the response of the loudspeaker, which is not an LTI-system due to non-linearities [21]. It is hard to characterize and equalize signal, went through a loudspeaker, especially using a FIR filter [33]. For improving the beamforming algorithm it is desired to have the best directivity as possible. Determining a linear system is easier than determining a non-linear system, the inverse of the acoustic system is assumed *linear* and time-invariant (LTI) from now on. Some infinite impulse response systems (IIR) have pretty good finite impulse response system (FIR) approximations. To determine whether such an approximation exists, two methods based on the method of *linear* least squares are going to be attempted. This will be discussed in the following sections.

4.2.1. Equalization theory

The sent output signal, made by the computer, is $s[n]$, the impulse response of the system is $g[n]$, the impulse response of the smartphone microphone is $p[n]$, the measured signal measured by the phone is $x_p[n]$ and by the microphone is $x_m[n]$. The desired signal is $x_r[n]$. If we assume:

$$
\begin{align}
x_m[n] &= (s * g)[n] \tag{4.3} \\
x_p[n] &= ((s * g) * p)[n] \tag{4.4} \\
x_r[n] &= (s * p)[n] \tag{4.5}
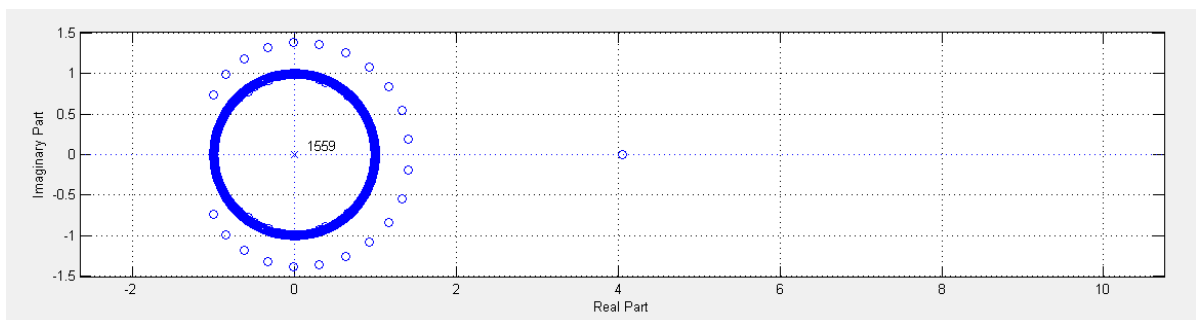\end{align}
$$

Figure 4.1: The pole-zero plot for the 1559 first entries of the impulse response of the system, measured by the microphone [29]. This plot is generated by the `pdatool` in MATLAB.

Because the effect of a filter can be undone by another filter $h[n]$, $\exists h[n]$ such that (4.6). From this follows (4.8), because the convolution is a commutative operation. When this filter $h[n]$ is used on the signal recorded with the smartphone $x_p[n]$, the impulse response of the smartphone microphone can be found by (4.9).

$$
\begin{aligned}
h[n] * x_m[n] &= s[n] & (4.6) \\
(h * (s * g))[n] &= (h * x_m)[n] & (4.7) \\
(h * g)[n] &= \delta(0) & (4.8) \\
(x_p * h)[n] &= (s * p)[n] & (4.9)
\end{aligned}
$$

Two methods to determine the inverse of the system are tested, they both compute (an approximation of) the least squares solution of the inverse, which will be elucidated in the next section.

4.2.2. Method of linear least squares

The method of linear least squares is a typical method of approximating a linear system [34, p. 580-584] [35, p. 483-525]. Because the system is assumed *linear* and time-invariant (LTI), the inverse system is assumed to be linear too. First an example of a least squares approximation will be given, after which the definition of a least squares solution will be discussed and in the following two sections, two solution methods will be presented.

The least squares approximation finds a curve that 'best fits' a set of data points. As example there are three data-points given: $(1,2)$, $(2,2)$ and $(3,4)$. Suppose there is a reason to assume these (x, y)-values are related by a linear function (just like the inverse filter problem), so there exists a line with the equation $y = ax + b$ that fits through these three given values, in other words equation (4.10). Unfortunately, these three points are not in one line, so this system is called inconsistent

$$
\exists a, b \text{ such that } A\mathbf{x} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} = \mathbf{b}. \tag{4.10}
$$

The least squares approximation finds the line that fits as close as possible through these three points. For any line, the vertical distance from each data point to the line will be measured (the error), and then the line will be chosen which minimizes the total error. For now, the errors are denoted as e_1, e_2, e_3 for the given three datapoints, together in the error-vector $\mathbf{e} = [e_1 \ e_2 \ e_3]^\top$. If $\mathbf{e}$ is desired to be as small as possible, $\|\mathbf{e}\|$ must be as close to zero as possible. The Euclidean norm is the best choice of norm to use [34, p. 581]. Here does the name least squares come from: the minimization of

$$
\|\mathbf{e}\| = \sqrt{\langle \mathbf{e}, \mathbf{e} \rangle} \sqrt{e_1^2 + e_2^2 + e_3^2} \text{ or equivalently } \|\mathbf{e}\|^2 = e_1^2 + e_2^2 + e_3^2,
$$

$$
\text{with (in the example)} \quad e_1 = 2 - (a + b \cdot 1), \quad e_2 = 2 - (a + b \cdot 2), \quad e_3 = 4 - (a + b \cdot 3),
$$

with $\|\mathbf{e}\|$ called the least squares error of the approximation. In this example the line $y = 2/3 + x$ is the best fit, with an error of $\|\mathbf{e}\| = \sqrt{2/3}$.

The error vector $\mathbf{e} = \mathbf{b} - A\mathbf{x}$, so for larger systems the following definition [34, p. 583] of the least squares solution of a problem is used: If A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$, the least squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that $\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \ \forall \mathbf{x} \in \mathbb{R}^n$.

There are multiple ways to solve such problems. The first solution uses the pseudo-inverse of a matrix, of which the following property is used: The pseudoinverse $D^\dagger$ of a (nonsquare) matrix D is $D^\dagger = (D^\top D)^{-1} D^\top$, for D of full column rank. The least squares problem $D\mathbf{y} = \mathbf{z}$ has a unique least squares solution $\hat{\mathbf{y}}$ of minimal length given by $\hat{\mathbf{x}} = D^\dagger \mathbf{z}$ [34, p. 594]. The other solution concerns a recursive method to approach the least squares solution.

4.2.3. Toeplitz method

So first the solution using the pseudo-inverse will be addressed. Remember the description of the system given in Section 4.2.1. Working with digital systems, vectors are used. The convolution can be done in many ways, one way is to use Toeplitz matrices.

Assume two signals which are convolved: $\mathbf{a}$ (length n) and $\mathbf{b}$ (length m) (both columnvectors). A and B are their relative Toeplitz-matrices, the example of the Toeplitz-matrix of $\mathbf{a}$ can be found in (4.12). A Toeplitz-matrix can be made as wide as needed. Because the convolution is commutative, (4.11) follows. A Toeplitz matrix can have as many columns as needed, in this case A has m columns so it is multiplicatable with $\mathbf{b}$,

$$A\mathbf{b} = \mathbf{a} * \mathbf{b} = \mathbf{b} * \mathbf{a} = B\mathbf{a}. \tag{4.11}$$

Here, the matrices A and B need to be as wide as the length of the signal they are multiplied with. Hence equation (4.9) can be written as $X_p\mathbf{h} = \mathbf{x}_r$ and therefore $X_p^{\dagger}X_p\mathbf{h} = X_p^{\dagger}\mathbf{x}_r = \hat{\mathbf{h}}$, with $X_p^{\dagger}$ the pseudoinverse of X_p (see end of Section 4.2.2), and $\hat{\mathbf{h}}$ the least squares approximation of $\mathbf{h}$. If the least squares approximation of $\mathbf{h}$ is good enough, this filter $\hat{\mathbf{h}}$ can be applied on the signals recorder by the smartphones to reverse the influence of the system on the recorded signal.

One of the larger disadvantages of this method is the size of the Toeplitz-matrices. A vector of a recording has about 32000 entries, when a relatively large filter is taken into account, this gives large Toeplitz-matrices which also need to be inverted. This is very time consuming. Before implementing this method on the recordings, some experiments have been done with a MATLAB generated signal and filter to conclude whether to try this method for larger signals.

The results are shown in Figure E.1 in Appendix E. The Toeplitz-method for the inversion works fine, if the filter does not cause too much delay. If there is too much delay caused by the filter, the inverse Toeplitz does not compute into a working inverse filter. The suspect of this problem is believed to be the high number of zeros at the start of the Toeplitz matrix which needs to be inverted. For the unknown delay in the recorded signals, the Toeplitz method has not been tested any further and the recursive least squares method is tried.

$$
\text{Toeplitz matrix} \quad A =
\begin{bmatrix}
a_1 & 0 & 0 & \cdots & 0 \\
\vdots & a_1 & 0 & \cdots & 0 \\
a_n & \vdots & \ddots & \ddots & \vdots \\
0 & a_n & \vdots & \ddots & 0 \\
\vdots & 0 & \ddots & \vdots & a_1 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & a_n
\end{bmatrix}
\quad \text{for the vector} \quad \mathbf{a} =
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_{n-1} \\
a_n
\end{bmatrix}
\tag{4.12}
$$

4.2.4. Recursive least squares (RLS)

The RLS algorithm is a recursive implementation of the method of linear least squares, which starts with known initial conditions and uses the information contained in new data samples to update old estimates [35, p. 562-570]. Here a little overview of the algorithm will be given, because implementation is already available in MATLAB it will not be discussed as detailed as implementation level.

Because the algorithm is recursive in time, next to a vector index, there will also be a time index in the vectors $\mathbf{e}$ (which will be written as $e_M(i,n)$)[1] and $\mathbf{h}$ (written as $h_M(i,n)$), the vector with the

---

[1] The subscript M denotes the length of the vector.

filter-coefficients [19, p. 866-877]. The input-vector $\mathbf{X}(n)$, derived $\mathbf{x}_p$ (the measured results) will also depend on time, but a little different than the others:

$$\mathbf{X}_M(n) = \begin{bmatrix} x_p[n] \\ x_p[n-1] \\ \vdots \\ x_p[n-M+1] \end{bmatrix}$$

Here the assumption is made that $x_p[n] = 0 \ \forall n < 0$, this is called prewindowing of the input data [19, p. 867].

The recursive least squares problem is formulated using the cost-function (4.13), with the observed vectors $\mathbf{X}_M(i)$, for $i = 0, 1, \ldots, n$, and the filter-coefficient vector $\mathbf{h}_M(n)$. The cost function minimizes the weighted sum of magnitude-squared errors (4.14) with a forgetting factor $0 < w < 1$ (if $w = 1$ there would be an infinite memory, so that would create an IIR filter) [35, p. 564]. This w weights recent data points more heavily than older ones. The error (4.14) is defined as the difference between the desired sequence $d(i)$ and the estimate $\hat{d}(i, n)$.

$$
\begin{aligned}
\zeta_M &= \sum_{i=1}^{n} w^{n-i} |e_M(i, n)|^2 & (4.13) \\
e_M(i, n) &= d(i) - \hat{d}(i) = d(i) - \mathbf{h}_M^\top[n]\mathbf{X}_M(i) & (4.14)
\end{aligned}
$$

On this equation the matrix inversion lemma is applied [35, p. 565]. The theory of this lemma will not be discussed here for the lack of results of RLS algorithm, discussed at the end of this section. The recursive part of the algorithm is a result of this lemma, which gives the recursive equation (4.15) for the filter after n iterations $\mathbf{h}_m(n)$, where $\mathbf{K}_M(n)$ is computed with signal is the nth iteration [35, p. 569] [19, p. 870].

$$\mathbf{h}_m(n) = \mathbf{h}_M(n-1) + \mathbf{K}_M(n)\mathbf{e}_M(n) \tag{4.15}$$

The RLS method has also been tested using MATLAB. In MATLAB there has been made use of the function `dsp.RLSFilter` [36, 37]. The results of these test are in Figure E.2 in Appendix E. With this approximation of the least squares solution more progress was made, however the inverse filter still did not result in anything useful. This is because of the response of the acoustic system being non-minimum phase, which makes it impossible to invert as an LTI.

