

# Harmonic Encoding in Cochlear Implants

Tyler Ganter

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Science in Electrical Engineering

University of Washington

2016

Reading Committee:

Les Atlas, Chair

Jay Rubinstein

Kaibao Nie

Program Authorized to Offer Degree:  
Electrical Engineering

University of Washington

## **Abstract**

Harmonic Encoding in Cochlear Implants

Tyler Ganter

Chair of the Supervisory Committee:  
Professor Les Atlas  
Electrical Engineering

Today's standard in cochlear implant (CI) signal processing is based on incoherent feature extraction to acquire temporal envelopes and fine structure. Incoherent envelopes are sufficient for the baseline task: speech recognition in quiet, however current efforts to improve secondary tasks such as speech recognition in noise, lexical tone discrimination and music appreciation are fundamentally limited by this processing.

Harmonic signals are ubiquitous in speech and music. This thesis argues for the benefits of coherent extraction of harmonic envelopes and temporal fine structure. By taking harmonic structure into account when designing a feature extraction system, processing artifacts can be minimized and signals can be represented more efficiently with the limited data rate of cochlear implants. Furthermore, the proposed method will open up more possibilities for improved cochlear implant encoding.

This thesis is a guide to developing a coherent feature extraction strategy. Incoherent and coherent extraction systems are evaluated and a generalized method is defined. This method is then applied to harmonic signal encoding. Performance metrics are defined and evaluated and best designs are suggested.

# TABLE OF CONTENTS

	Page
List of Figures . . . . .	iii
Chapter 1: Introduction . . . . .	1
1.1 Overview . . . . .	1
1.2 Survey of Literature . . . . .	3
1.3 Contents of Thesis . . . . .	4
1.4 Notational Conventions . . . . .	4
Chapter 2: Cochlea Implant Processing . . . . .	6
2.1 DSP Algorithms . . . . .	12
Chapter 3: Envelope Extraction Methods . . . . .	18
3.1 Incoherent Methods . . . . .	18
3.2 Coherent Methods . . . . .	22
3.3 Coherent Phase Encoding . . . . .	24
3.4 The Relationships . . . . .	28
3.5 Summary . . . . .	35
Chapter 4: Harmonic Envelope Extraction . . . . .	36
4.1 Steady-State Analysis . . . . .	37
4.2 Steady-State Metrics . . . . .	45
4.3 Induced VS Explicit Temporal Modulation . . . . .	47
4.4 Time-Varying F0 . . . . .	50
4.5 Transients . . . . .	50
4.6 Evaluation of Strategies . . . . .	51
4.7 Non-ideal Pitch Estimators . . . . .	65
4.8 Mapping and Selection . . . . .	72

4.9 Implementation Considerations . . . . .	73
4.10 Summary . . . . .	75
Chapter 5: Conclusion . . . . .	76
5.1 Summary . . . . .	76
5.2 Future Work . . . . .	76
Bibliography . . . . .	77
Appendix A: Derivations . . . . .	79

## LIST OF FIGURES

Figure Number		Page
1.1	Classification Scheme for CI Processing Strategies . . . . .	2
2.1	CA vs CIS . . . . .	7
2.2	Channel Vocoder Processing . . . . .	9
2.3	Signal Flow in CI . . . . .	11
2.4	Transformation from DSP output to Electrical Signal, the DSP output is compressed, then applied an electric pulse train to generate the final modulated electric pulse train . . . . .	11
2.5	ACE Flow Diagram . . . . .	13
2.6	condensed ACE Flow Diagram . . . . .	13
2.7	F0mod Flow Diagram . . . . .	14
2.8	Induced vs Explicit Temporal Modulations Example . . . . .	15
2.9	HSSE Flow Diagram . . . . .	15
2.10	HSSE (with coherent phase encoding) Flow Diagram . . . . .	16
3.1	Incoherent vs Coherent Envelope Extraction . . . . .	19
3.2	Magnitude Only VS Coherent Phase Encoding Block Diagrams . . . . .	24
3.3	Cello Example, top left: $m_{k,harmonic}^{(2)}[n]$ , top right: $m_{k,harmonic}^{(1)}[n]$ , bottom left: $y_k^{(2)}[n]$ , bottom right: 55 . . . . .	26
3.4	Method Comparison: magnitude spectrum at each step . . . . .	34
4.1	Magnitude of spectrum for equations (4.5) - (4.8) . . . . .	38
4.2	(a) – (c): magnitude of spectrum for equations (4.19) - (4.21), (d): time-reversal of (c), (e): magnitude of spectrum for (4.22), (f): contributions of separate components of (e) . . . . .	41
4.3	Envelope Estimate $-2F_0$ Component . . . . .	42
4.4	Envelope Estimate $-F_0$ Component . . . . .	42
4.5	Envelope Estimate Baseband Component . . . . .	42

4.6	(a) – (c): magnitude of spectrum for equations (4.23) - (4.25), (d): time-reversal of (c), (e): magnitude of spectrum for (4.26), (f): contributions of separate components of (e) . . . . .	44
4.7	Clarinet vs Saxophone Harmonic Components . . . . .	48
4.8	Clarinet vs Saxophone Envelope Estimates . . . . .	49
4.9	$G_k$ Downshift Quantization . . . . .	52
4.10	$G_k$ filter order . . . . .	53
4.11	$G_k$ filter design . . . . .	54
4.12	$G_k$ variation across harmonics . . . . .	55
4.13	$SIR_k$ filter order and quantization . . . . .	56
4.14	$SIR_k$ filter design . . . . .	57
4.15	$SIR_k$ variation across harmonics with hanning filter . . . . .	58
4.16	$SIR_k$ variation across harmonics with rectangular filter . . . . .	58
4.17	$D_{k,i}$ filter design and order . . . . .	59
4.18	$D_{k,i}$ downshift quantization . . . . .	60
4.19	$D_{k,1}$ across harmonics . . . . .	60
4.20	$D_{k,i}$ at rate of $iF_0$ . . . . .	61
4.21	Transient Step Response, order = 128, 256, 512 (increasing order corresponds to longer reponse time) . . . . .	62
4.22	Transient Input/Output Change . . . . .	63
4.23	Transient Distortion for Common Instruments . . . . .	64
4.24	$SIR_k$ , hanning filter and pitch estimate quantization . . . . .	68
4.25	$SIR_k$ , adaptive filter and pitch estimate quantization . . . . .	68
4.26	$SIR_k$ , hanning filter, pitch estimate quantization and $\pm 5\text{Hz}$ estimation error . . . . .	69
4.27	$SIR_k$ , adaptive filter, pitch estimate quantization and $\pm 5\text{Hz}$ estimation error . . . . .	69
4.28	$D_{k,1}$ , hanning filter and pitch estimate quantization . . . . .	70
4.29	$D_{k,1}$ , hanning filter and pitch estimate quantizationr . . . . .	71
4.30	$D_{k,1}$ , hanning filter, pitch estimate quantization and $\pm 5\text{Hz}$ estimation error . . . . .	71
4.31	$D_{k,1}$ , hanning filter, pitch estimate quantization and $\pm 5\text{Hz}$ estimation error . . . . .	72

## ACKNOWLEDGMENTS

I want to express my sincere appreciation to Les Atlas, Kaibao Nie and Jay Rubinstein for their guidance. I would also like to thank Brad Ekin, Scott Wisdom, Tommy Powers, Eldridge Alcantara and Dave Dolengewicz for the insightful discussions. I couldn't have accomplished any of this without the professors and staff at the University of Washington that have supported me all along.

## **DEDICATION**

to the one and only G-ma



## Chapter 1

# INTRODUCTION

### 1.1 Overview

Cochlear implant technology has come a long way. From a rudimentary system where communication was little more than sound means “yes”, no sound means “no”, cochlear implant patients can now hold conversations over the phone. That being said, there is still a large gap between cochlear implant (CI) and normal hearing subjects on a diverse set of auditory tasks.

Fundamental to cochlear implant signal processing is acoustic feature extraction. Early methods extracted explicit features characteristic to speech signals such as formants, (resonant peaks that differentiate vowels)[14] [12]. This was later phased in out in favor of implicit methods based on the channel vocoder.

These implicit methods split the audio input into bandpass components and extract temporal envelopes and fine structure for each component. Since the introduction of the continued interleaved sampling strategy (CIS)[17], implicit encoding was readily adopted over the explicit approach due to dramatic improvement in speech recognition. Incremental improvements have been made since then; Cochlear Ltd’s advance combination encoder (ACE) strategy [18] has become a clinical standard commonly compared against. However, the foundational signal processing has not changed in over 20 years!

Strategies such as ACE have achieved high performance on speech recognition in quiet, however performance on more difficult tasks is fundamentally limited. These strategies use incoherent processing, and as a result subjectively similar signals may produce radically different outputs depending on their interaction with the time-invariant filters.

Alternatively, coherent feature extraction acquires temporal cues using knowledge about

the signal structure to avoid artifact-driven variations in output.

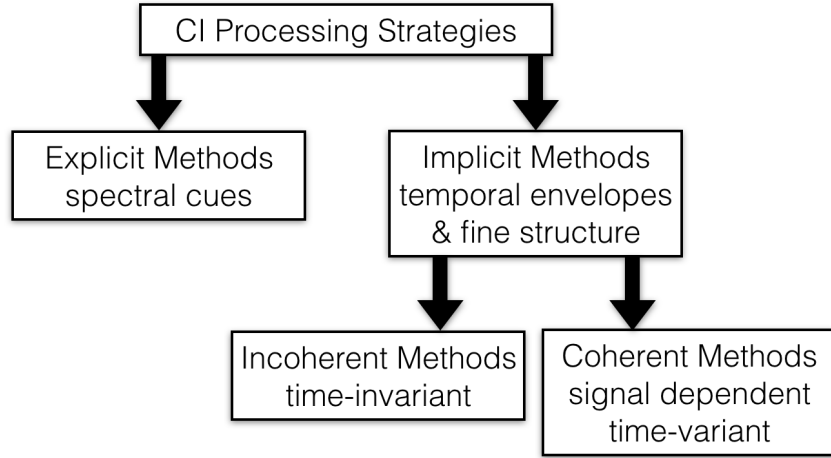


Figure 1.1: Classification Scheme for CI Processing Strategies

A natural application of coherent processing is harmonic signals. Harmonic signals are structured such that energy is focused in narrowband components centered around integer multiples of a fundamental frequency, denoted  $F_0$ .

Coherent processing has two possible advantages. The first is the quality of features extracted, where quality means the reliability of extracted features with changes in the input. For example, if the same note is played on two different instruments, same duration and volume, only the timbre should change. It could be argued that to some extent this is a subjective measure, (psychoacoustic studies have shown that for  $F_0$  differences of over an octave, there is a dependence of timbre perception on pitch [10]), however within reasonable conditions there is a clear difference between perceptual changes due to psychoacoustics and changes due to signal processing distortions.

The second advantage coherent processing may hold is the way in which features are transmitted to the cochlea. Electric hearing has a much lower data rate than acoustic. As a naive but simple explanation of this, consider that the cochlea has approximately 1500 hair

cells transmitting information to the brain. In contrast a typical cochlear implant has only 8-22 electrodes! This clearly demands a data compression scheme. Coherent processing can be used to intelligently select features in advantageous ways.

Quality of feature extraction and feature selection are two potentially independent benefits of coherent processing. An incoherent method could extract distortion free envelopes, but then select which envelopes to transmit suboptimally. Alternatively the envelopes selected could be the same as in a coherent method, however distortions are induced in the envelopes.

In this thesis, incoherent and coherent feature extraction methods are evaluated for the application of encoding harmonic signals. The primary focus is on quality of feature extraction, however, some time is spent considering feature selection as well.

## 1.2 *Survey of Literature*

Separately, there is a great deal of research in cochlear implant processing strategies and harmonic signal processing, but there is limited literature that investigates the two together.

Harris [2] investigated effects of discrete Fourier transform (DFT) window design on isolating harmonics in the presence of nearby strong harmonic interference. Liguori et al. [9] designed an intelligent fast Fourier transform (FFT)-analyzer that interpolates bins to minimize harmonic interference. Alternative to DFT analysis, Li and Atlas [6] used an extended least-squares harmonic model to estimate harmonic features of a signal with time-varying  $F_0$ .

Nogueira et al. [11] developed the MP3000 CI strategy which uses psycho-acoustic masking to more efficiently represent the same acoustic information. Lai and Dillier [4] investigated musical instrument discrimination with MP3000 and found no improvement over ACE. This strategy can eliminate the redundant representation of a single harmonic, however the envelopes are extracted incoherently and suffer the same artifacts as ACE.

Laneau et al. [5] proposed F0mod which explicitly modulates envelopes at a rate of  $F_0$ . Improved pitch discrimination was observed for some conditions. Vandali and van Hoesel

[16] developed an enhanced-envelope-encoder (eTone) strategy that temporally modulates envelopes at a depth proportional to the harmonic probability, (probability the incoherent envelope is from a signal harmonic). Furthermore, eTone attempts to coherently improve signal representation by biasing channel selection toward channels with higher harmonic probabilities. Neither F0mod nor eTone attempt to modify the incoherent envelope extraction method of ACE.

Li et al. [7] developed a harmonic single sideband encoder (HSSE) that uses a pitch estimator to coherently extract harmonic features. Tests on music perception [8] showed significant improvement on timbre recognition for CI users.

### **1.3 Contents of Thesis**

This thesis is organized as follows. In chapter 2 background information and three relevant signal processing strategies are reviewed. In chapter 3 incoherent and coherent methods of signal analysis are compared. In chapter 4 a generalized coherent analysis method is applied to extraction of harmonic features. Finally, chapter 5 takes into account practical implementation considerations and concludes this thesis.

### **1.4 Notational Conventions**

In order to remain consistent throughout this document, the following notational conventions will be defined.  $F_0$  is fundamental frequency,  $F_1$  is first harmonic,  $F_1 = 2F_0$  and the  $k$ th harmonic is  $F_k = (k + 1)F_0$ .

Specific conventions are itemized in the following list:

$j$  - the imaginary unit,  $\sqrt{-1}$

$\mathcal{K}$  - number of envelopes per frame

$\mathcal{M}$  - number of electrode channels

$\mathcal{N}$  - number of electrodes stimulated per frame

$x[n]$  - a time series, (digitally sampled signal)

$x_k[n]$  -  $k$ th subband of  $x[n]$

$m_k[n]$  - envelope of  $x_k[n]$

$c_k[n]$  - carrier of  $x_k[n]$

$\mathcal{H}\{\cdot\}$  - Hilbert Transform

$\mathcal{F}\{\cdot\}$  - Fourier Transform

$\hat{x}[n]$  - the Hilbert Transform of  $x[n]$

$x^+[n] = x[n] + j\hat{x}[n]$  - the analytic signal of  $x[n]$

$X(f)$  - discrete-time Fourier transform (DTFT) of  $x[n]$

$X[k]$  - discrete Fourier transform (DFT) of  $x[n]$

$X[n, f]$  - continuous-frequency short time Fourier transform (STFT) of  $x[n]$

$X[n, k]$  - short time Fourier transform (STFT) of  $x[n]$

$h[n]$  - filter impulse response

$\theta$  - angle

$\phi$  - phase

$f$  - frequency

## Chapter 2

### COCHLEA IMPLANT PROCESSING

Human hearing is tonotopic, that is, starting in the cochlea and through the rest of processing in the brain, sounds far apart in frequency are processed separately. The cochlea is spatially arranged; As a sound propagates through the basilar membrane the different frequencies are amplified or suppressed such that they stimulate locations physically far apart in the cochlea.

In a cochlear implant an array of electrodes is inserted into the cochlea. This array is intentionally designed to have a tonotopic organization. When current is sent to the most deeply inserted (apical) electrodes, neurons associated with low frequency sounds are stimulated. Conversely, current at a basal electrode will stimulate neurons associated with high frequencies.

Early cochlear implant strategies, under the category compressed-analog (CA), delivered band-specific analog signals to each electrode. By using bandpass filters and an electrode array the implant emulates the tonotopic organization of acoustic hearing.

Current processing strategies use feature extraction to achieve much higher performance on speech recognition. From each bandlimited signal a slow-time-varying envelope is extracted and the extra information is discarded [15]. The envelopes are amplitude compressed and then used to modulate continuous bipolar pulse trains on each electrode channel.

These strategies all stem from an original parent, continuous-interleaved-sampling (CIS). CIS is a solution to the problem of electric field interaction. By interleaving pulse-trains there is minimal interaction between electrodes.

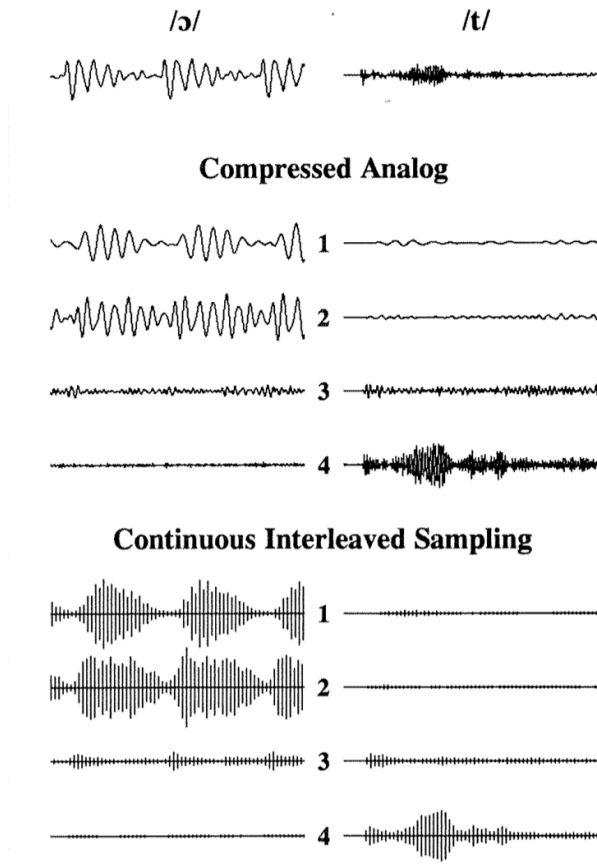


Figure 2.1: CA vs CIS

### 2.0.1 Sum-of-Products Model

We have now laid out enough background information to introduce a mathematical model for audio signals called the sum-of-products model.

In this model, a digitally sampled audio signal  $x[n]$  is composed of bandpass components  $x_k[n]$ . In each bandpass component a slow-time-varying envelope  $m_k[n]$  multiplies a quickly-oscillating carrier  $c_k[n]$ .

$$x[n] = \sum_k x_k[n] = \sum_k m_k[n] c_k[n] \quad (2.1)$$

Although there are infinite ways to decompose a signal into a sum-of-products, the model stems from real-word signals. To gain some intuition consider, for example, a voiced vowel. The vocal tract can be thought of as generating the carriers,  $c_k[n]$ . Without changing the position of the mouth, one can change the pitch of a note. The mouth then changes the temporal envelopes,  $m_k[n]$ . As the mouth changes shapes it changes the formant structure structure. Equivalently, it adjusts the relative amplitude of each bandpass component  $x_k[n]$ .

As another example, the pitch and timbre of musical instruments. The pitch is characterized by the carriers but the timbre which is predominantly characterized by the attack time and spectral centroid [3] will be encoded in the rise time and relative amplitude of the envelopes.

### *2.0.2 Why Envelopes?*

One of the motivations for this approach is the limited ability to perceive temporal modulations in electric hearing. In acoustic hearing modulations up to a few kHz may be perceptible, however cochlear implant envelope extraction techniques are designed to limit modulations, typically to around 160 to 320 Hz, which is closer to the range perceptible in electric hearing.

Modulation rates are also limited by pulse rate. Although there isn't a quantitative value analogous to Nyquist rate, modulations at rates higher than a certain percentage of the constant pulse rate will not be represented accurately by the modulated pulse train. That being said, cochlear implants today support modulations typically upwards of 2000pps (pulses per second), which should be sufficient provided modulations limited to about 320Hz.

### *2.0.3 The Channel Vocoder*

To gain some intuition as to how and why CIS processing works, consider a closely related system, the channel vocoder. Vocoding is a method of signal analysis and synthesis initially designed for audio data compression in telecommunication. As of the mid 70's the vocoder has gained widespread familiarity via the music industry as a funky voice effect. It is most well known for the signature robot voice heard in hits such as Kraftwerk's song "The Robots"



or Styx’s “Mr. Roboto”. In its application to music, the vocoder extracts the bandlimited envelopes of one source, (typically vocal), and applies them to each subband component of a second source.

What’s interesting is that this second source can be essentially any arbitrary broadband signal and yet we still understand speech from the first source. In this way the vocoder acts as a form of lossy data compression; the low data-rate envelopes are extracted and they may be later applied to, for example, white-noise.

This tells us that speech information is predominantly contained in the bandlimited envelopes, and thanks to the incredible robustness of speech to distortion, an estimated envelope is sufficient for speech comprehension.

