Measuring Loudspeaker Distortion and Room Reverberation Time Using a Speakerphone

Viktor Wilhelmsson
Abstract

This master thesis project was carried out during the spring semester of 2016 at the company Limes Audio. The company specializes in making software, electronics, industrial design and mechanics with the aim to improve audio quality in loudspeaking communication systems.

The performance of audio conferencing systems may degrade if there are distortions, or if the acoustical properties of the room are unfavorable. To ensure that the system works optimally, any unwanted effects must first be identified.

This thesis will cover the necessary theory of acoustical systems and measurements. The implementations of three different measurement sequences are presented. The measurements are evaluated on three different speakerphone units and three venues to assess the accuracy of the presented methods.

The results indicate that it is possible to use a single measurement signal, the exponential sine sweep, to measure both room acoustical reverberation time and loudspeaker distortion for a speakerphone setup.
## Contents

1 Introduction ........................................ 4

2 Linear Systems ..................................... 5

3 Acoustics ........................................... 7

4 Impulse Response Measurement .................. 8
   4.1 Introduction .................................. 8
   4.2 Exponentially Swept Sine Method .......... 9
   4.3 Reverberation time .......................... 11

5 Distortion Measurement .......................... 13
   5.1 Introduction .................................. 13
   5.2 Exponentially Swept Sine Method .......... 13
   5.3 Silence Sweep Method ......................... 13
   5.4 Multitone Method ............................ 14
   5.5 Total Harmonic Distortion .................. 14

6 Implementation ................................. 16
   6.1 Exponential Sine Sweep ..................... 16
      6.1.1 Generating the Signal ................. 16
      6.1.2 Evaluating the Impulse Response .... 18
   6.2 Silence Sweep .................................. 19
   6.3 Multitone ...................................... 19
      6.3.1 Optimizing the Phases ................. 20
      6.3.2 Analyzing the Response ............... 20

7 Result ........................................... 21
   7.1 Measured Impulse Response ................. 21
   7.2 Measured Distortion .......................... 23

8 Conclusion ...................................... 27
   8.1 Further Work .................................. 27

9 References ...................................... 28
1 Introduction

Any loudspeaking communication device, for instance a speakerphone, needs to limit the echo traveling from the speaker to the microphone. Acoustic echo cancellation (AEC) is a method that can be used for this purpose. For most AEC algorithms to work optimally, the echo needs to be without distortion and not too delayed. A damaged loudspeaker will produce distortion, and a large room gives longer echoes. This master thesis is about measuring these two factors, using a speakerphone, and determining when something is inadequate.

The echo path can be thought of as a system which takes an input signal, plays it out through the loudspeaker, the sound interacts with the room and the enclosure before reaching the microphone, forming the systems output signal.

Many acoustical systems can be approximated as linear time-invariant systems. This type of system can be completely specified by its impulse response. However, due to a number of different physical effects, non-linearities arise in the loudspeaker and enclosure. No practical system is truly linear so we complement the impulse response with additional measures of the non-linearities.

Chapter 2 will go through the general theory about linear systems. After that, chapter 3 will shortly apply this theory to acoustical systems. Chapters 4 and 5 describe methods to measure the impulse response and distortions of acoustical systems. The implementation of these methods is described in more detail in chapter 6. The result from evaluating the methods are found in chapter 7. Finally, chapter 8 contains conclusions and remarks on further work.
2 Linear Systems

We can model any input-output system as a black-box that takes an input signal $x(t)$ and produces an output $y(t)$. Naming this system $H$ we could describe this relation as $y(t) = H\{x(t)\}$. The system $H$ acting on a signal $x(t)$ and producing the output $y(t)$. We can also represent this visually as the figure below.

\[ x(t) \rightarrow H \rightarrow y(t) \]

A special subset of these systems are called linear and time-invariant (LTI). For a system to be LTI it needs to fulfill two properties:

1. Linearity: $\alpha y_1(t) + \beta y_2(t) = H\{\alpha x_1(t) + \beta x_2(t)\}$
2. Time-invariance: $y(t - t_0) = H\{x(t - t_0)\}$

A very practical feature of LTI systems is that they can be fully described by a single measure, its impulse response. The impulse response, often designated $h(t)$, is the output of the system after an instantaneous impulse $\delta(t)$, i.e. $h(t) = H\{\delta(t)\}$, where $\delta(t)$ is the Dirac delta function.

\[ \delta(t) \rightarrow H \rightarrow h(t) \]

If the impulse response is known, we can calculate the output of the system for any input signal by convolving the input with the impulse response,

\[ y(t) = x * h(t) = \int_0^t x(\tau)h(t - \tau) \, d\tau. \]

Linear systems can be further divided into dynamic or static systems. In a static system the output at some time $t_0$ only depends on the instantaneous input at that time $x(t_0)$; static systems are also called memoryless. A dynamic system is said to posses memory and the output $y(t_0)$ depends not only on the input at time $t_0$ but
also on previous values of the input. Another way to see this distinction is that
dynamic systems are frequency dependent, whereas static systems are not.