4.3. Log-equalizer

As the literature predicted, the attempts to build a FIR filter using the above two methods did not deliver any usable result. To conclude: there is no linear least squares approximation to invert the acoustic transfer function of the acoustic system on the directivity of the microphone, probably because of non-linearities in the loudspeaker [21]. Due to time constrains a suitable (non-linear) equalizer cannot be implemented, instead a Log-equalizer is implemented, which does not compensate for phase shifts due to the acoustic system, but it can compensate for increased or decreased gain. This compensation for different gain is especially prominent in the impulse response plots. The results of the equalized data on the beamforming are discussed later on.

The method of the Log-equalizer is explained by Thomas [21]. And here it will be introduced to the reader with the adaptation that in the presented Log-equalizer the original phase of the signal is kept. A Log-equalizer only corrects for the different gains per frequency, so it is an equalizer in the frequency-domain. This is done by setting a perfect frequency response and determining the difference between the recorded microphone signal and the perfect frequency response. This is displayed the best in the decibel-domain, because a multiplication becomes an addition in the
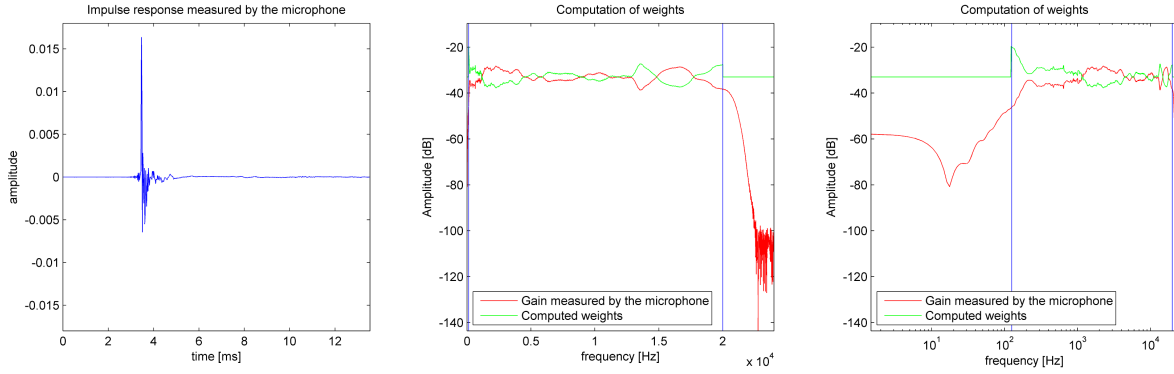
Figure 4.2: Computation of the weights. The left plot contains the first 650 samples of the impulse response measured by the microphone.
The two right plots both contain the same data, plotted linear and logarithmic. The computed weights are added to the mean value to show when the weights are added to the gain there will be a flat frequency response between f_{min} and f_{max} (given by the blue lines).

logarithmic domain [1], therefore this is called a Log-equalizer. An overview of the equalizer is given in Figure 4.3.

The measured impulse responses by the microphone will be used as a reference. As stated in section 3.5 the impulse response of the frequencies smaller than $f_{\text{min}} = 125$ Hz and larger then $f_{\text{max}} = 20$ kHz are not representable for the system. This will be the boundaries of the frequencies to equalize. When frequencies outside these boundaries are equalized, they are not representable for the system any more, because they were not representable for the system in the first place. The gain measured by the microphone will be given as $\Gamma_{\text{mic}}[\omega]$. This vector had n entries in total and n_{b} entries between f_{min} and f_{max}.

To determine the gain of desired impulse response (4.16), the mean value of the measured impulse response within the boundaries will be used. This will cause the smallest absolute largest weight to be added. For the values outside the boundaries only this mean value will be used as weight, for values within the boundaries, the difference between the value and the mean value will be used as weight (4.17). This is summarized in the given equations and graphically shown in Figure 4.2

$$\Gamma_{\text{desired}} \quad = \quad \frac{1}{n_{\text{b}}} \sum_{f_{\text{min}}}^{f_{\text{max}}} \Gamma_{\text{mic}}[\omega], \tag{4.16}$$

$$w[i] \quad = \quad \begin{cases} \Gamma_{\text{desired}} & i < f_{\text{min}} \\ \Gamma_{\text{desired}} & i > f_{\text{min}} \\ \Gamma_{\text{desired}} - \Gamma_{\text{mic}}[i] & \text{elsewhere.} \end{cases} \tag{4.17}$$

Because these weights are for given frequencies, they need to be added per frequency. Assume the signal $x_{\text{phone}}[n]$, first it will be transformed to the Fourier domain to X_{phone}. In the signal X_{phone} is information about the gain and the delay for a given frequency. When the gain of this signal is computed, the weights per frequency are added and the gain plots can be made.

For further use in the beamformer however, the phase information is rather important. If the gain of the signal is changed to decibel ($a_{\text{dB}} = 20\log_{10}|a|$, with $a \in \mathbb{C}$) and transformed back $a_{\text{mag}} = \left(10^{\frac{a_{\text{db}}}{20}}\right)$, with $a_{\text{mag}} \in \mathbb{R}$) the result of the decibel to magnitude operation is a real (and positive) number, but the original signal was complex. So, if we only apply the weights, as for the plots, the original

---

[1] $\ln(x \cdot y) = \ln(x) + \ln(y)$

phase information will be lost. To keep the original phase information, the ratio between the magnitudes of the equalized and the original data will be computed and this ratio, α ($\in \mathbb{R}\setminus[-\infty, 0\rangle$), will be multiplied with the original signal, to retain phase information.[1] An example with $b \in \mathbb{C}$, with j the imaginary unit can be found in equation (4.18)

$$
\begin{aligned}
b &= |b|e^{j\,\arg(b)} \\
&= |b|e^{j\phi}, \\
\alpha \cdot b &= \alpha \cdot |b|e^{j\phi} \\
&= |\alpha b|e^{j\phi}.
\end{aligned}
\tag{4.18}
$$

The weights contain some 'noise'-like pattern on the overall line. These are artefacts from the acoustic system and the microphone, which are not representable for the system. To not let this influence the Log-equalizer, a low-pass filter is applied on the weights: $w * \frac{1}{7}[1\ 1\ 1\ 1\ 1\ 1\ 1] = w_{\text{smooth}}$. The smoothened weights are thereafter used. The removed noise pattern is shown with the blue line in Figure 4.4a on page 28, where one also sees the flattening working of the inverse filter on the gain measured by the microphone.

4.4. Results after equalization

Applying the Log-equalizer to the phone recordings gives us the result in Figure 4.4b on page 28. Applied on all phone recordings, the difference between the result after Chapter 3 and the equalized result is given in Figure 4.5, on page 29 (the settings stated in Section 4.1 are also applied on the equalized figures). In Figure 4.5d the equalized data is displayed on a logarithmic frequency-axis so it is better comparable to the results of Gaubitch et al. [2] in Figure 3.3 (page 17).
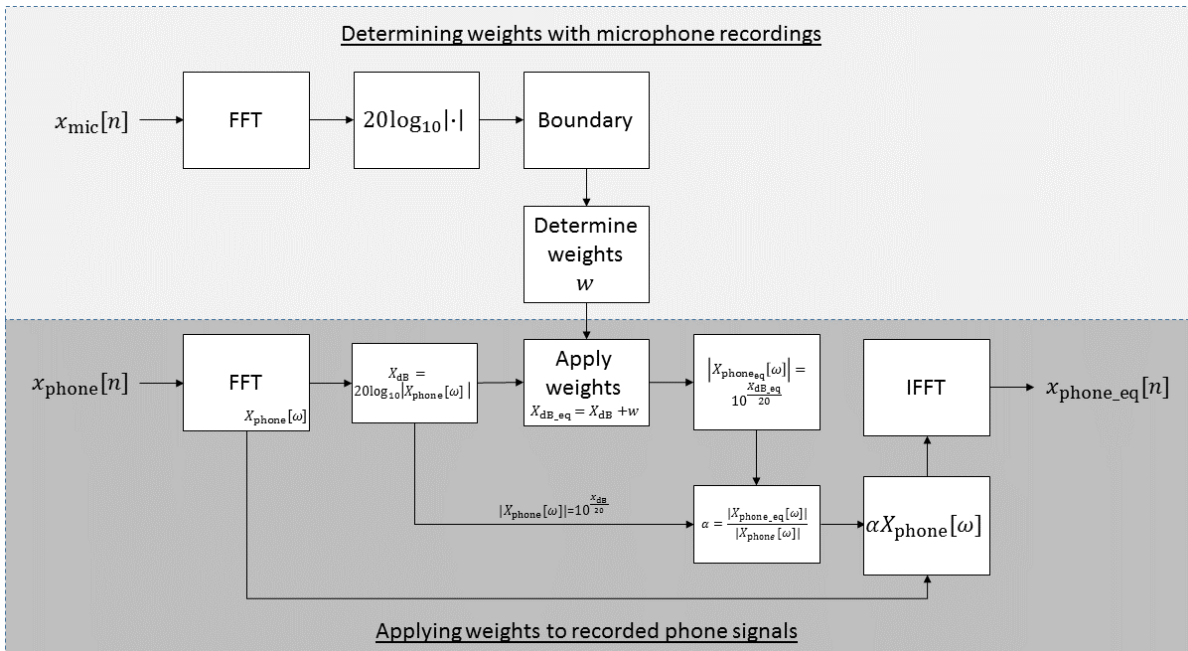

Figure 4.3: Overview of the Log-equalizer, with preserved phase (based on [21])

---

[1]This can also be done by not taking the log of the signal and just use the ratio between the length of the signal before after applying the weights. The weights can be determined as ratios too.

The equalized gain is much better comparable to the results of Gaubitch et al. and therefore the Log-equalizer is used on all the smartphone recordings. For different days of measurement, different microphone recording are available to determine the weights for the Log-equalizer.

Equalization - Conclusion

After eliminating measurement errors caused by the difference in start time of playing and recording the sound signal, a reference point of no gain (0 dB) has been chosen (the equalized datapoint at $f = 200$ Hz, $\phi = 90°$, $\theta = 0°$). The next part of this chapter consisted of the design of the equalizer, the solutions to solve the inverse filter if it is assumed *linear* and time-invariant. The assumption that the inverse filter is LTI is rejected and a Log-equalizer design is presented to equalize the data-sets from the measurements.

Figure 4.4: The result of adding the determined weights on two signals: the microphone and a smartphone measurement.