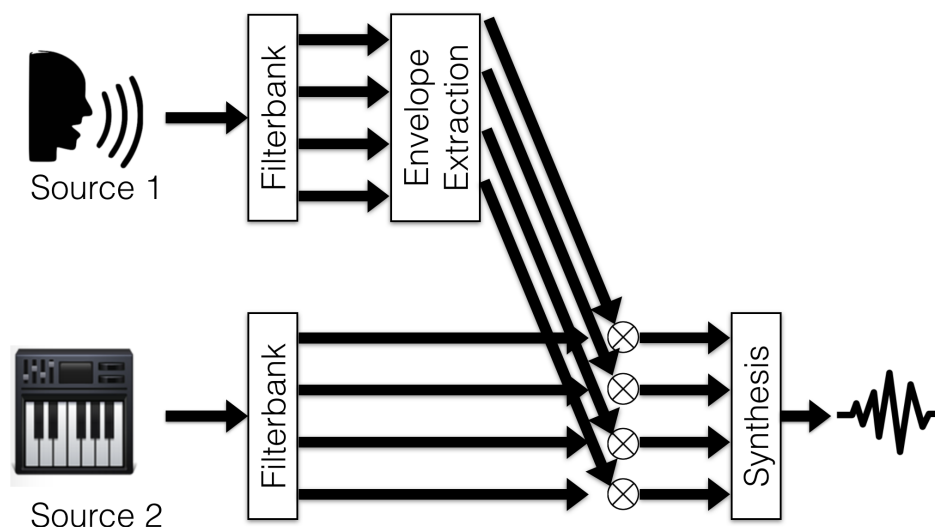


Figure 2.2: Channel Vocoder Processing

It should be noted that the second source is typically chosen to be a broadband stationary signal. If the signal is non-stationary it will have time-varying envelopes of its own which will interact with the envelopes of the first source. In terms of the sum-of-products model, the second source,  $x^{(2)}[n]$ , is a sum of time-invariant subbands and envelopes from the first source,  $x^{(1)}[n]$  are applied.

$$x^{(1)}[n] = \sum_k x_k^{(1)}[n] = \sum_k m_k^{(1)}[n] c_k^{(1)}[n] \quad (2.2)$$

$$x^{(2)}[n] = \sum_k x_k^{(2)}[n] \quad (2.3)$$

$$y[n] = \sum_k m_k^{(1)}[n] x_k^{(2)}[n] \quad (2.4)$$

Linking back, cochlear implant envelope extraction strategies do the same thing as vocoder signal analysis, as seen in figure 2.2, however rather than using a second source to synthesize a new sound, the envelopes directly modulate electrical pulse trains.

#### 2.0.4 Temporal Fine Structure

A major drawback to this method of encoding is the loss of temporal fine structure. Recall that the extracted envelopes are transmitted and carrier information is discarded.

When using a vocoder, vocals sung at different pitches generate roughly the same output,  $y[n]$ . Similarly in cochlear implants temporal fine structure that encodes pitch, as well as other signal characteristics, is lost in processing.

The previously statements don't paint the entire picture though. Some temporal fine structure information may be transmitted if carrier information leaks into the estimated envelope. Some strategies take advantage of this and intentionally allow for carrier leakage in the envelope. ACE is an example of this approach.

#### 2.0.5 Processing Blocks

The main blocks of cochlear implants processing are visualized in figure 2.3. While at every stage adjustments can be made, for the purpose of comparing DSP algorithms, all other stages will be assumed constant throughout this work unless otherwise specified.

In this thesis, the output of the DSP stage will be a strictly positive signal used to amplitude modulate a constant bipolar pulse train. T/C (threshold and comfort) Level Mapping is a logarithmically-compressed mapping from amplitude to current level.



Figure 2.3: Signal Flow in CI

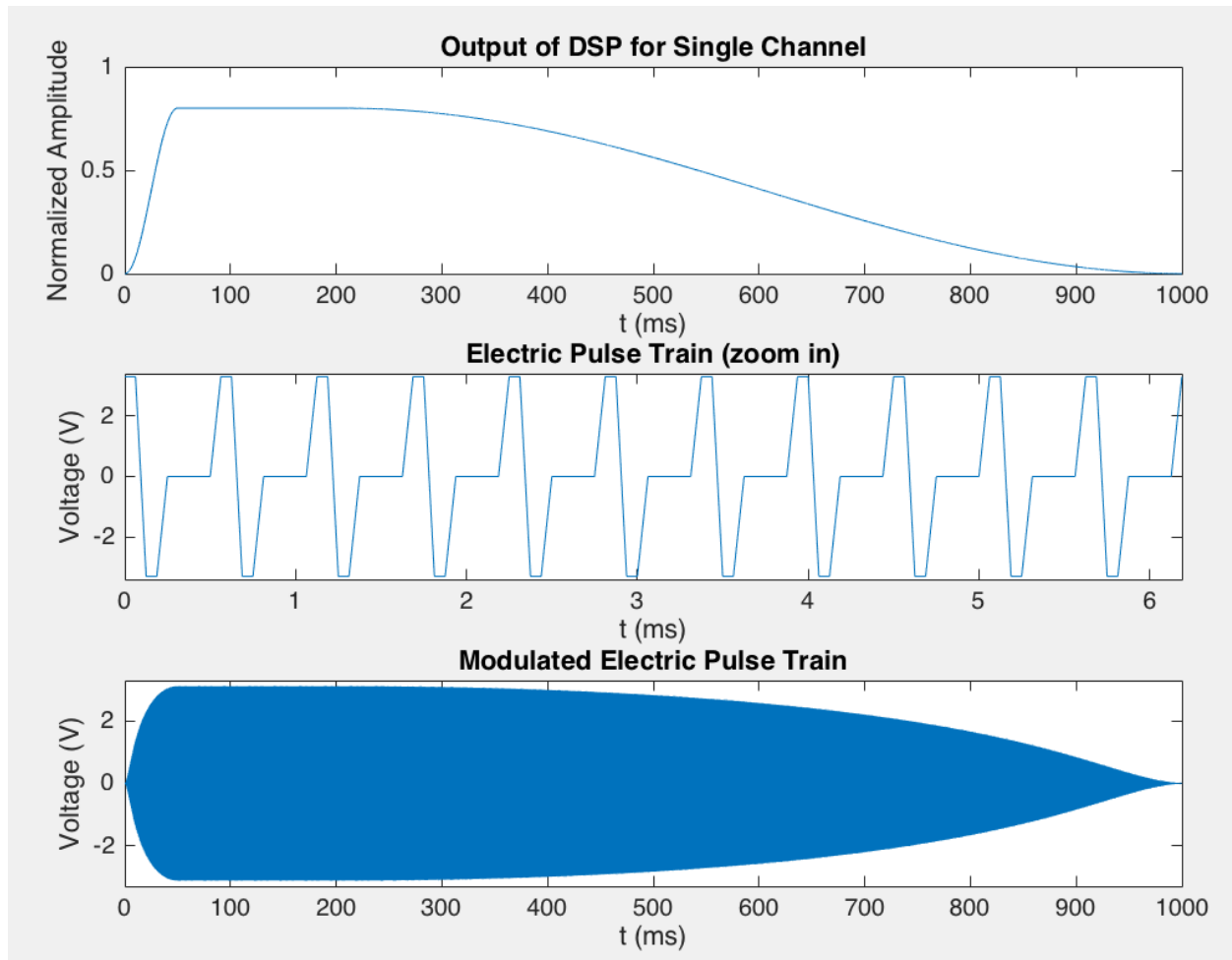


Figure 2.4: Transformation from DSP output to Electrical Signal, the DSP output is compressed, then applied an electric pulse train to generate the final modulated electric pulse train

To summarize up to this point, a cochlear implants has an array of tonotopically organized electrodes. On each electrode a electric pulse train is transmitted and that pulse train is

modulated by a temporal envelope corresponding to a subband of the acoustic input signal.

## 2.1 DSP Algorithms

In order to gain insight into how to encode harmonic signals, in this section we will look inside the “DSP Algorithm” box; three specific strategies, ACE, F0mod, and HSSE will be compared with the goals of evaluating the pros and cons of each and considering how to optimize performance for harmonic encoding.

### 2.1.1 ACE

The simplest of the considered strategies is the Advanced Combination Encoder (ACE). ACE is Cochlear Ltd’s instance of the auditory community’s generalized category of  $\mathcal{N}$ -of- $\mathcal{M}$  strategies. In these strategies,  $\mathcal{K}$  extracted envelopes are allocated to  $\mathcal{M}$  channels corresponding uniquely to electrodes. During each processing frame a subset,  $\mathcal{N}$ -of- $\mathcal{M}$ , channel envelopes is selected for stimulation on the internal implant. In the case that more than one envelope is allocated to a channel, the allocation stage must make a decision to select or combine envelopes in some way.

$\mathcal{K}$  - number of envelopes per frame

$\mathcal{M}$  - number of electrode channels

$\mathcal{N}$  - number of electrodes stimulated per frame

$$\mathcal{K} \geq \mathcal{M} \geq \mathcal{N}$$

A block diagram of ACE is visualized in figure 2.5 and an equivalent condensed notation is shown in figure 2.6. This condensed notation will be carried through to the other strategies analyzed.

While ACE does a sufficient job for many CI users in speech recognition tasks, a large gap remains between normal hearing and cochlear implants in many tasks such as pitch

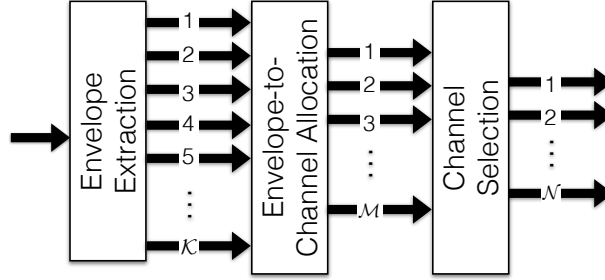


Figure 2.5: ACE Flow Diagram



Figure 2.6: condensed ACE Flow Diagram

discrimination. This is largely attributed to the lack of temporal fine structure information in this envelope encoding strategy.

ACE does, however, provide limited temporal modulations via beat frequencies. Through intentional processing artifacts, beat-frequencies will be induced in the processing of harmonic signals at a rate of the difference between the two harmonic frequencies, i.e.  $F_0$ . Typically these modulations are not full depth and are usually limited to under 320Hz.

In this thesis, these artifact based modulations are termed induced modulations. Looking at the flow diagram of figure 2.6 it is not apparent that temporal modulations are contained in the processing path, however these modulations are encoded in the envelope itself.

Induced modulations are complementary to explicit modulations, used in F0mod and HSSE. Explicit modulations are extracted from the signal separate from envelopes, and later applied to the final outputs.

### 2.1.2 $F_0\text{mod}$

To get at the problem of pitch discrimination, Laneau et al. [5] developed a new research strategy,  $F_0\text{mod}$ .  $F_0\text{mod}$  provides the same processing as ACE with one important change, explicit carrier modulation. It achieves this by adding a pitch estimator into the processing.

Once a fundamental frequency is acquired, all output envelopes are modulated by a raised sinusoid at a rate of  $F_0$ .  $F_0$  is used because of the limitations on ability to perceive high modulation rates with a CI.

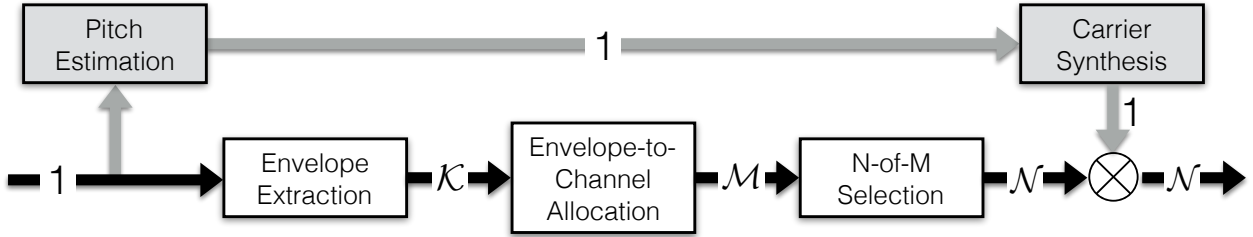


Figure 2.7:  $F_0\text{mod}$  Flow Diagram

This raised sinusoid, defined in (2.5), is constant modulation depth, (full dynamic range), and same across channels, (phase aligned). An example comparing this to induced modulations is shown in figure 2.8.

$$c[n] = 0.5 + 0.5\cos(2\pi F_0 n) \quad (2.5)$$

### 2.1.3 $HSSE$

Looking for a novel approach to improved pitch perception and, more broadly, music perception, Li et al. [7] developed Harmonic Single Sideband Encoder ( $HSSE$ ). There are two different versions of  $HSSE$ . We will start by describing the version most similar to  $F_0\text{mod}$ .

In this version, coherent demodulation extracts harmonic envelopes. These harmonic envelopes are then combined into channels based on the harmonic index and  $F_0$ . Just as

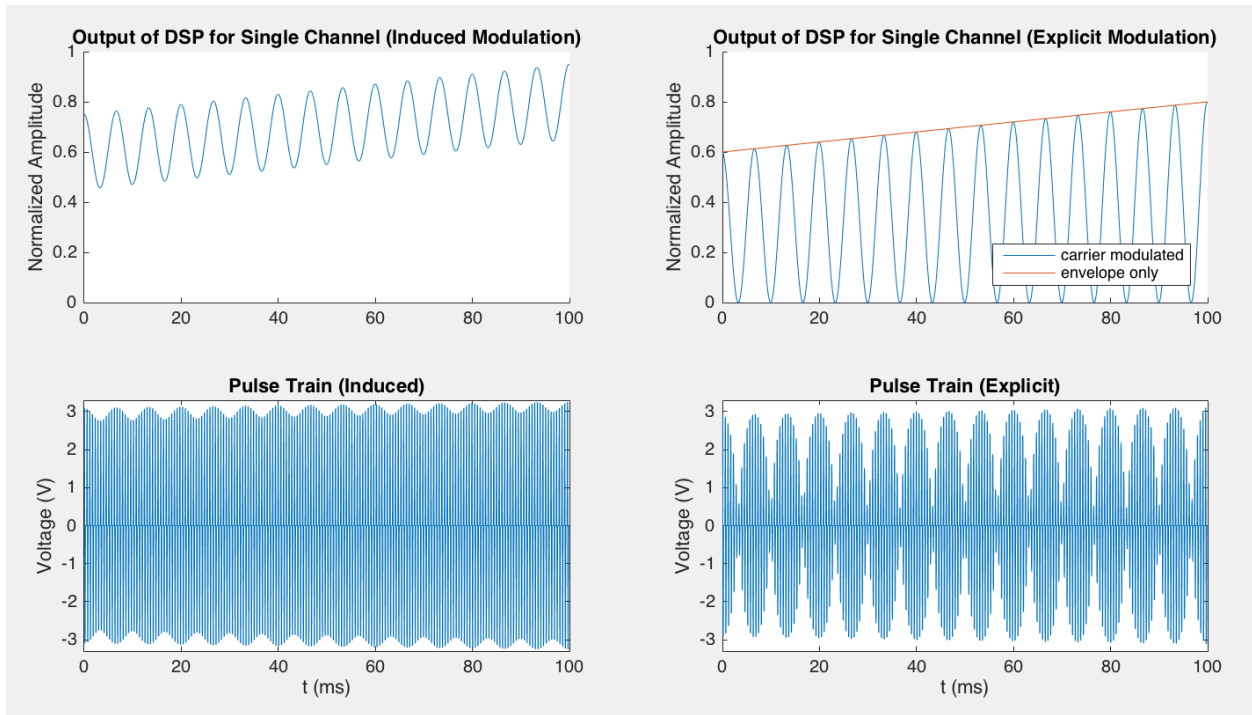


Figure 2.8: Induced vs Explicit Temporal Modulations Example

in F0mod a subset is selected for stimulation and then these envelopes are combined with carrier modulators.

$$\mathcal{K}, \mathcal{M} \geq \mathcal{H}$$

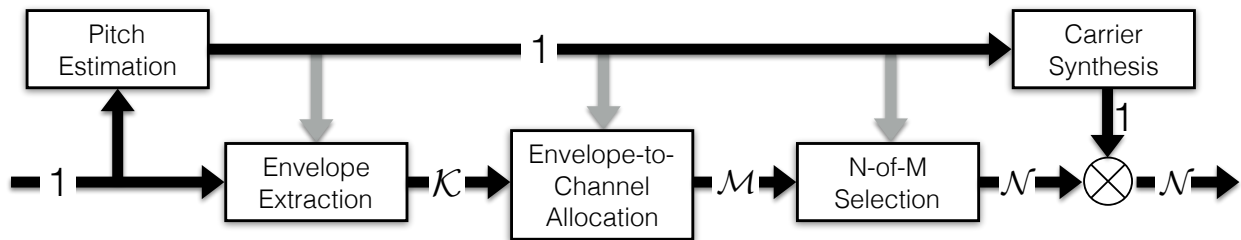


Figure 2.9: HSSE Flow Diagram

The key differences between this and F0mod can be summarized quite simply: every stage of typical ACE processing is now done coherently using  $F_0$  information.

It should be noted that it is not necessarily true that  $\mathcal{K} \geq \mathcal{M}$ . In the case that no envelopes are allocated to a channel that channel is ruled out during the selection stage.

In the second version, more information about the carriers is retained than just the fundamental frequency. This puts some restrictions on the type of carrier than can be used, however it encodes time varying phase information which is unique to each envelope.

Because of the unique characteristics of each carrier, the carrier synthesis block must be moved to an earlier point in the processing. First, complex envelopes containing phase information are extracted. These envelopes are then combined with a common carrier at a rate of  $F_0$ , however, each output, which we will call a modulator, will be unique and time-varying in both magnitude and phase. This version, which will be termed HSSE with coherent phase encoding, is visualized in figure 2.10

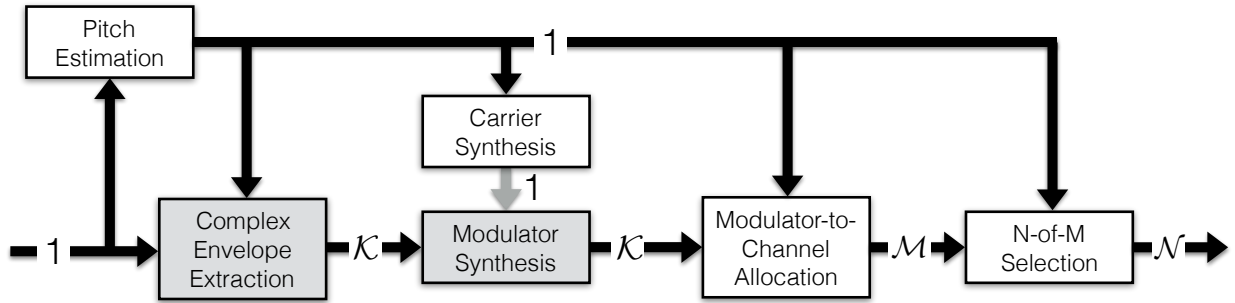


Figure 2.10: HSSE (with coherent phase encoding) Flow Diagram

#### 2.1.4 Summary

Comparing these strategies, the differences may be summarized as:

- 1) Envelope Extraction Method (not discussed yet)
- 2) Temporal Encoding Method



- a) induced vs explicit
- b) phase encoding (explicit only)
- c) modulation waveform (explicit only)
- 3) Envelope-to-Channel Allocation and Channel Selection

We will start by investigating 1 and 2(a,b). Some considerations for 2(c) and 3 will be brought up upon concluding this thesis, however, the primary focus will be on 1 and 2(a,b).

Chapter 3 will discuss mathematical methods to envelope extraction as well as phase preservation since phase is extracted at the same time. As a result we will generalize 1 and we will answer 2(b). Chapter 4 will evaluate design considerations for 1 and in doing so, answer 2(a). Chapter 5 will briefly discuss 2(c) and 3.

## Chapter 3

### ENVELOPE EXTRACTION METHODS

This chapter will define the methods used to extract bandlimited temporal envelopes. These methods fall under a general signal processing category of analysis-synthesis systems. A signal is decomposed into its envelopes and carriers. Then the envelopes and/or carriers are manipulated individually before recombination.

One of the major focuses of research in this area is the evaluating the amount of distortion induced by the system. For example, Ghitza's test is a way of measuring the out-of-band distortion of a modulation filtering system [1].

Cochlear implant processing is a special case in that the final output is not an audio signal. This means that only the first half, the analysis step, is applicable. This is critical to understand when considering methods, as all of the considerations related to synthesis or full-system distortion are no longer relevant.

This chapter is organized as follows. First, the incoherent and coherent envelope extraction methods will be defined. Then an aside will be taken to consider the efficacy of coherent phase encoding. Finally the extraction methods will be compared and a generalization defined.

#### **3.1 Incoherent Methods**

The difference between incoherent and coherent is actually quite simple. Consider a system  $T\{\cdot\}$ . If the system is time-invariant then it is incoherent. If it is time-varying and the way in which it varies is a function of the input, it is called a coherent system. This is visualized in figure 3.1. In coherent methods the input not only passes through the system, it changes the system.

