MASTER

Quantization of sinusoidal parameters of audio and speech

Goossens, T.L.J.

Award date:
2004

Link to publication
Quantization of Sinusoidal Parameters of Audio and Speech

by T.L.J. Goossens

Master of Science thesis

Project period: july-2003 – april 2004
Report Number: 07-04

Commissioned by:

Supervisors:
Dr.ir. S.L.J.D.E. van de Par (Philips Research)
Dr.ir. L.L.M. Vogten (TU/e)

Additional Commission members:
Prof.dr.ir. J.W.M. Bergmans (TU/e)
Prof.dr. A.G. Kohlrausch (Philips Research / TU/e)
TECHNISCHE UNIVERSITEIT EINDHOVEN
Department of Electrical Engineering

MASTER'S THESIS
Quantization of Sinusoidal Parameters of Audio and Speech Signals

by

T.L.J. Goossens

Supervisors
: dr. ir. S.L.J.D.E. van de Par (Philips Research)
: prof. dr. A.G. Kohlrausch (Philips Research / TU/e)
: prof. dr. ir. J.W.M. Bergmans (TU/e)
: dr. ir. L.L.M. Vogten (TU/e)

Eindhoven, 12th April 2004
"Now what is "music" — a sequence of vibrations in the air, or a succession of emotional responses in a brain? It is both."

Douglas R. Hofstadter in Gödel, Escher, Bach: An Eternal Golden Braid, 1979
Summary

In the context of a graduate research project executed at the Philips Research Laboratories in Eindhoven, this thesis aims to give an answer to the question: How can a frequency-dependent quantization of sinusoidal parameters in audio and speech be implemented in such a way that the least amount of bits is used without reducing audio quality?

In this thesis a frequency dependent quantization method for sinusoidal parameters is presented of which the design is based on the results of psychoacoustical experiments. These experiments include threshold measurements of quantization detection for frequency, amplitude and phase parameters using pure tones in the range of 25 Hz up to and including 10 kHz.

The frequency dependent quantization algorithm is subjectively evaluated and shows a significant improvement over the existing quantization method as present in a sinusoidal coder named SiCAS [4] which has been developed at Philips Research.

A second experiment was conducted with single-formant Klatt shaped harmonic tone complexes with a fundamental frequency of 100 Hz and a formant center frequency in the range of 200 Hz up to and including 10 kHz. These experiments indicate that additional increases in coding efficiency may be obtained by creating a frequency dependent quantization that in addition is signal dependent.
Preface

In front of you lies the result of my nine months graduation research that I have conducted at Philips Natlab in order to receive my Master of Science degree in Electrical Engineering from the University of Technology in Eindhoven (TU/e). The project has been initiated by the Auditory and Multi-Sensory Perception cluster in the Digital Signal Processing (DSP) group of Philips Natlab and was accepted under the responsibility of the Signal Processing Systems group in the sub-department Measurement and Control Systems at the TU/e.

The reason for me to decide to do this particular project is my passion for music and thus for sound. The project stood out to me because it allowed me to not only work with sound but with the perception of it as well. It turned out to be a good decision for I have really enjoyed my time at the Natlab.

This enjoyment was on the one hand due to the assignment that has made me go through the different steps in a scientific research project. At first, reading and trying to get a hold of the matter. Then devising and conducting the experiments. Subsequently analyzing the results, trying to understand them to be able to use the results for implementation in a new algorithm. And finally to evaluate the implementation and to discover whether or not all the efforts have been worthwhile.

On the other hand the enjoyment was due to the people involved in this with me. Therefore, in this preface, I would like to express my gratitude to my supervisors dr. ir. Steven van de Par and prof. dr. Armin Kohlrausch at the Natlab and to dr. ir. Leo Vogten and prof. dr. ir. Jan Bergmans at the TU/e for guiding and supporting me on my first scientific steps of which hopefully a lot more will follow. I also would like to thank my colleagues of the DSP group, especially the ones that have been so kind to ‘put their ears at stake’ when participating in my listening experiments. And last, but you know how they say not least, thanks to the students in the ‘students room’ making my time at the Natlab such a fun time (listening to Dub on the friday afternoon and walking in large processions towards the coffee machine) which I will enjoy looking back to for a long time.
# Contents

Summary i
Preface ii

1. Introduction 1
   1.1. Sinusoidal coding 2
   1.2. Sinusoidal analysis and synthesis 3
   1.3. Research question 7
   1.4. Approach 7
   1.5. Outline of this thesis 8

2. Pure Tone Listening Test 9
   2.1. Parameter Quantization 11
   2.2. Method 12
      2.2.1. Stimuli and Conditions 12
      2.2.2. Procedure and subjects 13
      2.2.3. Setup 14
   2.3. Experimental Results 15
      2.3.1. Frequency Quantization of Pure Tones 15
      2.3.2. Amplitude Quantization of Pure Tones 15
      2.3.3. Phase Quantization of Pure Tones 18
   2.4. Analysis 18
      2.4.1. Frequency Quantization of a single sinusoid 20
      2.4.2. Amplitude Quantization of a single sinusoid 21
      2.4.3. Phase Quantization of a single sinusoid 23
   2.5. Discussion 24

3. Tone Complex Listening Test 27
   3.1. Method 28
      3.1.1. Stimuli and Conditions 28
      3.1.2. Procedure and subjects 29
      3.1.3. Setup 30
   3.2. Experimental Results 30
      3.2.1. Frequency Quantization of Complex Tones 32
      3.2.2. Amplitude Quantization of Complex Tones 32
3.2.3. Phase Quantization of Complex Tones ............................................ 32
3.3. Discussion .................................................................................. 32

4. Quantization Scheme ..................................................................... 35
   4.1. Implementation ....................................................................... 35
   4.2. Evaluation ............................................................................. 36
       4.2.1. Method ....................................................................... 36
       4.2.2. Results ...................................................................... 37
   4.3. Discussion ............................................................................. 38

5. Conclusions and Recommendations ............................................... 41
   5.1. Conclusions .......................................................................... 41
   5.2. Recommendations ................................................................. 41

Bibliography .................................................................................... 43

A. Appendix A - Mathematical Analysis ............................................ 45
   A.1. Frequency Quantization of a single sinusoid ........................... 45
   A.2. Amplitude Quantization of a single sinusoid .......................... 46
   A.3. Phase Quantization of a single sinusoid ................................ 47

B. Appendix B - Sound excerpt descriptions .................................... 49

C. Appendix C - List of Abbreviations ................................................ 51
1. Introduction

Audio coding algorithms are omnipresent in modern day audio applications, and their utilization is still increasing. Bandwidth and storage limitations combined with high quality expectations of users, call for better audio quality obtained with lower bit-rates and of course with as little computational complexity (read: costs) as possible. Therefore a lot of research effort has been put in developing audio coders that try to meet these demands. These efforts have resulted in different classes of coding algorithms [12] amongst which:

- Subband coding;
- Transform coding;
- Linear prediction based coding;
- Parametric coding.

The first three classes are so-called waveform coders. The first class, subband coding, is a coding technique where the audio signal is filtered into a number of different subbands, hence the name. For every individual subband the permitted noise-floor is determined on the basis of a perceptual model. By specifying the number of bits to be spent for each particular band, the noise-floor can be spectrally shaped in such a way that the noise will always be masked by the audio signal once the signal is reconstructed using a synthesis filter-bank. Transform coders do basically the same thing with the difference that they do (part) of the analysis/synthesis in another domain than the time domain (e.g. frequency domain).

Whereas Subband- and Transform coding algorithms code the received waveform, Linear Prediction based Coding (LPC) algorithms in general use a source model that aims to reconstruct the acoustical signal by modelling an excitation signal and the acoustical channel that shapes its spectrum.

Parametric audio coding is a technique where the audio signal is modelled with a number of auditory objects each of which require certain parameters to specify the content of the object. Examples of such objects are sinusoids, transients and noise. In this way a very low bit-rate representation of audio signals can be obtained, much lower than for example with waveform coding based techniques like the popular MP3 [6].

Sinusoidal coding, the topic on which this thesis focuses, is a parametric coding technique that uses sinusoids as auditory objects. This technique was, until recently, mostly used for speech coding and therefore relatively little research was reported on these kinds of coders in the context of audio coding. But research on this topic in an audio coding context is intensifying
1.1. Sinusoidal coding

because the waveform coders are approaching their fundamental limits in coding efficiency and new paradigms need to be explored to obtain even higher compression rates.

In this work we focus on the parameters defining the sinusoids in a sinusoidal coding scheme i.e. frequency, amplitude and phase parameters. These need to be quantized to allow useful bitstream encoding. This has to be done in such a way that quantization steps are as large as possible to minimize the encoded bit-rate, but without audible loss in quality. This research aims to find an answer to the question of how sinusoidal parameters should be quantized to fulfill the requirement stated before.

1.1. Sinusoidal coding

Figure 1.1 is a schematic overview of a sinusoidal codec. The left side of this figure represents the encoding section and the right side represents the decoding section of the codec. The information sent from the encoder to the decoder is contained in a bitstream. The purpose of the codec is to make the bitrate of this bitstream as low as possible. The encoding as well as the decoding section can again be subdivided into a section that is responsible for the sinusoidal coding task (the upper part) and a section that codes the residual signal (the lower part). At Philips Research Laboratories, a sinusoidal coder called SiCAS (Sinusoidal Coding of Audio and Speech) [4] with a similar architecture is available. Later on in this thesis it will be used to evaluate a renewed quantization method.

In the sine extraction block, parametric descriptions of the perceptually most relevant sinusoids in waveform $x$ are typically obtained every 20 ms. 20 ms is the time-interval during which the audio signal is usually quasi stationary so it is fairly safe to treat it as such when parameterizing. A typical number of sinusoidal components to extract from such a 20 ms section is 30 components. The extracted sinusoids are parameterized by describing the frequency, amplitude and phase of each sinusoid.

After subtracting these sinusoidal components from the original waveform, a noiselike residue $r$ is obtained. This residue is then passed on to the residual coder which encodes it into a useful representation before being transmitted to the decoder. This representation can be anything but is frequently a waveform coder, a codebook or a parametric noise description.

The sinusoidal parameters need to be quantized in order to reduce the amount of information that is being sent to the encoder. This is done in the quantizer (Q in the codec scheme) right after the sinusoid extraction.

Finally, before being transmitted, the sinusoidal parameters are linked and entropy coded in the parameter encoding block. Linking is an operation that exploits correlation between sinusoids across time and/or frequency. This correlation stems from the fact that in a physical signal it is very common for a tonal component to be present for far longer than one frame length. This is for instance the case in a constant tone or even a slowly varying tone in an instrument or in speech. It is therefore often advantageous to link the sinusoidal components in subsequent frames and only code the relative parameter distances instead of coding the absolute values. After linking, lossless entropy encoding is performed on the parameters to remove redundancies in the final bitstream e.g. by using Huffman tables.

After quantization, linking and entropy coding, sinusoidal and residual parameters are trans-
mittmed to the decoding section. The transmission of the information contained in these parameters can be realized in storage media like CD-rom, solid state memory, hard disc, etc. or by direct transmission methods like GSM, copper wire or any other means.

The decoding section now reconstructs both the sinusoidal and the residual part of the signal. The sum of these parts should be the perceptual equivalent of the original. Note the word should, for perceptual equivalency is in this case dependent on factors like the number of sinusoids used, the residual representation, quantization of the parameters and the perceptual model used for sinusoidal extraction. This investigation is mainly concerned with sinusoidal quantization and will not elaborate on the residual coding part of the codec.