(a) Equalizing the mean of the gain of 40 microphone measurements.
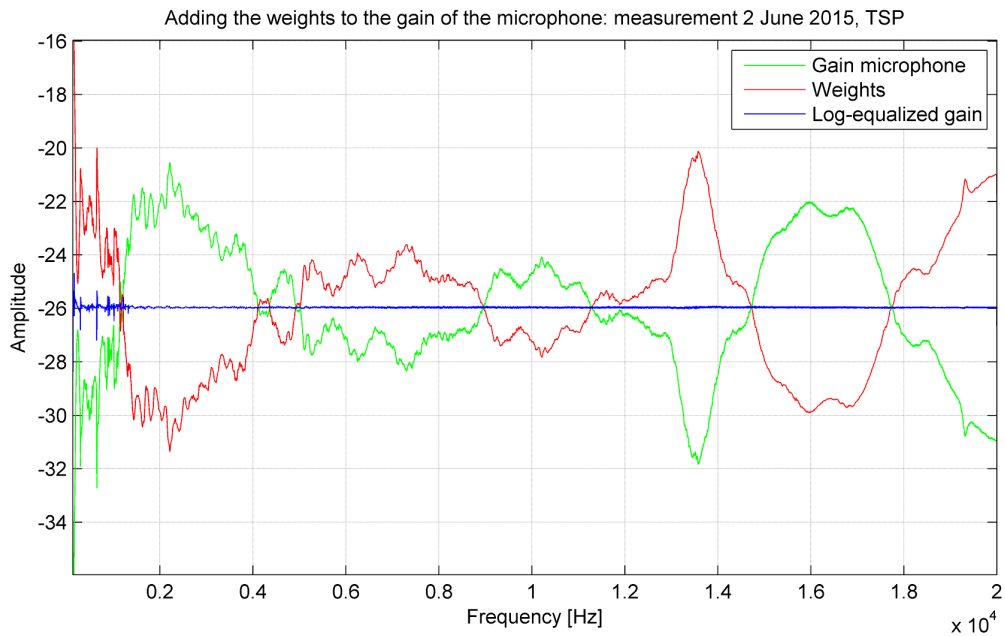
(b) Equalizing the gain of Google's Nexus 5™ mobile phone, labelled with number 6, at $\phi = 90°$ and $\theta = 0°$
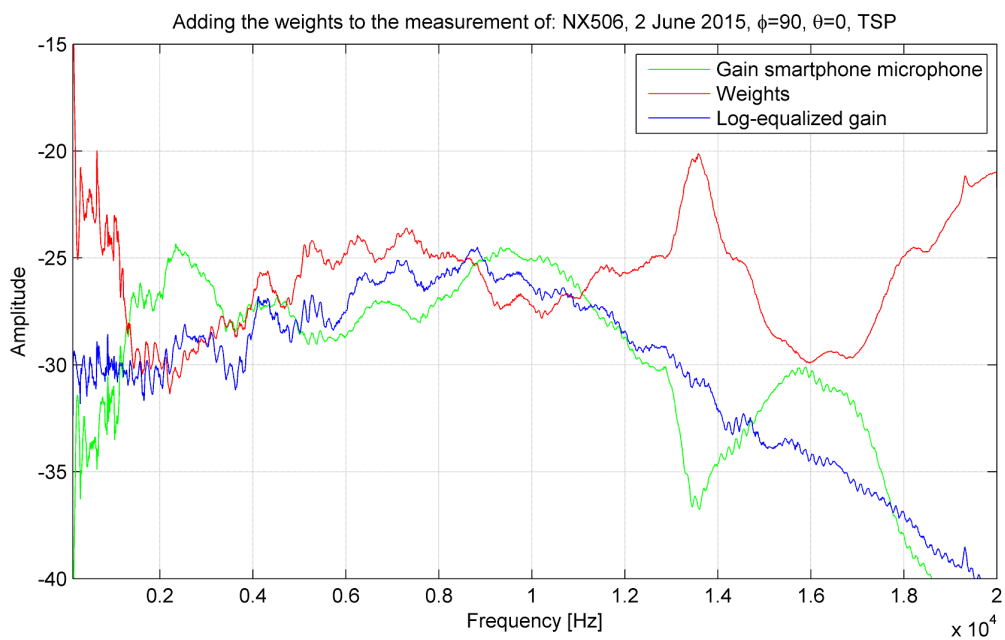
Figure 4.5: Preliminary measurement results of Google's Nexus 5™ mobile phone, labelled with number 6, at ϕ = 90 degrees, with limited frequency axis (from 125 Hz to 20 kHz) for the TSP measurement, non-equalized versus equalized.

(a) Non-equalized data on a linear frequency-scale
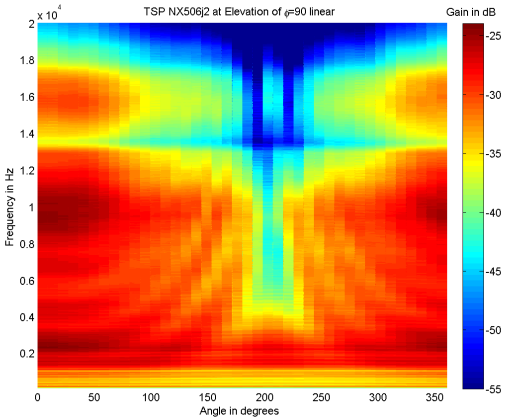
(b) Equalized data on a linear frequency-scale
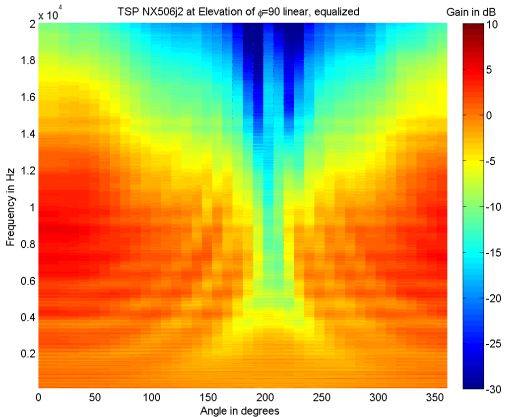
(c) Non-equalized data on a logarithmic frequency-scale
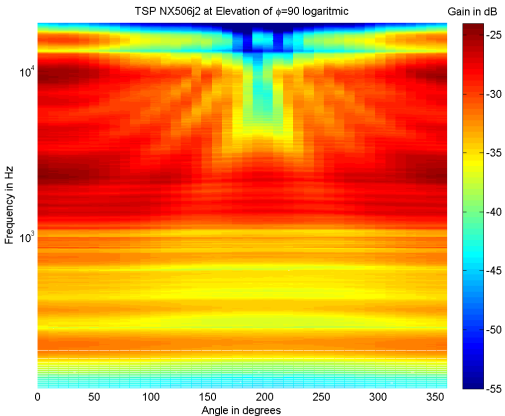
(d) Equalized data on a logarithmic frequency-scale
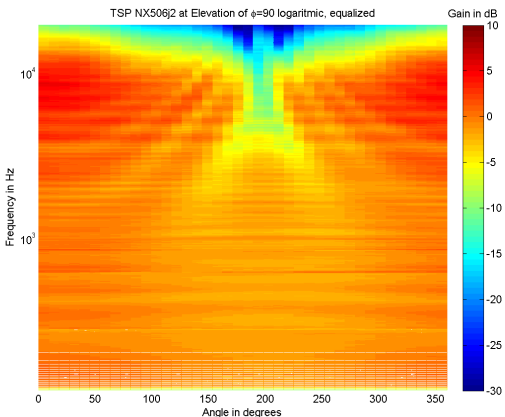
Chapter 5

Experimental results

In this chapter, the final results will be presented first (a few results can be found on page 33 to 36). Thereafter the results are discussed: are there notable results, which do or do not match with the expectations (Section 3.4). A little conclusion will be provided at the end of this chapter.

5.1. Results

After measuring, processing and equalizing, some of the results are presented at the end of this chapter (from page 33 to 36). For there are conducted a lot of measurements, only a handful of result is displayed in this thesis for comparison. For more results, the writers can be contacted.

The results that will be discussed are from four different measurement set-ups:

- Google's Nexus 5™ mobile phone, labelled with number six in mid-air, for a full sphere (Figure 5.1);

- Google's Nexus 5™ mobile phone, labelled with number six face up on the surface, for a sphere with $\phi \in [0, 135]°$ (Figure 5.2);

- Google's Nexus 5™ mobile phone, labelled with number six face down on the surface, for a sphere with $\phi \in [0, 135]°$ (Figure 5.3); and,

- Google's Nexus 5™ mobile phone, labelled with number one in mid-air, due to time constrains only measured for a sphere with $\phi \in [0, 90]°$ (Figure 5.4).

There are a few things to notice from these given results. When the smartphone microphone is turned away from the loudspeaker, the gain is less then when the smartphone microphone is turned toward the loudspeaker (Figure 5.1a). For elevations higher or smaller than 90°, this effect is less (Figure 5.1b, 5.1c).

The addition of the surfaces changes the reception of the signals, especially the signals from below the surface (Figure 5.2c, 5.3c).

Between the two phones there is a visible difference between the behaviour of the directivity (Figure 5.4), the gain of the the phone labelled with number one seems higher than the gain of the phone labelled with number six. In the next section, these results will be placed in different perspectives.

5.2. Discussion

In this chapter, the data will be looked at and compared. This will be mostly done by the Figures from page 33 to page 36. One of the first things that stands out is that the directivity is non-symmetric in the θ-axis (Figure 5.1). This can be explained by the lack of symmetry in Google's Nexus 5™ mobile phone. Its microphone is not positioned in the middle of the smartphone, but 2 cm to the right, which causes the non-symmetrical behaviour.

The microphone also seems to lack quality, comparing the equalized response of the smartphone's microphone in Figure 4.4b, to the response of the other microphone measured with (Figure D.2), there is a lot of difference. Of course the microphones are of different quality (and costs), but the response of the smartphone microphone nowhere seems to be flat and has large variations of response, even in the speechdomain ($f \in [125, 8000]$ Hz). When recording signals containing frequencies larger than 1 kHz, the response drops down pretty fast, which does not make this smartphone's microphone suitable for for instance recording music.

This response thus gives an idea of the quality of the microphone and the costs and production process of this part of the smartphone. This raises questions about the comparability of the directivity of two different Google's Nexus 5™ mobile phone smartphones, which will be answered in Section 5.2.2. Before this, the results of the addition of the surface will be discussed and at last the improvement of the beamformer will be addressed.

In the ϕ-direction, the directivity seems to be symmetric around the equator of the sphere (Figure 5.1e). This means it measures signals from the direction from the screen as good as signals comming from the back. In one way this would seem logical: a phone with which will be filmed, needs to get signals from the backside of the phone. When calling however, it is less desirable, since it gives more noise to the speech.

There are two more things to notice. The first one is the different sphere grid between the Nexus in mid-air (Figure 5.1e) and the other sphere grids. This was a mistake which happened during the measurements: instead of moving the loudspeaker with 9°, the loudspeaker was moved 8°, which resulted in a slightly different samplingscheme, with $\phi \in \{0, 9, 18, 26, 34, 42, 50, 58, 66, 74, 82, 90, 98, 106, 114, 122, 130, 138, 146, 154, 162, 171, 180\}$. The second one is the border in Figure 5.1e and 5.1f. In the first measurements, conducted in May, there were some changed settings which were unnoticed at that time. This were the setting for the middle three ϕ-measurements ($\phi \in \{82, 90, 98\}$). So these measurements have been done again in June, with different settings and a different equalizer. Also the measurement of $\phi = 74°$ has been conducted again, to equalize this with the old measurement at $\phi = 74°$. Although a lot is done to fit these measurement sin, the difference is pretty visible in the result. This will also be addressed in section 5.2.2.