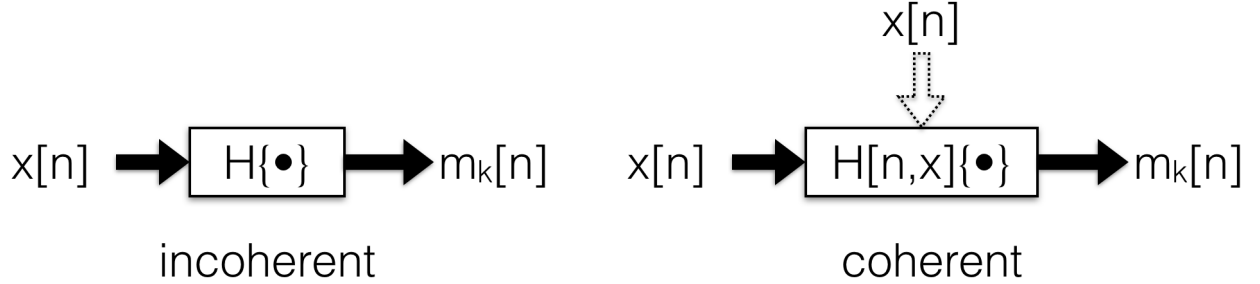


Figure 3.1: Incoherent vs Coherent Envelope Extraction

In all considered methods, the input is a real digitally sampled audio waveform,  $x[n]$  bounded in the normalized range  $[-1, 1]$ . For incoherent methods the output will be  $\mathcal{K}$  real digital waveforms,  $m_k[n]$ , in the range  $[0, 1]$ . All filters considered will be finite impulse response (FIR).

### 3.1.1 Continuous Interleaved Sampling (CIS)

This method is specifically implemented by the CIS strategy. The input is first bandpass filtered, where  $h_k[n]$  is a bandpass filter and  $k$  has arbitrary limits. The subband is then full-wave rectified, (magnitude operation), and lowpass filtered.  $h_{lp}[n]$  is a lowpass filter, typically with a cutoff around 200-400Hz.

$$m_{k,CIS}[n] = |x[n] * h_k[n]| * h_{lp}[n] \quad (3.1)$$

In this method the number of filters is usually the same as the number of electrodes,  $\mathcal{M} \approx 8$  to 22, making  $\mathcal{K}$  to  $\mathcal{M}$  a one-to-one mapping. These filters are thus non-uniform bandwidth and center frequency, with increasing bandwidth and wider spacing at higher frequencies.

### 3.1.2 Hilbert Envelope

The Hilbert envelope is a method of decomposition applied far more broadly than the field of cochlear implants. Despite only retaining the envelope, we look at the carrier to gain insight as to how the signal  $x[n]$  is represented in the decomposition. The analytic bandpass signal,  $x_k^+[n]$  is computed as (3.2). The envelope is defined as the magnitude, (3.3).

$$\begin{aligned} x_k^+[n] &= x[n] * h_k[n] + j\mathcal{H}\{x[n] * h_k[n]\} \\ &= x_k[n] + j\mathcal{H}\{x_k[n]\} \end{aligned} \quad (3.2)$$

$$m_{k,hilbert}[n] = |x_k^+[n]| \quad (3.3)$$

$$c_{k,hilbert}[n] = \cos(\angle x_k^+[n]) \quad (3.4)$$

Intuitively, if the filterbank  $[h_1[n], h_2[n], \dots, h_K[n]]$  has a flat total response, all of the information of the original signal is contained in the envelopes and carriers, and thus it is possible to reconstruct the input.

### 3.1.3 Short Time Fourier Transform (STFT)

The short-time Fourier transform (STFT) is not commonly associated with envelope extraction with respect to its prevalence in signal processing, however through analysis we will see that it fits the sum-of-products model.

The STFT has two classic interpretations: a series of windowed Fourier transforms, each at a different time instant, or a collection of uniform bandpass filters, each at a different center frequency. The latter is more directly applicable to envelope extraction.

An STFT bin at discrete time  $n$  and discrete frequency  $k$  is defined as

$$X[n, k] = \sum_{r=-\infty}^{\infty} x[r]w[r-n]e^{-j\frac{2\pi}{N}kr}, \quad 0 \leq k < N \quad (3.5)$$

where  $N$  is the DFT order, not to be confused with electrodes stimulated per frame,  $\mathcal{N}$ . Defining a new variable  $r' = r - n$  and defining the window such that  $w[n] = 0$  for  $n < 0$  or  $N \leq n$ ,

$$\begin{aligned} X[n, k] &= \sum_{r'=0}^{N-1} x[n + r'] w[r'] e^{-j \frac{2\pi}{N} k(n+r')} \\ &= e^{-j \frac{2\pi}{N} kn} \sum_{r'=0}^{N-1} x[n + r'] w[r'] e^{-j \frac{2\pi}{N} kr'}. \end{aligned} \quad (3.6)$$

Let  $X[n, k]$  be represented in polar form as

$$X[n, k] = |X[n, k]| e^{j \angle X[n, k]}. \quad (3.7)$$

Assuming the window  $w[n] \neq 0$  for  $0 \leq n \leq N - 1$  the inverse may be solved as

$$\begin{aligned} x[n + r'] &= \frac{1}{N w[r']} \sum_{k=0}^{N-1} X[n, k] e^{j \frac{2\pi}{N} k(n+r')} \\ &= \frac{1}{N w[r']} \sum_{k=0}^{N-1} |X[n, k]| e^{j(\frac{2\pi}{N} k(n+r') + \angle X[n, k])} \end{aligned} \quad (3.8)$$

$$x[n] = \sum_{k=0}^{N-1} \frac{1}{N w[0]} |X[n, k]| e^{j(\frac{2\pi}{N} kn + \angle X[n, k])} \quad (3.9)$$

(3.8) simplifies to (3.9) when the STFT hop-factor is one sample, which can be assumed without loss of generality. For greater hop factors the inverse can always be computed from (3.8). Of course, if the hop factor is greater than  $N$  the original signal cannot be fully reconstructed. This is especially noted because the factor  $w[0]$  will be seen recurrently throughout this thesis.

The sum-of-products becomes clear in (3.9).

$$m_{k,STFT}[n] = \frac{1}{N w[0]} |X[n, k]| \quad (3.10)$$

$$c_{k,STFT}[n] = e^{j(\frac{2\pi}{N} kn + \angle X[n, k])} \quad (3.11)$$

The STFT can be thought of as a series of  $N$  linear time-invariant (LTI) systems that each downshift the input signal, then lowpass filter. This can be seen mathematically by rewriting (3.5) as

$$\begin{aligned} X[n, k] &= \sum_{r=-\infty}^{\infty} x[r] e^{-j\frac{2\pi}{N}kr} w[-(n-r)] \\ &= x[n] e^{-j\frac{2\pi}{N}kn} * w[-n]. \end{aligned} \quad (3.12)$$

The STFT envelope has a similar form to the other methods after plugging (3.12) into (3.10).

$$m_{k,STFT}[n] = \frac{1}{Nw[0]} \left| x[n] e^{-j\frac{2\pi}{N}kn} * w[-n] \right|, \quad 0 \leq k \leq \frac{N}{2} \quad (3.13)$$

Also note that, due to symmetry of the Fourier transform, envelopes are only valid for indices between 0 and  $\frac{N}{2}$ .

### 3.2 Coherent Methods

Due to their LTI nature, incoherent methods fail to explicitly represent time varying characteristics like fundamental frequency or formant structure [17]. Alternatively, coherent methods will adapt to represent specific characteristics.

#### 3.2.1 Spectral Center-of-Gravity

One coherent method is the spectral center-of-gravity (COG) [?]. Similar to the previously described incoherent methods, spectral COG uses a fixed number of filters. The key difference lies in the center frequencies of each of these filters, which adapt over time as a function of the spectral distribution within predefined band limits.

Spectral COG certainly has some advantages of better representation of the signal in comparison to incoherent methods, however it still does not escape the limitation of fixed

and pre-determined band limits that each filter operates within. Spectral COG will not be investigated further.

### 3.2.2 Harmonic

Li et al. [7] proposed a harmonic method which uses knowledge of the structure of common audio signals to decompose the signal in a less arbitrary way. The first step is to compute a pitch estimate,  $F_0[n]$ , of the signal.  $\mathcal{K}$  complex carriers are defined. There is a hard limit as a function of Nyquist sampling rate,  $\mathcal{K} \leq \lfloor \frac{F_s}{2F_0} \rfloor$  [?].

$$c_{k,harmonic}[n] = e^{jk\phi_0[n]} \quad (3.14)$$

where

$$\begin{aligned} \phi_0[n] &= \frac{2\pi}{F_s} \sum_{p=0}^n F_0[p] \\ &= \phi_0[n-1] + 2\pi \frac{F_0[n]}{F_s} \\ \phi_0[-1] &= 0 \end{aligned} \quad (3.15)$$

As mentioned earlier there are two versions of HSSE. One uses a real non-negative envelope, the other uses a complex envelope. The envelope of the first method is defined as

$$\begin{aligned} m_{k,harmonic}^{(1)}[n] &= \left| x[n] c_{k,harmonic}^*[n] * h[n, F_0[n]] \right| \\ &= \left| x[n] e^{-jk\phi_0[n]} * h[n, F_0[n]] \right| \end{aligned} \quad (3.16)$$

where  $h[n, F_0[n]]$  is a lowpass filter that may vary as a function of  $F_0[n]$ . Note that it is possible to have a different LPF for each  $k$  however since the carriers of a harmonic signal are linearly spaced it is natural to keep  $h[n, F_0[n]]$  consistent over  $k$ .

The second, complex envelope is the same as the first but without the final magnitude operation.

$$m_{k,harmonic}^{(2)}[n] = x[n]e^{-jk\phi_0[n]} * h[n, F_0[n]] \quad (3.17)$$

### 3.3 Coherent Phase Encoding

As mentioned earlier, the final DSP output is a set of real non-negative signals. We take a short aside to compare the two coherent harmonic methods, one of which, due to it's complex output, cannot be evaluated as an envelope extraction method independent of temporal fine structure encoding.

The two approaches are visualized in figure 3.2. For the magnitude-only case, this can be thought of as a restriction on what the carrier can be. Since the envelope is already real non-negative the  $Re\{\cdot\}$  and half-wave rectification stages don't change anything. Passing a complex exponential through these two operations before multiplying the envelope is equivalent to defining the carrier as a half-wave rectified sinusoid and thus we have equivalent processing blocks as a single envelope of 2.9

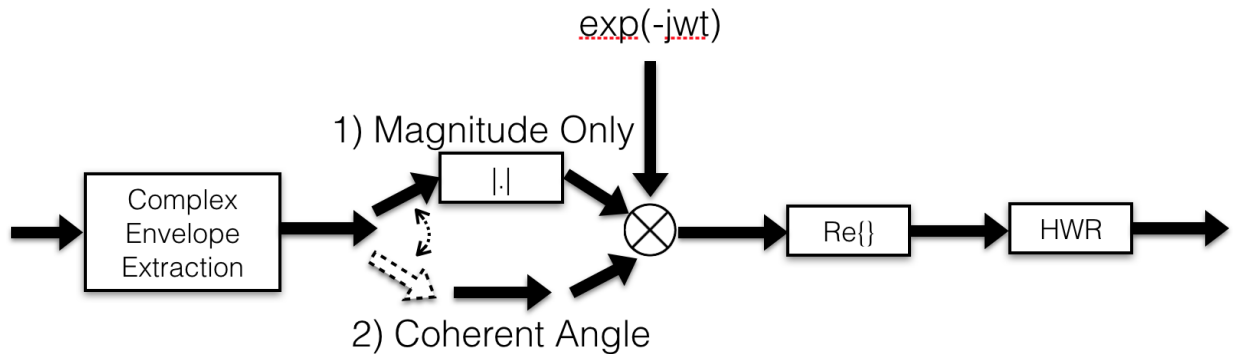


Figure 3.2: Magnitude Only VS Coherent Phase Encoding Block Diagrams

Consider a signal where the  $k$ th subband is of the form



$$x_k[n] = A_k[n] \cos(2\pi k F_0 n + \phi_k[n]) \quad (3.18)$$

$$BW \leq F_0$$

where  $A_k[n]$  represents a real nonnegative amplitude and  $BW$  is the signal's bandwidth. We may assume  $F_0[n] = F_0$  is constant, without loss of generality, so long as  $F_0[n]$  is roughly constant within each processing frame.

For this example, the filter is an ideal brick-wall filter:

$$H(f) = \begin{cases} 1, & |f| < \frac{F_0}{2} \\ 0, & else \end{cases}$$

The coherent harmonic envelopes for each method will be

$$m_{k,harmonic}^{(1)}[n] = A_k[n] \quad (3.19)$$

$$m_{k,harmonic}^{(2)}[n] = A_k[n] e^{j\phi_k[n]} \quad (3.20)$$

Let  $Rect\{y_k[n]\}$  be the half-wave rectified carrier-modulator signal which is the final output. Using the first harmonic method

$$\begin{aligned} y_k^{(1)}[n] &= m_{k,harmonic}^{(1)}[n] \cos(2\pi F_0 n) \\ &= A_k[n] \cos(2\pi F_0 n) \end{aligned} \quad (3.21)$$

Alternatively, the second method results in

$$y_k^{(2)}[n] = Re\{2m_{k,harmonic}^{(2)}[n] e^{j2\pi F_0 n}\} \quad (3.22)$$

$$\begin{aligned} &= Re\{2A_k[n] e^{j(2\pi F_0 n + \phi_k[n])}\} \\ &= A_k[n] \cos(2\pi F_0 n + \phi_k[n]) \end{aligned} \quad (3.23)$$

It is clear from (3.21) and (3.23) that the difference between  $y_k^{(1)}[n]$  and  $y_k^{(2)}[n]$  is simply the extra term,  $\phi_k[n]$ . What this means in terms of information delivered to the user may be best shown by example.

Comparing figure 3.3 top left and top right makes it clear that taking the magnitude forces symmetry about 0Hz. The bottom left,  $y_k^{(2)}[n]$ , better represents the blue than does the bottom right,  $y_k^{(1)}[n]$ , because the spectral asymmetries manifest themselves in the phase, not magnitude. It is unnatural and certainly won't happen in real world scenarios that a subband signal will be symmetric about the downshift frequency, however magnitude only methods force this symmetry.

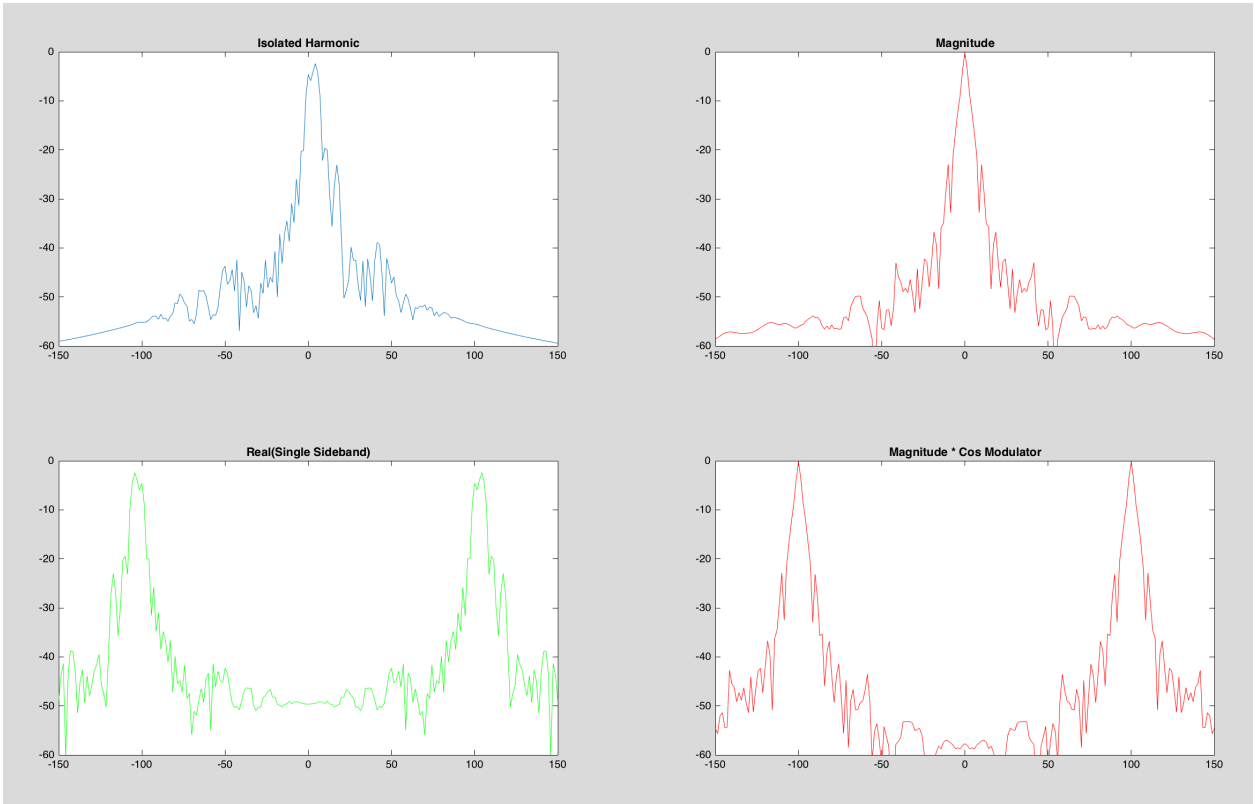


Figure 3.3: Cello Example, top left:  $m_{k,harmonic}^{(2)}[n]$ , top right:  $m_{k,harmonic}^{(1)}[n]$ , bottom left:  $y_k^{(2)}[n]$ , bottom right: 55

### 3.3.1 Appropriate Scaling

Despite better representing the signal, there is still an issue with  $y_k^{(2)}[n]$ . A more correct method is actually

$$m_{k,harmonic}^{(3)}[n] = A_k[n]e^{j\frac{1}{k}unwrap(\phi_k[n])} \quad (3.24)$$

$$\begin{aligned} y_k^3[n] &= Re\{2m_{k,harmonic}^{(3)}[n]e^{j2\pi F_0 n}\} \\ &= A_k[n]\cos\left(2\pi F_0 n + \frac{1}{k}unwrap(\phi_k[n])\right) \end{aligned} \quad (3.25)$$

Why is the  $\frac{1}{k}$  term necessary? Consider an example where our true pitch estimate is actually  $F_{0,groundtruth} = F_0 + F_{err}$ . So,

$$x_k[n] = A_k[n]\cos\left(2\pi k(F_0 + F_{err})n + \phi_k[n]\right) \quad (3.26)$$

In this case,

$$y_k^{(2)}[n] = A_k[n]\cos\left(2\pi(F_0 + \textcolor{red}{k}F_{err})n + \frac{1}{k}unwrap(\phi_k[n])\right) \quad (3.27)$$

$$y_k^{(3)}[n] = A_k[n]\cos\left(2\pi(F_0 + F_{err})n + \frac{1}{k}unwrap(\phi_k[n])\right) \quad (3.28)$$

Essentially the term  $\phi_k[n]$  may be thought of as the deviation from  $kF_0$ . If the signal is downshifted such that  $kF_0 \rightarrow F_0$  then it is appropriate that  $\phi_k[n]$  is scaled similarly. This is not the case for (3.27).

### 3.3.2 Efficacy

So what is the efficacy of 3.19 versus 3.24? One hypothesis is that  $\phi_k[n]$  may encode the noise-like characteristics of a signal, in which case it would remain constant for a pure sinusoid and fluctuate randomly for noise. Put to test, the harmonic phase preservation did little to affect the signal and this was confirmed by testing varying filter bandwidths as well. In comparison

of a toy experiment, the choice of filter bandwidth dominated noise-like qualities, with wider bandwidth capturing more of the variations.

Since the term  $\phi_k[n]$  does not distinguish noise-like signals from narrowband sinusoidal signals, it is only preserving phase alignment. But this begs the question, what does it mean to preserve the phase of a harmonic when downshifted to  $F_0$ ? It is questionable as to whether this even has any logical meaning.

Furthermore, it has been suggested [5] that phase alignment is important for pitch perception in CIs. By using a magnitude-only method we guarantee alignment across channels.

These preliminary tests suggest that phase encoding is not a path worth further investigating, and thus for the rest of this thesis explicit temporal modulation will be evaluated as two independent blocks: envelope extraction and carrier synthesis, each being a strictly real nonnegative signal.

### 3.4 The Relationships

All envelope extraction methods are summarized in table 3.1. We now consider the relationships between each of these.