In a static system, the impulse response is simply $h(t) = \alpha \delta(t)$, i.e. a single peak at
$t = 0$ and zero for every other $t$. For a (causal) dynamic system the impulse response
can be non-zero for all $t > 0$.

A system which is not linear often gives rise to distortions in the output signal.
Distortions are any kind of alteration of the signal, often unwanted. In acoustical
systems distortions can introduce undesired frequencies in the output signal that
degrade the perceived sound quality. Distortion could also be desired in for instance
electric guitar amplifiers where it gives a characteristic sound. However, in speaker
phone systems any loudspeaker distortion is detrimental.
3 Acoustics

A physical system is never fully linear nor time-invariant, but is often approximated as such to simplify the analysis. An acoustical system may violate the linearity assumption in two ways: It could fail to be scale invariant, i.e. \( H\{ax(t)\} \neq ay(t) \), because louder volumes often causes harmonic distortion. The system might also not be commutative, i.e. \( H\{x_1(t) + x_2(t)\} \neq y_1(t) + y_2(t) \), for instance due to what is called intermodulation distortion. Acoustical systems are seldom time-invariant since, for instance, temperature affects both the speaker properties and also the propagation of sound in the room.

When an acoustical system is analyzed it is often approximated as linear to examine impulse response or transfer function, but later considered non-linear when studying distortions.

Sound that is emitted from a loudspeaker will radiate outwards until it hits a wall. Some of the energy will be absorbed by the wall and some will reflect back into the room. Some of the reflected sound will reach the microphone as first order echo, but some will instead reflect of another wall, and so on. The number of different path the sound can take to reach back to the microphone quickly grows. It is not possible to distinguish all the different levels of echo, instead there is only a heightened level of sound energy at the microphone which diminishes with time. This phenomena is known as reverberation.

For each time the sound strikes a wall, some of the energy is lost. The amount that will be absorbed is highly dependent on the material of the wall and the frequency of the sound. The effect is that the echoes, or reverberation, become fainter and fainter.
4 Impulse Response Measurement

4.1 Introduction

Acoustical impulse response has historically been measured by recording an approximate impulse generated by gunshot or by popping a balloon. Current methods often rely on noise signals or swept tones as stimulus which by mathematical transformations can yield the impulse response. Some advantages of these methods are; significantly improved signal-to-noise ratios and reliable repeatability.

The method used here for estimating the impulse response relies on a stimulus signal \( x(t) \) and what is called an inverse filter \( \tilde{x}(t) \). If \( x(t) \) is the input signal to the system, and we let \( \tilde{x}(t) \) be the inverse defined as:

\[
\tilde{x} \ast x(t) \equiv \delta(t),
\]

then we can extract the impulse response from the output \( y(t) \) as:

\[
\tilde{x} \ast y = \tilde{x} \ast x \ast h = \delta \ast h = h.
\]

Alternatively, we can formulate the same relations in the frequency space where convolutions are substituted by multiplications:

\[
\tilde{X}(f)X(f) \equiv 1
\]

\[
\tilde{X}Y = H.
\]

In theory, any wide band signal could be used for determining the impulse response of a dynamical system. However, as we will see, there are certain methods that are more suitable for determining the impulse response of acoustical systems.
4.2 Exponentially Swept Sine Method

A method initially proposed in (Farina 2000) and further improved by (Novak et al. 2015) utilizes an exponentially swept sine signal as the excitation (sometimes called logarithmic sweep or logarithmic chirp). The impulse response is acquired after linear convolution with the so-called inverse filter $\tilde{x}(t)$. This method separates the linear impulse response and passes most of the harmonic distortion into the non-causal part, which is easy to separate.