In some spherical plots there seems to be an outlier: a point which has much higher or lower gain than its neighbouring points. This is interpreted as a false measurement. These false measurements could be the result of different reasons: most likely something has gone wrong with the impulse response measurement. It could be that the first of last TSP pulse has not been recorded, which influences the result of the MATLAB analysis. All measurements have been checked by hand, but some mistakes could have slipped through, there are a lot of measurements.

5.2.1. Adding the surface

The addition of the surface[1] results in reflection patterns and high gain loss for high ϕ, which is consistent with our expectations. Signals from above the surface however, are more amplified because of these reflection, resulting in higher gain for smaller ϕ (Figure 5.2e) in comparison with the mid-air case (Figure 5.1e).

---

[1]The face up and down measurements have been conducted with the same settings as the mid-air case in Figure 5.1.

The face up and face down measurements also show differences in behaviour, although this seems to be for frequencies larger than 10 kHz, outside the speechdomain.

The results of the addition of the surface is not directly comparable to the mid-air measurements, but for frequencies below 10 kHz it merely seems to be a difference in gain, which is pretty easy to compensate for. Lets assume there will be some difference in phase shift, due to the reflections of the surface. For the real phase shifts are unknown, they are not compared in this thesis. For signals coming from below the surface this is less applicable, but in the usual situation (like a conference room) not so much sounds will originate from below the surface.

5.2.2. Comparing two phones

For the other phone, labelled with number one (Figure 5.4), there are less measurements available. These also have been conducted on an other day, with other settings, so with other equalization. Therefore the difference between the phone labelled with number one and number six (Figure 5.1), which mostly seems to be a difference in gain and not in behaviour, cannot be compared that easy. Especially because of the three ϕ-measurements of the phone labelled with number six in June, which do not really seem to fuse in with the data of the measurements with the same phone in May. This could be the result of a lot of different thing: next to different settings, different cables have been used and therefore there is no way to be sure the two measurements behave the same. Next to that the phone could have fallen or something, which could influence its behaviour too. There can be concluded that the pattern for the upper half sphere behaves the same, but for the comparison between different phones of the same model it is best to see [2].

Experimental outcome - Conclusion

In this chapter the outcome of the measurements is presented (page 33 to 36) and discussed. In the discussion different questions are addressed: what are noticeable similarities or differences with the expectations, how does the addition of the surface influence the directivity and what is the difference of the behaviour of two phones of the same type? For the answers of the main questions of this thesis, the reader is referred to Chapter 7.

Figure 5.1: Google's Nexus 5™ mobile phone, labelled with number 6, measurements in mid-air, equalized

(a) $\phi = 90°$

(b) $\phi = 42°$

(c) $\phi = 138°$

(d) North pole: $\phi = 0°$

(e) Full sphere $f = 10000$ Hz, from the left

(f) Full sphere $f = 10000$ Hz, from the right
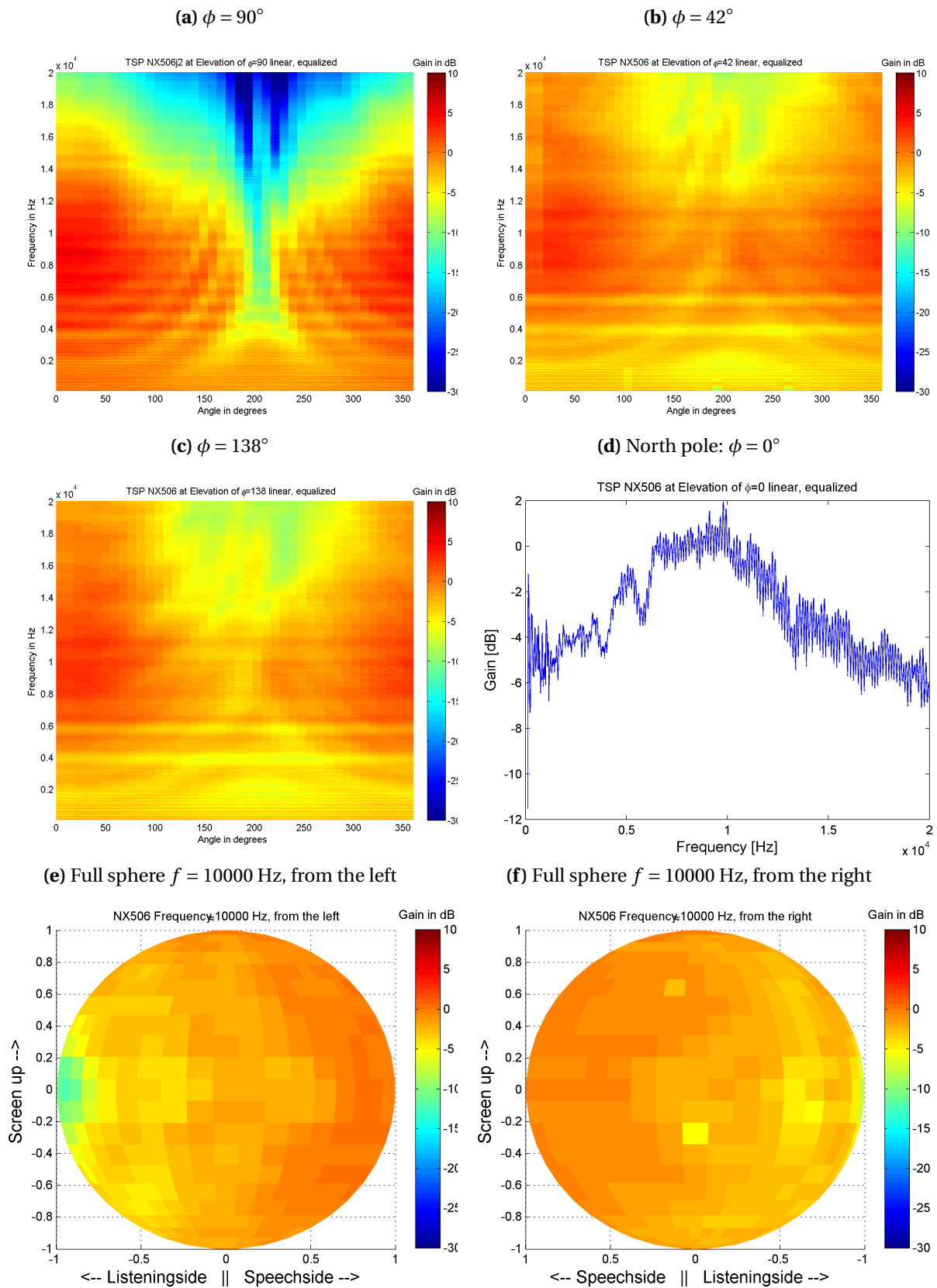
Figure 5.2: Google's Nexus 5™ mobile phone, labelled with number 6, measurements in face up position, equalized

(a) $\phi = 90°$

(b) $\phi = 45°$

(c) $\phi = 135°$

(d) North pole: $\phi = 0°$

(e) Full sphere $f = 10000$ Hz, from the left

(f) Full sphere $f = 10000$ Hz, from the right
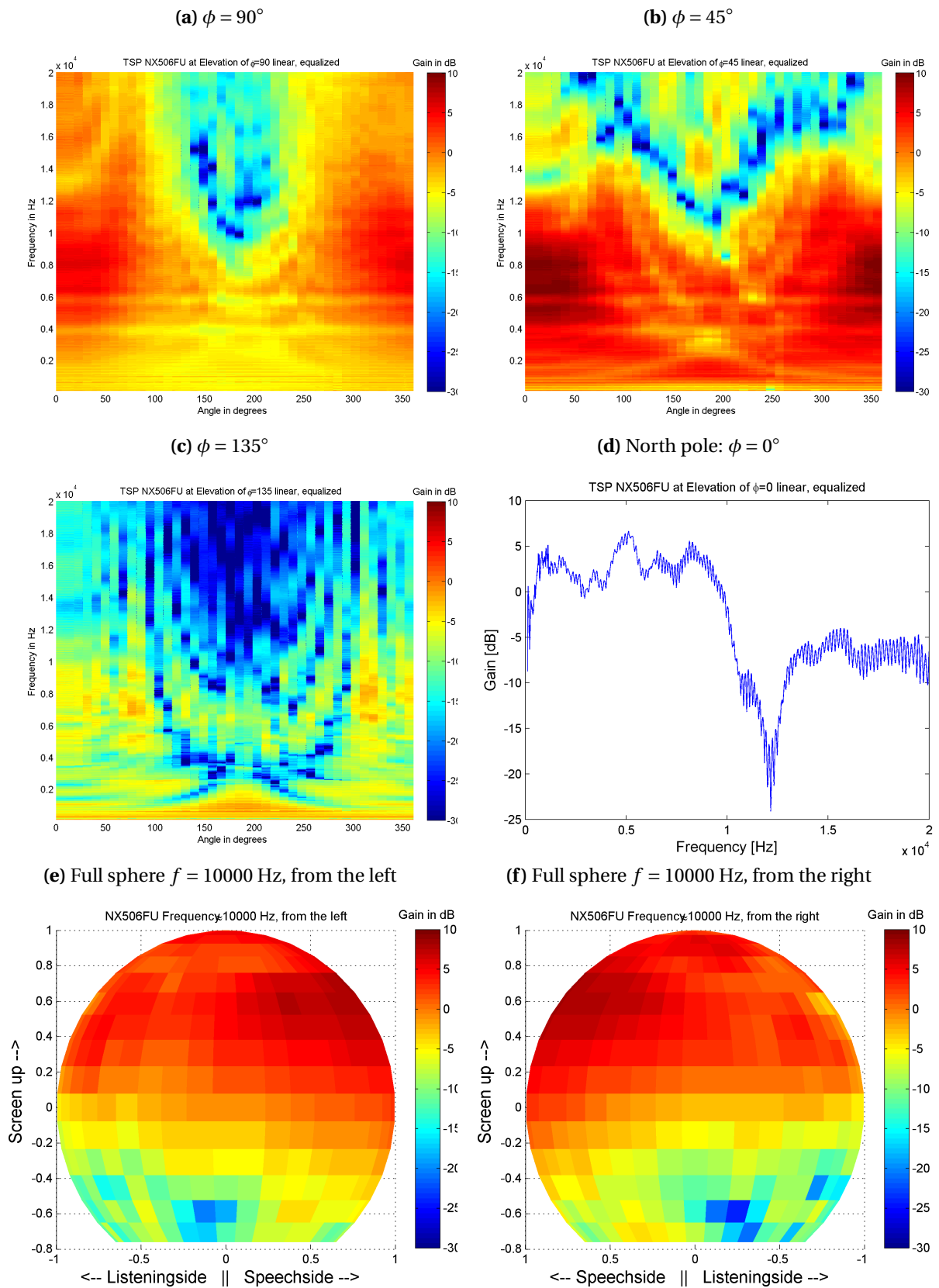
Figure 5.3: Google's Nexus 5™ mobile phone, labelled with number 6, measurements in face down position, equalized