#### 3.4.1 Hilbert VS STFT

Using the property that the Hilbert transform of a convolution is the convolution of the Hilbert transform on either factor:

$$\begin{aligned}
 x_k^+[n] &= x[n] * h_k[n] + jH\{x[n] * h_k[n]\} \\
 &= x[n] * h_k[n] + x[n] * jH\{h_k[n]\} \\
 &= x[n] * h_k^+[n]
 \end{aligned} \tag{3.29}$$

Defining the filter specifically as

Method	$m_k[n] =$
CIS	$ x[n] * h_k[n]  * h_{lp}[n]$
Hilbert	$ x_k^+[n]  =  x[n] * h_k[n] + j\mathcal{H}\{x[n] * h_k[n]\} $
STFT	$\frac{1}{Nw[0]}  x[n]e^{-j\frac{2\pi}{N}kn} * w[-n] $
Harmonic Coherent	$ x[n]e^{-jk\phi_0[n]} * h[n, F_0[n]] $ , $\phi_0[n] = \frac{2\pi}{F_s} \sum_{p=0}^n F_0[p]$

Table 3.1: Envelope Extraction Methods

$$h_k[n] = \frac{1}{Nw[0]} w[-n] \cos(\frac{2\pi}{N}kn) \quad (3.30)$$

if the sidelobes of  $w[n]$  roll-off sufficiently fast in relation to the center-frequency  $\frac{2\pi k}{N}$ , the Hilbert transform of the filter may be approximated as

$$\begin{aligned} \mathcal{H}\{h_k[n]\} &\approx \frac{1}{Nw[0]} w[-n] H\{\cos(\frac{2\pi}{N}kn)\} \\ &= \frac{1}{Nw[0]} w[-n] \sin(\frac{2\pi}{N}kn) \end{aligned} \quad (3.31)$$

$$h_k^+[n] \approx \frac{1}{Nw[0]} w[-n] e^{j\frac{2\pi}{N}kn} \quad (3.32)$$

Plugging (3.32) into (3.29) results in

$$\begin{aligned}
x_k^+[n] &\approx x[n] * \frac{1}{Nw[0]} w[-n] e^{j\frac{2\pi}{N}kn} \\
&= \frac{1}{Nw[0]} \sum_{r=-\infty}^{\infty} x[n-r] w[-r] e^{j\frac{2\pi}{N}kr}
\end{aligned}$$

Let  $r' = -r$

$$\begin{aligned}
&= \frac{1}{Nw[0]} \sum_{r'=0}^{N-1} x[n+r'] w[r'] e^{-j\frac{2\pi}{N}kr'} \\
&= \frac{1}{Nw[0]} \left[ e^{-j\frac{2\pi}{N}kn} \sum_{r'=0}^{N-1} x[n+r'] w[r'] e^{-j\frac{2\pi}{N}kr'} \right] e^{j\frac{2\pi}{N}kn} \\
&= \frac{1}{Nw[0]} X[n, i] e^{j\frac{2\pi}{N}kn} \\
&= \left( \frac{1}{Nw[0]} x[n] e^{-j\frac{2\pi}{N}kn} * w[-n] \right) e^{j\frac{2\pi}{N}kn} \tag{3.33}
\end{aligned}$$

This relates the envelopes as

$$\begin{aligned}
m_{k,hilbert}[n] &\approx m_{k,STFT}[n] \left| e^{j\frac{2\pi}{N}kn} \right| \\
&= m_{k,STFT}[n] \tag{3.34}
\end{aligned}$$

We come to the conclusion that a filter bank with  $\frac{N}{2} + 1$  filters may designed, (3.35), such that these two methods are equivalent.

$$h_k[n] = w[-n] \cos\left(\frac{2\pi}{N}kn\right), \quad 0 \leq k \leq \frac{N}{2} \tag{3.35}$$

What this tells us is that the Hilbert decomposition may be viewed as a superset of the STFT method that is not constrained to uniform bandwidth linearly spaced filters.

### 3.4.2 STFT vs Harmonic

Following a similar approach, consider a harmonic coherent filter to be time-invariant and defined as

$$h[n, F_0[n]] = \frac{1}{Nw[0]}w[-n] \quad (3.36)$$

where  $w[n]$  is a lowpass filter and

$$w[n] \begin{cases} \neq 0, & 0 \leq n < N \\ = 0, & \text{else} \end{cases}$$

In this case,

$$\begin{aligned} m_{k,harmonic}[n] &= \left| x[n]e^{-jk\phi_0[n]} * \frac{1}{Nw[0]}w[-n] \right| \\ &= \frac{1}{Nw[0]} \left| x[n]e^{-jk\phi_0[n]} * w[-n] \right| \end{aligned} \quad (3.37)$$

This bears striking resemblance to equation 3.13. In the case that  $F_0[n] = \frac{F_s}{N}$ ,

$$m_{k,harmonic}[n] = m_{k,STFT}[n] \quad (3.38)$$

More generally, for any window of time  $n$  to  $n + N - 1$  where  $F_0[n]$  is constant

$$\begin{aligned} m_{k,harmonic}[n] &= \frac{1}{Nw[0]} \left| X[n, NF_0[n]k] \right| \\ &= \frac{1}{Nw[0]} \left| X[n, \lambda[n]k] \right| \end{aligned} \quad (3.39)$$

where  $\lambda[n] = NF_0[n]$ . The “)” denotes that (3.39) is a DTFT.

It is important to note that in practice  $\lambda[n]$  is not a continuous variable. It is constrained by the quantization of the implemented pitch tracker. Provided this quantization it is possible to compute any term  $X[n, \lambda[n]k]$  by zero-padding the DFT.

This implies that in practice,  $m_{k,harmonic}[n]$  can be approximated using  $F_0[n]$  and a zero-padded STFT under the assumptions:

- 1)  $F_0[n]$  is quantized
  - 2)  $F_0[n]$  is roughly constant withing a time window of  $\frac{N}{F_s}$  seconds
- and the restriction:
- 3)  $h[n, F_0[n]]$  is time-invariant, i.e.  $h[n, F_0[n]] = h[n]$

### 3.4.3 CIS VS Hilbert

Stemming from the CIS and Hilbert envelope equations, consider the following two functions.

$$Y_{k,Hilbert}(f) = \mathcal{F} \left\{ \left| x_k^+[n] \right|^2 \right\} \quad (3.40)$$

$$Y_{k,CIS}(f) = \mathcal{F} \left\{ \left| x_k[n] \right|^2 \right\} \quad (3.41)$$

$Y_{k,Hilbert}(f)$  is the DFT of the squared Hilbert envelope.  $Y_{k,CIS}(f)$  is equivalent to the DFT of the squared CIS envelope if the final lowpass filter is not applied.

Provided an ideal brick-wall filter defined as

$$H_k(f) = \begin{cases} 1, & f_k - \frac{1}{2}f_{bw} < |f| < f_k + \frac{1}{2}f_{bw} \\ 0, & \text{else} \end{cases} \quad (3.42)$$

(A.15) and (A.16) are only nonzero within subbands:

$$Y_{k,Hilbert}(f) = \begin{cases} X_k^+(f) * X_k^{*+}(-f), & |f| < f_{bw} \\ 0, & |f| \geq f_{bw} \end{cases} \quad (3.43)$$

$$Y_{k,CIS}(f) = \begin{cases} 2Y_{k,Hilbert}(f), & |f| < f_{bw} \\ 0, & f_{bw} \leq |f| \leq 2f_k - f_{bw} \\ X_k(f) * X_k^*(-f), & 2f_k - f_{bw} < |f| < 2f_k + f_{bw} \\ 0, & |f| \geq 2f_k + f_{bw} \end{cases} \quad (3.44)$$

Lowpass filtering  $Y_{k,CIS}(f)$  by a filter defined



$$H_{lp}(f) = \begin{cases} \frac{1}{2}, & |f| < f_{bw} \\ 0, & 2f_k - f_{bw} < |f| < 2f_k + f_{bw} \end{cases} \quad (3.45)$$

then

$$Y_{k,CIS}(f) = Y_{k,Hilbert}(f) \quad \forall f \quad (3.46)$$

Thus provided the proper filter designs

$$\left| x_k[n] \right|^2 * h_{lp}[n] \approx \left| x_k^+[n] \right|^2 \quad (3.47)$$

Things to consider are delay and non-ideal filters, however provided the distance between baseband and the  $\pm 2f_k$  terms in (3.44) a sufficient filter is feasible in practice.

Now the relationship between  $m_{k,CIS}[n]$  and  $m_{k,Hilbert}[n]$  is muddled by the nonlinear square root operation, however the nonlinearities induced won't be noticeably distorted by  $h_{lp}[n]$ . In practice, the only significant difference will be the added delay from the final lowpass filter in the CIS method.

#### 3.4.4 Abstract Interpretation

One of the easier ways to interpret the methods is through a frequency domain analysis. Figure 3.4 shows an abstract view for a simple two harmonic example. For mathematical convenience the output (orange) is actually the squared envelope. At each step a new operation is applied. This abstract analysis ignores scale factors that can always be modified by scaling filter coefficients.

First note that there are two paths for STFT. This is because there is an ambiguity in the order of operations. This can be seen mathematically in (3.48).

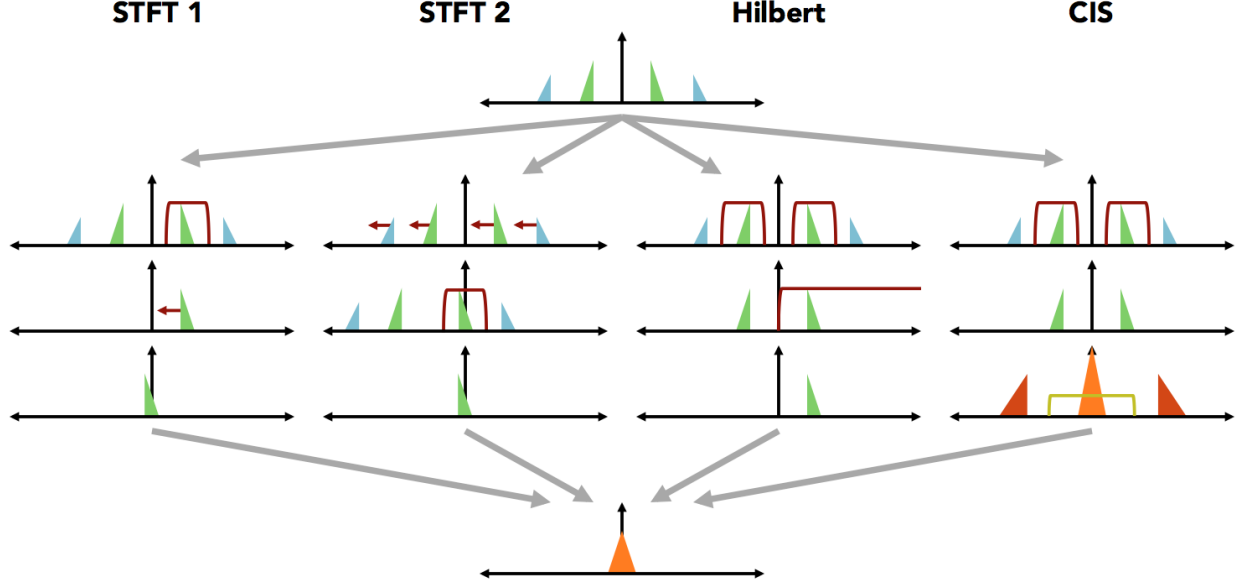


Figure 3.4: Method Comparison: magnitude spectrum at each step

$$e^{-j\frac{2\pi}{N}kn} \left( x[n] * \left( w[-n] e^{j\frac{2\pi}{N}kn} \right) \right) = \left( x[n] e^{-j\frac{2\pi}{N}kn} \right) * w[-n] \quad (3.48)$$

The left side of (3.48) corresponds to the STFT 1 path. First an analytic bandpass filter centered at radian frequency  $\frac{2\pi k}{N}$  is applied. The output of that is then downshifted to baseband. The right side of (3.48) corresponds to the STFT 2 path. The signal is first downshifted by radian frequency  $\frac{2\pi k}{N}$ , then lowpass filtered. For both STFT 1 and STFT 2 the final operation is a magnitude squared.

The harmonic coherent method is missing from figure 3.4. This is because, ignoring exact details of downshift frequency and filter coefficients, it is actually the same as the STFT method: downshift followed by lowpass filter.

Moving on the the Hilbert envelope, in figure 3.4 the signal is first bandpass filtered, the analytic signal is acquired, which is equivalent to setting the negative frequencies to zero.

The final operation is to take the magnitude squared, which is invariant to frequency shifts. Because the magnitude is invariant to frequency shifts, this result should be the same as the STFT method.

For CIS, taking the magnitude squared of the real bandpass signal causes double frequency terms, and the baseband term is scaled by a factor of 2. The final filter operation (yellow) rescales the baseband term and eliminates the double frequency terms.

### 3.5 Summary

So what are the differences? To come to the conclusions made, some assumptions had to be made. We found that the Hilbert and CIS methods are approximately the same. STFT decomposition is a subset of the Hilbert method where the filterbank is comprised of uniform-bandwidth linearly spaced filters. Coherent harmonic is an expansion of STFT decomposition using the fundamental frequency of a signal to adaptively change downshift frequency and filter bandwidth.

Excluding CIS, the other three methods can all be derived from the generalized equation (3.49).  $h_k[n, F_0[n]]$  is a function of  $k$  allowing for non-uniform bandwidths and a function of  $F_0[n]$ , allowing for coherent filter adaptation. Similarly,  $\omega_k[F_0[n]]$  is a function of  $F_0[n]$ , allowing for coherent downshift frequencies.

$$m_k[n] = \left| x[n] e^{-j\omega_k[F_0[n]]n} * h_k[n, F_0[n]] \right| \quad (3.49)$$

In the next chapter we will investigate encoding harmonics in cochlear implants using this generalized envelope extraction equation.

## Chapter 4

### HARMONIC ENVELOPE EXTRACTION

The objective of this chapter is to design an envelope extraction system that best represents harmonic signals. To do this there must be an ideal envelope to aim for. A harmonic signal is modeled as a restricted sum-of-products model. The carriers are sinusoids centered at multiples of  $F_0$ . In this representation  $x_0[n]$  is the fundamental centered at  $F_0$ ,  $x_1[n]$  is the 1st harmonic centered at  $2F_0$ , etc. Without loss of generality, the analytic signal will be considered,  $x^+[n]$ .

$$\theta_k[n] = 2\pi(k+1)\frac{F_0[n]}{F_s}n + \phi_k[n] \quad (4.1)$$

$$x^+[n] = \sum_{k=0}^K x_k^+[n] = \sum_{k=0}^K m_k[n] e^{j\theta_k[n]} \quad (4.2)$$

We change our notation slightly from chapter 3. In this chapter  $m_k[n]$  is the unknown desired envelope, and  $\tilde{m}_k[n]$  is our extracted envelope estimate.

$$\tilde{m}_k[n] = \left| x[n] e^{-j\omega_k[F_0[n]]n} * h_k[n, F_0[n]] \right| \quad (4.3)$$

Provided our envelope extraction equation, 4.3, the goal is to best represent the desired  $m_k[n]$ .

The design can be summarized by two things:

- downshift frequency,  $\omega_k[F_0[n]]$
- lowpass filter,  $h_k[n, F_0[n]]$

If  $w_k[\cdot]$  and  $h_k[\cdot]$  are functions of  $x^+[n]$  this is coherent envelope extraction. If they are time-invariant, its incoherent extraction.  $\omega_k[n]$  is defined in (4.4) such that it is equivalently represented by  $F_{ds,k}[n]$ , the downshift frequency in Hz.

$$\omega_k[n] = 2\pi \frac{F_{ds,k}[n]}{F_s} \quad (4.4)$$

#### 4.1 Steady-State Analysis

The simplest scenario is when  $x^+[n]$  is a steady-state signal. The conditions required for this are:

- constant pitch:  $F_0[n] = F_0$
- narrowband modulator:  $m_k[n]$  is slow-time-varying, i.e.  $m_k[n] \approx \text{constant}$  over very short periods of time
- constant phase term:  $\phi_k[n] = \phi_k$ , it is assumed  $\phi_k[n] = 0$  for simplicity however this is not necessary

##### 4.1.1 3 Harmonic Example: Desired Envelope

The frequency domain for a signal with three harmonics ( $K = 2$ ) is visualized in figure 4.1. For this example, the considered envelope is for the 1st harmonic ( $k = 1$ ), centered at  $2F_0$ .

Figure 4.1(d) is the spectrum of the squared envelope,  $\mathcal{F}\{m_1^2[n]\}$ . This relationship is shown in (4.8)

$$(a) \quad x^+[n] \Longleftrightarrow X^+[n, f] \quad (4.5)$$

$$(b) \quad x_1^+[n] \Longleftrightarrow X_1^+[n, f] \quad (4.6)$$

$$(c) \quad x_1^{*+}[n] \Longleftrightarrow X_1^{*+}[n, -f] \quad (4.7)$$

$$(d) \quad m_1^2[n] = x_1^+[n]x_1^{*+}[n] \Longleftrightarrow X_1^+[n, f] * X_1^{*+}[n, -f] \quad (4.8)$$

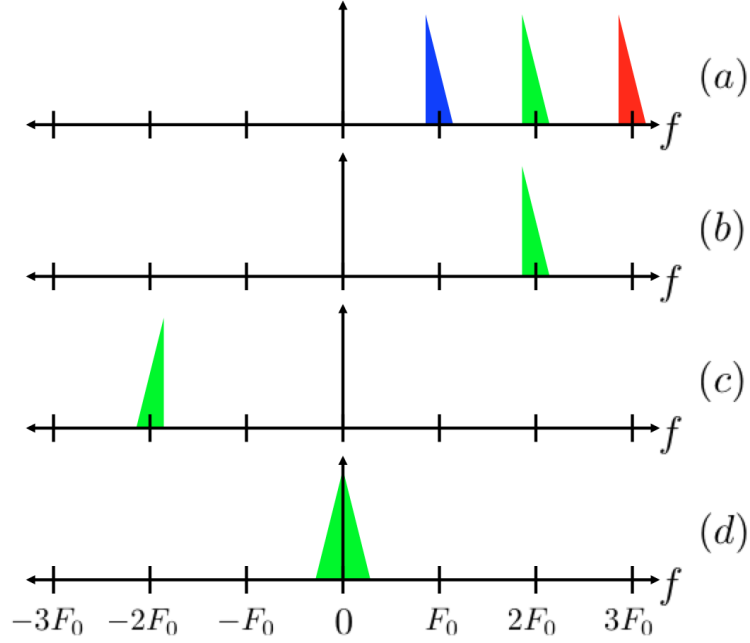


Figure 4.1: Magnitude of spectrum for equations (4.5) - (4.8)

The envelope can always be acquired from the squared envelope by a final square root operation. This operation introduces nonlinearities at multiples of  $F_0$  that are difficult to analyze. For mathematical convenience, during analysis only the squared envelope will be considered. This final square root operation will remain constant across all examples.

$$m_1[n] = \left| x_1^+[n] \right| = \left[ x_1^+[n] x_1^{*+}[n] \right]^{\frac{1}{2}} \quad (4.9)$$

#### 4.1.2 Estimated Envelope

The estimated envelope is acquired using (4.3). As stated above, we consider the squared envelope. (4.2) is substituted into (4.10). The approximation in (4.12) comes from the assumption that  $m_k[n] \approx \text{constant}$  over short time windows.  $f_{k,l}$  is defined as the downshifted center frequency of the  $l$ th harmonic for the estimate of the  $k$ th envelope.

$$\tilde{m}_k^2[n] = \left| x^+[n] e^{-j\omega_k n} * h_k[n] \right|^2 \quad (4.10)$$

$$= \left| \sum_{l=0}^K m_l[n] e^{j(\theta_l[n] - \omega_k[n])} * h_k[n] \right|^2 \quad (4.11)$$

$$\approx \left| \sum_{l=0}^K m_l[n] \left( e^{j(\theta_l[n] - \omega_k[n])} * h_k[n] \right) \right|^2 \quad (4.12)$$

$$= \left| \sum_{l=0}^K m_l[n] e^{j2\pi \frac{f_{k,l}}{F_s} n} H_k(f_{k,l}) \right|^2 \quad (4.13)$$

$$\begin{aligned} f_{k,l} &= \frac{F_s}{2\pi} (\theta_l[n] - \omega_k[n]) \\ &= (l+1)F_0 - F_{ds,k} \end{aligned} \quad (4.14)$$

Expanding (4.13) results in

$$\begin{aligned} \tilde{m}_k^2[n] &= \sum_{l=0}^K \sum_{i=0}^K m_l[n] m_i^*[n] e^{j2\pi \frac{(l-i)F_0}{F_s} n} H_k(f_{k,l}) H_k^*(f_{k,i}) \\ &= \sum_{l=0}^K \left| m_l[n] \right|^2 \left| H_k(f_{k,l}) \right|^2 \\ &\quad + e^{-j2\pi \frac{F_0}{F_s} n} \sum_{l=0}^{K-1} m_l[n] m_{l+1}^*[n] H_k(f_{k,l}) H_k^*(f_{k,l+1}) \\ &\quad + e^{j2\pi \frac{F_0}{F_s} n} \sum_{l=1}^K m_l[n] m_{l-1}^*[n] H_k(f_{k,l}) H_k^*(f_{k,l-1}) \\ &\quad + e^{-j2\pi \frac{2F_0}{F_s} n} \sum_{l=0}^{K-2} m_l[n] m_{l+2}^*[n] H_k(f_{k,l}) H_k^*(f_{k,l+2}) \\ &\quad + e^{j2\pi \frac{2F_0}{F_s} n} \sum_{l=2}^K m_l[n] m_{l-2}^*[n] H_k(f_{k,l}) H_k^*(f_{k,l-2}) \\ &\quad + \dots \\ &\quad + e^{-j2\pi \frac{KF_0}{F_s} n} m_0[n] m_K^*[n] H_k(f_{k,0}) H_k^*(f_{k,K}) \\ &\quad + e^{j2\pi \frac{KF_0}{F_s} n} m_K[n] m_0^*[n] H_k(f_{k,K}) H_k^*(f_{k,0}) \end{aligned} \quad (4.16)$$

From (4.16),  $\tilde{m}_k[n]$  can be thought of as a sum of terms each centered at  $iF_0$  where the magnitude of each term is

$$\left| \tilde{m}_{k,i}[n] \right| = \left[ \sum_{l=0}^{K-|i|} \left| m_l[n] \right| \left| m_{l+i}[n] \right| \left| H_k(f_{k,i}) \right| \left| H_k(f_{k,l+i}) \right| \right]^{\frac{1}{2}}, \quad -K \leq i \leq K \quad (4.17)$$

Evaluated at DC:

$$\left| \tilde{m}_{k,0}[n] \right| = \left[ \sum_{l=0}^K \left| m_l[n] \right|^2 \left| H_k(f_{k,l}) \right|^2 \right]^{\frac{1}{2}} \quad (4.18)$$

#### 4.1.3 3 Harmonic Example: Estimated Envelope

Returning to the three harmonic example, the goal is to acquire the 1st harmonic,  $m_1[n]$  (green).