An excitation signal of length $T$ between angular frequencies $\omega_1$ and $\omega_2$ can be expressed as

$$x(t) = \sin \left( \frac{\omega_1 T}{\ln(\omega_2/\omega_1)} \exp \left( \frac{t}{T} \ln(\omega_2/\omega_1) \right) \right), \quad t = [0, T]. \quad (4.1)$$

The instantaneous frequency is defined as the time derivative of the phase, which yields

$$\omega(t) = \frac{d\phi}{dt} = \omega_1 \exp \left( \frac{t}{T} \ln(\omega_2/\omega_1) \right).$$

We can see that the frequency is exponentially increasing and has the expected end-point values,

$$\omega(0) = \omega_1, \quad \omega(T) = \omega_2.$$  

A spectrogram of $x(t)$ can be seen in figure 4.1.

The inverse of $x(t)$ is denoted $\tilde{x}(t)$ and is simply the time-reversal of $x(t)$ with a six decibel per octave decreasing amplitude (Farina 2000). A full derivation of this inverse signal can be found in (Novák et al. 2010).

By playing the sine sweep signal while recording the response, we end up with results like those shown in figure 4.2. Note the offset distortion lines above and to the left of the linear response. After convolving with the inverse filter, each line in the spectrogram is tilted to vertical, as can be seen in figure 4.3. The main line is now the linear impulse response of the system, while all the others are distortion effects that are easy to crop away. The linear impulse response can be further analyzed for determining room acoustical properties, as we will see in the next section. We can also analyze the distortion for determining loudspeaker attributes, which we will see in section 5.2.
Figure 4.1: Spectrogram of the excitation signal $x(t)$

Figure 4.2: Spectrogram of recorded sine sweep measurement

(a) Medium to low distortion       (b) High distortion with spurious noise effects
For a system where sound propagates in a room, there exists multiple relevant measures extractable from the impulse response. Perhaps the most important parameter of any acoustical space is the reverberation time. Reverberation is a form of diffuse echo which makes any sound persist in the room before diminishing. The time it takes for a sound to decay 60 dB from its original power is called the reverberation time, denoted $RT$. This parameter is often estimated from an impulse response and a common procedure to do this is described in the international standard ISO 3382-1:2009.

The energy versus time of an impulse response from an exponential sine sweep measurement looks approximately like in figure 4.4. After cropping of distortion to the left of the main peak we end up with the linear impulse response depicted in figure 4.5. The rate of the decay is what is interesting for room acoustical measurements.

4.3 Reverberation time
Figure 4.4: Impulse response from exponential sine sweep measurement. The main peak to the right shows the linear impulse response and the peaks to the left show harmonics.

Figure 4.5: Linear impulse response with a noise floor
5 Distortion Measurement

5.1 Introduction

Acoustical systems often exhibit some type of distortion of the signal, especially at higher amplitudes. One common type of distortion is harmonic distortion, which introduces tones at multiples of the input frequencies. Another related form is intermodulation distortion which gives rise to tones at sums and differences of the input frequencies and their multiples. Acoustical systems can also have random distortions which are not deterministic and give rise to noise or otherwise distort the sound. This chapter will describe three different methods of estimating distortion.

5.2 Exponentially Swept Sine Method

The same method as presented in section 4.2 for estimating the impulse response and the reverberation time, can also be used for determining the level of harmonic distortion in the system. In section 4.2 the linear response was extracted and the distortion trimmed off. To measure the distortion, the linear response is trimmed off and the distortion (together with noise) is left for analysis.

Instead of lumping together all distortion, it is also possible to extract each harmonic for individual analysis. This is method is described in (Farina 2000), but not further investigated here.

5.3 Silence Sweep Method

A new method was proposed by Farina (Farina 2009), which is quite related to the exponential sine sweep method. This method is able to detect not only the harmonic distortion but also intermodulation and other types of distortion. The idea is to excite the system with full-band noise, while leaving a small section of the spectrum free for studying the leaked distortion. This suppressed part of the spectrum is then swept over time from low to high frequencies.
To generate the signal, one creates a white noise signal with a short silent part in the middle. Convolving the noise signal with an exponential sine sweep, the group delay 'tilts' the signal so that the silent part is swept through the frequencies, see figure 5.1. Since the sine sweep frequency is exponentially increasing, the energy in each octave is halved (−6 dB/octave). When the white noise signal is convolved with the sine sweep, the spectrum changes to −6 dB/octave, this kind of spectrum is also called a pink spectrum.

\[ f_t \ast f_t = f_{t+} \]

Figure 5.1: Simplified spectrograms with logarithmic frequency axis showing the generation of a silence sweep signal. The signal can be trimmed along the dashed lines.