1.2. Sinusoidal analysis and synthesis

One way to perform the analysis of the signal is to use an algorithm called matching pursuit [10] which decomposes a given signal into a linear expansion of sinusoids. The parameters of these sinusoids are iteratively selected such that for each iteration the residual energy in the signal is reduced with the largest amount possible. Another analysis algorithm is the psychoacoustic matching pursuit [5] which performs the analysis in a similar way but selects the perceptually relevant sinusoids by reducing the perceptual distortion instead of the residual energy for each iteration. This perceptual distortion is measured using a perceptual model developed by van de Par et al. [13].

At the decoder, synthesis of sinusoids is in general realized using an overlap-add algorithm. This algorithm combines the synthesized segments by applying a windowing function and adding the windowed segments with the appropriate overlap. A typical window that is used for this procedure is the Hanning window [2] which is amplitude complementary when added with 50% overlap.

To formally describe the audio signal using sinusoidal parameters we will first define the notations given in table 1.1.

Figure 1.1.: Decomposition of a sinusoidal codec.
1.2. Sinusoidal analysis and synthesis

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>time</td>
</tr>
<tr>
<td>$h_k(t)$</td>
<td>Hanning window of segment $k$</td>
</tr>
<tr>
<td>$x(t)$</td>
<td>audio signal</td>
</tr>
<tr>
<td>$r(t)$</td>
<td>residual signal</td>
</tr>
<tr>
<td>$\omega_h$</td>
<td>angular frequency of the Hanning window</td>
</tr>
<tr>
<td>$f_{k,i}$</td>
<td>frequency parameter of sinusoidal component $i$ in segment $k$</td>
</tr>
<tr>
<td>$a_{k,i}$</td>
<td>amplitude parameter of sinusoidal component $i$ in segment $k$</td>
</tr>
<tr>
<td>$p_{k,i}$</td>
<td>phase parameter of sinusoidal component $i$ in segment $k$</td>
</tr>
<tr>
<td>$ns$</td>
<td>number of segments</td>
</tr>
<tr>
<td>$nc$</td>
<td>number of components</td>
</tr>
</tbody>
</table>

Table 1.1.: Notations.

The definition of the Hanning window associated with segment $k$ is:

$$h_k(t) = \begin{cases} \frac{1}{2} \left( 1 - \cos \left( \omega_h \left( t - \frac{\pi k}{\omega_h} \right) \right) \right) & \text{when } \frac{\pi(k-1)}{\omega_h} \leq t \leq \frac{\pi(k+1)}{\omega_h}; \\ 0 & \text{elsewhere}, \end{cases} \quad (1.1)$$

The segment index $k$ is used to time-shift the Hanning window such that subsequent segments have 50% overlap. This time shift is half of the segment length (i.e. $\frac{2\pi}{\omega_h}$) multiplied by the segment index $k$. The parametric representation of the original audio signal is:

$$x(t) = \sum_{k=0}^{ns-1} \sum_{i=0}^{nc-1} a_{k,i} \cos \left( 2\pi f_{k,i}(t - \frac{\pi k}{\omega_h}) + p_{k,i} \right) \cdot h_k(t) + r(t) \quad (1.2)$$

where $r(t)$ is the residual part of the signal that is not contained in the sinusoidal part of the representation. Note that the sinusoidal components in segment $k$ are time-shifted with the same amount as the Hanning window of segment $k$.

Let us study the effects of quantization by considering the synthesis described above. In Figure 1.2 the Hanning window is compared with rectangular windowing. One can see that the spectra of the sinusoids multiplied by a Hanning window (see Figure 1.3) have a spectral decay that is much steeper and more persistent than the spectral decay of the sinusoids multiplied by a rectangular window. In addition, the spectral shape of the Hanning windowed sinusoid is more comparable to the shape of the human auditory filters which also makes it more suitable for analysis/synthesis purposes.

Moore [11] describes the auditory filter as: “One of an array of bandpass filters that are assumed to exist in the peripheral human auditory system” and explains masking as “When trying to detect a signal in a noise background, the listener is assumed to make use of a filter with a centre frequency close to that of the signal. This filter passes the signal but removes a great deal of the noise. Only the components in the noise which pass through the filter have any effect in masking the signal.”

The thick lines in Figure 1.3 indicate approximations of the shape of the auditory filters as modelled by a 4th order gamma tone filter [13], a commonly used model in psychoacoustics.
Confidential

Chapter 1. Introduction

Auditory filters are given at 100, 250 and 1000 Hz in panels a, b and c respectively. The figure shows that the auditory filters at the low frequencies have a spectral resolution comparable to that of the hanning window. As a consequence the side lobes may be resolved in auditory filters adjacent to the filter centered at the main lobe. In that case the side lobes may be detected by the auditory system. At high frequencies where auditory filters are wider, adjacent filters will not resolve the side-lobes independently from the main lobe and as a consequence the main lobe will effectively mask the side-lobes. Therefore it is expected that the side-lobes may play a role at frequencies below 1 kHz where the spectral resolution of the human auditory system is high enough.

Adding Hanning windowed sinusoids with a 50% overlap results in a flat temporal envelope, as shown in Figure 1.4, provided that the phases of the sinusoids are equal at the transition from one window to another. Adding an infinite sequence of Hanning windowed sinusoids of the same frequency with 50% overlap and starting phases that are properly fitted to one another
1.2. Sinusoidal analysis and synthesis

results in an infinitely long sinusoid with no envelope modulation. In the frequency domain the spectral side-lobes of all the Hanning windowed sinusoids are cancelled due to the addition and time shifting in the temporal domain, and only a dirac pulse remains at the sinusoidal frequency. This demonstrates that under ideal circumstances, without quantization, spectral splatter resulting from the windowing operation can be removed by the overlap add procedure.

![Figure 1.4: Overlap add using Hanning windows. For illustration the Hanning windows in this figure are not multiplied with sinusoidal components.](image)

The sinusoidal parameters resulting from sinusoidal analysis, are at first floating point representations of amplitude, frequency and phase. To store or transmit these parameters as such, would not result in the information reduction desired in an audio coding context. Therefore, as stated in the beginning of this chapter, the parameters describing the sinusoids need to be quantized to allow efficient bitstream encoding. Quantization reduces the amount of information in the coded signal. This is the desired effect, but it also reduces the accuracy of the parameters describing the audio signal, which can in some cases result in audible artifacts. Think of the infinite sinusoid for which the phases had to be properly fitted to each other in order to cancel the spectral side-lobes caused by the Hanning windows. By reducing the accuracy of the phase parameters we also reduce the effectiveness of the spectral cancellation and, as a result, introduce modulations and spectral splatter. Therefore we need to find out how quantization can be done such that quantization steps are as large as possible without introducing audible artifacts.

Quantization of frequency, amplitude and phase can cause two different types of auditory cues, temporal modulations and spectral side-lobes. In fact, modulations and side-lobes are two interpretations of the same phenomenon. In psychoacoustics, side-lobes are usually referred to when the spectral splatter lies in other auditory filters than the filter centered at the carrier frequency. When the spectral splatter predominantly lies within one auditory filter, this is referred to as being a modulation. The distinction between side-lobes and modulations is made in psychoacoustics because the detection of these auditory cues are fundamentally different.

Side-lobes are a low frequency problem because this is where the auditory filters are more
narrow and side-lobes tend to be in different auditory filters. Modulations in frequency and amplitude on the other hand are expected to be auditory cues in the higher frequencies. Hence, already before doing experiments we can expect some kind of frequency dependency in the sensitivity to sinusoidal parameter quantization due to the presence of two different types of cues at high and low frequencies.

1.3. Research question

Given the expectation that different auditory cues play a role at low and high frequencies, we want to conduct listening experiments to obtain information on how these cues manifest themselves in sinusoidal parameter quantization.

The central question is: how can a frequency dependent quantization of sinusoidal parameters in audio and speech be implemented in such a way that the least amount of bits is used without reducing audio quality? Stated otherwise: What is the most efficient quantization of sinusoidal parameters of audio and speech signals?

In order to find an answer to this question we have raised two subquestions:

1. What are the detection thresholds of quantization steps for frequency, amplitude and phase parameters as a function of frequency in tonal and harmonic signals obtained with listening tests?

2. How can the threshold data from the listening tests be incorporated in a newly designed quantization scheme for a sinusoidal coder that is available at Philips Research?

1.4. Approach

A first step in implementing an appropriate quantization algorithm could be to take data from psychoacoustic literature about just noticeable differences (JNDs) in frequency and amplitude and incorporate them in the quantization procedure. These JNDs are determined by comparing one tone with another tone that has a small difference in one of the parameters. The smallest difference that can still be detected by a test-subject is the just noticeable difference (e.g. a 1000 Hz pure tone can just be discriminated from a 1003 Hz pure tone at 80 dB SPL in which the 0 dB reference intensity is $2 \cdot 10^{-5} N/m^2$). However, quantization in a sinusoidal coder is not performed on sinusoids but on parameters that describe sinusoids in an overlap add scheme. This leads to different types of artifacts as we will discuss in more detail in Chapter 2.

To really determine when quantization becomes audible the approach in this thesis is to use psychoacoustic listening tests to obtain quantization limits for sinusoidal parameters. There is a vast amount of conceivable paradigms for such a listening test. This thesis focuses on two of them which are believed to cover the most relevant and critical situations that may occur in natural audio signals. In addition, both listening tests use stimuli that are generated in a similar way as a signal could be synthesized in a decoder.

The first listening test uses single sinusoids or pure tones at various frequencies to test the frequency dependency of the quantization limits. Pure tones are used for two reasons. One reason is that pure tones are critical because side-lobes produced by quantization errors are not
1.5. Outline of this thesis

The outline of the rest of this thesis follows the chronological order in which the research was conducted. In Chapter 2 the method, results and analysis of the pure tone experiments are described. Chapter 3 focusses on the method and results of the tone complex experiments. Chapter 4 describes the implementation and subjective evaluation of a new quantization algorithm based on the listening experiments. Finally, in Chapter 5 conclusions are drawn and recommendations for future work are given.
2. Pure Tone Listening Test

As said in the introduction, the first listening test is based on pure tones to test the frequency dependency for the quantization limits for reasons of simplicity. Simplicity allows us to analyse the stimuli analytically which again enables us to interpret the experimental results and to compare them to literature data. The other reason to use pure tones is that a tone with quantized parameters can behave like a modulated carrier which can be distinguished very well from a non-modulated sinusoid. In addition, when the carrier is sinusoidal there will be less masking of the side-lobes and therefore a quantized sinusoid is expected to provide a critical test stimulus.

In order to produce the relevant types of artifacts the stimuli are synthesized in a similar way as would be the case in a sinusoidal codec.

![Stimulus generation diagram](image)

Figure 2.1.: Stimulus generation.

Figure 2.1 shows the outline of the scheme used to synthesize the stimuli. It starts with generating the sinusoidal parameters for each segment in the stimulus. Subsequently, these parameters are quantized. In fact, quantization errors are added to the parameters generated in the first block to simulate the quantization process. In the next section we will discuss what types of errors are added to the parameters. The quantized parameters are then used to synthesize a sinusoid for each segment. The stimulus generation is now completed by overlap adding the segments.

Generating the parameters for a single sinusoid is pretty straightforward for amplitude \(a\) and frequency \(f\) parameters. They remain constant during the entire stimulus. The parameter generation block simply assigns the correct frequency and amplitude to all the segments in the stimulus. The phase \(p\) however, does not remain constant during the stimulus except for the rare occasion that a discrete number of sinusoidal periods fit precisely in one segment.

The phase of a segment can be defined at the beginning of the segment, but may also be defined at other places in the segment provided that the encoder and decoder assume the same definition. We defined the phase parameter in the middle of the segment because this leads to the smallest phase error when subsequent segments have frequencies which are rounded off in opposite directions (as illustrated in Figure 2.2).
An alternative method to reduce phase errors between subsequent frames is phase prediction which is sometimes also called phase continuation. This method assigns the phase of the sinusoid in the next segment such that the phases of the sinusoids the current and the next segment are equal at the transition. Doing this results in little phase error at transition but may result in propagation of phase errors.