Spectra of (4.19) -(4.22) are visualized in figure 4.2. In this example  $F_{ds,k} = 2F_0$ .

$$x^+[n] \Longleftrightarrow X^+[n, f] \quad (4.19)$$

$$x^+[n]e^{-j2\pi\frac{2F_0}{F_s}n} \Longleftrightarrow X^+[n, f - 2F_0] \quad (4.20)$$

$$x^+[n]e^{-j2\pi\frac{2F_0}{F_s}n} * h_2[n] \Longleftrightarrow X^+[n, f - 2F_0]H_1(f) \quad (4.21)$$

$$\tilde{m}_1^2[n] \Longleftrightarrow X^+[n, f - 2F_0]H_1(f) * X^{*+}[n, -f + 2F_0]H_1^*(-f) \quad (4.22)$$

The interesting part of figure 4.2 is (f). The desired green component is present, however there are a whole lot of other things present.

Figure 4.1(d) is equivalent to the green component of figure 4.2(f) if the filter  $|H_1(f)| = 1$  when  $f \approx 0$ .

The other components come from interactions with the unwanted harmonics that were not completely filtered out. For clarity the convolution is visualized in figures 4.3, 4.4, 4.5. Positive and negative components of figure 4.2 (f) are mirror images, so the positive components are not explicitly visualized.



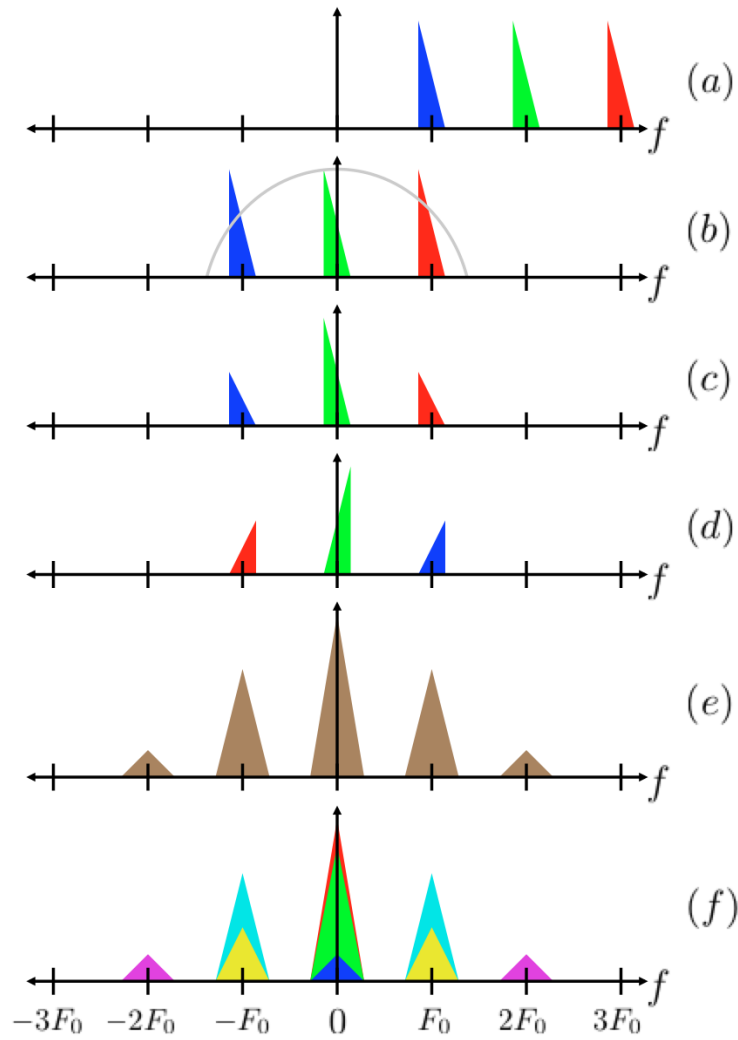


Figure 4.2: (a) – (c): magnitude of spectrum for equations (4.19) – (4.21), (d): time-reversal of (c), (e): magnitude of spectrum for (4.22), (f): contributions of separate components of (e)

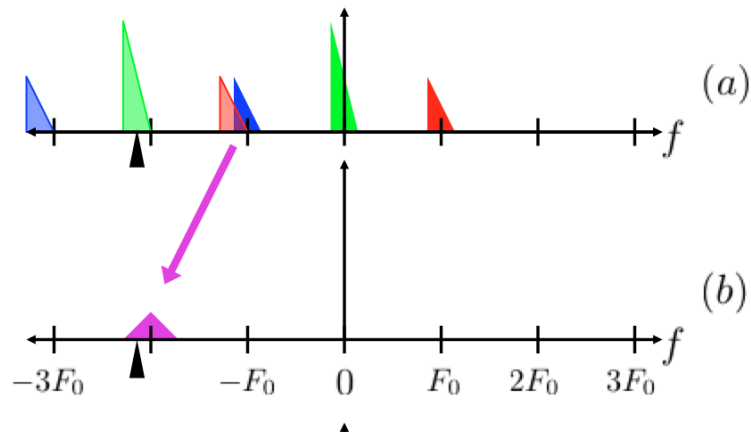
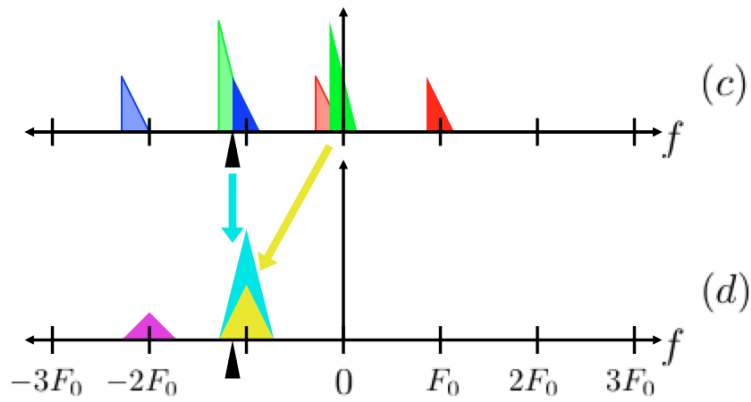
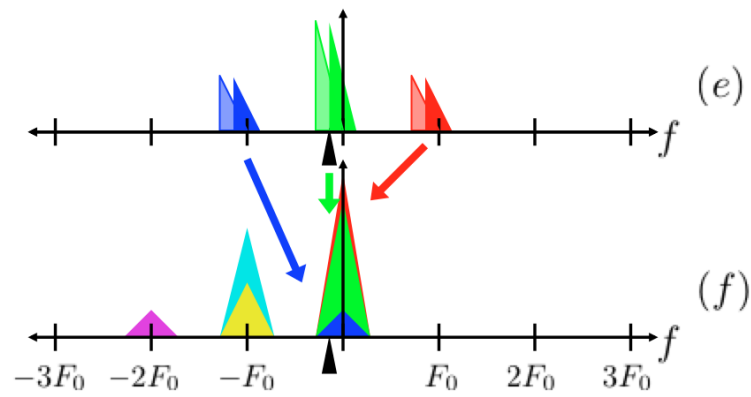
Figure 4.3: Envelope Estimate  $-2F_0$  ComponentFigure 4.4: Envelope Estimate  $-F_0$  Component

Figure 4.5: Envelope Estimate Baseband Component

#### 4.1.4 3 Harmonic Example: Estimated Envelope with downshift error

The 3 harmonic example has demonstrated how a non-ideal filter causes distortions in the estimate. Now consider the effects of downshift frequency:  $F_{ds,k} = kF_0 + F_{err}$ .

$$x^+[n] \Longleftrightarrow X^+[n, f) \quad (4.23)$$

$$x^+[n]e^{-j2\pi\frac{2F_0+F_{err}}{F_s}n} \Longleftrightarrow X^+[n, f - 2F_0 - F_{err}) \quad (4.24)$$

$$x^+[n]e^{-j2\pi\frac{2F_0+F_{err}}{F_s}n} * h_2[n] \Longleftrightarrow X^+[n, f - 2F_0 - F_{err})H_1(f) \quad (4.25)$$

$$\tilde{m}_1^2[n] \Longleftrightarrow X^+[n, f - 2F_0 - F_{err})H_1(f) * X^{*+}[n, -f + 2F_0 + F_{err})H_1^*(-f) \quad (4.26)$$

In figure 4.6 the non-ideal downshift affects the relative amplitudes of the desired harmonic and interference harmonics when filtering.

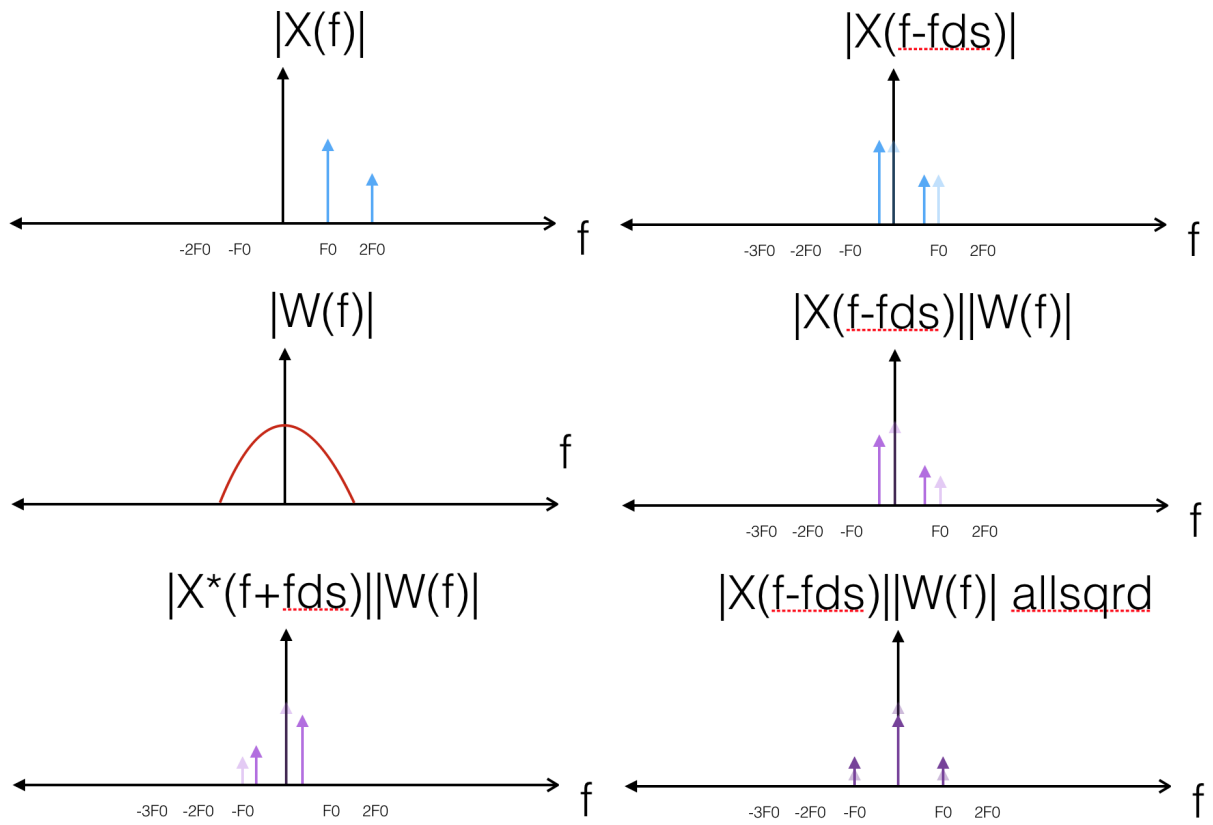


Figure 4.6: (a) – (c): magnitude of spectrum for equations (4.23) – (4.25), (d): time-reversal of (c), (e): magnitude of spectrum for (4.26), (f): contributions of separate components of (e)

## 4.2 Steady-State Metrics

In considering how well our envelope  $\tilde{m}_k[n]$  estimates  $m_k[n]$  there are three important metrics. Each will now be discussed in detail.

### 4.2.1 Coherent Gain

Coherent gain is defined as the gain of the harmonic of interest,  $k$ .

$$G_k = \left| H_k(f_{k,k}) \right| \quad (4.27)$$

Recalling (4.14), if  $F_{ds,k} = (k+1)F_0$  then,  $w_{k,k} = 0$  and the coherent gain is simply the DC gain of the filter.

$$G_k = \left| H_k(0) \right| = \sum_n h_k[n] \quad (4.28)$$

This may be further simplified by normalizing the filter such that  $\left| H_k(0) \right| = 1$ . Of course, the downshift frequency won't be ideal in real systems. Factors to consider include the quantization of  $F_{ds,k}$  and the accuracy of  $F_0$  estimation.

A similar metric, discussed [2] is termed scalloping loss, or picket-fence effect. This is the effect of the harmonic falling in between filter centers, where the gain reduction is greatest.

### 4.2.2 Harmonic SIR

Continuing a focus on the baseband, another question is: what is the contribution of the target harmonic relative to the others? The baseband component is contributed to by spectral leakage due to non-ideal filters. This is visualized as the red and blue in figure 4.5(f). The harmonic signal-to-interference-ratio (SIR) quantifies the ratio of target harmonic to spectral leakage.

$$SIR_k = \frac{|H_k(f_{k,k})|}{\left[ \sum_{l=0}^K |H_k(f_{k,l})|^2 \right]^{\frac{1}{2}}} \quad (4.29)$$

(4.29) is derived from (4.18) by setting  $|m_k[n]|^2 = 1, \quad \forall k$ . Harmonic SIR does not describe the true signal-dependent SIR, as varying envelope magnitudes across harmonics will change this, however it does provide an objective measure of the quality of a system to arbitrary harmonic inputs.

The terms will roll off as the harmonic center frequencies get further away from  $F_{ds,k}$ , so typically  $SIR_k$  is sufficiently described by only one or two harmonics on either side of the  $k$ th, i.e.  $k-2 \leq l \leq k+2$ .

#### 4.2.3 Modulation Depth

The final metric quantifies the magnitude of each bandpass component relative to baseband. These terms appear in the envelope estimate as modulations at rates that are multiples of  $F_0$ . Because of the forced symmetry of the real envelope, this metric is sufficiently described by only positive frequencies,  $iF_0$ .

$$D_{k,i} = \frac{\left[ \sum_{l=0}^{K-i} |H_k(f_{k,l})| |H_k(f_{k,l+i})| \right]^{\frac{1}{2}}}{\left[ \sum_{l=0}^K |H_k(f_{k,l})|^2 \right]^{\frac{1}{2}}}, \quad 1 \leq i \leq K \quad (4.30)$$

(4.30) is derived from (4.17). Typically, the largest value and for that reason most important value is  $D_{k,1}$ , the modulation depth at  $F_0$ . However, depending on filter design  $F_0$  could be near a filter zero, resulting in  $D_{k,1} < D_{k,2}$ .

### 4.3 Induced VS Explicit Temporal Modulation

The three metrics are coherent gain, harmonic SIR and modulation depth. We aim for a coherent gain of  $G_k = 1$  and maximized harmonic SIR.

In section 2.1.1 it was mentioned that temporal modulations are either induced or explicit. For explicit modulation systems the goal is to minimize envelope modulation depth. For induced that is not as clear.

In this document we argue that the latter, explicit modulation option is better. The reasoning is best shown by a motivational example.

Consider a single note played by two different instruments: clarinet and saxophone. In this example  $F_0 = 261Hz$ . The clarinet is an interesting instrument in that it only has energy at even harmonics,  $k = 0, 2, 4...$

We attempt to estimate the 2nd harmonic,  $m_2[n]$ . We first downshift by  $-3F_0$ , then lowpass filter. The spectrum of each signal at this stage is visualized in figure 4.7. The top panel shows the output of a sufficiently narrow filter where the 3rd harmonic is isolated. The bottom panel shows a different filter design that intentionally allows the two adjacent harmonics to pass through. The problem starts to become apparent here. Despite the wide filter bandwidth, there is (almost) no energy around  $\pm F_0$  for the clarinet because of the harmonic structure. (There is something present however it's down 30dB.)

Figure 4.8 shows the time-domain envelopes resulting from this processing. The input signals were normalized such that the top panel shows the same signal power for both instruments. The problem is clearly represented in the bottom panel, where there is a very large  $F_0$  modulation in the saxophone envelope but little to no change in the clarinet. The result is that there is a much stronger temporal pitch cue as well as louder overall volume to the saxophone.

Spectral leakage into other harmonic envelopes is not natural. It forces the envelope to modulate as a function of the adjacent harmonics which, as we just saw, is signal dependent.

Beyond this example, explicit modulation decouples  $F_0$  and modulation depth. This way,

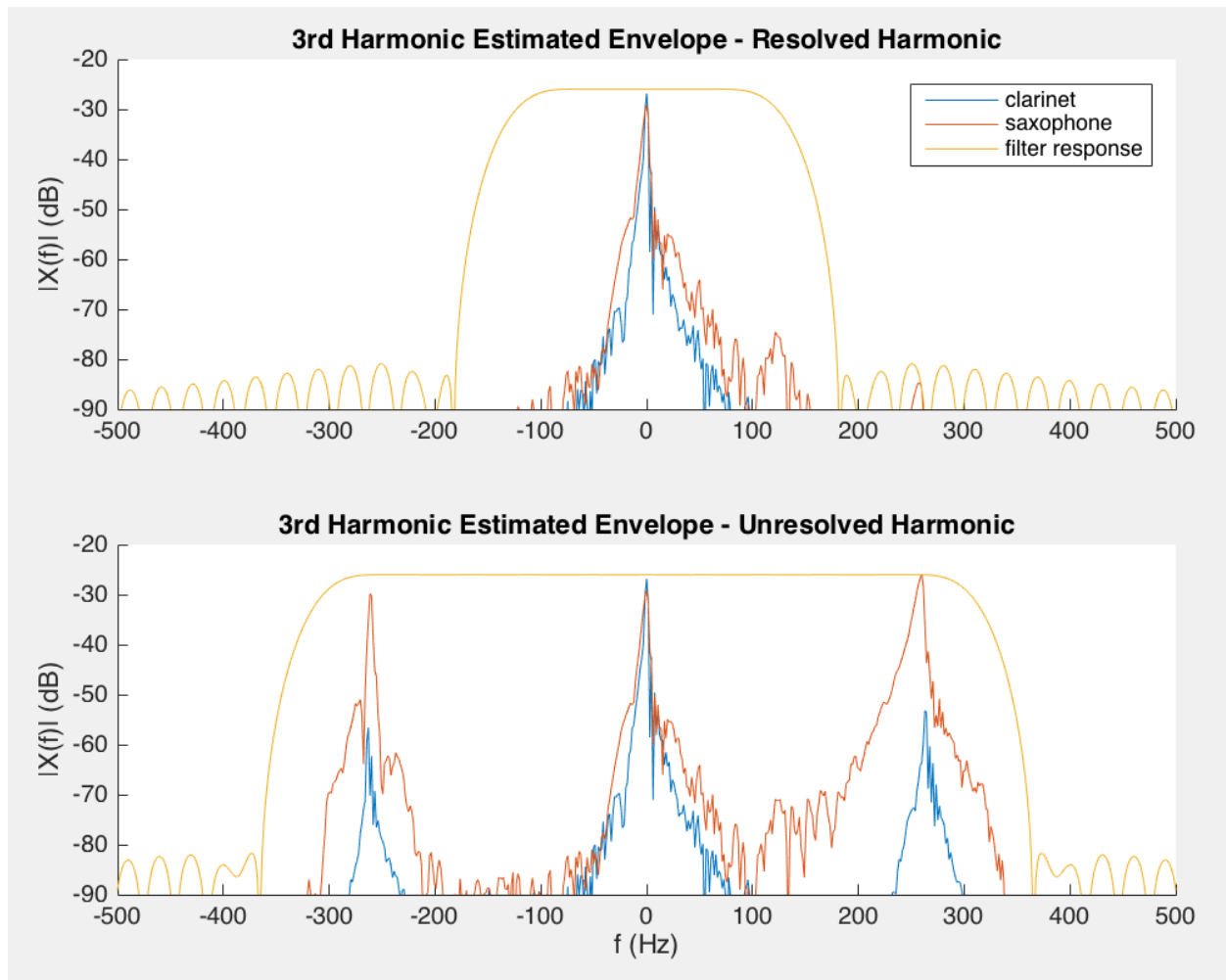


Figure 4.7: Clarinet vs Saxophone Harmonic Components

during system design there is much more control over modulation depth while still making optimal design decisions for envelope extraction. For example, modulation depth can be determined as a function of how harmonic the signal is. The eTone strategy [16] uses a harmonic probability metric to do just that.