Any sound recorded within the silent band is due to noise or distortion. Since the signal is effectively fullband noise, the distortion is not only harmonic but takes into account other form of distortion as well.

### 5.4 Multitone Method

A quick way to estimate the distortion is to play the sum of several tones with frequencies, \( f_n \), and at the same time measure the amount of energy produced in all frequencies except \( f_n \). This method is designed to measure both harmonic distortion and intermodulation distortion. The tones must be chosen so that they are periodic in the recorded time. They must also be chosen so that the frequencies are not multiples of each other, since they would then overlap with the harmonic distortion we want to measure (Czerwinski et al. 2001).

### 5.5 Total Harmonic Distortion

The most common measure of distortion is Total Harmonic Distortion (THD) or Total Harmonic Distortion and Noise (THD+R). This measure is a ratio between the energy of a fundamental tone to all harmonics (and noise).
\[ THD+N = \frac{\text{distortion + noise}}{\text{fundamental + distortion + noise}} \]

The fundamental tone can be fixed at for instance 1 kHz or stepped at multiple different frequencies. In the sine sweep measurement we can continuously sweep the fundamental tone to cover the whole frequency bandwidth of the unit.
6 Implementation

6.1 Exponential Sine Sweep

6.1.1 Generating the Signal

In practice, we make some modifications to the definition of the sine sweep in section 4.2. The implementation described here closely matches the implementation in (Vetter and Rosario 2011).

A direct reformulation of equation (4.1) into discrete time gives

\[ x(k) = \sin \left( \frac{\omega_1 K}{\ln(\omega_2/\omega_1)} \exp \left( \frac{k}{K} \ln(\omega_2/\omega_1) \right) \right) , \quad k = 0, \ldots, K. \]  

(6.1)

An important modification we need to make is to force the sweep to begin and end in zero phase. This minimizes any ripples due to the boundary effects at the start- and end-points. To achieve this, we make two adjustments. The first adjustment is to fix the start- and end-frequencies. Let the sweep end in exactly the Nyquist frequency, \( \omega_2 = \pi \) and let the sweep span a fixed number of octaves, i.e. \( \omega_1 = \pi/2^P \), where \( P \) is the number of octaves. With these changes, the start and end phase now becomes

\[ \phi(0) = \frac{\pi K}{2^P \ln 2^P}, \quad \phi(K) = \frac{\pi K}{\ln 2^P} \]  

(6.2)

Choosing \( K = M 2^P + 1 \ln 2^P \), where \( M \) is an integer ensures that both endpoints will have the correct phase. However, since \( k \) is discrete and integer, we will need to have two different values for \( K \), one ideal according to the previous criterion, and one rounded to the nearest integer. Our signal definition is thus

\[ x(k) = \sin \left[ \frac{\pi K_{\text{ideal}}}{2^P \ln 2^P} \exp \left( \frac{k}{K} \ln 2^P \right) \right] , \]  

(6.4)
where
\[ K_{\text{ideal}} = \text{floor} \left( \frac{K_{\text{max}}}{2^{P+1} \ln 2^P} \right) 2^{P+1} \ln 2^P, \]
and
\[ K = \text{round} \left( K_{\text{ideal}} \right). \]

It is thus possible, when generating the signal, to enter in the maximum number of samples of the sweep \( K_{\text{max}} \), which will then be slightly altered to fit the criteria. Examining the instantaneous frequency of this signal we see that the effect of having two different \( K \) is a frequency stretch by the factor \( K_{\text{ideal}}/K \),
\[ \omega(k) = \frac{d\phi}{dk} = \frac{\pi K_{\text{ideal}}}{2^P K} \exp \left( \frac{k}{K} \ln 2^P \right). \quad (6.5) \]

This factor is often very close to unity so the frequency stretch is not significant. The table below shows typical numbers for a 10 second sweep with sample rate 48 kHz and a sweep length of 10 octaves (23.4 Hz to 24 kHz). The error is typically on the order of \( 10^{-6} \sim 10^{-7} \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_s )</td>
<td>48 000 000</td>
</tr>
<tr>
<td>( K_{\text{max}} )</td>
<td>480 000</td>
</tr>
<tr>
<td>( P )</td>
<td>10</td>
</tr>
<tr>
<td>( K_{\text{ideal}} )</td>
<td>468 456.591</td>
</tr>
<tr>
<td>( K )</td>
<td>468 457</td>
</tr>
<tr>
<td>( \frac{K_{\text{ideal}}}{K} )</td>
<td>0.999 999 126</td>
</tr>
</tbody>
</table>

In this implementation, it is possible to provide a maximum sweep length \( K_{\text{max}} \). A longer sweep will give a better signal to noise ratio but there is also an increased risk of some interrupting impulsive noise disturbing the measurement (such as doors slamming, coughing or noisy traffic). Shorter measurements are easier to redo and can be averaged together for a better signal to noise ratio.