In general the sinusoidal parameters of a nearly stationary sinusoid can start jumping due to quantization in subsequent segments. Suppose a parameter in an encoded audio signal lies close to a round off decision border. Minor variations in the magnitude of this parameter can cause the parameter to switch between the higher and lower quantization levels.

In the parameter generation block the phases of the segments are calculated by adding the phase difference between the middles of the previous and the current frame to the phase of the previous frame \( \phi_k = \phi_{k-1} + \frac{1}{2} \omega c T h \), where \( \omega_c \) is the pure tone or carrier angular frequency, \( k \) is the segment number and \( T h = \frac{2 \pi}{\omega h} \) is the Hanning window time period). This way the phases of the sinusoids in the successive frames are exactly the same at the segment transition. Note that these are the original unquantized phases, in contrast to predicted phases which will be discussed later on.

![Figure 2.2.](image)

Figure 2.2.: (a) Two sinusoids with a different frequency. Their phases are the same at the beginning of the window and therefore the phase difference is zero at the start and large at the end of the window. (b) Here the phases are defined in the middle of the window. The largest phase difference occurs at both ends now but is only half the difference that is seen in plot a.

The parameters \( f, a \) and \( p \) coming from the first block are quantized in the second block. This is where the audible artifacts are caused. What types of artifacts can be expected will be discussed this in Chapter 2.1.

Wave synthesis takes the sinusoidal parameters of each segment and transforms these into the sinusoidal wave form of segment length. The segment is multiplied with a Hanning window of segment length (Equation 1.1) to facilitate smooth transition from one segment to another.

Finally, in order to obtain the entire stimulus, the Hanning windowed segments are added.
together in the overlap add block. It adds the subsequent segments together with a 50% overlap. If the parameters are unquantized, this will result in a pure tone, or sinusoid. If however these parameters are quantized, the resulting tone might be modulated or changed in frequency or amplitude.

2.1. Parameter Quantization

At this point we will define quantization of sinusoidal parameters as it is treated in this thesis.

**Definition 2.1 (Sinusoidal parameter quantization)** Sinusoidal parameter quantization is the process whereby the continuous range of values of the sinusoidal parameters is divided into nonoverlapping subranges that together cover all possible parameter values. For each of these subranges a discrete value is defined which typically lies within the subrange. Parameters falling within a given subrange will be assigned with the discrete value.

Suppose we have a sinusoid with an infinitely small amplitude modulation. This amplitude modulation causes the amplitude of the sinusoid to fluctuate, with an infinitely small amount, above and below a decision threshold in the quantization procedure. This causes the amplitude of the sinusoid to alternately fall within one of two adjacent subranges (mentioned in the definition). Therefore, parameters describing the amplitude of this sinusoid in subsequent segments will jump from the quantization level below the decision threshold ($\hat{a}^-$) to the quantization level above this decision threshold ($\hat{a}^+$) and back. This is shown in the upper part of Figure 2.3 where the abscissa shows the segment index and the ordinate shows the parameter value (not a waveform sample value). The curly solid line is the envelope of the amplitude modulated sinusoid. The points marked with a circle are the actual parameters after quantization. In this situation, which we will refer to as Alternating Quantization Error (AQE), the parameters alternate between two adjacent quantization levels.

Figure 2.4 shows what happens to a signal's envelope when subsequent segments have alternating amplitudes.

In the lower part of the figure a similar situation is given but now all amplitude parameters are rounded to the quantization level above the threshold. We will refer to this as Unidirectional Quantization Error (UQE). This would cause the amplitude of the stimulus to increase when quantizing the amplitude, or it would cause the starting phase of the stimulus to shift in the case of phase quantization. For frequency quantization the consequences are more complex because if no further precautions are taken, the starting phases of the segments will as a consequence no longer be correct. The segment phases can either be left unaltered or they can be recalculated (original phase and phase prediction respectively). Phase prediction will not cause any extra artifacts besides the shift in frequency in this pure tone situation. If the phases are left unaltered, spectral spatter will be introduced due to phase mismatches between consecutive segments.

There are countless conceivable variations on both previously mentioned artifacts. Stochastic distribution of the quantization error or unidirectionally rounding the parameters downwards to name but a few. Stochastic quantization is harder to analyse analytically than is the case for alternating errors. Unidirectionally rounding downward does not fundamentally differ from unidirectionally rounding upward. Therefore, in the experiment we will simulate the artifacts in
2.2. Method

2.2.1. Stimuli and Conditions

As the research question states, there is special interest in the spectral behavior of the quantization limits. A logarithmic-like frequency behavior can be expected because the ear also behaves in an, by approximation, logarithmic fashion. This is shown by Equation 2.1 \[3\].

\[
\text{Number of ERBs} = 21.4 \log_{10}(4.37f + 1), \tag{2.1}
\]

where \(f\) is frequency in kHz and \(\text{Number of ERBs}\) or \(\text{ERB rate}\) is a unit that relates frequency to the ear’s critical band number. The critical bandwidth is a measure for the bandwidth of the human auditory filter at a certain frequency and is therefore a measure for the frequency selectivity or resolution of the human ear at a certain frequency. The critical bandwidth is often expressed as the Equivalent Rectangular Bandwidth (ERB). Because of the logarithmic behavior of the ERB-rate, the conditions for the psychoacoustic tests are placed in the spectrum such that the observations are approximately equidistant on a logarithmic scale. The measurements are placed at 25, 50, 100, 250, 500, 1000, 4000 and 10000 Hz.

Measurements are done for frequency, amplitude and phase parameters using both AQE and UQE hypotheses to alter the parameters in the quantization procedure. Frequency measurements include conditions for original phase as well as phase prediction.
All stimuli were presented with a sound level of 70 dB SPL except for the 25 Hz stimuli which were presented at 80 dB SPL. The window length of the Hanning windows was 23.2 ms (equivalent to 1024 samples at a sample frequency of 44.1 kHz).

2.2.2. Procedure and subjects

The procedure used for the quantization threshold measurements is the Three Interval Forced Choice (3IFC) procedure. This is a commonly used method in psychoacoustics to determine threshold values.

In the 3IFC procedure the subject’s task is to identify the quantized stimulus interval out of three presented stimuli per trial where the other two stimulus intervals are unquantized. Since the position of the quantized interval is not known a priori by the subject this procedure prohibits the experimental results of being influenced by the subjects’ interpretations. The stimulus durations are 400 ms and the duration of the silence intervals between the three stimuli are 100 ms. After two subsequent correct identifications, the quantization error used for generating the quantized interval is decreased. Every single incorrect answer results in an increase of the quantization error. This is called two-down/one-up adaptive tracking [9].

Figure 2.5 shows a typical adaptive track for the measurement of a threshold value. The abscissa features the trial number. The ordinate features the quantization error size in dBs that is decreased or increased after each answer. This quantization error size can be $20 \log_{10} \frac{\Delta f}{f}$, $20 \log_{10} \frac{\Delta a}{a}$ or $20 \log_{10} \pi \Delta P$ depending on the parameter that is quantized. Observe that after changes in the direction of the track, called reversals, the amount with which the quantization error size changes is decreased with a factor two. This amount will continue to decrease until it reaches a minimum value of 2 dB. Once the minimum value is reached the measurement phase starts. From here on the next eight subsequent reversals, indicated with circles in the figure, count as measurements for which the median is calculated. This median, indicated with the dashed line in the figure, is the threshold value for which the subject correctly identifies the
quantized stimulus with a probability of 71%. Each threshold measurement was repeated three times.

![Graph](image)

Figure 2.5.: Two-down/one-up adaptive track (solid line) and median of the last 8 reversals (circular markers) of the adaptive track (dashed line).

The subjects received feedback after each single trial about the correctness of the given answer. Three subjects participated in this test. All three subjects had normal hearing and participated voluntarily (not payed) in the experiments. Their ages lie between 26 and 37 and they had experience in psychoacoustic measurements. One of the subjects was the author of this thesis.

Due to the extensiveness of the test we have omitted the UQE conditions in original phase paradigm for frequency quantization as well as amplitude quantization for one test subject. The motivation is that for frequency quantization these conditions did not show significantly different behavior from the AQE conditions in original phase paradigm measured by the other subjects. The motivation for omitting these particular conditions in the amplitude quantization measurements is that these limits were all less critical than AQE conditions for the other subjects and therefore do not contribute to knowledge about the worst case situation relevant for audio coding.

2.2.3. Setup

Stimuli were generated digitally, 16 bits per sample at a sampling frequency of 44.1 kHz, and presented to the subjects diotically through Beyerdynamic DT990 headphones in an acoustically isolated listening room.
2.3. Experimental Results

2.3.1. Frequency Quantization of Pure Tones

Figure 2.6 shows the experimental results for the frequency quantization measurements of the pure tone test for all three test subjects. The abscissa shows the pure tone frequency $f$ and the ordinate shows $\Delta f$ where $\Delta f$ is the size of the quantization error in Hz and $f$ is the frequency of the unquantized tone also in Hz. All conditions were measured three times. The markers in the figures are placed at the mean of the three measurements and the error bars indicate the standard deviations. The trends of the data for the test subjects are consistent. Subject SP and TG have both measured some extra conditions, primarily to see if the trend at these data points was convex or concave. These extra conditions are original phase measurements at 35, 75 and 150 Hz. When interpreting the figures the reader can keep in mind that doubling the threshold at a certain frequency yield a reduction of roughly one bit needed to store parameters at that frequency. This is indicated in the figures with the horizontal grid-lines that lie at distances of doubled thresholds (6 dB distances). Higher thresholds cost less in terms of bandwidth/storage than lower thresholds.

The solid lines show the thresholds for the original phase measurements. In these measurements there were no significant differences in thresholds for AQE (solid line, triangular markers) and UQE (solid line, square markers) condition. Both trends start at 25 Hz with a value of roughly 0.02 climbing up to 0.05 at 50 Hz. Subsequently the curves drop to 0.005 at 250 Hz and from there on they fluctuate around 0.0025. At 10 kHz, subjects tend to be more sensitive to these quantization artifacts and the curves go down some more.

A similar trend is to be seen in the phase prediction measurements for the AQE conditions (dashed line, triangular markers), only the thresholds lie somewhat higher. Notice the anomaly in the thresholds for the first test subject (OS), the 10 kHz threshold does not go downwards but goes upward instead.

Finally the dashed line with the square markers shows the UQE conditions in the phase prediction paradigm. The curve starts at approximately 0.07 at 25 Hz to gradually descend to a value of 0.0025 at 1 kHz. From there on it starts climbing again to reach a value of about 0.015 at 10 kHz.

2.3.2. Amplitude Quantization of Pure Tones

The results from the amplitude quantization measurements of the pure tone test are shown in Figure 2.7. Again, the abscissa shows the pure tone frequency. The ordinate now shows $\Delta a$ where $\Delta a$ is the size of the quantization error proportional to sound pressure in $N/m^2$ and $a$ is the amplitude of the quantized tone also proportional to sound pressure. All conditions were measured three times. The markers in the figures are placed at the mean of the three measurements and the error bars indicate the standard deviation. Doubling the threshold at a certain frequency still yields a reduction of roughly one bit needed to store parameters at that frequency. This is again indicated with the horizontal grid-lines.