(a) $\phi = 90°$

(b) $\phi = 45°$

(c) $\phi = 135°$

(d) North pole: $\phi = 0°$

(e) Full sphere $f = 10000$ Hz, from the left

(f) Full sphere $f = 10000$ Hz, from the right
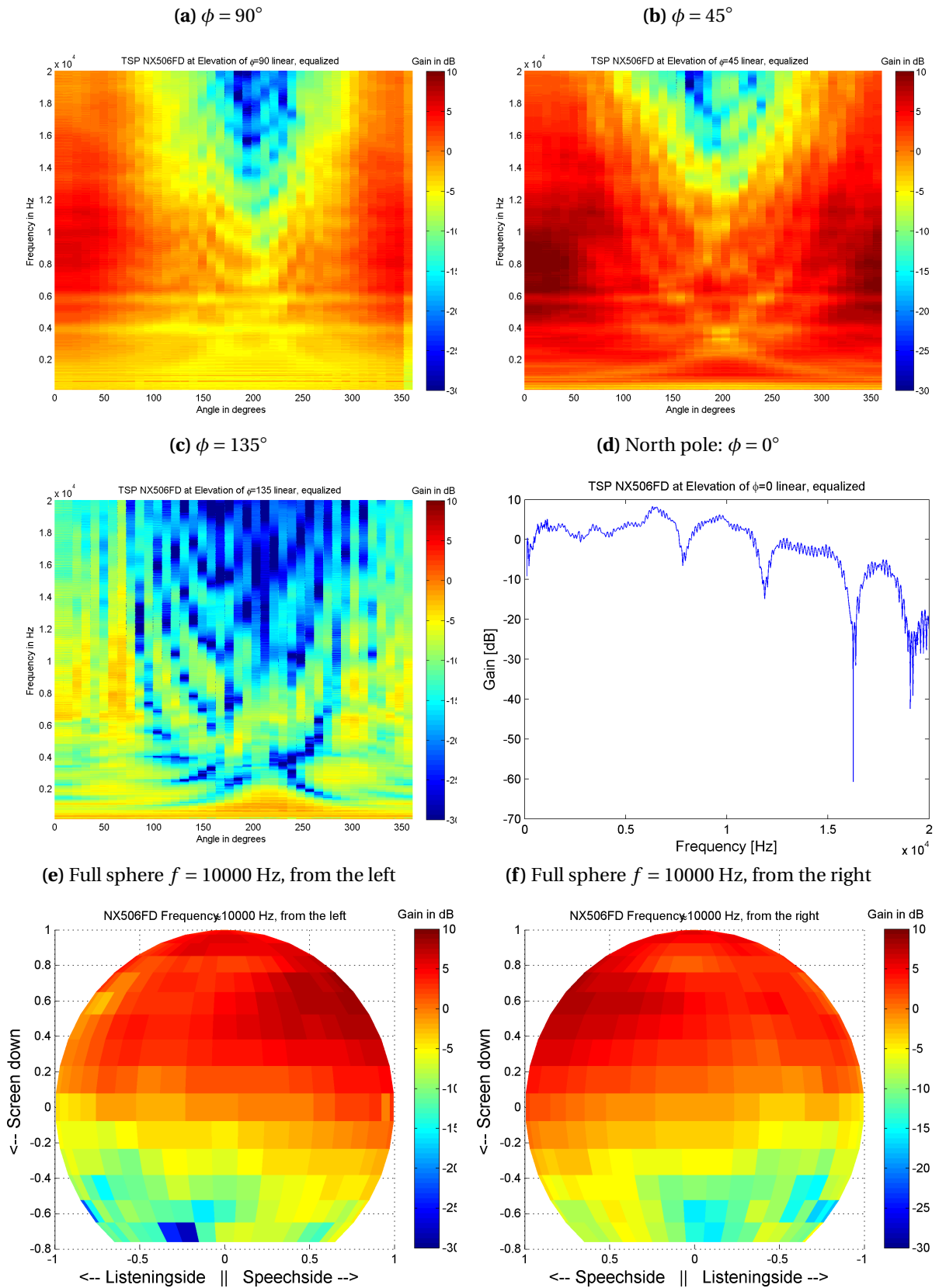
Figure 5.4: Google's Nexus 5™ mobile phone, labelled with number 1, measurements in mid-air, equalized

(a) $\phi = 90°$

(b) $\phi = 45°$

(c) North pole: $\phi = 0°$

(d) Upper half sphere $f = 10000$ Hz, from the left

(e) Upper half sphere $f = 10000$ Hz, from the right
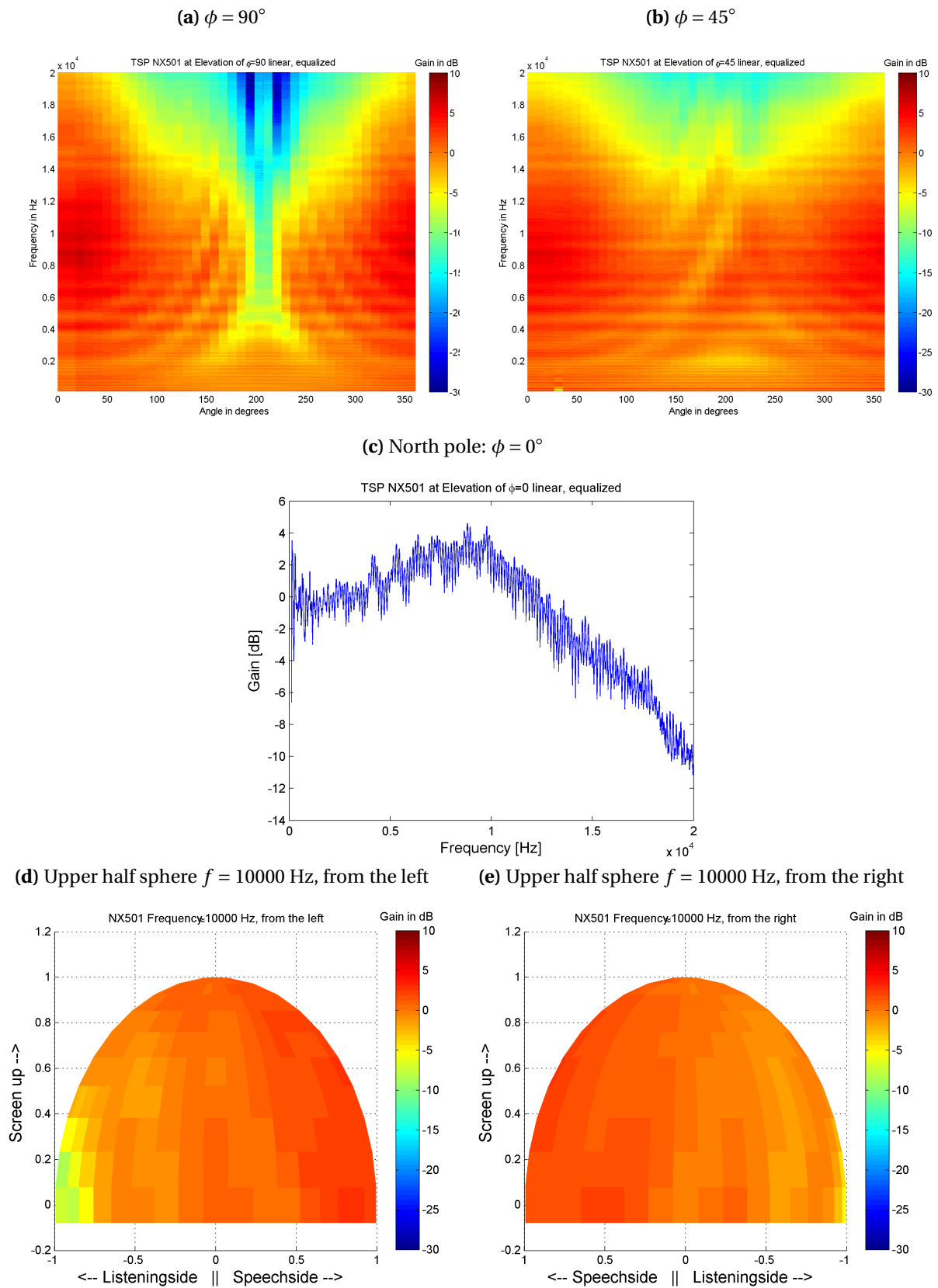
Implementation in the beamforming algorithm

In this chapter the function which bridges between the experimental outcome and the beamforming algorithm will be described shortly. At first the choice of returning impulse responses with a length of 1800 datapoints is discussed, where after the determination of the angles θ and ϕ between two places in space is explained. With this functionality experiments have been conducted [6], which results will be addressed shortly in the conclusion (Chapter 7).

6.1. Functionality

For the implementation of the directivity in the beamformer algorithm, the algorithm will call a function `directivity`, which will give an impulse response as result for given spacial locations. This function is written in MATLAB, to easily implement it in the beamforming algorithm, also written in MATLAB [6]. This impulse response will be transformed to the frequency domain by the algorithm, from which it will receive a gain for different frequencies. This makes it easier for the algorithm to zero-pad or cut-off and so find the complex gains for the right frequencies.

The conducted measurements resulted in an impulse response vector with a length of approximately 32000 datapoints. For the function to be quick, it needs as little memory communication as possible, next to that, the beamforming algorithm uses much less different frequency points than the 16000 which can be computed with the result (about 2000 [6]). Next to that, it is easy for the algorithm to use more different frequencies using zero-padding.

The powerspectrum of impulse responses for two different directions is given in Figure 6.1. This impulse response is normalized to the maximum absolute value of the largest peak in all impulse responses for all directions for the Google's Nexus 5™ mobile phone labelled with number 6. For the impulse response is a signal on which a convolution is performed, it can be read as all different Delta-peaks: they compute a delay and an attenuation of the original signal.

Very small Delta-peaks do not contribute to the impulse response, they are considered as noise: an attenuation of approximately -60 dB is inaudible by the human ear, when a normal sound level of a conversation of 60 dB is assumed [30], therefore all peaks smaller than -60 dB can be ignored. In the figure can be seen that after approximately 1500 datapoints (or 0.03 seconds) both signals are below the -60 dB line, datapoint from this point are insignificant. For some certainty 1800 datapoints are outputted to the beamforming system.

6.2. Determining angles

To meet the requirements given in Appendix A, the function should give an impulse response for a given θ and ϕ. For convenience and to avoid mistakes in different spherical coordinate-systems, another function is written, which has as input two places in space (the place of the smartphone ($\mathbf{p}_p$) and the place of the source ($\mathbf{p}_s$)) and which solves the problem of determining θ and ϕ.

Using the inner product [34, p. 21] (6.1), the angle between two vectors can be determined. Taking the phone as the origin of the Cartesian coordinate system, the angle between the x-axis and the (x, y)−vector pointing from the phone towards the source results in θ (6.3). And in the same

way, the angle between the z-axis and the (x,z)-vector pointing from the phone towards the source results in ϕ (6.4).