The signal \( x(k) \) has a pink spectrum, i.e. 6 dB damping for each octave. The inverse signal is the reverse of \( x(k) \) with a compensation for the pink spectrum. This compensating factor is derived in (Vetter and Rosario 2011) and gives a formula for the inverse filter
\[ \tilde{x}(k) = \frac{P \ln 2}{1 - 2^{-P}} \left( \frac{2^P}{K} \right)^{-k} x(K - k). \quad (6.6) \]
6.1.2 Evaluating the Impulse Response

After convolving with the inverse filter and cropping away the distortion, as described in section 4.2, we end up with an impulse response like figure 6.1. The procedure used for estimating the reverberation is outlined in the following list:

1. Filter the impulse response to divide it into one-third octave bands.

2. For each band:
   a) Calculate the power envelope by squaring the signal and fitting splines between peaks.
   b) Estimate the noise floor level by first performing a (kernel) density estimation and then choosing the maximum of the density estimate as the noise floor level.
   c) Smooth the envelope to remove noise peaks in the decay.
   d) Perform a linear regression \( y = ak + b \) between the peak and 6 dB above the noise floor.
   e) Estimate the reverberation time, for this band, as \( RT = \frac{-60}{af_s} \), where \( f_s \) is the sample frequency.
6.2 Silence Sweep

Generation of the silence sweep signal was described in section 5.3. The noise signal $n(k)$ can be created as

$$n(k) = \begin{cases} 
\varepsilon(k) & \text{for } 0 \leq k < 2N_s \\
0 & \text{for } 2N_s \leq k < 2N_s + N_g \\
\varepsilon(k) & \text{for } 2N_s + N_g \leq k < 4N_s + N_g 
\end{cases} \quad (6.7)$$

where $\varepsilon$ is normally distributed noise, $N_s$ is the length of the sweep and $N_g$ is the length of the silence gap. If $s(k)$ is an exponential sine sweep signal with length $N_s$ as defined in section 6.1.1 we construct the silence sweep signal as:

$$x(k) = (n * s)(N_s + k), \quad k = 0, \ldots, 3N_s + N_g. \quad (6.8)$$

Analysis of the response is done by convolving with the inverse sweep and extracting the silent region, which will contain the distortion.

6.3 Multitone

The multitone signal can be expressed as

$$x(k) = \sum_{n=1}^{N} \sin(\omega_n k + \phi_n), \quad k = 1, \ldots, K \quad (6.9)$$

where $N$ is the number of tones, $\omega_n$ is the frequency of each tone, and $\phi_n$ the phase. The length of the signal, in samples, is exactly one FFT analysis length. The frequencies must be chosen such that the period is a multiple of the signal length $K$. In the implementation, a desired lowest ($f_l$) and a highest frequency ($f_h$) is prescribed. The other $N-2$ frequencies are approximately logarithmically spaced between them. The following snippet of Matlab-code describes the generation of the frequencies.

```matlab
m1 = fLow/fs * nFFT;
m2 = fHigh/fs * nFFT;
toneFreqs = fs/nFFT * floor(10.^linspace(m1, m2, nTones));
```
6.3.1 Optimizing the Phases

To maximize the signal to noise ratio in the recordings, we want the test signal to have a high average energy. We cannot let the peaks of the signal overload the device, since this will result in clipping or unnecessary distortion. The ratio between the highest peak and the RMS level of a signal is called the crest factor

\[
CF = \frac{\max_{k} x(k)}{\sqrt{\frac{1}{K} \sum x^2(k)}}.
\]

We can minimize the crest factory of the test signal by varying the phases of the constituent tones. To minimize the crest factor it is sufficient to minimize the numerator since the denominator is constant with respect to the phases. We can formulate this as a minimization problem

\[
\begin{align*}
\text{minimize} & \quad \max_{\phi_n} \left[ \sum_{n=1}^{N} \sin(\omega_n k + \phi_n) \right] \\
\text{subject to} & \quad 0 \leq \phi_n \leq 2\pi, \quad n = 1, \ldots, N.
\end{align*}
\]

However, it is not necessary to guarantee the global minimum. Using a constrained non-linear minimization algorithm (such as SQP) yields adequate results in a short time.