Phase prediction and original phase do not influence the amplitude quantization measurements because frequency as well as phase are left unaltered and therefore no separate measurements were done for both conditions. The curves in Figure 2.7 indicate UQE and AQE...
2.3. Experimental Results

Figure 2.6.: (a), (b) and (c) Experimental results of the pure tone frequency quantization threshold measurements for all three test subjects. (d) Cross subject means. Triangular markers: Alternating Quantization Error (AQE). Square markers: Unidirectional Quantization Error (UQE). Solid lines: original phase. Dash-dotted lines: phase prediction. The solid horizontal lines indicate the SiCAS quantization.
Figure 2.7.: (a), (b) and (c) Experimental results of the pure tone amplitude quantization measurements for all three test subjects. (d) Cross subject means. Triangular markers: Alternating Quantization Error (AQE). Square markers: Unidirectional Quantization Error (UQE). The solid horizontal lines indicate the SiCAS quantization. The curve with circular markers in figure (c) indicates extra conditions with a window length of 5.8 ms in contrast to the normal window length of 23.2 ms.
2.4. Analysis
measurements, shown by square and triangular markers respectively. The AQE curve starts at approximately 0.01 at 25 Hz climbs to a value of about 0.1 at 100 Hz. After this peak the curve descends to roughly 0.045 at 250 Hz. From there it shallowly descends to about 0.035 at a frequency of 10 kHz. The UQE curve with the square markers is the equivalent of an amplitude JND. At 25 Hz this curve has a value of about 0.3 dropping to 0.25 at 50 Hz. From there on it drops to about 0.05 at 4 kHz. At 10 kHz it slightly rises again to about 0.06.

The curve marked with the circles is an extra set of measurements only executed by one subject (TG). The conditions are precisely the same as the conditions for the curve with the square markers with the only difference being that the window length was made 4 times as short. The window length used in these measurements was 5.8 ms. One can see that this curve is significantly different for frequencies up to 1 kHz.

2.3.3. Phase Quantization of Pure Tones

In the phase quantization measurements no UQE and no phase prediction conditions were measured. UQE would only result in a different starting phase for the target stimulus and this is not audible for diotic pure tones. Phase quantization is not applicable to a phase prediction situation. The only conditions left are the AQE conditions in the original phase paradigm. The results of the measurements are shown in Figure 2.8 and again the abscissa shows the pure tone frequency. The ordinate shows \( \pi \Delta \phi \) where \( \Delta \phi \) is the size of the quantization error in radians. All conditions were measured three times. The markers in the figures are placed at the mean of the measurements and the error bars indicate the standard deviation. Also in these plots, doubling the threshold at a certain frequency yields a reduction of roughly one bit needed to store parameters at that frequency.

At 25 Hz there is a relatively large deviation between the subjects. The lowest threshold at 25 Hz is 0.003 and the highest threshold is 0.01. The curve climbs to a value of approximately 0.03 at 100 Hz then drops to about 0.014 at 250 Hz. From there on it starts climbing to about 0.1 at 10 kHz.

2.4. Analysis

In this chapter we derive analytical expressions of the amplitude and frequency modulation for AQE stimuli with original phase parameters. The amplitude and frequency modulations at the audibility threshold for quantization that are present within stimuli, are determined using these equations to be able to compare them to literature data concerning amplitude and frequency modulation detection thresholds.

The general approach is to formulate the subsequent Hanning windows \( h_1 \) and \( h_2 \) which are multiplied with carrier \( c_1 \) and \( c_2 \) respectively. Carrier \( c_1 \) and \( c_2 \) are quantized versions of the original pure tone \( c \) in Equation 2.5. Their frequency, phase and amplitude properties depend on the quantization paradigm and are formulated accordingly in the subsections below. Equation 2.2 and 2.3 formulate the Hanning windows. Equation 2.4 follows trivially from Equation 2.2
Figure 2.8.: (a),(b) and (c) Experimental results of the pure tone phase quantization measurements (AQE) for all three test subjects. (d) Cross subject means. The solid horizontal lines indicate the SiCAS quantization.
2.4. Analysis

and 2.3.

\[ h_1(t) = \frac{1}{2}(1 + \cos(\omega_h t)) \quad (2.2) \]

\[ h_2(t) = \frac{1}{2}(1 - \cos(\omega_h t)) \quad (2.3) \]

\[ h_1(t) + h_2(t) = 1 \quad (2.4) \]

\[ c(t) = ae^{i(\omega_c t)} \quad (2.5) \]

where \( \omega_h \) equals \( 2\pi \) over the time period of the Hanning window (in practice \( \omega_h = 271 \text{ rad} \cdot \text{s}^{-1} \)), \( t \) equals time and \( \omega_c \) is the carrier or pure tone angular frequency.

For convenience the problem is stated in terms of complex functions. In the final solutions we are only interested in the real part. Note that we neglect the negative frequency part of the carrier. This means that spectral components which could normally be introduced into the positive part of the spectrum by multiplication with the negative frequency part of the carrier will not be present in the calculated signal. We assume that this will only affect our predictions at low frequencies where modulations are not expected to be important anyway.

2.4.1. Frequency Quantization of a single sinusoid

Equations 2.6 and 2.7 formulate the quantized sinusoids in two subsequent segments where \( \Delta \omega \) is the quantization error which is alternately added and subtracted from the angular carrier frequency \( \omega_c \).

\[ c_{f,1}(t) = ae^{i(\omega_c t + \Delta \omega t)} \quad (2.6) \]

\[ c_{f,2}(t) = ae^{i(\omega_c t - \Delta \omega (t + \frac{\pi}{\omega_h})} \quad (2.7) \]

\[ 0 \leq t \leq \frac{\pi}{\omega_h} \quad (2.8) \]

The term \( \frac{\pi}{\omega_h} \) in Equation 2.7 is a time shift that accounts for the fact that the phase of a sinusoid is defined in the middle of the segment. The stimulus is now constructed using the following equation:

\[ s_f(t) = h_1(t)c_{f,1}(t) + h_2(t)c_{f,2}(t). \quad (2.9) \]

This equation can be transformed into Equation 2.10. See appendix A.1 for this transformation.

\[ s_f(t) = c_f(t)A_f(t)p_f(t), \quad (2.10) \]

in which \( c_f \) (Equation 2.11) is the sinusoidal carrier with a starting phase that is dependent on the magnitude of the quantization error. The stimulus is a mixed modulated sinusoid. \( A_f \) is the stimulus’ amplitude modulation (Equation 2.12) and \( p_f \) is its phase modulation (Equation 2.13).

\[ c_f(t) = ae^{i(\omega_c t + \frac{\Delta \omega \pi}{2\omega_h})} \quad (2.11) \]

\[ A_f(t) = \sqrt{\cos^2(\Delta \omega t - \frac{\Delta \omega \pi}{2\omega_h}) + \cos^2 \omega_h t \sin^2(\Delta \omega t - \frac{\Delta \omega \pi}{2\omega_h})} \quad (2.12) \]

\[ p_f(t) = e^{i \arctan(\cos \omega_h t \tan(\Delta \omega t - \frac{\Delta \omega \pi}{2\omega_h}))} \quad (2.13) \]
Confidential

Chapter 2. Pure Tone Listening Test

The frequency modulation can now be determined by taking the first derivative of the exponent of \( p_f \).

\[
 f_f(t) = \frac{-\omega_h \tan \left( \Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h} \right) \sin \omega_h t + \Delta \omega \cos \omega_h t \left( 1 + \tan^2 \left( \Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h} \right) \right)}{1 + \cos^2 \omega_h t \tan^2 \left( \Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h} \right)}, \tag{2.14}
\]

Both the amplitude modulation \( A_f \) as well as the frequency modulation \( f_f \) in which we are interested, are dependent on the magnitude of the quantization error \( \Delta \omega \) and duration of the Hanning window expressed in its angular frequency \( \omega_h \). Note that neither the amplitude nor the frequency modulation is sinusoidal. The periodicity of the frequency modulation corresponds to 43 Hz and for amplitude modulation to 86 Hz.

Substituting threshold values from the psycho-acoustic experiments in Equation 2.12 and 2.14 gives the amplitude and frequency modulation as a function of time. The standard deviations or effective modulation depths are plotted in Figure 2.9b together with literature data of amplitude [8] (modulation frequency \( f_m = 50 \) Hz) and frequency [15] (\( f_m = 10 \) Hz) modulation detection thresholds. The standard deviations and not the peak-to-peak values of amplitude modulation (\( \Delta A \)) and frequency modulation (\( \beta \)) are plotted because the modulations are not necessarily sinusoidal. The cross subject means of the experimental data are shown in Figure 2.9a. The cross subject standard deviations are indicated with the error bars.

The target stimuli in the phase prediction UQE measurements (dash-dotted line, square markers) are only shifted in frequency and do not contain side-lobes. Therefore, this curve can be compared to frequency JND measurements. Comparable JND’s measured by Wier et al. [17] are plotted with dotted lines in Figure 2.9a. The shape of these curves are similar, but the data of Wier et al. are somewhat lower. The overall lower thresholds may be explained by the fact subjects in our test were presented stimuli with rather complex modulations while the subjects in Wier et al.’s tests could focus on frequency modulation JND alone.

The frequency modulations in Figure 2.9b calculated from the experimental data (dashed line, circular markers) coincide with the literature data from Sek and Moore [15]. It appeared that the subjects participating in our test were less sensitive to amplitude modulations (solid lines) at 4 and 10 kHz than literature data of Kohlrausch et al. [8]. However, note that cross subject standard deviations are fairly large at 10 kHz.

2.4.2. Amplitude Quantization of a single sinusoid

Modulations for amplitude quantization AQEs are found in a similar way. Equation 2.15 and 2.16 formulate the quantized sinusoids in subsequent segments where \( \Delta a \) is the amplitude quantization error which is alternately added and subtracted from the carrier amplitude \( a \).

\[
 c_{a,1}(t) = (a + \Delta a)e^{i \omega t}, \tag{2.15}
\]

\[
 c_{a,2}(t) = (a - \Delta a)e^{i \omega t}, \tag{2.16}
\]

\[
 t \in \mathbb{R}. \tag{2.17}
\]

The stimulus is now constructed using the following equation:

\[
 s_a(t) = h_1(t)c_{a,1}(t) + h_2(t)c_{a,2}(t). \tag{2.18}
\]
2.4. Analysis

Confidential

Figure 2.9.: (a) Frequency quantization: cross subject means of original phase AQE (solid line, triangular markers), phase prediction AQE (dash-dotted line, triangular markers) and phase prediction UQE (dash-dotted line, square markers). The dotted curves indicate frequency JNDs measured by Wier et al. [17] (square marker @ 80 dB SPL, triangular marker @ 40 dB SPL). The solid horizontal line indicates the SiCAS quantization. (b) Solid lines: standard deviations of amplitude modulation detection threshold (m). Dotted lines: standard deviations of frequency modulation detection threshold (β). Circular markers: experimental thresholds. Star markers: literature thresholds (amplitude [8], frequency [15]).

This equation can be transformed into Equation 2.19. See appendix A.2 for this transformation.

\[ s_a(t) = c_a(t)A_a(t), \]  
(2.19)

in which \( c_a \) (Equation 2.20) is the sinusoidal carrier and \( A_a \) is the stimulus' amplitude modulation (Equation 2.21). The stimulus is an amplitude modulated sinusoid.

\[
c_a(t) = ae^{i\omega_a t},
\]
(2.20)

\[
A_a(t) = 1 + \frac{\Delta a}{a}\cos(\omega_a t).
\]
(2.21)

Note that amplitude parameter quantization only causes amplitude modulation and does not cause the stimulus to be modulated in frequency. The amplitude modulation is dependent on the magnitude of the quantization error \( \Delta a \) and duration of the Hanning window expressed in its angular frequency \( \omega_h \). The amplitude modulation is sinusoidal with a periodicity of 86 Hz.

Substituting threshold values from the psycho-acoustic experiments in Equation 2.21 gives the amplitude modulation as a function over time. The standard deviations, and not the peak-peak values, of the amplitude modulations (m) are plotted in Figure 2.10b together with amplitude modulation detection thresholds according to Kohlrausch et al. [8] \( f_m = 50 \, \text{Hz} \). The cross subject means of the experimental data are plotted in Figure 2.10a.
Chapter 2. Pure Tone Listening Test

Figure 2.10: (a) Amplitude quantization: cross subject means of AQE (solid line, triangular markers) and UQE (solid line, square markers). The solid horizontal line indicates the SiCAS quantization. (b) Solid lines: standard deviations of amplitude modulation detection threshold (m). Circular markers: experimental thresholds. Star markers: literature thresholds (amplitude [8]).