After the determination of the angles, making use of the `load`-function in in MATLAB to only read one vector in the memory, the impulse response for the given places in space is returned.

$$\langle \mathbf{a}, \mathbf{b} \rangle \quad = \quad \|\mathbf{a}\| \|\mathbf{b}\| \cos\theta \qquad \text{with } \theta \text{ the angle between } \mathbf{a} \text{ and } \mathbf{b} \tag{6.1}$$

$$\mathbf{p} \quad = \quad \begin{bmatrix} p^x \\ p^y \\ p^z \end{bmatrix} = \mathbf{p}_s - \mathbf{p}_p \tag{6.2}$$

$$\theta \quad = \quad \arccos\left(\frac{\left\langle \begin{bmatrix} p^x \\ p^y \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle}{\left\| \begin{bmatrix} p^x \\ p^y \end{bmatrix} \right\|} \right) \tag{6.3}$$

$$\phi \quad = \quad \arccos\left(\frac{\left\langle \begin{bmatrix} p^x \\ p^z \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\rangle}{\left\| \begin{bmatrix} p^x \\ p^z \end{bmatrix} \right\|} \right) \tag{6.4}$$
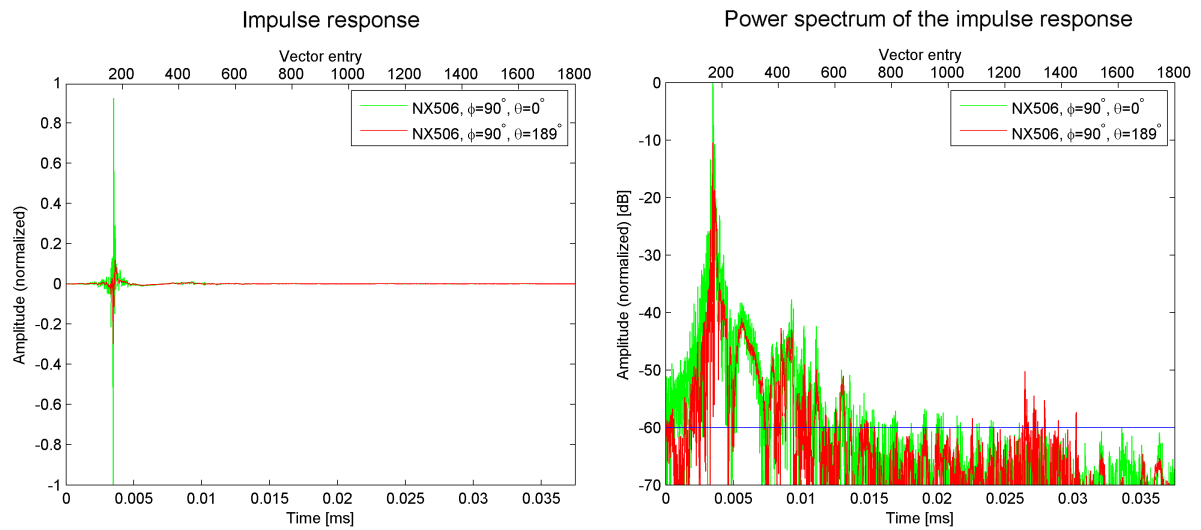

Figure 6.1: The impulse response of two different measurements in the dB-domain. The blue line denotes the -60 dB line from the highest point.

Conclusion and recommendations

The research of this thesis continues the work of Gaubitch et al. [2]. First some background (Chapter 2) was addressed. The working of a beamforming algorithm, with the choice of an ad-hoc near-field broadband beamformer was explained. After that different sampling schemes on a sphere were presented with their advantages and disadvantages. And the last part of background consisted of a comparison of three different impulse response determination methods, with the conclusion that the maximum length sequence (MLS) and time-stretched pulse (TSP) techniques are more appropriate for determining the directivity of a smartphone microphone than the sine sweep technique and therefore those two were explained in more detail.

This chapter was followed by the presentation of an equiangular sampling scheme (with latitudes and longitudes 9° apart) for determining the smartphone directivity. Measurements using this scheme were conducted using the TSP and MLS impulse response determination techniques (an overview of the measurement set-up can be found in Appendix C). Preliminary results, presented at the end of Chapter 3, showed unexpected frequency behaviour.

Different attempts and methods to equalize this unexpected frequency behaviour were described in Chapter 4. The approach of determining an inverse filter was based on the assumption the inverse filter to have a *linear* and time-invariant system approximation. Although the impulse response of the system was non-minimum phase, the Toeplitz method and recursive least squares method were tested, with poor results: which let to the rejection of the assumption that the inverse of the acoustic system has a linear approximation. Due to time constrains, equalizing the response of the acoustic system (a non-minimum phase system) was not possible, so a rather simple Log-equalizer was presented and applied to the data set.

In Chapter 5 the equalized results are presented and discussed, followed by Chapter 6, where in short the implementation in the beamforming algorithm is explained. In this last chapter, we will provide our conclusions based on the results and from this, we will present some recommendations for future work, for there is much room for improvement.

7.1. Conclusion

The directivity for Google's Nexus 5™ mobile phone has been determined, but only for one smartphone of its type. For another smartphone of the same type only measurements were done on the upper half of a sphere, with other settings, so they are not comparable. Of course their patterns can be compared, but not for the full sphere. Therefore there is not much to state about the differences of directivity between smartphones of the same type: it seems they follow the same patterns, which has also been concluded by Gaubitch et al. [2].

The addition of the surface does change the directivity of the smartphone as expected, but for audio signal coming from above the surface, the frequencies below 1 kHz are better received than the same signals for the smartphone with no surface. This is an interesting result, for the reception of signals in the speech domain from above the surface is not attenuated but enhanced by the surface. Also noticeable was the effect of face up or face down on the surface, something which changed the perception of the audio signals by the smartphone more than expected.

7.1.1. Beamforming algorithm improvement

For the title of this thesis is 'On determining smartphone microphone directivity with application to beamforming', the improvement of the beamforming algorithm is an essential part of the research. This part of the research has been done by Van Wijngaarden en Wouters [6], who did experiments using the data presented in this section with results. A little overview of the application of data to the algorithm is given in the Chapter 6.

The first results looked not so promising: no improvement of the beamforming algorithms has been observed. For further results, the reader is encouraged to read their work. The beamforming algorithm uses both gain and phase information of the determined directivity. Gain loss due to the acoustic system has been compensated, but phase shifts due to the acoustic system are uncorrected by the equalizer.

The lack of improvement of the beamforming algorithm could be caused by these uncorrected phase shifts. Also the difference between the directivity of the microphone on a surface with respect to mid-air measurements might be a cause. A final cause could be the low robustness of the beamforming algorithm, this might be limiting its performance [6].

Some performance improvement of the beamforming algorithm with equalized data against non-equalized data is expected, for it will be used for frequencies in the speech-domain: between 125 and 8000 Hz. As can be seen best in Figure 4.4b these frequencies are influenced by the equalizer and so the equalized data could improve the beamforming algorithm. Due to time-constrains, only the equalized data has been tested in the beamforming algorithm.

7.2. Recommendations for future work

For future work there are a lot of recommendations. To start with, of course, the impulse response of the acoustic system. The making of the filter has delayed our research and the disability of our filter compensate for phase shift due to the acoustic system have made the the results less usefull.

In addition, for the manufactures of smartphones, we recommend to use better microphones, for this not only would improve experience for calling and recording sounds, but a more flat frequency response is easier to process by the beamforming algorithm. Manufacturers also should give access via the software to other microphones on the smartphone. Most modern smartphones have multiple microphones, for instance one on the listeningside of the phone to suppress environmental noise when calling. If the directivity of this microphone also is known, its recording can be used in beamformer algorithms: more microphones improve the results of the algorithm. At last for the manufactures, the smartphone's orientation (and position) at this moment is not so accurate [3].

If the orientation of the smartphone is more accurately known, the directivity can be used more accurate, which may be the cause to implement an interpolation method to cover all angles of the sphere and not only the ones in the grid at this moment. Also considering different sampling schemes is a topic to have more research about. Already given in Chapter 2, Background, there are more ways to sample a sphere than equiangular and this could lower your number of sampling points very much.

For future work in the line of this thesis, the difference between multiple phones of the same type could be expressed explicitly and not only by comparing different patterns. What is the standard deviation between different phones of the same type. In addition, phones from other manufacturers than LG could be looked at: is the directivity of an Apple or a Samsung much alike an LG smartphone? If this would be so, it would simplify the product in mind, for the beamforming algorithm would only need one directivity for all phones.

In the same vein, the question rises how much the beamforming algorithms function is affected by placing the smartphones on a surface. For we know the directivity differs, how much would it the algorithm improve if there are different directivity given for different smartphone setting.

It looks like there is much more research needed to finish the product in mind, but with the results achieved, the goal is much closer than before.

Bibliography

[1] W. Zhang, M. Zhang, R. Kennedy, and T. Abhayapala, *On high-resolution head-related transfer function measurements: An efficient sampling scheme,* IEEE Transactions on Audio, Speech and Language Processing **20**, 575 (2012).

[2] N. Gaubitch, J. Martinez, W. Kleijn, and R. Heusdens, *On near-field beamforming with smartphone-based ad-hoc microphone arrays,* in *Acoustic Signal Enhancement (IWAENC), 2014 14th International Workshop on* (2014) pp. 94–98.

[3] S. Bosma and R. Smeding, *Smartphone Audio Acquisition and Synchronization Using an Acoustic Beacon,* Bachelor's thesis, Delft University of Technology (2015).

[4] B. D. Van Veen and K. M. Buckley, *Beamforming: A versatile approach to spatial filtering.* IEEE ASSP magazine **5**, 4 (1988).

[5] M. H. Hennecke, G. Fink, *et al.,* *Towards acoustic self-localization of ad hoc smartphone arrays,* in *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on* (IEEE, 2011) pp. 127–132.

[6] N. van Wijngaarden and E. Wouters, *Acoustic Enhancement via Beamforming using Smartphones,* Bachelor's thesis, Delft University of Technology (2015).

[7] *MW-BO329_nexus5_MG_20131031144916.jpg,* `http://s.marketwatch.com/public/resources/MWimages/MW-BO329_nexus5_MG_20131031144916.jpg`, accessed: 2015-06-14.

[8] Google, *Nexus 5 - Google,* `http://www.google.nl/intl/nl_nl/nexus/5/,` accessed: 2015-04-24.

[9] LG, *LG Nexus 5 D821 | Het beste van LG en Google | LG ELECTRONICS Benelux Nederlands,* `http://www.lg.com/nl/mobiele-telefoons/lg-D821-Nexus-5-zwart`, accessed: 2015-06-19.

[10] R. A. Mucci, *A comparison of efficient beamforming algorithms,* Acoustics, Speech and Signal Processing, IEEE Transactions on **32**, 548 (1984).

[11] J. Ryan and R. Goubran, *Near-field beamforming for microphone arrays,* in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97. IEEE International Conference on (Volume:1)* (1997) pp. 363–366.

[12] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications,* Digital Signal Processing - Springer-Verlag (Springer, 2001).

[13] S. Doclo and M. Moonen, *Design of far-field and near-field broadband beamformers using eigenfilters,* Signal Processing **83**, 2641 (2003).

[14] M. R. Thomas, J. Ahrens, and I. Tashev, *Optimal 3D beamforming using measured microphone directivity patterns,* in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on* (VDE, 2012) pp. 1–4.

[15] J. Marzo and B. Pridhnani, *Sufficient conditions for sampling and interpolation on the sphere,* Constructive Approximation (2014), 10.1007/s00365-014-9252-4.

[16] A. P. Bates, Z. Khalid, and R. A. Kennedy, *An optimal dimensionality sampling scheme on the sphere for antipodal signals in diffusion magnetic resonance imaging,* in *International Conference Acoustics, Speech and Signal Processing 2015 (ICASSP'2015)* (2015) pp. 94–98.

[17] Z. Khalid and R. A. Kennedy, *Spherical harmonic transform for minimum dimensionality regular grid sampling on the sphere,* (2015), `http://users.cecs.anu.edu.au/~rod/papers/2015/3819.pdf`.

[18] J. Goslinski, M. Nowicki, and P. Skrzypczynski, *Performance comparison of ekf-based algorithms for orientation estimation on android platform,* (2015).

[19] J. Proakis and D. Manolakis, *Digital Signal Processing* (Pearson Prentice Hall, New Jersey, 2007).

[20] B. Girod, R. Rabenstein, and A. Stenger, *Signals and systems* (John Wiley & Sons Inc, 2001).