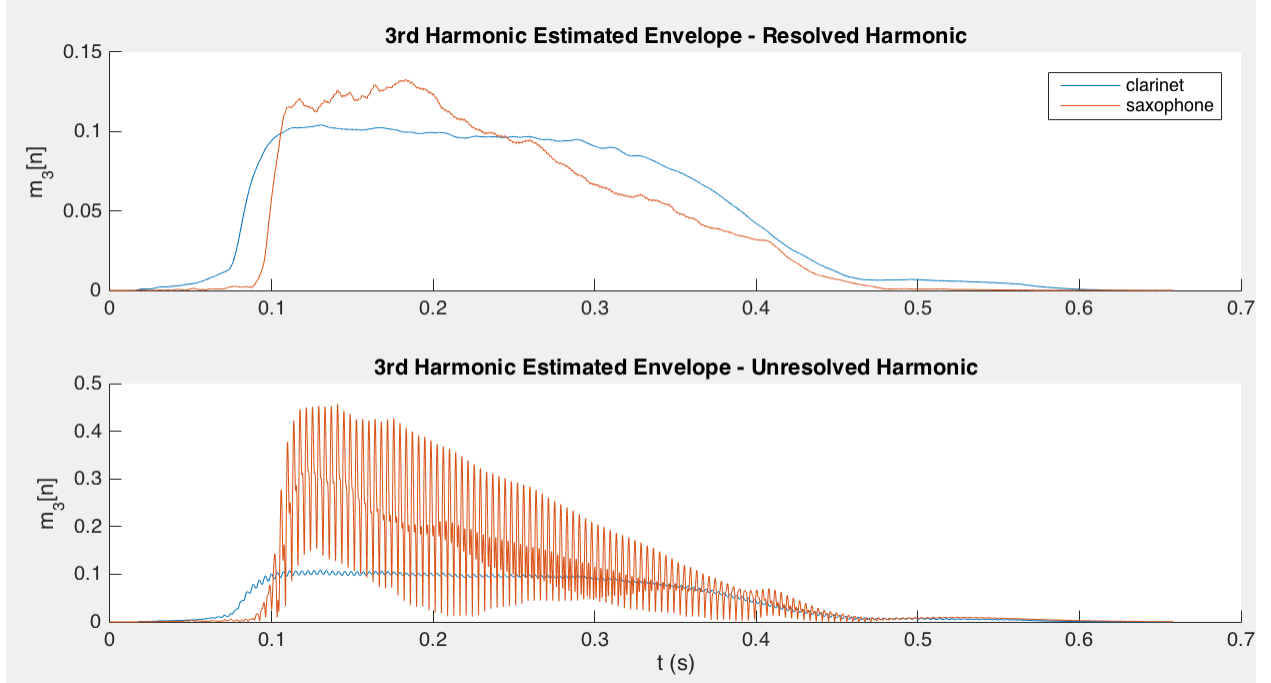


Figure 4.8: Clarinet vs Saxophone Envelope Estimates

#### 4.3.1 Followup Filter

Another thing to note is that regardless of downshift frequency, the harmonic envelope will always have its energy centered at baseband and integer multiples of  $F_0$ . An alternative way of eliminating induced modulations is to add a lowpass filter to the end of the processing chain. There are a handful of research strategies that have used this additional filter. eTone's envelope follower is an example of this.

The main improvement to a followup filter is that the system can guarantee to eliminate temporal modulations. This could also be achieved by designing a sufficiently narrow filter,  $h_k[n]$  however this brings about a tradeoff, where the narrower the filter is the more susceptible the system is to error in downshift frequency.

In terms of the three metrics, the followup filter will provide a robust coherent gain and guaranteed low modulation depth at the cost of lower harmonic SIR. Another point to consider is the cost of adding an extra processing stage. The additional stage means more

memory, clock cycles and processing delay.

#### 4.4 *Time-Varying $F_0$*

We shift focus to time-varying dynamics. Only continuous changes in  $F_0[n]$  are important to system design. Jumps would imply different harmonic envelopes.

$\tilde{m}_k[n]$  uses a window of samples of  $x^+[n]$ , equal to the length of  $h_k[n, F_0[n]]$ . If  $F_0[n]$  changes significantly within this window there will be problems with the estimate. That being said, the longest windows considered in this document are 32ms long. In terms of music, 32ms is equivalent to a sixty-fourth note at 120BPM (beats per minute), i.e. very very fast. We will consider this sufficient for typical rates of change of  $F_0[n]$ .

Even though the window is sufficiently short, the steady-state metrics are a function of  $F_0$  and thus if  $F_0$  changes with time, the metrics may change as well. The effects this has on system performance can be evaluated by simply looking at the continuous metrics as a function of  $F_0$ .

#### 4.5 *Transients*

Nearly everything considered so far has suggested the narrower the filter, the better. The problem with this is the time-domain response of filters with fast rolloffs. There is a tradeoff where the sharper a filter rolls off is, the more transient smearing will incur.

Studies on timbre perception [3] have suggested that for both acoustic and electric hearing humans hear changes in rise time in the log domain, i.e. the shorter a transient is, the more sensitive our perception is to smearing distortion.

Of course if the pre-processing smears the transients, a system is limited in how well it can perform. Most cochlear implants nowadays use pre-processor dynamic range compression. Some insight is gained from a study performed on hearing aids, which would use a similar compression system. “Almost all of the hearing aids tested have attack times less than or equal to 10 ms.” [?]. 1ms is faster than most classic instrument attack times, so transient smearing should be as little as possible in the envelope extraction processing.

All of this suggests filter bandwidth be as wide as possible without encompassing the other harmonics, which is a lowpass cutoff of  $\frac{F_0}{2}$ .

#### 4.6 Evaluation of Strategies

As stated above the design can be summarized by downshift frequency and lowpass filter. The ideal downshift frequency is simply  $(k+1)F_0[n]$ . The question is what degree of quantization is sufficient to estimate the harmonic signal. For filter design we need to consider bandwidth as a function of filter order and filter/window type. Ideally, the cutoff is somewhere below  $F_0$  but high enough to incorporate the bandwidth of  $m_k[n]$ .

The filters can be different as a function of  $k$ . This is a natural path to pursue, considering the critical bands of the cochlea. This will be discussed later in this thesis however for now,  $h_k[n] = h[n]$ . This is natural for harmonic envelopes as harmonics are linearly spaced.

The designs considered are:

downshift quantization,  $F_q$  - 1, 31, 63, 125Hz

filter order,  $N$  - 128, 256, 512

filter design - rectangular, hanning, adaptive hamming

$k$  - which harmonic, (is performance different for different  $k$ ?)

Adaptive hamming is an adaptive bandwidth filter with a lowpass cutoff (-6dB point) of  $\frac{F_0[n]}{2}$ . For practical considerations, maximum quantization =  $\frac{F_s}{N}$  which is the quantization of a order- $N$  DFT.

$$N = 256 \longleftrightarrow F_q \leq 63Hz$$

$$N = 512 \longleftrightarrow F_q \leq 31Hz$$

Only fundamental frequencies in the range of 50-550Hz are considered. This range encompasses the adults and children which are predominantly within 80-300Hz [15] and it provides some extra range for musical instruments.

The following subsections will evaluate each of the three steady-state metrics as well as transient response as functions of system design.

#### 4.6.1 Coherent Gain

We first look at different downshift quantizations, all else constant. This is visualized in figure 4.9. When  $F_0$  is exactly at a quantized value,  $G_k = 0\text{dB}$ , however the gain decreases as  $F_0$  drifts away until the worst case where it is exactly in between quantization values. Decreasing the quantization increases the number of dips and in turn improves the worst case  $G_k$ .

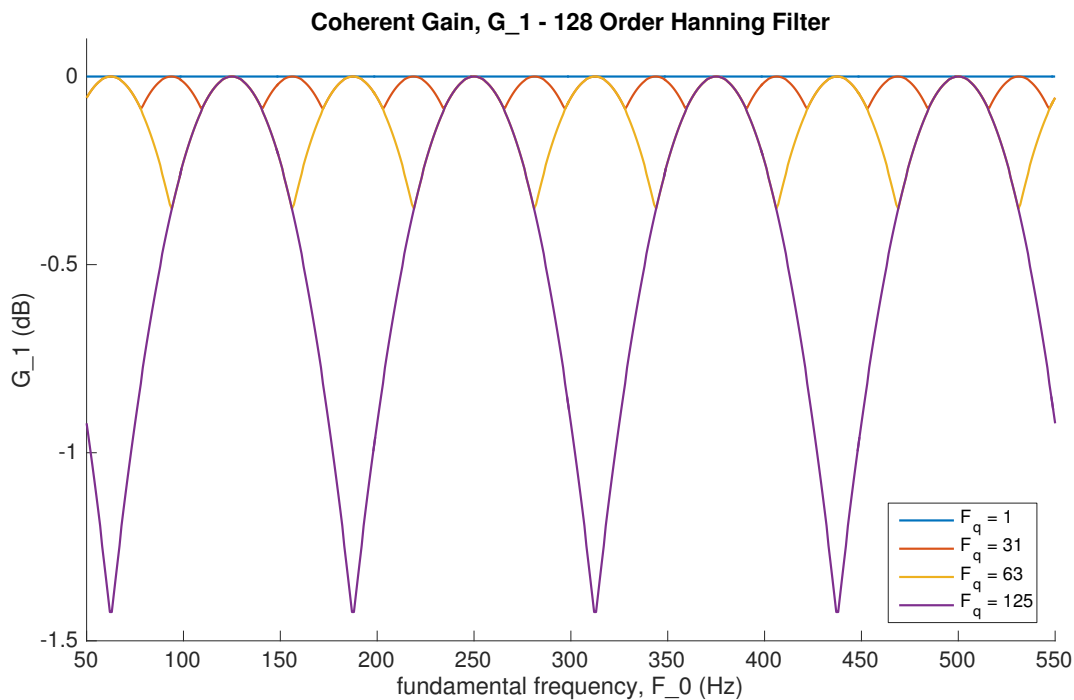


Figure 4.9:  $G_k$  Downshift Quatization

Figure 4.10 compares the three different filter orders. Using a hanning window, the lower order filters have slower rolloffs and better worst case  $G_k$ . This doesn't necessarily hold

true for adaptive filters. Provided a high enough desired cutoff that the 128 order filter can achieve this reasonably well, the order become irrelevant.

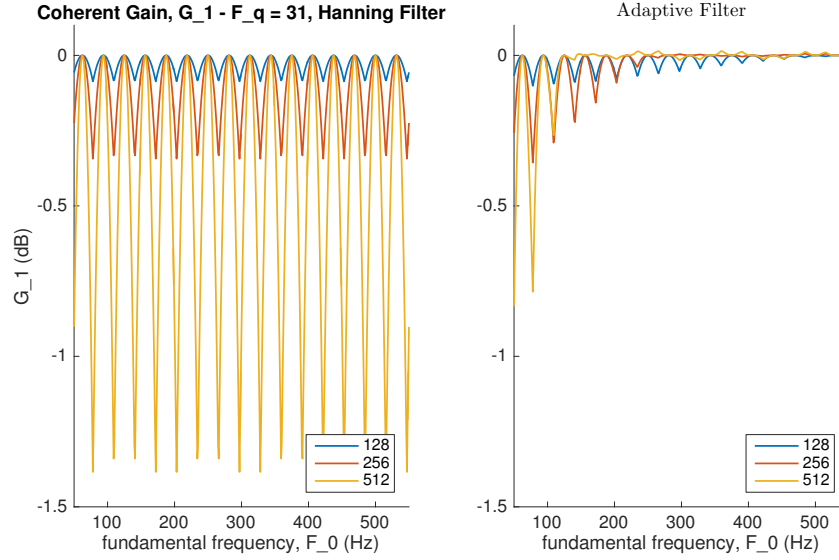


Figure 4.10:  $G_k$  filter order

Figure 4.11 compares the different filter designs. The wider bandwidth filters have smoother  $G_k$  across  $F_0$  and as a result the adaptive bandwidth becomes optimal at high  $F_0$ 's.

So lower quantization and wider bandwidth both improve  $G_k$ , but that's pretty intuitive. The interesting part here is the relationship between harmonics. Consider the first three harmonics; figure 4.12 shows that the number of dips is proportional to  $k$ . As a result,  $G_k$  varies more across  $k$  at certain values of  $F_0$ . For example, if  $F_0 = 1.5F_q = 188\text{Hz}$ , even harmonics will be at a minimum and odd harmonics will be at a maximum. This results in a distortion between harmonics where some are attenuated more than others.

It should be noted that pre-processing compression or automatic gain control (AGC) will cause harmonic distortions. This could arguably be used to either make the case that it is important to minimize further distortions, or alternatively that these further distortions are

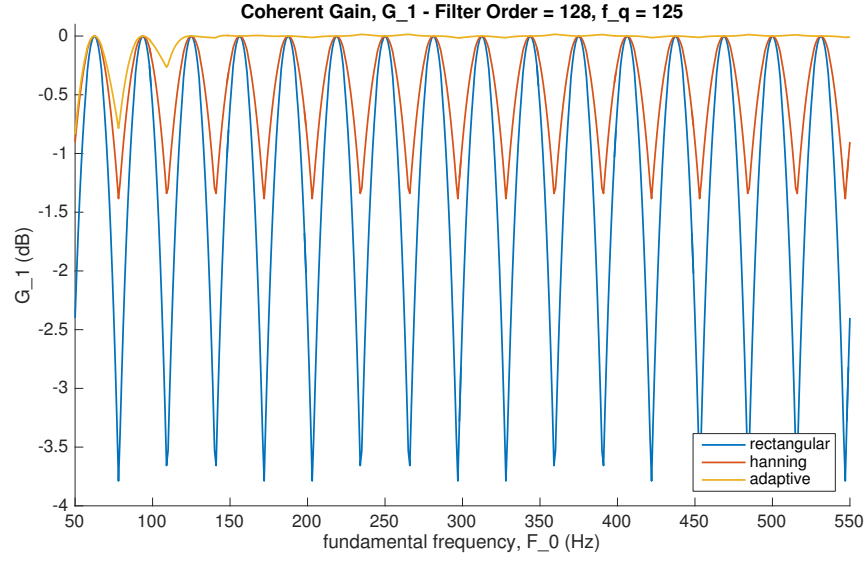


Figure 4.11:  $G_k$  filter design

minimal in comparison and thus shouldn't be over engineered.

Considering maximum quantization is  $\frac{F_s}{N}$  and hanning filter as a baseline, worst case:  $G_k \approx -1.5dB$ . Proportionally increasing the filter order and decreasing quantization increases the number of dips while keeping depth the same. The relationship between harmonics and the case of  $F_0[n]$  continuously changing over time put emphasis on minimizing the dynamic range of  $G_k$ .

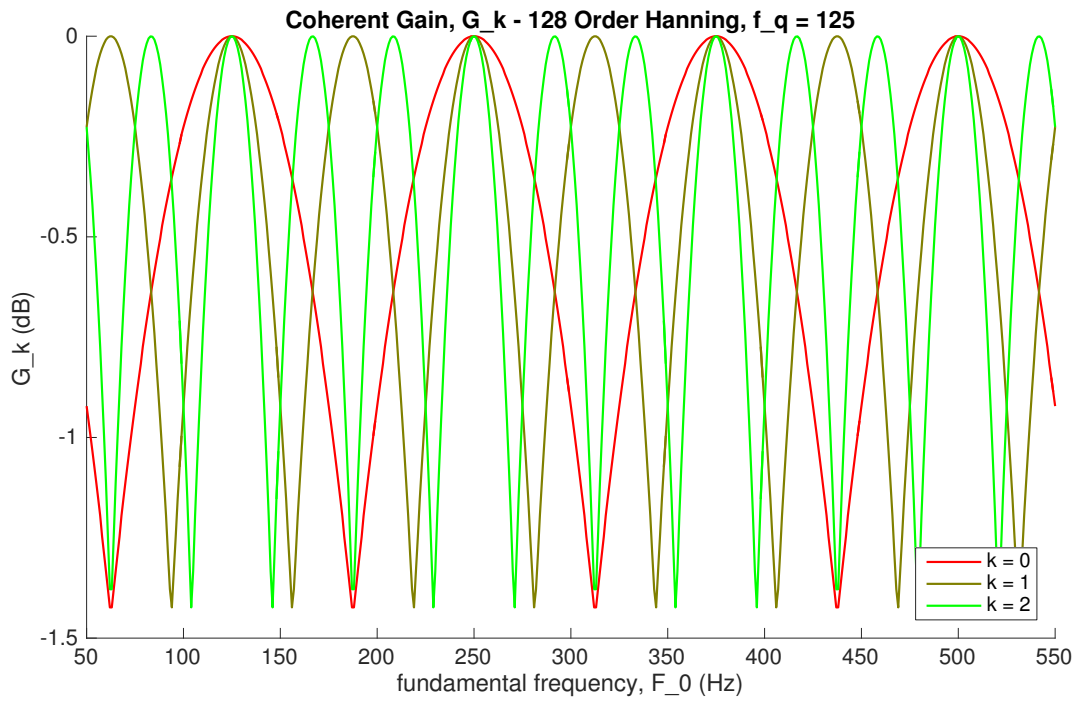


Figure 4.12:  $G_k$  variation across harmonics

#### 4.6.2 Harmonic SIR

Moving along to harmonic SIR, figure 4.13 compares all filter orders with and without quantization. The downshift quantization doesn't actually affect performance significantly. This can be seen in figure 4.13 by looking at the two lines corresponding to order = 128. Above  $F_0 = 250\text{Hz}$  the harmonics are spaced far enough apart that the quantization doesn't matter. Below  $F_0 = 130\text{Hz}$  the filter cutoff is not sharp enough to isolate the harmonic, in which case downshift quantization is irrelevant.

Also note that for order = 512 the cutoff is narrow enough that harmonic SIR is ideal over all  $F_0$ .

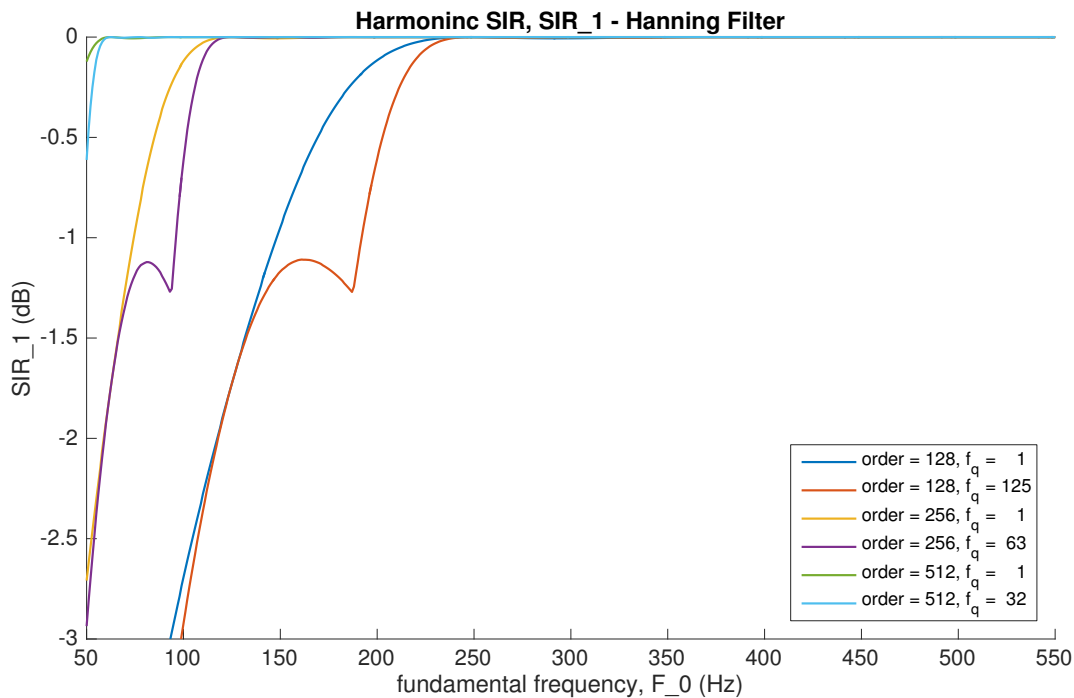


Figure 4.13:  $SIR_k$  filter order and quantization

Figure 4.14 compares filter design methods. Hanning and adaptive are essentially the same, showing that the limiting factor is still filter order. Rectangular provides a better



lower limit for what  $F_0$  the SIR breaks down at, and it does this at the cost of dips at higher frequencies. This agrees with the fact that rectangular windows have the sharpest rolloff at the expense of large sidelobes.