The modulation thresholds measured by Kohlrausch et al. [8] (star markers in Figure 2.10b) agree well with the modulations calculated from the experimental data (circular markers).

2.4.3. Phase Quantization of a single sinusoid

The amplitude and frequency modulation results from phase quantization in the condition AQE are found in a similar way as the modulations were found in the quantization conditions. Equations 2.22 and 2.23 formulate the quantized sinusoids in subsequent segments where $\Delta \phi$ is the phase quantization error which is alternately added to and subtracted from the carrier’s phase.

\begin{align*}
    c_{p,1}(t) &= ae^{i(\omega_c t + \Delta \phi)} & (2.22) \\
    c_{p,2}(t) &= ae^{i(\omega_c t - \Delta \phi)} & (2.23) \\
    t &\in \mathbb{R}. & (2.24)
\end{align*}

The stimulus is now constructed using the following equation:

\begin{equation}
    s_p(t) = h_1(t)c_{p,1}(t) + h_2(t)c_{p,2}(t). 
\end{equation}

This equation can be transformed into Equation 2.26. See appendix A.3 for this transformation.

\begin{equation}
    s_p(t) = c_p(t)A_p(t)p_p(t),
\end{equation}

in which $c_p$ (Equation 2.27) is the sinusoidal carrier. The stimulus is a mixed modulated sinusoid. $A_p$ is the stimulus’ amplitude modulation (Equation 2.28) and $p_p$ is its phase modulation.
2.5. Discussion

The frequency modulation can now be determined by taking the first derivative of the exponent of $p_p$.

$$f_p(t) = \frac{-\omega_h \tan \Delta \phi \sin \omega_h t}{1 + \cos^2 \omega_h t \tan \Delta \phi}.$$  \hspace{1cm} \text{(2.30)}

Both the amplitude modulation $A_p$ as well as the frequency modulation $f_p$, in which we are interested, are dependent on the magnitude of the quantization error $\Delta \phi$ and duration of the Hanning window expressed in its angular frequency $\omega_h$. Note that neither the amplitude nor the frequency modulation is sinusoidal. The periodicity of the frequency modulation corresponds to 43 Hz and for amplitude modulation to 86 Hz.

Amplitude and frequency modulations as a function of time at threshold values are found by substituting the experimental thresholds in Equation 2.28 and 2.30 respectively. The standard deviations or effective modulation depths of amplitude modulation ($\beta$) and frequency modulation ($\beta$) are plotted in Figure 2.11b together with literature data of amplitude [8] ($f_m = 50$ Hz) and frequency [15] ($f_m = 10$ Hz) modulation detection thresholds. The cross subject means of the experimental data are shown in Figure 2.11a.

The amplitude modulation thresholds calculated from the experimental data shown in Figure 2.11 do not become larger than the modulation thresholds measured by Sek and Moore although in the higher frequencies they become more similar. The experimental frequency modulation thresholds coincide with the literature thresholds measured by Kohlrausch et al. [8].

2.5. Discussion

Looking at the modulation thresholds for frequency and phase quantization it seems likely that the most important cue for detection of frequency and phase quantization is frequency modulation because the frequency modulations calculated from the experimental thresholds coincide with frequency modulation detection thresholds in literature. Whereas the amplitude modulations at frequencies below 4 kHz are at least one order of magnitude smaller than literature data on amplitude modulation detection.

The auditory cue for detection of amplitude quantization is very likely to be amplitude modulation of the signal. First of all, because amplitude quantization does not introduce frequency modulations. Besides that, the experimental data coincide with the literature data on amplitude modulations.

Let us consider the side-lobe producing conditions which are all conditions except frequency UQE with phase prediction (dashed line with square markers, Figure 2.6) and amplitude UQE resulting (solid line with square markers, Figure 2.7). They do not produce side-lobes because
frequency UQE with phase prediction results in a frequency shift and amplitude UQE in an amplitude increment.

In the Figures 2.6 – 2.8 we can observe that all these side-lobe producing quantization paradigms have curves that start ascending at 25 Hz, whereas the other curves start descending at 25 Hz. This may be explained by the fact that the carrier at 25 Hz is suppressed more than the target stimulus' right side-lobe due to filtering in the middle ear. Hereby skewing the relation between carrier and side-lobe which again makes it easier for the test subject to detect the quantization artifacts because they are less well masked.

Another observation is the impact that window length has on the quantization thresholds (extra measurements with circular markers in Figure 2.7c). A reduction of the window length (by increasing $\omega_h$) does increase the distance between side-lobes and carrier (see Equation 2.21). This larger distance will have the effect that the side lobes are masked less by the carrier at lower frequencies which leads to lower detection thresholds. In addition, changing the shape of the windows will also have impact on detection thresholds because by altering its temporal shape, its spectral shape will be altered as well and thus detection thresholds will be different.
3. Tone Complex Listening Test

In this second listening test we focus on quantization limits in a tone complex to be able to investigate the interaction of many sinusoids in a single signal. There were already many different possible conditions in the pure tone listening test, in a tone complex listening test the parameter space gets even larger. Here are a few parameters that enlarge the parameter space:

- fundamental frequency;
- number of harmonics;
- spectral envelope;
- starting phase conditions.

It is difficult to select a paradigm that is not arbitrary and that covers a relevant portion of the parameter space. We have chosen to use a single-formant harmonic complex. This is a vowel-like signal with a peak or formant in the spectral envelope.

After doing some pilot experiments we decided to measure Stochastic Quantization Errors (SQE, explained in § 3.1.1) with start phases equal to zero and to include extra conditions at 4 kHz to measure the effects of e.g starting phase. The 4 kHz frequency is interesting because there are several sinusoidal components that fall within one critical band at 4 kHz which may cause specific types of interaction effects. These extra conditions include AQE measurements for frequency and phase with random starting phases as well as starting phases equal to zero. Also a phase prediction condition was included at 4 kHz.

The stimuli in the second listening test are generated in very much the same way as in the first listening test. The block diagram in Figure 2.1 still holds for the tone complexes. The difference lies in the fact that instead of generating parameters for only one tonal component per window, parameters are now generated per window for all tonal components in the complex. In the quantization block, only the parameters within the critical band associated with the center frequency of the formant are being quantized, the other parameters are left untouched. The wave synthesis block now generates a sinusoid for every tonal component within a segment. These components are added together by means of superposition. The final overlap add block is precisely the same as in the first experiment, meaning that all segments are added with 50% overlap.
3.1. Method

3.1.1. Stimuli and Conditions

Parameters for the tone complex are generated by specifying the frequency amplitude and phase in the first segment. Parameters for the remaining segments are calculated from the values of the parameters in the first segment. The fundamental frequency is chosen to be 100 Hz because this is a typical value for the fundamental frequency of male speech. It is known that male speech is difficult to model in a sinusoidal coder. Moreover, such a low fundamental frequency results in several components falling within the critical band of an auditory filter and therefore the signal within this critical band will have a peaky temporal structure. Subjects may be extra sensitive to differences in this temporal structure.

The harmonic structure is produced by assigning an equal amplitude and a starting phase of zero to each harmonic of this fundamental up to half the sample frequency ($f_s = 44.1$ kHz). This harmonic structure is then passed through a second order digital filter, shown in Figure 3.1a, as has been used by Klatt [7] to model vocal formants. After this operation the amplitudes are no longer equal for all components but will have a peak at the resonator's center frequency. Also the starting phases will be altered by the resonator. The multiplication factors in Figure 3.1a are:

\[
A = 1 - B - C \tag{3.1}
\]
\[
B = 2e^{-\pi WT}\cos 2\pi FT \tag{3.2}
\]
\[
C = -e^{-2\pi WT} \tag{3.3}
\]

Figure 3.1.: (a) Digital resonator. (b) Frequency spectrum of a tone complex filtered by the resonator with a center frequency of 4 kHz and a bandwidth of 457 Hz. The dotted sinusoidal components are the components that will be quantized.
where $T$ is one over the sampling rate, $F$ is the formant’s center frequency and $W$ is the formant’s 3 dB bandwidth.

The transfer function of the digital resonator as used by Klatt can be expressed as:

$$ H_k(f) = \frac{A}{1 - Bz^{-1} - Cz^{-2}} $$

and

$$ z = e^{i2\pi fT}, $$

with $f$ the frequency in Hz.

Figure 3.1b shows a typical stimulus spectrum. The center frequency of the formant is 4 kHz and its 3 dB bandwidth equals 457 Hz which is the ERB width of the critical band at 4 kHz. Above the center frequency, the resonator has a slope of -12 dB/oct.

The dotted sinusoidal components in Figure 3.1b fall within this ERB width. These are the components that will be quantized in the target stimulus. The sound pressure level of the reference stimulus is set at 60 dB. The unquantized tonal components of the target stimulus get precisely the same amplitude as the corresponding components in the reference interval. The quantized parameters are the same as in the reference stimulus, except for the fact that quantization errors are made. Any SPL differences between target and reference interval are due to the quantization in the target interval and should be considered as an auditory cue.

In the tone complex listening test we have chosen not to use the UQE or AQE quantization error hypotheses shown in Figure 2.3 but to introduce a third kind of quantization error. Since no mathematical analysis of the complex tones was planned due to complexity, there is no need to keep the stimuli deterministic. A stochastic process was used to determine if the quantization error would be added or subtracted from the original parameter value. We will refer to this type of error as *Stochastic Quantization Error* (SQE). This is shown by formula 3.6 where $x$ is one of the sinusoid parameters.

$$ \dot{x} = x + \Delta x \cdot U $$

and

$$ U = \{-1, 1\}, \quad P(U = -1) = P(U = 1) = \frac{1}{2} $$

The reason to use this stochastic process is that pilot experiments have shown that SQE’s are more easily detected in the tone complex case than UQEs and AQEs are.

Table 3.1 shows the conditions measured in the listening test. Center frequencies for the formants in the conditions are placed equidistantly throughout the spectrum on a logarithmic scale. The accompanying 3 dB bandwidths are equal to the ERB width of the critical bandwidths at these center frequencies. All conditions have a sound pressure level of 60 dB, a fundamental frequency of 100 Hz and 220 harmonics. The phases are all original except for the last condition in the table, it uses phase prediction. The starting phases of the sinusoidal components are zero except in condition 17 and 19 where the starting phases are assigned random values between 0 and $2\pi$.