[21] M. R. P. Thomas, *A Novel Loudspeaker Equalizer,* Master's thesis, Imperial College London (2006).

[22] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, *Comparison of different impulse response measurement techniques,* AES: Journal of the Audio Engineering Society **50**, 249 (2002).

[23] J. Hee, *Impulse response measurements using MLS,* (2003).

[24] N. Aoshima, *Computer-generated pulse signal applied for sound measurement.* Journal of the Acoustical Society of America **69**, 1484 (1981).

[25] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, *An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,* Journal of the Acoustical Society of America **97**, 1119 (1995).

[26] *Turntable system - Brüel & Kjær,* `http://www.bksv.com/Products/transducers/ear-simulators/electroacoustic-accessories/turntable-type-9640` (), accessed: 2015-06-09.

[27] *RME: Fireface 800,* `http://www.rme-audio.de/en/products/fireface_800.php` (), accessed: 2015-06-09.

[28] *Tymphany 4" midrange, type number: M10MD-39-08,* Tymphany and Vifa (2007), `http://www.audiokit.it/ITAENG/Altoparlanti/VIFA/M10MD-39-08.pdf`.

[29] *Brüel & Kjær Free-field 1/2" Microphone Type 4190, Serial No: 2639635,* `http://www.bksv.com/Products/transducers/acoustic/microphones/microphone-preamplifier-combinations/4190-L-1` (), accessed: 2015-06-09.

[30] P. C. H. Intermountain Healthcare, *Let's talk about... hearing and testing,* (2013).

[31] P. C. Loizou, *Speech Enhancement: Theory and Practice* (CRC press, 2013).

[32] T. Ajdler, L. Sbaiz, and M. Vetterli, *Plenacoustic function on the circle with application to hrtf interpolation,* (2005) pp. III273–III276.

[33] M. Karjalainen and T. Paatero, *Equalization of loudspeaker and room responses using kautz filters: Direct least squares design,* EURASIP Journal on Applied Signal Processing **2007**, 185 (2007).

[34] D. Poole, *Linear Algebra: A Modern Introduction* (Thomson Brooks/Cole, 2006).

[35] S. Haykin, *Adaptive filter theory*, 3rd ed. (Prentice-Hall, NJ, USA, 1996).

[36] *RLS Adaptive Filters - MATLAB & Simulink - MathWorks Benelux,* `http://nl.mathworks.com/help/dsp/ug/rls-adaptive-filters.html` (), accessed: 2015-06-04.

[37] *Compute output, error and coefficients using Recursive Least Squares (RLS) algorithm - MATLAB - MathWorks Benelux,* `http://nl.mathworks.com/help/dsp/ref/dsp.rlsfilter-class.html` (), accessed: 2015-06-04.

Appendix A

Schedule of requirements

This product consists of two parts: a function which gives the impulse response in different directions for a smartphone microphone and a measurement set-up description to add new data to this function. This function will be part of a system which final objective is to enable crystal clear conference calls using an array of smartphones. For this products, measurements to determine the impulse response for different directions are conducted on a commercially available smartphone. These measurements need to be reproducible for any concerned smartphone, so manufacturers can add their own phones to the system, as input for the function. The measurements and analyses carried out, are an extension of previous research [2].

The final product, the conference calling system, is intended for the professional market: large companies which make use of conference calling systems. Next to that, smartphone manufacturers must be able to measure the impulse responses from their smartphones to apply this data to the function.

A.1. Requirements concerning the intended use

1. The function gives the impulse response for a given frequency and datapoint location.

2. The location of a datapoint is indicated by the angle between the projection of the vector pointing to this point on the x, y-plane and the x-axis (θ) and the angle between the projection of this vector on the x, z-plane and the z-axis (ϕ).

3. There must be an anechoic chamber available in order to do the measurements.

4. The measurements concerning the directivity must be replicable within three days.

5. The measurement results must be accurate enough to be used in a beamforming algorithm.

A.2. Requirements from an ecological point of view

There are no ecological requirements since the product solely consists of software.

A.3. Requirements concerning the system

A.3.1. Usage characteristics

1. The function returns the impulse response for a given datapoint location.

2. The location of a datapoint is indicated by the angle between the projection of the vector pointing to this point on the x, y-plane and the x-axis (θ) and the angle between the projection of this vector on the x, z-plane and the z-axis (ϕ).

3. The function must be quick enough to be used in a beamforming algorithm.

4. The function must have a clear overview provided with sufficient comments.

5. The user must be able to retrieve results for not directly measured datapoints.

6. There must be a function which generates a spherical plot showing the entire three dimensional directivity gain pattern (in decibels) of the concerned smartphone microphone.

7. The measurements could be done within three days.

A.3.2. Production and commissioning characteristics

1. The function must be usable by a user with MATLAB Student R2014a.

2. The measurements must be replicable by a customer with an anechoic chamber available.

A.3.3. Recycling features

Recycling is not an issue since the product solely consists of software.

A.4. Requirements regarding the production

1. MATLAB must be available to run the measurement and directivity software.

2. The function-code will be developed using MATLAB Student R2014a.

3. The measurement conducting system will be developed using MATLAB Student R2014a.

4. There must be an anechoic chamber available.

5. There must be a turntable available (for example: Brüel & Kjær, system type 9640).

6. There must be a loudspeaker available.

7. There must be an arc with a 1-meter radius available.

8. There must be an audio interface available.

A.5. Requirements concerning the recycling system

Recycling is not an issue since the product solely consists of software.

A.6. Requirements from a strategic, marketing and sales point of view

1. The product must be directly available after purchase.

Ethical considerations

This chapter in the thesis is not technically part of the thesis. It describes some ethical considerations related to (smartphone) microphone arrays and their applications. This chapter is included in all thesis works of the project group, e.g. also the work of Van Wijngaarden and Wouters [6] and Bosma and Smeding [3]. This chapter contains *opinions* of the authors and is not a technical document.

B.1. The need for ethical discussion

In creating the beamforming system, we envisioned a single purpose: speech enhancement in ad-hoc teleconference calls. Unfortunately, history teaches us some well-intended inventions turn out to be dangerous to the point where regulation is needed to ban it. Examples are numerous: tetraethyllead was added to gasoline starting in the 1920s but phased out worldwide from the 1970s because of its environmental impact. Radithor is another 1920s example; a "wonder medicine" in its heyday, it was subsequently discovered to be extremely toxic to the point that the main marketeer for the drugs' "jaw fell off". Although these examples are very extreme, it is worthwhile examining the potential influence on society of our system-to-be.

B.2. Unintentional use cases

The question we must ask ourselves thus becomes: what are some possible unforeseen use cases of our product? The possibility of a surveillance state (or company) using ad-hoc beamforming techniques to snoop on all its residents springs to mind. The revelations of NSA whistle blower Edward Snowden have documented just how far some American and European states are willing to go to keep tabs on their citizens. Moreover, they have revealed great levels of cooperation from tech-industrial giants such as Google, Facebook, Apple and others.

B.3. Beamformer contribution

But how much would a smartphone beamforming system contribute to any level of mass-surveillance? The truth is that beamformers only work when the locations of smartphones are known to extreme accuracy. Such accuracy is not achievable with sensors fitted on even the latest smartphones. Furthermore, indoor localization of smartphones has thus far always relied on either large, extra hardware (RF beacons) or a sonic beacon. Neither seem appropriate to a state wanting to covertly listen in on conversations.

B.4. Applications

Our prime focus has been with companies deploying the technology in teleconferencing situations. We do not seek out to be the next Skype, Hangouts or appear.in. A company using such software must be fully aware of the consequences of privacy invasion or potential information leakage caused by such software. Our solution would not be connected directly to the internet; it could, for example, serve as a virtual microphone input to the computer. What is done with the recorded audio is

completely up to the user.

B.5. Final remarks

Theoretically, if the technology becomes more mature, it may become a weapon in the arsenal of surveillance states. But the technology is not remotely there yet. Ad-hoc beamforming on smartphones is still relatively uncharted territory and the lack of silent, unnoticable synchronization options kills off a potential to listen in discretely. We honestly see no potential for abuse of the technology investigated in our research.

<div align="right">Appendix C</div>

Measurement set-up: overview

In this appendix an overview of the measurement set-up will be given for replication. The measurements were conducted in the anechoic room in the Physics department at the Delft University of Technology. The used software to generate the TSP and MLS sequences is made by Thomas [21]. First the used hardware and the connection between those will be addressed, after that the measurement procedure will be repeated in short and at the end of this appendix some pictures of the measurement set-up are given.

C.1. Used hardware

Smartphone and measurement microphone

- Google's Nexus 5™ mobile phone, with an application specifically made for this project (which implementation is detailed in [3]);

- Microphone: Brüel & Kjær Free-field microphone [29] (Figure D.2);

- WiFi-router to connect the smartphone to the computer;

Set-up hardware

- Aluminium arc, with a radius of 1.21 meter, with a reach of angles in $[-75, 75]°$, were the reference $0°$ is the horizon, or equivalently ($\phi \in [15, 165]°$);

- Turntable: the turntable which is part of the remote controllable Brüel & Kjær 9640 turntable system [26];

- A stick, with a length of 1.15 m, which fits in the turntable, with enough surface at the top to support the smartphone, which top was covered with sound absorbing foam;

- A stick, with a length of 1.15 m, which fits in the turntable, with $290 \times 240 \times 8$ mm plywood plate on top;

- Double-sided tape (1 mm thick);

- Tripod;

Acoustic system

- HP EliteBook 8540w personal computer with Windows 8 operating system, running MATLAB-Student R2014a;

- RME Fireface 800 [27];

- Audio amplifier: a custom-build 8-channels amplifier of 25W per channel with 4Ω, built with the IC Philips TDA8560Q;

A8

- Loudspeaker: Tymphany 4"Midrange loudspeaker, type number M10MD-39-08 [28] (Figure D.1);

  Connections

- Firewire cable to connect the computer to the Fireface;

- XLR-cable to connect the Fireface to the audio amplifier;

- Twist-lock Speaker cable to female-banana and banana cable to connect the amplifier to the loudspeaker;

- Coaxial cable, RCA cable and a connection between these two, to connect the measurement microphone to the Fireface.

C.2. Volume settings

Before the start of the measurements the settings of the equipment have to be adjusted such that:

- The Fireface does not clip (its GUI is very clear about clipping), this can be done in two ways:

  - Turning the volume down in Windows 8 (our settings: 5% volume);
  - Using a gain factor in MATLAB to lower the outputsignal (our settings: -7 dB).

- The loudspeaker signal is not too loud so it is distorted by the loudspeaker or the loudspeaker is affected by the signal, this can be done in two ways:

  - Changing the volume settings of Windows 8 and MATLAB;
  - Changing the volume settings of the Fireface in its GUI (our settings: -30 dB outputsignal).

C.3. Conducting the measurements

The measurements are conducted with a 9° equiangular sampling scheme as described in Section 3.2.1. For measurements in mid-air, the smartphone was placed on the stick covered with foam. For measurements $\phi \leq 90°$ the smartphone lay on its back, with its middle point in the middle of the stick, starting with the speechside toward the loudspeaker and the turntable was turned in positive direction in MATLAB (negative direction according to the turntable system). First all 41 measurements with the TSP sequence where conducted, where after the MLS sequences was used to measure. For measurements $\phi > 90°$ the smartphone was flipped over so it laid on its screen, and the turntable was instructed to turn in the other direction, so the the movement of the microphone with respect to the speaker is the same as for the measurements $\phi \geq 90°$.