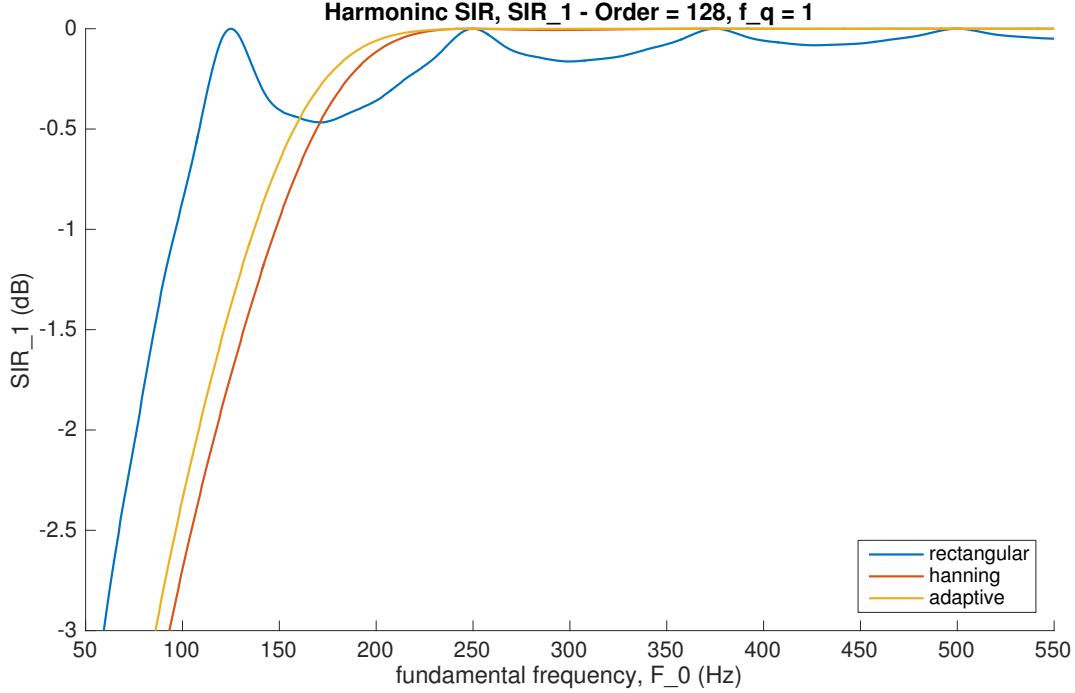


Figure 4.14:  $SIR_k$  filter design

The higher order harmonics are compared in figures 4.15 and 4.16. A pattern emerges, similar to figure 4.12 where the number of dips is proportional to  $k$ . These figures reinforce that improvement from decreasing quantization,  $F_q$ , is bounded.

For hanning the incremental 1dB of improvement is arguably not important. For rectangular there is a significant improvement in the 80-130Hz region for  $k > 3$ .

Filter order is certainly the dominant factor for harmonic SIR. For  $N = 128$ , it starts to break down for  $F_0 \approx 220$  Hz and degrades as  $F_0$  decreases. For  $N = 256$ , it starts to break down for  $F_0 \approx 110$  Hz. For  $N = 512$  the harmonic SIR performance is essentially optimal across all values of  $F_0$ .

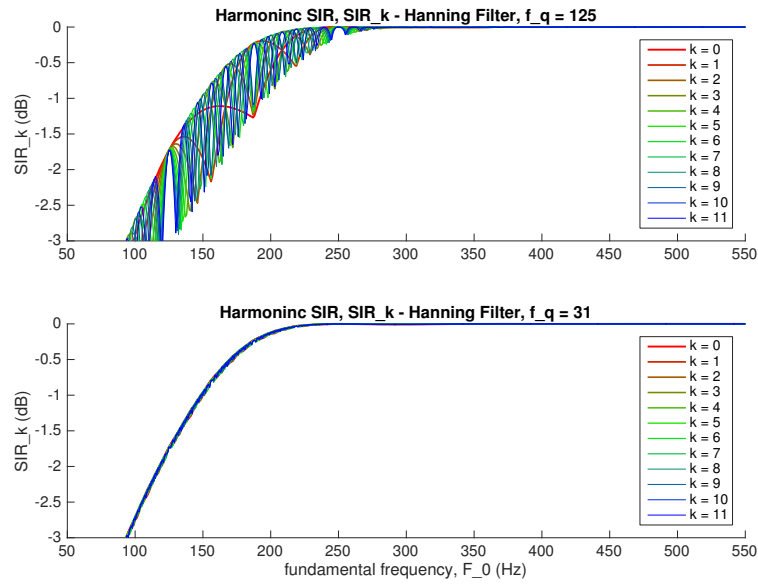


Figure 4.15:  $SIR_k$  variation across harmonics with hanning filter

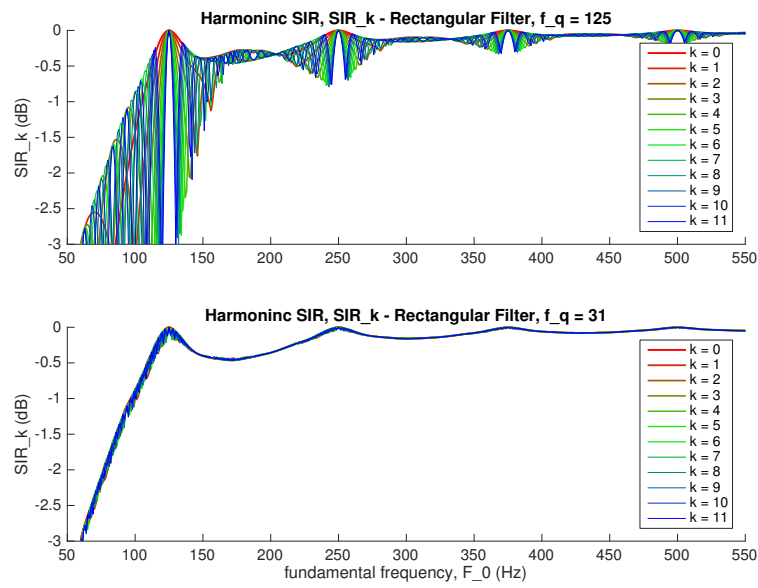


Figure 4.16:  $SIR_k$  variation across harmonics with rectangular filter

#### 4.6.3 Modulation Depth

It was decided in section 4.3 that explicit modulations are less artifact prone than induced, in which case the design should have minimal modulation in the extracted envelope.

Figure 4.17 compares each filter design method at the different filter orders. For all orders rectangular windows do a poor job of suppressing modulations due to high sidelobe amplitude. Hanning and adaptive show similar responses. For these two filter designs, the dominant variation is the depth at low  $F_0$  as a function of filter order.

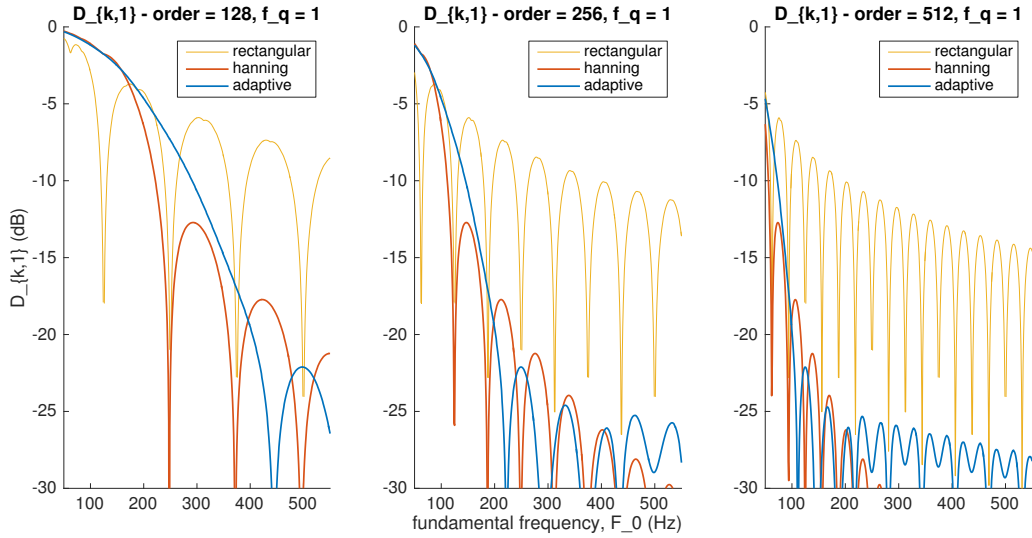
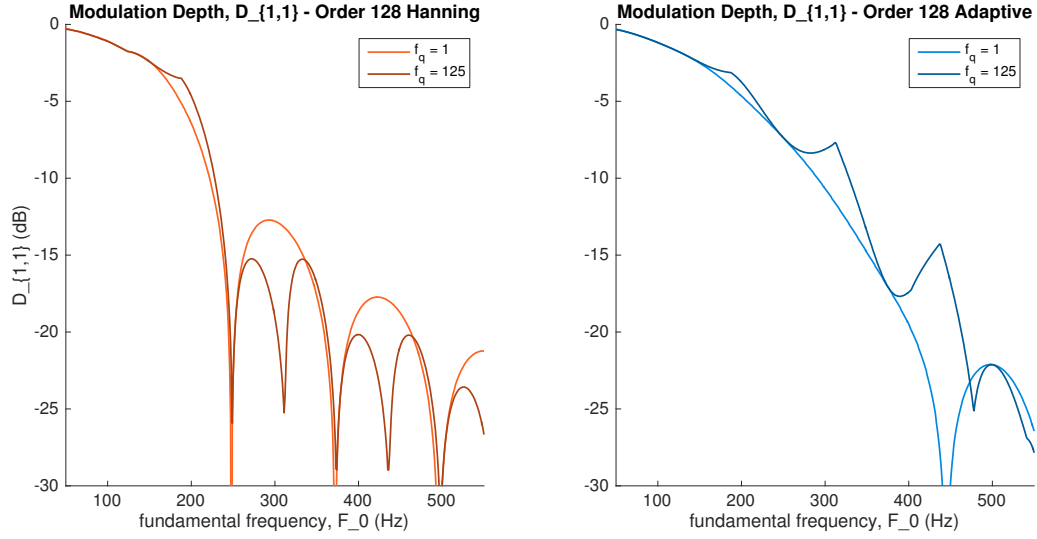
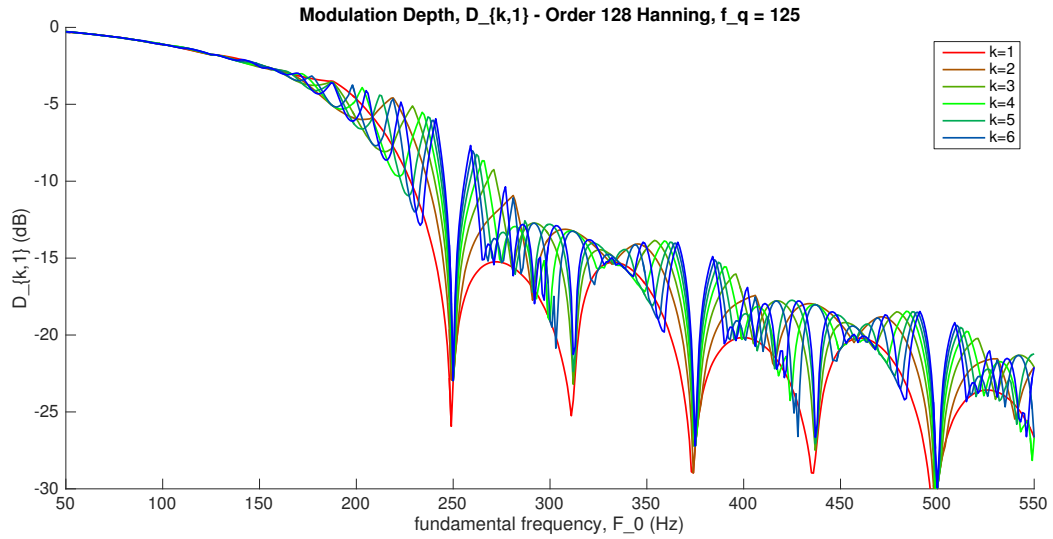


Figure 4.17:  $D_{k,i}$  filter design and order

Downshift quantization shows little affect on modulation depth. This is shown for both hanning and adaptive filter designs in figure 4.18.

Provided no downshift quantization, modulation depth won't change as a function of  $k$ . Figure 4.19 shows this variation, when  $F_q = 125\text{Hz}$  however it has minimal impact.

Recall  $D_{k,i}$  is the modulation depth of the estimate of the  $k$ th harmonic at a rate of  $iF_0$ . As  $i$  increases,  $iF_0$  moves further away from baseband and the filter does a better job of eliminating modulations. This is verified in figure 4.20.

Figure 4.18:  $D_{k,i}$  downshift quantizationFigure 4.19:  $D_{k,1}$  across harmonics

These results suggest that  $D_{k,1}$  is the most important measurement, and that hanning and adaptive filter designs achieve approximately the same performance. At low  $F_0$ 's filter order plays a large roll in modulation depth.

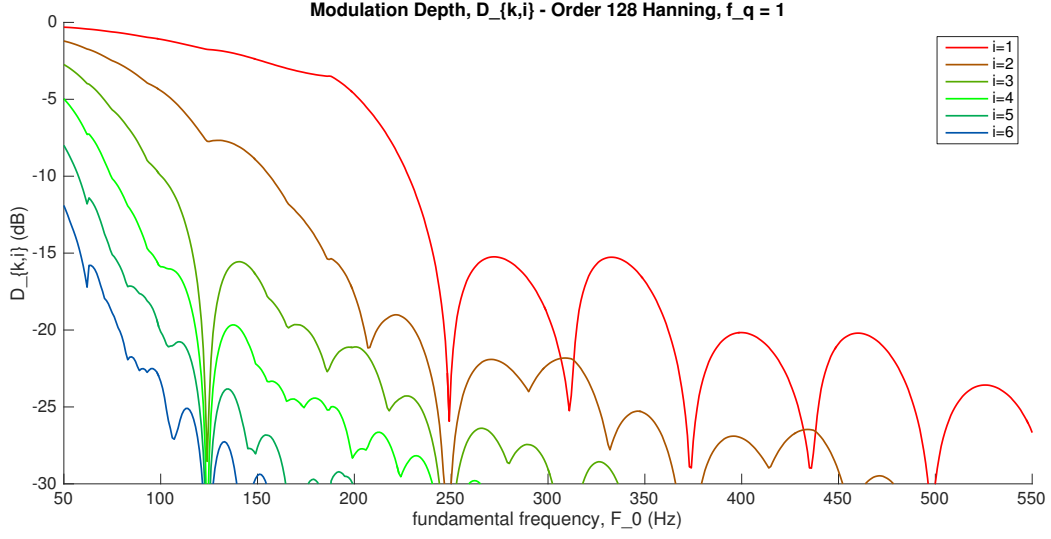


Figure 4.20:  $D_{k,i}$  at rate of  $iF_0$

Psychophysical studies have found that for reliable pitch discrimination amplitude-modulations of approximately 10% to 40% are required on average [15].

$$10\% \rightarrow D_{k,1} = -20\text{dB}$$

$$40\% \rightarrow D_{k,1} = -8\text{dB}$$

This implies that depending on the user:

- order 128 breaks down at  $F_0 \approx 240$  to  $400\text{Hz}$
- order 256 breaks down at  $F_0 \approx 120$  to  $200\text{Hz}$
- order 512 breaks down at  $F_0 \approx 60$  to  $100\text{Hz}$

In the best case, order 512 is sufficient for all  $F_0$ . In the worst case, order 128 will have artifacts across almost the entire  $F_0$  range.

#### 4.6.4 Transients

Time-responses are a bit more difficult to analyze, since there is no standard measurement like decibels that we are familiar with. We will consider transient responses of the different filter designs and filter orders with three different analyses.

The first is the unit step response, shown in figure 4.21. Latency on the order of 15ms isn't of much concern. The more important difference is the rise time. The 10-90% rise times are displayed in table 4.1.

The adaptive filters all have the same rise time at high enough  $F_0$  however the lower order filters are fundamentally constrained on how slow the rise time can be. The rectangular window is the worst of them all.

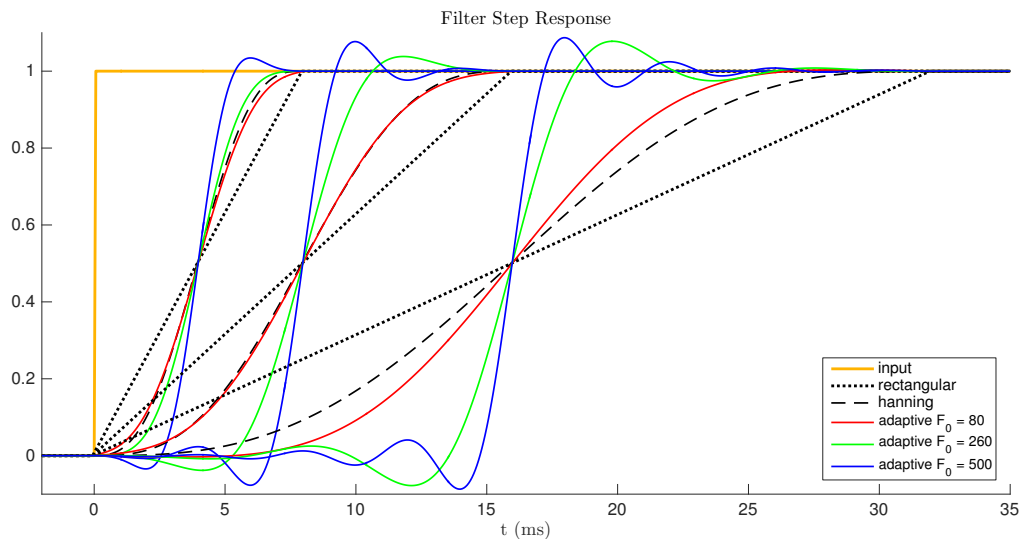


Figure 4.21: Transient Step Response, order = 128, 256, 512 (increasing order corresponds to longer reponse time)

An alternative view is shown in figure 4.22. For typical attack times in the range of 5-200ms an input-to-output change in attack time is plotted. As mentioned in section 4.5 humans hear transient changes in the log domain, and thus the axes are log scaled. The output rise time is computed as

	rectangular	hanning	adaptive 80	adaptive 260	adaptive 500
Order	Rise Time (ms)				
128	7	4	4	3	2
256	13	8	8	4	2
512	26	16	12	4	2

Table 4.1: filter rise times

$$risetime_{out} = risetime_{in} + risetime_{system} \quad (4.31)$$

For the worst case, rectangular order 512, more than half the dynamic range is lost due to smearing.

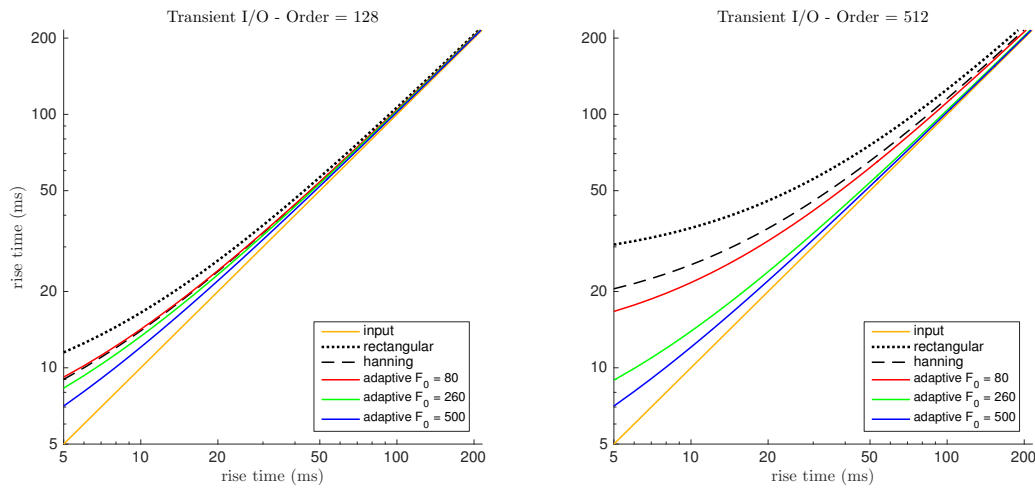


Figure 4.22: Transient Input/Output Change

As a final perspective on transients, we consider typical instrument attack times. Figure 4.23 shows the shifted attack times of twelve instruments typical attack times. The vertical scale has no meaning, it is simply for visual clarity.