### 3.1.2. Procedure and subjects

Also in this listening experiment a three interval forced choice (3IFC) procedure using two-down/one-up adaptive tracking was used to determine the thresholds. The subjects received...
3.2. Experimental Results

<table>
<thead>
<tr>
<th>Exp</th>
<th>F [Hz]</th>
<th>W [Hz]</th>
<th>quantized harmonic(s)</th>
<th>start phases</th>
<th>quantization type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>6</td>
<td>2</td>
<td>zero</td>
<td>frequency SQE</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>68</td>
<td>4</td>
<td>zero</td>
<td>frequency SQE</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>133</td>
<td>10</td>
<td>zero</td>
<td>frequency SQE</td>
</tr>
<tr>
<td>4</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>zero</td>
<td>frequency SQE</td>
</tr>
<tr>
<td>5</td>
<td>10000</td>
<td>1104</td>
<td>95-105</td>
<td>zero</td>
<td>frequency SQE</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>6</td>
<td>2</td>
<td>zero</td>
<td>amplitude SQE</td>
</tr>
<tr>
<td>7</td>
<td>400</td>
<td>68</td>
<td>4</td>
<td>zero</td>
<td>amplitude SQE</td>
</tr>
<tr>
<td>8</td>
<td>1000</td>
<td>133</td>
<td>10</td>
<td>zero</td>
<td>amplitude SQE</td>
</tr>
<tr>
<td>9</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>zero</td>
<td>amplitude SQE</td>
</tr>
<tr>
<td>10</td>
<td>10000</td>
<td>1104</td>
<td>95-105</td>
<td>zero</td>
<td>amplitude SQE</td>
</tr>
<tr>
<td>11</td>
<td>200</td>
<td>6</td>
<td>2</td>
<td>zero</td>
<td>phase SQE</td>
</tr>
<tr>
<td>12</td>
<td>400</td>
<td>68</td>
<td>4</td>
<td>zero</td>
<td>phase SQE</td>
</tr>
<tr>
<td>13</td>
<td>1000</td>
<td>133</td>
<td>10</td>
<td>zero</td>
<td>phase SQE</td>
</tr>
<tr>
<td>14</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>zero</td>
<td>phase SQE</td>
</tr>
<tr>
<td>15</td>
<td>10000</td>
<td>1104</td>
<td>95-105</td>
<td>zero</td>
<td>phase SQE</td>
</tr>
<tr>
<td>16</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>random</td>
<td>phase AQE</td>
</tr>
<tr>
<td>17</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>random</td>
<td>phase AQE</td>
</tr>
<tr>
<td>18</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>zero</td>
<td>frequency AQE</td>
</tr>
<tr>
<td>19</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>random</td>
<td>frequency AQE</td>
</tr>
<tr>
<td>20</td>
<td>4000</td>
<td>457</td>
<td>38-42</td>
<td>zero</td>
<td>frequency AQE</td>
</tr>
</tbody>
</table>

Table 3.1.: Conditions for the tone complex listening experiment.

feedback after each trial about the correctness of the given answer. Stimulus duration was 400 ms and the duration of the silence intervals between the three stimuli were 100 ms. The window length of the Hanning windows was 23.2 ms. Three subjects have participated in the pure tone test. All three subjects had normal hearing and participated voluntarily (not payed) in the experiments. Their ages lie between 26 and 37 and they have experience in psychoacoustic measurements. One of the subjects was the author of this thesis. The threshold measurements were repeated three times.

3.1.3. Setup

Stimuli were generated digitally, 16 bits per sample at a sampling frequency of 44.1 KHz, and presented to the subjects diotically through Beyerdynamic DT990 headphones in an acoustically isolated listening room.

3.2. Experimental Results

Let us have a look at the experimental results shown in Figure 3.2. The first three plots are means per subject where the error bars indicate the standard deviation. The last plot shows the means and standard deviations across all three subjects. The abscissa represents the center frequency of the Klatt synthesizer (in Hz on a logarithmic scale) which coincides with the center frequency of the band that is quantized. The ordinate shows the threshold value. The observations were fairly consistent for the three subjects, no real anomalies were observed.
Chapter 3. Tone Complex Listening Test

Figure 3.2.: Frequency (solid lines, threshold = $\Delta f$), amplitude (dashed, threshold = $\Delta A$) and phase (dash-dotted, threshold = $\pi \Delta p$) quantization limits. Horizontal lines indicate SiCAS quantization. Plots a,b and c are individual measurements per subject. Plot d is the cross subject average. See Table 3.1 for more information about the conditions.
3.2.1. Frequency Quantization of Complex Tones

The circles in Figure 3.2 indicate the mean thresholds for frequency quantization SQEs (conditions 1-5). The dash-dotted horizontal line indicates the quantization error used in the current SiCAS codec. The ordinate shows the threshold value for the frequency quantization conditions. The threshold values are shown as $\frac{\Delta F}{F}$ where $F$ is the Klatt resonator center frequency in Hz and $\Delta f$ is the quantization error in Hz.

The curve starts at approximately 0.02 at 200 Hz and gradually drops to 0.01 at 1 kHz. Subsequently it drops to a value of 0.0004 at 4 kHz from where it descends even further to 0.0002 at 10 kHz. Extra conditions have been placed at 4 kHz for AQE, SQE with random start phase and SQE with phase prediction (left triangle, five pointed star and right triangle respectively). Compared to SQE artifacts the test subjects were less sensitive to the first two artifacts. Subjects appeared to be more sensitive to the phase prediction SQE artifacts than to the original phase SQE artifacts. We will discuss why SQE is more critical in Chapter 3.3.

3.2.2. Amplitude Quantization of Complex Tones

Thresholds for amplitude quantization SQEs are indicated with the upward triangles (conditions 6-10) in Figure 3.2. The quantization error used in the SiCAS coder is indicated with the horizontal dashed line. For amplitude quantization the threshold is plotted as $\frac{\Delta a}{a}$ where $\Delta a$ is the quantization error and $a$ is the amplitude proportional to the sound pressure level in Pa.

This curve starts at a value of 0.1 at 200 Hz and, like the frequency curve, stays fairly flat until 1 kHz where the threshold value is about 0.15. At 4 kHz and 10 kHz the threshold values are lower, 0.025 and 0.045 respectively. No extra conditions were measured in the amplitude quantization measurements.

3.2.3. Phase Quantization of Complex Tones

The phase quantization thresholds, plotted with the six pointed stars (conditions 11-15) in Figure 3.2, follow a similar trend as the amplitude quantization thresholds do except that their values are lower (roughly 10 dB). Here the threshold value are shown as $\pi \Delta p$ where $\Delta p$ is the phase quantization error in radians. Note that the threshold is not divided by another kind of phase and is therefore not dimensionless. The solid horizontal line represents the quantization as performed in SiCAS.

At 200 Hz the curve has a value of 0.035 ascending to about 0.1 at 1 kHz after which the curve drops to 0.01 at 4 kHz and then climbs to 0.015 at 10 kHz. Extra conditions have been placed at 4 kHz for AQEs with zero startphases (square marker) and SQEs with random start phase (diamond marker). Subjects were less sensitive to both these artifacts compared to the SQE artifacts.

3.3. Discussion

One observation that can be made from the experimental results is that the thresholds above 1 kHz are lower than the thresholds up to and including 1 kHz. If we consider that the conditions at 4 kHz and 10 kHz are the only conditions that include more than one quantized sinusoidal
component, then it is arguable that these lower thresholds can be explained by the change in temporal structure at the output of the auditory filter in which the quantized components reside. Due to the highly peaked structure of the unquantized stimulus (see fig. 3.3 for the full bandwidth stimulus) quantization will result in a modification of this peaked structure, e.g. a filling of the valleys between the peaks. Subjects may be highly sensitive to such modifications. In contrast, when only a single sinusoid falls within an auditory filter, there are no valleys, nor peaks, that will be affected by quantization.

The tone complex produces a structure in the time domain that is very 'peaky', see Figure 3.3, because of the phase relation of the sinusoidal components. This phase relation is important for the temporal structure of the wave. By degrading this phase relation, either by altering phases or frequencies, the 'peakiness' of the waveform is degraded as well. In the coder this phase relation is degraded by quantizing the frequency, amplitude and phase parameters but the degradation is different in the different quantization hypotheses.

![Hilbert envelope of a reference stimulus at 4 kHz.](image)

In the experimental results it appeared that subjects were more sensitive to stochastic quantization errors at 4 kHz (SQE, condition 4 and 9) than to alternating quantization errors at 4 kHz (AQE, condition 16 and 18). This can be explained by the fact that a relatively large portion of the energy added to the reference signal due to quantization is located at the peaks of the reference stimulus when using AQE. In contrast, the extra energy in SQE is spread more over time and therefore fills the gaps between the peaks where it is more easily noticed by the subject.

Using random start phases (condition 17 and 19) entails a reference interval that already has little phase relation. Because of this the energy of the wave is smeared over the entire period. As a result, test subjects have a hard time detecting the quantization compared to the zero starting phase conditions because there is no real phase relation that can be disturbed by quantization.

The extra measurement with phase prediction (condition 20) has a very low threshold compared to the other measurements. This is due to the fact that the sinusoidal components are shifted in frequency by quantization, so they are not really harmonic anymore. They are almost harmonic and hence, the phase relation between the components will change over time. This
causes the temporal structure of the stimulus to change in time as well which is relatively well audible.

According to Edler and Purnhagen [1] “Subjective evaluations have shown that the relevancy of phase parameters generally is so low that they do not need to be transmitted. However in this case the temporal structure must be maintained by using an appropriate transient model and phase continuity must be guaranteed by frame to frame tracking.” Our observations have also shown that in absence of strong temporal structure less phase information needs to be transmitted. The reason why Edler and Purnhagen do not transmit phase data is probably due to the fact that they aim for very low bit-rates (as low as 6-8 kbit/s) resulting in fairly poor audio quality whereas SiCAS [4] aims for better audio quality at somewhat higher bit-rates (approximately 24-32 kbit/s).
4. Quantization Scheme

In this chapter we describe how the data from the psychoacoustical measurements are incorporated into a quantization method. For this purpose we needed to make decisions about which data to use and how to translate these into an algorithm. Subsequently the new algorithm will be evaluated by means of a subjective listening test. This subjective listening test uses encoded audio excerpts to compare the new quantization method to the quantization method currently used in the SiCAS codec mentioned in the introduction.

4.1. Implementation

The implementation of the quantization algorithm is based on the pure tone experiment data. The reason will be discussed in Chapter 4.3. The algorithm is a Frequency Dependent Quantization (FDQ) that divides the spectrum into five sub-bands. In each sub-band the sinusoidal parameters are quantized with a quantization step suitable for that particular frequency range according to the pure tone quantization limits.

The quantization steps per frequency range are given in Table 4.1 for frequency, amplitude and phase parameters. These settings have been chosen such that the quantization error does not become larger than quantization limits determined in the pure tone experiments, illustrated with the solid lines in Figure 4.1.

Quantization of the parameters is done logarithmically for frequency and amplitude parameters (Equation 4.1, 4.2) and uniformly for phase parameters (Equation 4.3). In these equations the square brackets are round off operators and \( \Delta_f, \Delta_a \) and \( \Delta_p \) are the allowed quantization errors.

\[
\hat{f} = (1 + 2\Delta_f) \left[ \frac{\log(f)}{\log(1 + 2\Delta_f)} \right] \\
\hat{a} = (1 + 2\Delta_a) \left[ \frac{\log(a)}{\log(1 + 2\Delta_a)} \right] \\
\hat{p} = 2\Delta_p \cdot \left[ \frac{p}{2\Delta_p} \right] .
\]

The subdivision into frequency bands is done on basis of the frequency parameter of the sinusoidal component that is being quantized. We first quantize the frequency parameter of the component with the quantization step belonging to the frequency of the original parameter. The appropriate quantization step for quantizing the amplitude and phase parameter of the component is determined by the quantized frequency parameter.
4.2. Evaluation

Table 4.1.: Quantizer settings.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Amplitude</th>
<th>Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>band</td>
<td>type</td>
<td>f range [Hz]</td>
</tr>
<tr>
<td>I</td>
<td>log</td>
<td>0-50</td>
</tr>
<tr>
<td>II</td>
<td>log</td>
<td>50-160</td>
</tr>
<tr>
<td>V</td>
<td>log</td>
<td>7200-∞</td>
</tr>
</tbody>
</table>

Figure 4.1.: Solid lines: quantizer settings for frequency (a), amplitude (b) and phase (c) parameters. Dashed lines: pure tone quantization detection thresholds.

4.2. Evaluation

4.2.1. Method

To evaluate the quantization scheme devised in Section 4.1 a subjective comparison experiment was set up. In this experiment the subject’s task was to express his/her preference for one of two presented sound excerpts per trial. Both excerpts were generated from the same source file with the difference that one of the excerpts was quantized with the current SiCAS quantization and the other excerpt was quantized with the newly designed FDQ method. The excerpts were presented through loudspeakers in an acoustically isolated room. All subjects participated voluntarily (not payed) and were experienced listeners. The experimental data does not include the author’s nor his supervisor’s results because of possible biases.