For angles $\phi \in [0, 15\rangle \cup \langle 165, 180]°$ the turntable was positioned in an angle and raised to place the phone in the same position in space. For stability the smartphone was taped to the stick with double-sided tape.

The measurements with the phone placed on a surface were conducted using the stick, with the plywood attached to it. These were only measured for $\phi \in [0, 135]°$, once with the screen on the surface (face down) and once with the back of the phone on the surface (face up). For $\phi \in [0, 15\rangle$ also the double-sided tape is used. For these measurements the turntable always turned in positive direction following the MATLAB-code.

C.3.1. Measurements with the microphone

For equalization these measurements were also conducted with the B&K microphone. The microphone was placed on a tripod on top of the turntable, also with a distance of 1 meter from the loudspeaker (positioned at $\phi = 90°$), and the same code as used for the smartphone measurements was run, except the control-functionality of the turntable was turned off, so the microphone stayed in the same place.

Pictures of the measurement set-up can be found on the next page.

C.4. Pictures

(a) Measurement at $\phi = 63°$

(b) Measurement at $\phi = 9°$

(c) Measurement at $\phi = 0°$

(d) Measurement with the B&K microphone

(e) Measurement at $\phi = 90°$, face up on the surface
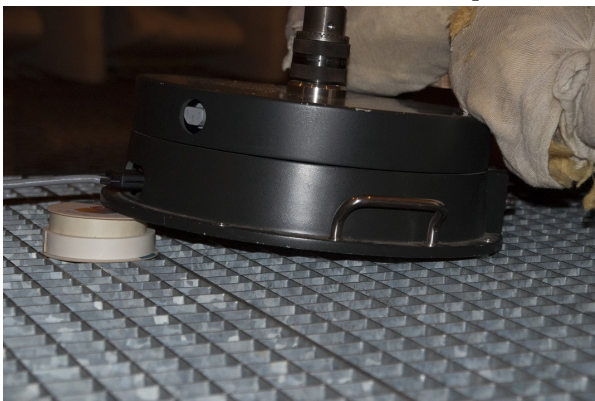
(f) Placement of turntable an an angle of 6°

(g) Connection between the hardware

Figure C.1: Pictures of the measurement set-up

Hardware Data

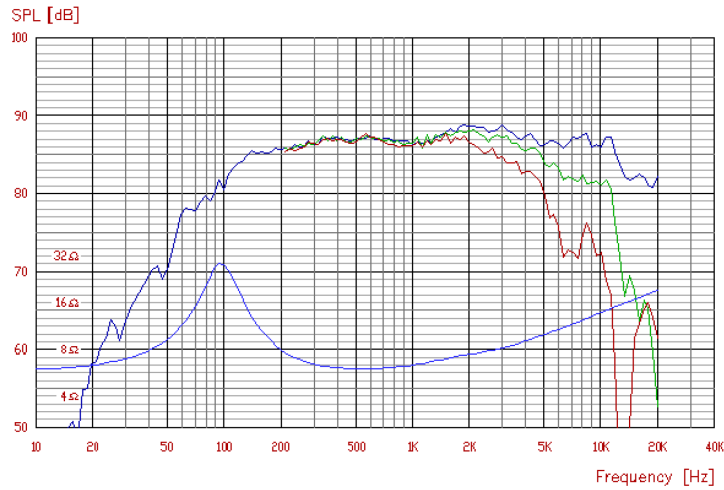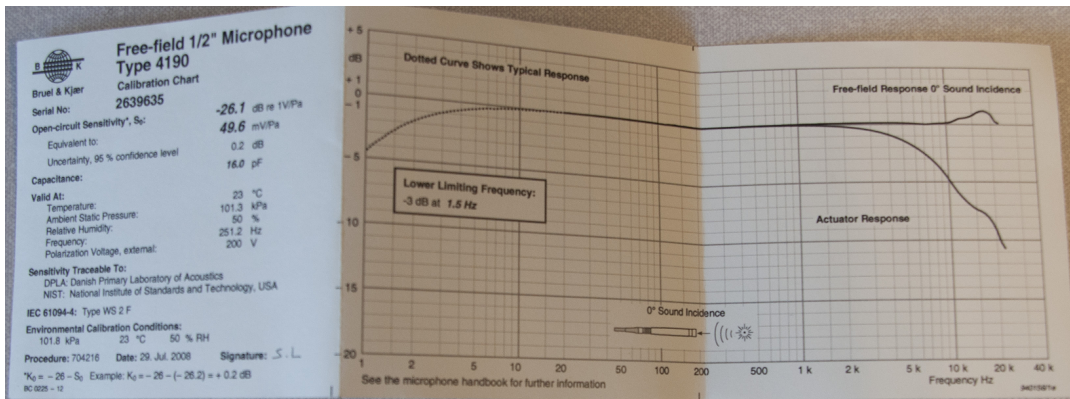**Figure D.1:** The frequency response of the loudspeaker [28].


Figure D.2: The frequency response of the microphone [29], a picture taken from its manual.

Testing Toeplitz and RLS methods

Figure E.1: The results of finding a inverse filter using the Toeplitz-method (described in Section 4.2.3). A simple inputsignal has been chosen $\left(s_{\text{input}} = \sin(6\pi t) + 0.25\sin(80\pi t)\right)$ and as a simple filter the Delta-function at different times has been chosen. In **(a)** the resulting inverse filter gives a good approximation of s_{input} (green line), in **(b)** however, the result of the use of the inverse filter doesn't give a good approximation at all.

(a) Used filter: a Delta-pulse at 10 ms
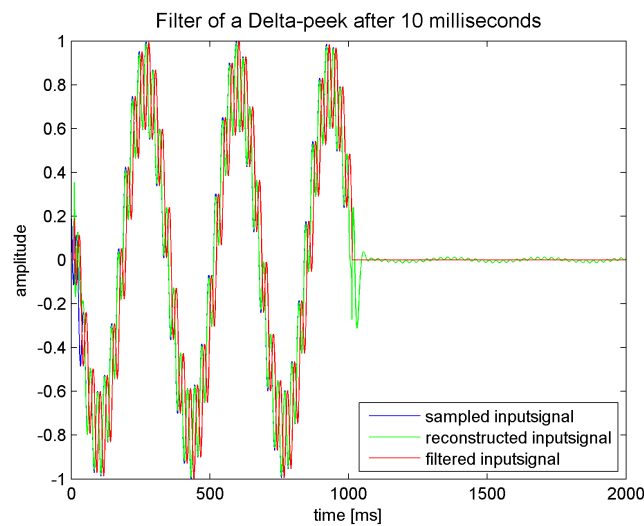
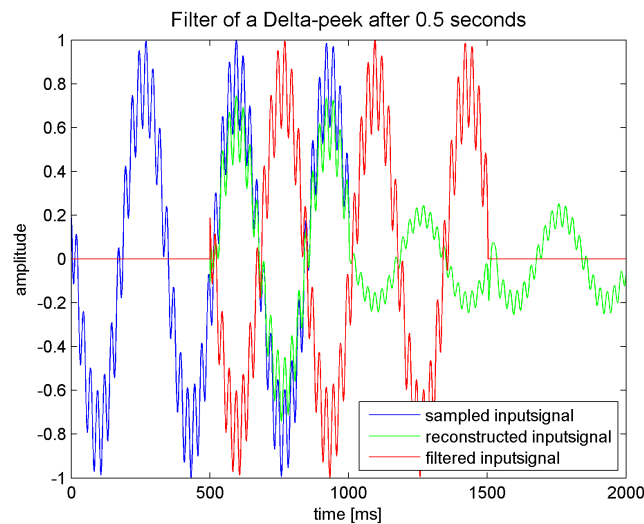**(b)** Used filter: a Delta-pulse at 500 ms, or 0.5 s

Figure E.2: The results of finding a inverse filter using the RLS-method (described in Section 4.2.4).
The test with simple input signals and filters worked very good, so here the results of the next step:
the recorded signal (one TSP pulse) from the microphone. In **(a)** the response of the soundsystem
has been caught in a FIR-filter with a length of 500 and the approximation has good results (the
error is given by the red line), although a high forgetting factor is required.
In **(b)** the same has been tried for the inverse filter, with very poor result: a very high error between
the estimation and the original system.

(a) Finding a FIR-filter for the soundsystem. Used RLS-settings: a filterlength of 500 and a forgetting factor of
0.999999.
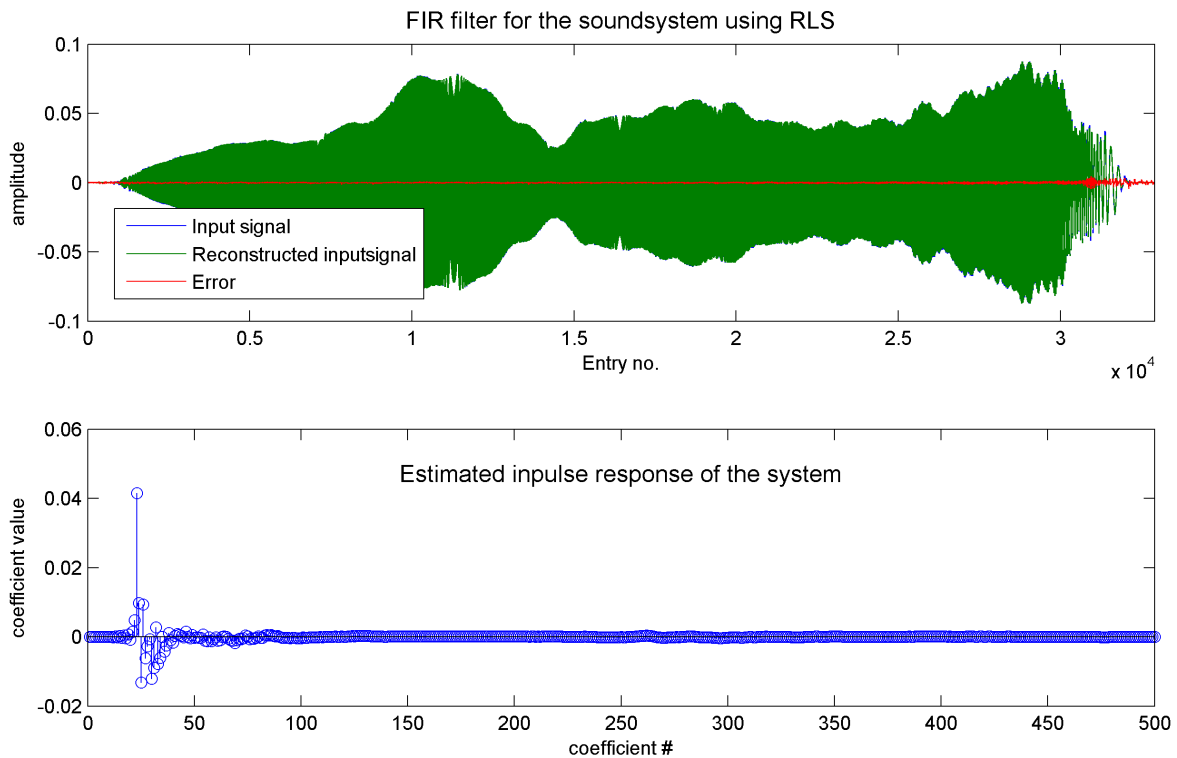
(b) Finding a FIR-filter for the inverse of the soundsystem. Used RLS-settings: a filterlength of 500 and a
forgetting factor of 0.999999. (Other settings did not have better results and took a lot of time to compute.
Therefore the inverse filter with the same RLS-settings is shown.)