6.3.2 Analyzing the Response

To reduce noise, several sections of length \(K\) can be recorded and averaged (in the time domain). This is possible since the signal is periodic with \(K\), while random noise is not.

From the (possibly averaged) response, a FFT of length \(K\) will yield peaks at the stimulus frequencies. The peaks will have minimum spectral leakage since the tones are periodic with length \(K\). Any other part of the frequency response will be distortion and noise.
7 Result

To evaluate the proposed methods, tests were performed on three similar speaker phone units and three different rooms. The units are described in the following list:

- **Unmodified unit — Reference**
- **Rattling screw —** A small loose screw was placed inside the speaker phone enclosure. This introduces spurious noises and distortion for certain frequencies and amplitudes.
- **Loose element —** The loudspeaker element was not fastened properly to the baffle. This introduces severe audible distortions, especially at lower frequencies.

The rooms used in the tests were:

- **Conference room —** A medium sized conference room (16 m²) with carpet floors, one glass wall, and windows on one wall.
- **Lunchroom —** 70 m², wood floor, two of the walls lined with windows
- **Semi-anechoic room —** A small (7 m²) acoustically isolated room with low noise and acoustically damped walls

7.1 Measured Impulse Response

The reverberation time was estimated in one-third octave bands from 23 Hz to 24 kHz, for a total of thirty bands. A single band of the impulse response can be seen in figure 7.1 along with a linear regression line used for estimating the reverberation time. In figure 7.2, we can see the reverberation time measurement estimate for all three devices in the same conference room. Finally, figure 7.3 shows the mean reverberation time for all three rooms.
Figure 7.1: Example of one 1/3-octave band impulse response

Figure 7.2: Measured reverberation time in the conference room
7.2 Measured Distortion

The distortion estimated from the exponential sine sweep measurement can be found in figure 7.4. The mean and maximum THD is summarized in table 7.1. Note that the distortion is high for the ‘Loose element’ device, especially in the beginning of the sweep, which corresponds to lower frequencies (the sweep goes from low to high frequency). The reference device showed a high peak in distortion corresponding to a fundamental tone of about 500 Hz. This distortion in the unmodified device was not apparent when listening to normal material such as speech.
The responses from multitone measurements are shown in figure 7.5. The distortion is visibly higher for the 'Loose element' unit, and the linear response is also weaker for high frequencies. The two other units are difficult to tell apart from these measurements.

The response from the silence sweep measurement is shown in figure 7.6. The supposedly silent part is dominated by distortion, as can be seen on the red line, way above the yellow background noise. The distortion is almost on the same level as the linear response. As in the multitone measurement, we can see that the 'Loose element' unit has a lower linear response in the high frequencies, but otherwise it is very difficult to distinguish the different units.
Figure 7.5: Response from a multitone measurement. The black lines show the linear response to the stimulus, red line is distortion and noise, blue line is noise.
Figure 7.6: Estimated distortion from a silence sweep measurement
8 Conclusion

The results in chapter 7 indicate that it is possible to estimate both room impulse response and loudspeaker distortion from a single measurement, using the exponential sine sweep method. This could be used in speakerphone applications to assess if the room is unsuitable, or if the device is faulty.

The exponential sine sweep method was initially only supposed to be used for estimating the impulse response but proved to be the most effective method for analyzing the distortion.

Both the multitone and silent sweep methods proved difficult to extract quantitative values of the distortion. The measurements should preferably be simple to transform into a scalar value such as THD or similar to be able to easily assign criterion for pass/fail.

8.1 Further Work

With further testing on more devices, threshold limits for pass/fail could be set up. This would enable fully automatic testing. The next step after that would be to incorporate the test signal and analysis algorithm into a speaker phone to enable it to perform the tests without connecting to a computer.

Speaker phones often incorporate several directed microphones. In the normal case, the phone mixes the microphone signals and sends out the signal which best picks up the person talking. The units used in these tests have four microphones but the automatic mixing has been bypassed and only one microphone signal has been used. A significant improvement would be to record all four microphone signals simultaneously. This could reduce noise and give better estimates. The internal software on the phones did not allow sending several channels simultaneously without major modifications, which were out of scope for this work.
9 References