Table 4.2 shows the bit rates of the excerpts quantized with the SiCAS quantization (R_{SiCAS}) and the FDQ method (R_{FDQ}). The bit rates are determined by calculating the entropy of the sinusoidal parameters divided by the excerpt length. The SiCAS excerpts were generated using 60 sinusoidal components. Due to the somewhat more expensive quantization, FDQ excerpts were generated using 54 sinusoids hereby making sure that none of the FDQ bit rates exceeds the SiCAS bit rate (R_{SiCAS} - R_{FDQ} = ΔR ≥ 0).

Note that these values are only presented here to illustrate that the entropies of the FDQ
excerpts did not exceed those of the SiCAS excerpts to make sure that if FDQ performs better, it is due to better bit assignment and not due to more information. The values in table 4.2 are not meant to provide true bit-rates of the coder. In these entropies we have not accounted for the reduction in bit-rate that can be obtained by using variable segment size, variable number of sinusoids per segment or linking of sinusoids [4].

All six excerpts were presented to the listener eight times in a balanced random manner. Which means that all 48 trials were placed in a random order but such that every excerpt trial was presented four times in AB as well as in BA order. This way, if the subject was to answer randomly, the expectation would be that there is no preference for one of the two quantizations.

Before doing the actual experiment, the subjects participated in a training experiment meant for the subject to get used to detecting the quantization artifacts. This experiment was set up exactly the same as the actual experiment with the difference that there were only two excerpts, which are different from the ones used in the actual test, that were presented to the subject four times. One excerpt was coded with the SiCAS quantization and the other one with FDQ. In the training experiment the two quantization algorithms were not presented versus one another but versus the unquantized excerpt. Besides that, the quantization artifacts were exaggerated a little by making the quantization error larger.

### 4.2.2. Results

In this section we will analyse the experimental results of the subjective comparison experiment within the framework of a hypothesis test [14]. Therefore we will first formulate a null hypothesis which represents the pattern of the outcome when the test results would be due to random answers from the test subjects. Furthermore we will formulate an alternative hypothesis which we will accept if the test results are not a result of randomness but contain a certain non random relationship. We will attempt to prove the null hypothesis in the hope that in the end we will have to reject it in favor of the alternative hypothesis which states that in fact there is a pattern in the test results.

If none of the quantization algorithms is superior then the probabilities for preferring SiCAS ($P_{FDQ}$) is equal to the probability of preferring FDQ ($P_{SiCAS}$). With these probabilities we can formulate the hypotheses:
4.3. Discussion

Null hypothesis: $H_0 = \text{There is no preference. } P_{\text{SiCAS}} = P_{\text{FDQ}} = \frac{1}{2}$

Alternative hypothesis: $H_1 = \text{FDQ or SiCAS performs better. } P_{\text{SiCAS}} \neq P_{\text{FDQ}}$

$p$ is the chance to incorrectly reject $H_0$. The desired confidence or significance level $\alpha = 0.05$ is the threshold value under which we will reject $H_0$. Thus, we will reject $H_0$ when $p \leq \alpha$.

$p$ can be determined using a cumulative binomial function:

$$p = P(U \geq k) = P(U \leq n - k) = \text{CumulativeBinomial}(n - k, n, P = 0.5).$$

Where $n$ is the number of trials and $k = \max(k_{\text{FDQ}}, n - k_{\text{FDQ}})$ in which $k_{\text{FDQ}}$ is the number of times FDQ is preferred. Our stochastic $U$ equals the number of times that FDQ is preferred over SiCAS (or visa versa if $k_{\text{FDQ}} < n - k_{\text{FDQ}}$). The results of the observations are summarized in Table 4.3.

4.3. Discussion

One could, with good reason, wonder why the tone complex data were not incorporated in the quantization algorithm. The reason is that pure tones, as stated in Chapter 1.4, are still considered being critical signals in many circumstances in terms of quantization detection. Even so, in our tone complex measurements some conditions proved to be more critical than pure tones. This seems to be a result of the fact that the tone complexes used, have highly structured waveforms. To reduce the structure in the complex, for instance by adding randomness to the start phases, is to reduce the sensitivity to quantization artifacts. It does not make sense to find the very worst case tone complex scenario and to implement a straight forward frequency dependent quantization based on this scenario if the signal is not highly temporally structured in nature. The worst case scenario would not satisfactorily reduce the information in the parameters. A more intelligent quantization, exploiting not only knowledge about how accurate to quantize a highly temporally structured audio signal but also exploiting knowledge about how inaccurate to quantize an audio signal with little temporal structure, would be a better solution. This was unfortunately not feasible within this graduate research but the tone complex measurements have been very useful by providing this insight.

Instead, a frequency dependent quantization based on the pure tone measurements was implemented. Data show that rather large quantization errors are allowed at low frequencies for frequency quantization. Although this can be used to improve efficiency, the relatively small bandwidth associated with these low frequencies indicates a relatively small gain in bit-rate. Unfortunately this gain is not sufficient to compensate for the extra bits needed to more precisely quantize the phase parameters. To compensate for this increase of information, for FDQ less sinusoids were used compared to the SiCAS excerpt in the evaluation experiment.

Initially the evaluation experiment was set up using sound excerpts containing 30 sinusoids. This is a typical value used for such a codec but listening tests produced results that were insufficiently consistent for subjects less trained to the artifacts that sinusoidal parameter quantization produces. This was due the fact that audio signals decoded with 30 sinusoids suffer
### Table 4.3: Subjective comparison results per subject and grand total. Significant results ($p \leq 0.05$) are marked with a *.

<table>
<thead>
<tr>
<th>Subject</th>
<th>sicas</th>
<th>fdq</th>
<th>pref</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trumpets</td>
<td>1</td>
<td>7</td>
<td>fdq</td>
<td>0.04*</td>
</tr>
<tr>
<td>German Male</td>
<td>5</td>
<td>3</td>
<td>sicas</td>
<td>0.36</td>
</tr>
<tr>
<td>Harpsichord</td>
<td>0</td>
<td>8</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
<tr>
<td>Vega</td>
<td>1</td>
<td>7</td>
<td>fdq</td>
<td>0.04*</td>
</tr>
<tr>
<td>Bagpipes</td>
<td>5</td>
<td>3</td>
<td>sicas</td>
<td>0.36</td>
</tr>
<tr>
<td>Bells</td>
<td>2</td>
<td>6</td>
<td>fdq</td>
<td>0.14</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>34</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>sicas</th>
<th>fdq</th>
<th>pref</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trumpets</td>
<td>4</td>
<td>4</td>
<td>none</td>
<td>0.64</td>
</tr>
<tr>
<td>German Male</td>
<td>5</td>
<td>3</td>
<td>sicas</td>
<td>0.36</td>
</tr>
<tr>
<td>Harpsichord</td>
<td>4</td>
<td>4</td>
<td>none</td>
<td>0.64</td>
</tr>
<tr>
<td>Vega</td>
<td>3</td>
<td>5</td>
<td>fdq</td>
<td>0.36</td>
</tr>
<tr>
<td>Bagpipes</td>
<td>1</td>
<td>7</td>
<td>fdq</td>
<td>0.04*</td>
</tr>
<tr>
<td>Bells</td>
<td>2</td>
<td>6</td>
<td>fdq</td>
<td>0.14</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>29</td>
<td>fdq</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>sicas</th>
<th>fdq</th>
<th>pref</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trumpets</td>
<td>4</td>
<td>4</td>
<td>none</td>
<td>0.64</td>
</tr>
<tr>
<td>German Male</td>
<td>6</td>
<td>2</td>
<td>sicas</td>
<td>0.14</td>
</tr>
<tr>
<td>Harpsichord</td>
<td>3</td>
<td>5</td>
<td>fdq</td>
<td>0.36</td>
</tr>
<tr>
<td>Vega</td>
<td>5</td>
<td>3</td>
<td>sicas</td>
<td>0.36</td>
</tr>
<tr>
<td>Bagpipes</td>
<td>6</td>
<td>2</td>
<td>sicas</td>
<td>0.14</td>
</tr>
<tr>
<td>Bells</td>
<td>4</td>
<td>4</td>
<td>none</td>
<td>0.64</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>20</td>
<td>sicas</td>
<td>0.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>sicas</th>
<th>fdq</th>
<th>pref</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trumpets</td>
<td>4</td>
<td>4</td>
<td>none</td>
<td>0.64</td>
</tr>
<tr>
<td>German Male</td>
<td>5</td>
<td>3</td>
<td>sicas</td>
<td>0.36</td>
</tr>
<tr>
<td>Harpsichord</td>
<td>0</td>
<td>8</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
<tr>
<td>Vega</td>
<td>2</td>
<td>6</td>
<td>fdq</td>
<td>0.14</td>
</tr>
<tr>
<td>Bagpipes</td>
<td>3</td>
<td>5</td>
<td>fdq</td>
<td>0.36</td>
</tr>
<tr>
<td>Bells</td>
<td>4</td>
<td>4</td>
<td>none</td>
<td>0.64</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>33</td>
<td>fdq</td>
<td>0.01*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>sicas</th>
<th>fdq</th>
<th>pref</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trumpets</td>
<td>5</td>
<td>3</td>
<td>sicas</td>
<td>0.36</td>
</tr>
<tr>
<td>German Male</td>
<td>3</td>
<td>5</td>
<td>fdq</td>
<td>0.36</td>
</tr>
<tr>
<td>Harpsichord</td>
<td>0</td>
<td>8</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
<tr>
<td>Vega</td>
<td>1</td>
<td>7</td>
<td>fdq</td>
<td>0.04*</td>
</tr>
<tr>
<td>Bagpipes</td>
<td>3</td>
<td>5</td>
<td>fdq</td>
<td>0.36</td>
</tr>
<tr>
<td>Bells</td>
<td>1</td>
<td>7</td>
<td>fdq</td>
<td>0.04*</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>35</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grand Total</th>
<th>sicas</th>
<th>fdq</th>
<th>pref</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trumpets</td>
<td>32</td>
<td>32</td>
<td>none</td>
<td>0.55</td>
</tr>
<tr>
<td>German Male</td>
<td>33</td>
<td>31</td>
<td>sicas</td>
<td>0.45</td>
</tr>
<tr>
<td>Harpsichord</td>
<td>10</td>
<td>54</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
<tr>
<td>Vega</td>
<td>16</td>
<td>48</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
<tr>
<td>Bagpipes</td>
<td>29</td>
<td>35</td>
<td>fdq</td>
<td>0.27</td>
</tr>
<tr>
<td>Bells</td>
<td>19</td>
<td>45</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
<tr>
<td>Total</td>
<td>139</td>
<td>245</td>
<td>fdq</td>
<td>0.00*</td>
</tr>
</tbody>
</table>
from many inherent artifacts, even in the unquantized excerpts, the more because in these tests no residual coding was included. The solution was to increase the number of sinusoids to 60, thereby reducing the inherent artifacts and increasing the consistency of the test results.

To further increase the consistency of the results the training experiment was introduced. In the training experiment both quantizations were compared to unquantized excerpts. This enabled test subjects to 'hear out' the quantization artifacts from the inherent artifacts present in both the quantized as well as the unquantized excerpt.

Table 4.4 gives an summary of how many times we can reject our null hypothesis $H_0$ in favor of the alternative hypothesis $H_I$ counted over all subjects. The table also indicates which quantization is significantly preferred in case of rejecting $H_0$ and does not incorporate the data of the grand total. The first observation we can draw from these results is that five out of eight

<table>
<thead>
<tr>
<th>significance count</th>
<th>sicas</th>
<th>fdq</th>
</tr>
</thead>
<tbody>
<tr>
<td>trumpets</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>german male</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>harpsichord</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>vega</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>bagpipes</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>bells</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>subject total</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4.4.: Number of significant results counted per excerpt and per subject total.

subjects show significant preference for the new FDQ algorithm based on the subject totals. Of the three remaining subjects, one showed no preference at all, one subject showed an insignificant $(p \approx 0.16)$ preference for for the SiCAS quantization and one subject showed an insignificant $(p \approx 0.1)$ preference for for the FDQ method.

Another observation is that only for one single excerpt, the current SiCAS quantization method is significantly preferred over the FDQ method. Whereas three other subjects showed significant preference for FDQ regarding the same excerpt.

The grand total in Table 4.3 leads us to accept the null hypothesis $H_0$ for three out of six excerpts, meaning that there was no significant preference for either of the two quantization algorithms. For the remaining three excerpts, $H_0$ can be rejected and the alternative hypothesis $H_I$ is accepted in favor of the new FDQ algorithm. Besides that, the overall total tells us that there is highly significant $(p \approx 3 \cdot 10^{-7})$ subjective preference for the new FDQ algorithm.

This means that the new quantization performs better or similar to the current SiCAS quantization. Moreover, it tells us that listeners prefer to have (slightly) less sinusoidal components in the audio signal rather than artifacts like modulations and side-lobes due to quantization.
5. Conclusions and Recommendations

5.1. Conclusions

In this conclusion we will revert to the research question raised in the introduction and answer it. The research question was divided into two subquestions of which the first was:

"What are the detection thresholds of quantization steps for frequency, amplitude and phase parameters as a function of frequency in tonal and harmonic signals obtained with listening tests?"

Frequency dependent detection thresholds were found for pure tones. However, this research did not find a straightforward answer to this question for tone complexes because during the research it appeared that detection thresholds do not only depend on the signal’s spectrum but on its temporal properties as well. Therefore quantization limits in complex signals should be found in the spectro-temporal domain rather than the spectral domain only. This brings us to the second subquestion:

"How can the threshold data from the listening tests be incorporated in a newly designed quantization scheme for a sinusoidal coder that is available at Philips Research?"

Because we have no method to make use of the signal properties to determine quantization step sizes, we have decided to use the data from experiment one to improve existing quantization schemes (e.g. [4]) which are based on essentially frequency independent quantization.

Indeed, according to the subjective listening used to evaluate the frequency dependent quantization presented in this thesis, a method based on the measured quantization detection thresholds in pure tones gives significantly better results than a quantization based on amplitude and frequency JNDs.

The results of experiment two show that there is a potential for further improvement of sinusoidal quantization schemes when signal properties are taken into account.

5.2. Recommendations

Apparently there is frequency dependency in quantization limits for frequency, amplitude and phase parameters but there is also signal dependency. Future work should find answers to the question on how to exploit this signal dependency.

In the tone complex experiments we found that signals with strong phase relations between components within an auditory filter are more sensitive to quantization than signals that lack these phase relations. Therefore a possible future algorithm could find phase relations, or their absence, within an array of critical bands and assign bits accordingly.
Another spectro-temporal algorithm could be one that uses a suitable psychoacoustical model, fit for temporal tasks, to iteratively optimize the perceptual distortion due to parameter quantization. It would then enlarge the quantization step where distortion is smaller and vice versa where distortion is larger than a certain threshold level in each iteration.
Bibliography


    Proc. IEEE ICASSP-02 2, 1805-1808, 2002

    http://www.outopia.org/teach/istat/STAT-05L.pdf


A. Appendix A - Mathematical Analysis

Definition of the windows used for synthesis of segment 1 \(h_1\) and segment 2 \(h_2\):

\[
\begin{align*}
  h_1(t) &= \frac{1}{2}(1 + \cos(\omega_h t)) \\
  h_2(t) &= \frac{1}{2}(1 - \cos(\omega_h t)) \\
  h_1(t) + h_2(t) &= 1
\end{align*}
\]

A.1. Frequency Quantization of a single sinusoid

Frequency quantized sinusoids in segment 1 \(c_{f,1}\) and segment 2 \(c_{f,2}\):

\[
\begin{align*}
  c_{f,1}(t) &= ae^{i(\omega t + \Delta \omega t)} \\
  c_{f,2}(t) &= ae^{i(\omega t - \Delta \omega (t - \frac{\pi}{\omega_h}))} \\
  0 &\leq t \leq \frac{\pi}{\omega_h}
\end{align*}
\]

The signal after overlap add is now defined as:

\[
\begin{align*}
  s_f(t) &= h_1(t)c_{f,1}(t) + h_2(t)c_{f,2}(t) \\
 &= h_1(t)ae^{i(\omega t + \Delta \omega t)} + h_2(t)ae^{i(\omega t + \Delta \omega (t - \frac{\pi}{\omega_h}))} \\
 &= ae^{i(\omega t + \frac{\Delta \omega \pi}{2 \omega_h})}(h_1(t)e^{i(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h})} + h_2(t)e^{-i(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h})}) \\
 &= ae^{i(\omega t + \frac{\Delta \omega \pi}{2 \omega_h})}(h_1(t)\cos(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h}) + h_2(t)\cos(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h}) \ldots \\
 &\quad + ih_1(t)\sin(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h}) - ih_2(t)\sin(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h})) \\
 &= ae^{i(\omega t + \frac{\Delta \omega \pi}{2 \omega_h})}(\cos(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h}) + i(h_1(t) - h_2(t))\sin(\Delta \omega t - \frac{\Delta \omega \pi}{2 \omega_h}))
\end{align*}
\]
A.2. Amplitude Quantization of a single sinusoid

Amplitude quantized sinusoids in segment 1 \( (c_{a,1}) \) and segment 2 \( (c_{a,2}) \):

\[
\begin{align*}
c_{a,1}(t) &= (a + \Delta a) e^{i \omega_c t} \\
c_{a,2}(t) &= (a - \Delta a) e^{i \omega_c t} \\
t &\in \mathbb{R}
\end{align*}
\]

The signal after overlap add is now defined as:

\[
\begin{align*}
s_{a}(t) &= h_1(t)c_{a,1}(t) + h_2(t)c_{a,2}(t) \\
&= \frac{1}{2} (1 + \cos (\omega_h t))(a + \Delta a) e^{i \omega_c t} + \frac{1}{2} (1 - \cos (\omega_h t))(a - \Delta a) e^{i \omega_c t}
\end{align*}
\]
A.3. Phase Quantization of a single sinusoid

Phase quantized sinusoids in segment 1 \((c_{p,1})\) and segment 2 \((c_{p,2})\):

\[
\begin{align*}
c_{p,1}(t) &= a e^{i(\omega t + \Delta \phi)} \\
c_{p,2}(t) &= a e^{i(\omega t - \Delta \phi)} \\
t &\in \mathbb{R}
\end{align*}
\]

The signal after overlap add is now defined as:

\[
\begin{align*}
s_p(t) &= h_1(t)c_{p,1}(t) + h_2(t)c_{p,2}(t) \\
&= h_1(t)ae^{i(\omega t + \Delta \phi)} + h_2(t)ae^{i(\omega t - \Delta \phi)} \\
&= ae^{i\omega t}(h_1(t)e^{i\Delta \phi} + h_2(t)e^{-i\Delta \phi}) \\
&= ae^{i\omega t}(h_1(t)\cos \Delta \phi + h_2(t)\cos \Delta \phi + ih_1(t)\sin \Delta \phi - ih_2(t)\sin \Delta \phi) \\
&= ae^{i\omega t}(\cos \Delta \phi + i(h_1(t) - h_2(t))\sin \Delta \phi) \\
&= ae^{i\omega t}\sqrt{\cos^2 \Delta \phi + (h_1(t) - h_2(t))^2 \sin^2 \Delta \phi} \cdot e^{i\arctan \frac{h_1(t) - h_2(t)}{\cos \Delta \phi}} \\
s_p(t) &= c_p(t)A_p(t)p_p(t) \\
c_p(t) &= ae^{i\omega t} \\
A_p(t) &= \sqrt{\cos^2 \Delta \phi + (h_1(t) - h_2(t))^2 \sin^2 \Delta \phi} \\
&= \sqrt{\cos^2 \Delta \phi + \left(\frac{1}{2}(1 + \cos \omega t) - \frac{1}{2}(1 - \cos \omega t)\right)^2 \sin^2 \Delta \phi} \\
&= \sqrt{\cos^2 \Delta \phi + \cos^2 \omega t \sin^2 \Delta \phi} = \sqrt{\cos^2 \Delta \phi + (1 + \cos^2 \omega t) \sin^2 \Delta \phi} \\
&= \sqrt{\cos^2 \Delta \phi + \sin^2 \Delta \phi - \sin^2 \Delta \phi(1 - \cos^2 \omega t)}
\end{align*}
\]
A.3. Phase Quantization of a single sinusoid

\[
\begin{align*}
\sqrt{1 - \sin^2 \Delta \phi (1 - \cos^2 \omega_h t)} &= \sqrt{1 - \sin^2 \Delta \phi (1 - \frac{1}{2} (1 + \cos 2\omega_h t))} \\
&= \sqrt{1 - \frac{1}{2} \sin^2 \Delta \phi (1 - \frac{1}{2} \cos 2\omega_h t)} = \sqrt{1 - \frac{1}{2} \sin^2 \Delta \phi + \frac{1}{2} \sin^2 \Delta \phi \cos 2\omega_h t} \\
&= \sqrt{1 - \frac{1}{2} \sin^2 \Delta \phi} \sqrt{1 + \frac{\sin^2 \Delta \phi}{2 - \sin^2 \Delta \phi} \cos 2\omega_h t}
\end{align*}
\]

\[p_p(t) = e^{i \arctan \frac{(h_1(t) - h_2(t)) \sin \Delta \phi}{\cos \Delta \phi}} = e^{i \arctan ((\frac{1}{2} (1 + \cos \omega_h t) - \frac{1}{2} (1 - \cos \omega_h t)) \tan \Delta \phi)} = e^{i \phi(t)}\]

\[f_p(t) = \frac{d}{dt} \phi(t) = \frac{d}{dt} \arctan (\cos \omega_h t \tan \Delta \phi) = -\omega_h \tan \Delta \phi \sin \omega_h t \]

\[= \frac{-\omega_h \tan \Delta \phi \sin \omega_h t}{1 + \cos^2 \omega_h t \tan^2 \Delta \phi}\]
B. Appendix B - Sound excerpt descriptions

The following table describes the audio excerpts used in Chapter 4:

<table>
<thead>
<tr>
<th>Excerpt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>trumpets</td>
<td>Trumpet playing a classical tune with a string section in the background.</td>
</tr>
<tr>
<td>german male</td>
<td>Low male voice talking in German.</td>
</tr>
<tr>
<td>harpsichord</td>
<td>Harpsichord playing single notes as well as chords.</td>
</tr>
<tr>
<td>vega</td>
<td>Female voice singing a melody with large tonal intervals without accompanying instruments (voice only).</td>
</tr>
<tr>
<td>bagpipes</td>
<td>Bagpipes playing a bass note and a melody simultaneously.</td>
</tr>
<tr>
<td>bells</td>
<td>Several highly harmonic bell sounds playing a melody.</td>
</tr>
</tbody>
</table>
C. Appendix C - List of Abbreviations

3IFC - Three Interval Forced Choice
AQE - Alternating Quantization Error
ERB - Equivalent Rectangular Bandwidth
FDQ - Frequency Dependent Quantization
JND - Just Noticeable Differences
LPC - Linear Predictive Coding
MP3 - Mpeg 1 layer 3
SiCAS - Sinusoidal Coding of Audio and Speech
SPL - Sound Pressure Level
SQE - Stochastic Quantization Error
UQE - Unidirectional Quantization Error