What's interesting is that on a log scale, the instruments appear to bunch into two groups. The slow attack-time group is robust to the distortions of any of these filters. On the other hand the fast-attack time instruments change dramatically. For the narrow bandwidth 512 order filters, the smeared guitar output is closer in attack-time to an English horn than itself!

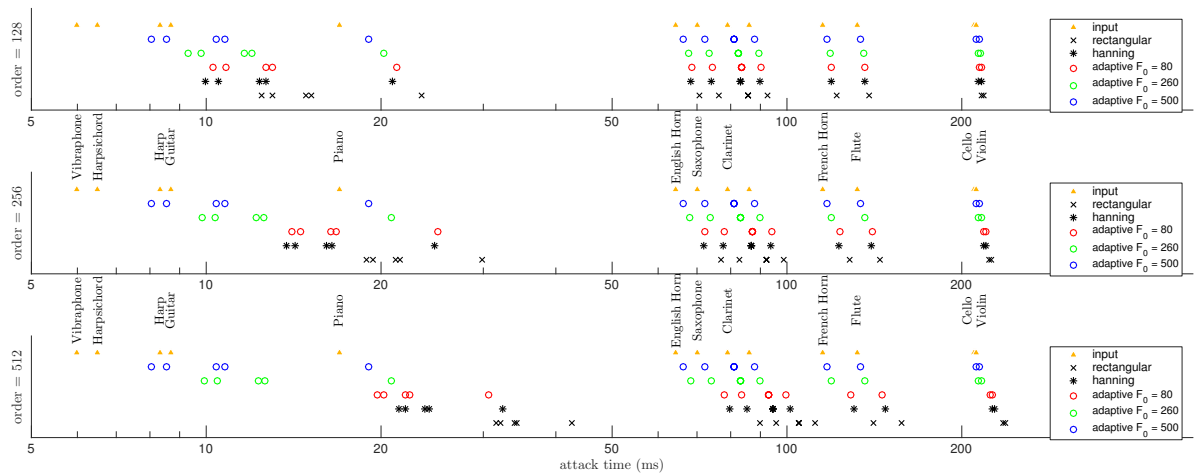


Figure 4.23: Transient Distortion for Common Instruments



#### 4.6.5 Summary

For the most part the hanning and adaptive filters outperformed rectangular. The rectangular window's performance on modulation depth makes it essentially unusable.

For coherent gain the worst case was roughly  $G_k \leq -1.5\text{dB}$ . From this result it doesn't appear to be an overly critical design consideration.

For harmonic SIR and modulation depth the critical performance variable was filter order. To very loosely summarize, order 128 fails for  $F_0 < 240\text{Hz}$ , order 256 fails for  $F_0 < 120\text{Hz}$  and 512 does sufficiently well for the full range considered.

Downshift quantization also did not seem to play a prominent role. This is in part affected by the restriction that quantization can't be worse than  $\frac{F_s}{N}$ .

There is clearly an envelope bandwidth tradeoff where the wider a filter is the less transients are smeared but the more the other harmonics interfere in the estimated envelope.

The sharp-cutoff order 512 filters smear the fast transients a significant amount, however the adaptive bandwidth filters seem to do well at smearing as little as possible while still achieving good performance on the other metrics. This could be the best compromise to the posed bandwidth tradeoff.

### 4.7 Non-ideal Pitch Estimators

The critical assumption thus far has been accurate pitch estimates. One problem to consider is error in the pitch estimator. The other analyzed in this thesis is pitch estimator quantization.

We consider a specific pitch estimator that uses autocorrelation. To summarize this method, first an autocorrelation is performed on the windowed input. A maxima is selected from this autocorrelation and the fundamental frequency is computed from the index of the maxima.

$$R_{xx}[n, \tau] = x_{windowed}[r] * x_{windowed}[-r] \quad (4.32)$$

$$\tilde{F}_0[n] = F_s \left[ \arg \max_{\tau} R_{xx}[n, \tau] \right]^{-1} \quad (4.33)$$

This can be implemented efficiently using the fast-autocorrelation method

$$R_{xx}[n, \tau] = \mathcal{F}^{-1} \left\{ X[n, k] X^*[n, k] \right\} \quad (4.34)$$

Defining the FFT order as  $N$ , for this method the possible values of  $F_0$  are

$$F_0 = \frac{F_s}{\tau}, \quad 1 \leq \tau \leq \frac{N}{2} \quad (4.35)$$

Since the considered  $F_0$  range is bounded to 50-550Hz we better resolution can be achieved by resampling the signal such that more values of  $F_0$  fall within these bounds:

$$\max\left(\frac{2F_s}{N}, 50\right) \leq F_0 \leq \min\left(\frac{F_s}{2}, 550\right) \quad (4.36)$$

Choosing  $F_s$  is important, since the quantization of  $F_0$  is not linearly spaced and becomes worse at higher values of  $F_0$ .

To be clear that this different sampling rate is only relevant to pitch estimation and not any of the other envelope extraction process, the pitch estimator sampling rate is defined:  $F_{s,p}$ . Having  $N$  as the filter orders previously considered,  $F_{s,p}$  is selected for maximal number of values of  $F_0$  within the region of interest. The designs are shown in table 4.2.

With this design each  $N$  covers approximately the same range, however the high orders have 2 or 4 times as many samples as  $N = 128$ . This is especially important at high values of  $F_0$  where the quantization is the worst.

The next subsections revisit harmonic SIR and modulation depth with non-deal pitch estimates. Downshift quantization is assumed:  $F_q = \frac{F_s}{N}$ .

Order ( $N$ )	$F_{s,p}$	min $F_0$	max $F_0$	best quantization	worst quantization
128	4kHz	63Hz	500Hz	1Hz	56Hz
256	8kHz	63Hz	533Hz	1Hz	33Hz
512	16kHz	63Hz	533Hz	1Hz	17Hz

Table 4.2:  $F_0$  estimate quantization

#### 4.7.1 Harmonic SIR

Harmonic SIR is visualized for two different filter design methods in figures 4.24 and 4.25. The pitch quantization, which is worse for lower orders, causes harmonic SIR to degrade for higher harmonics. This makes sense as the quantization error will be scaled by harmonic index  $k$ .

The hanning filter performs slightly at high  $F_0$ s better due to narrower filter bandwidth. Depending on the desired performance, harmonic indices above a certain threshold will no longer provide accurate harmonic envelopes. This threshold is slightly lower for adaptive filters than hanning filters and it is significantly lower for lower order filters.

Provided the same designs but with  $\pm 5$ Hz pitch estimation error, the worst case  $SIR_k$  is shown for hanning filter in figure 4.26 and for adaptive filter in figure 4.27.

The error degrades performance in two dimensions. Similar to quantization error, the performance degrades proportional to  $k$ . The other problem is at low values of  $F_0$ , where harmonics are more closely spaced. Take the right plot in figure 4.26 as an example. For the first 3 harmonics good harmonic SIRs are good for  $F_0 > 80$ Hz, however for  $k = 3, 4$  this increases to roughly  $F_0 > 180$ Hz and the even higher harmonics never achieve satisfactory SIR.

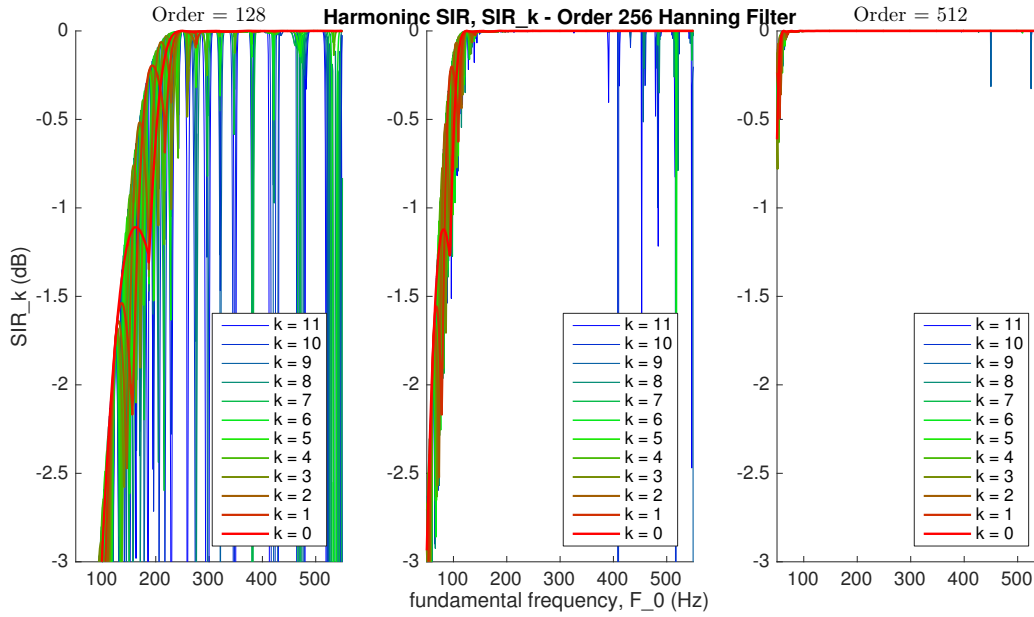


Figure 4.24:  $SIR_k$ , hanning filter and pitch estimate quantization

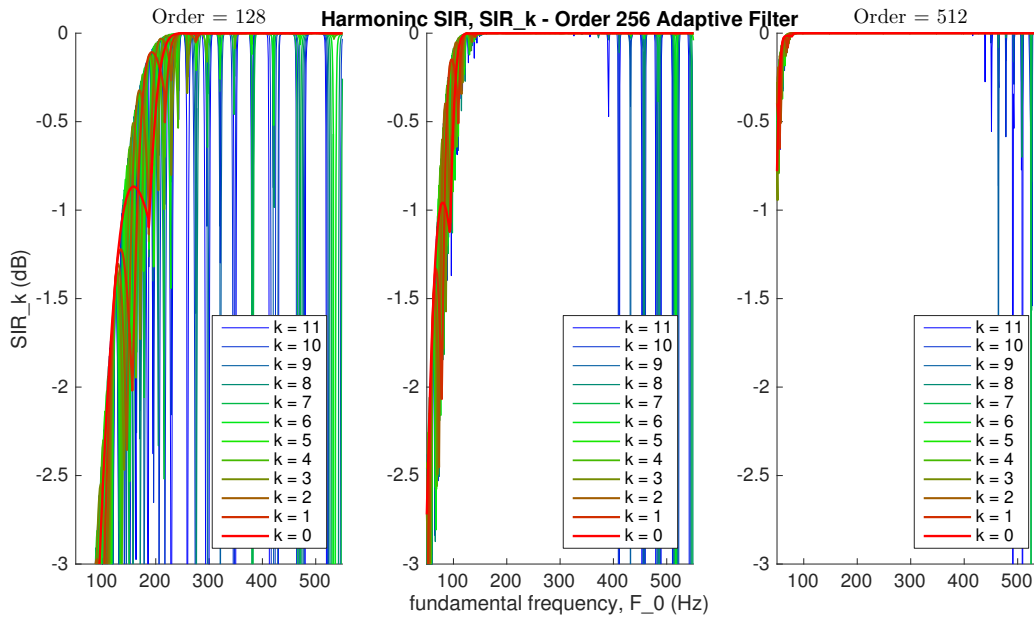


Figure 4.25:  $SIR_k$ , adaptive filter and pitch estimate quantization

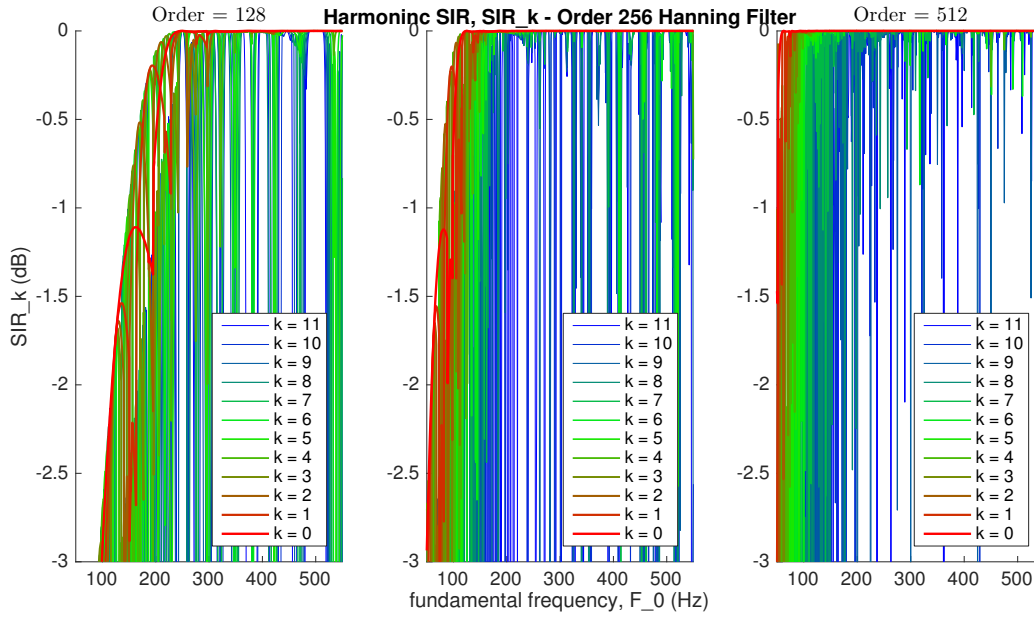


Figure 4.26:  $SIR_k$ , hanning filter, pitch estimate quantization and  $\pm 5$ Hz estimation error

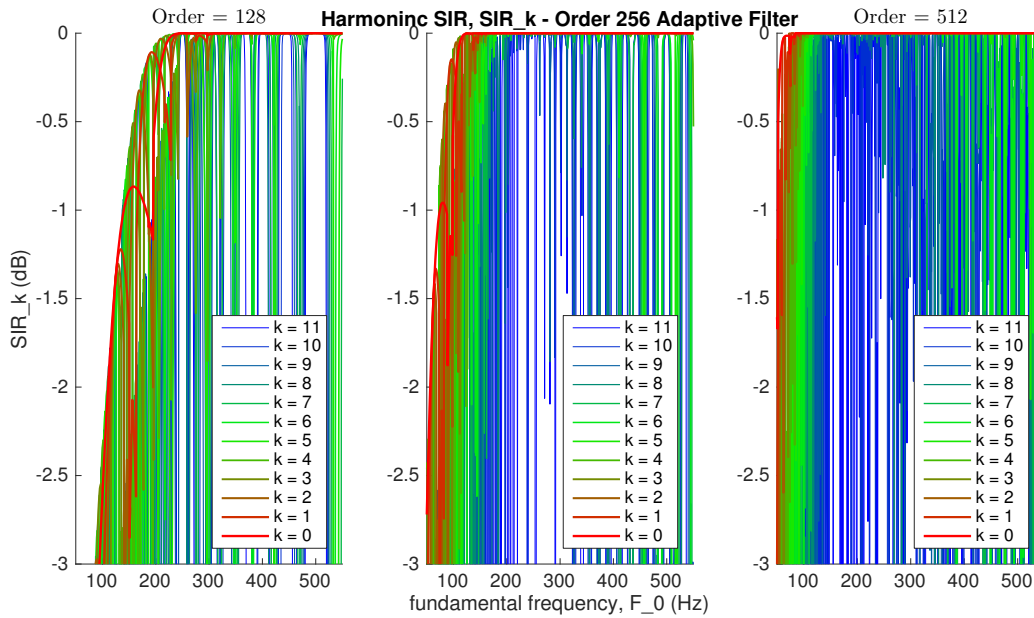


Figure 4.27:  $SIR_k$ , adaptive filter, pitch estimate quantization and  $\pm 5$ Hz estimation error

#### 4.7.2 Modulation Depth

The same analysis now repeated for modulation depth. Looking at figure 4.28, with hanning filter and pitch estimate quantization, high harmonics have very high modulations. Around the 6th harmonic, ( $k = 5$ ), we start to see big spikes in modulation depth at high  $F_0$ 's. Interestingly the same harmonics have poor performance regardless of  $N$ , however there is a far broader region of failure for lower  $N$ .

In figure 4.29 there is much better performance for  $N = 512$  in comparison to the hanning filter. This is because despite having wider bandwidth at high  $F_0$ 's, the sidelobes are much lower than the hanning filter. The first hanning sidelobe has a gain of -31dB, whereas for  $F_0 = 500\text{Hz}$  the adaptive filter has a first sidelobe gain of -56dB.

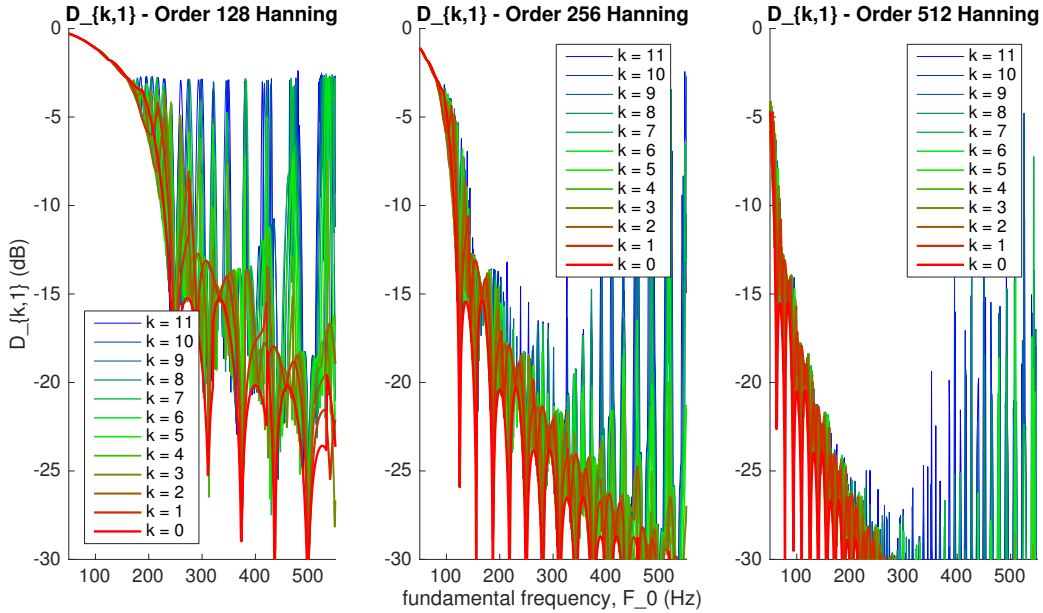


Figure 4.28:  $D_{k,1}$ , hanning filter and pitch estimate quantization

Now considering  $\pm 5\text{Hz}$  estimation error, figures 4.30 and 4.31 show the same shift right where higher harmonics at low  $F_0$ 's perform worse. The adaptive order 512 filter performs the best, being very robust error.

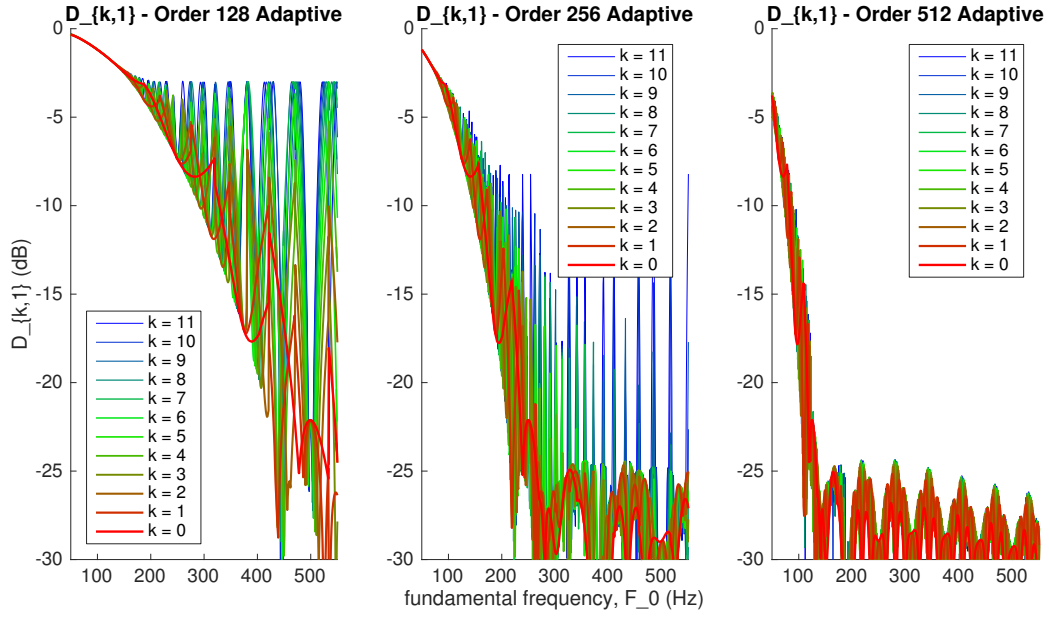


Figure 4.29:  $D_{k,1}$ , hanning filter and pitch estimate quantization

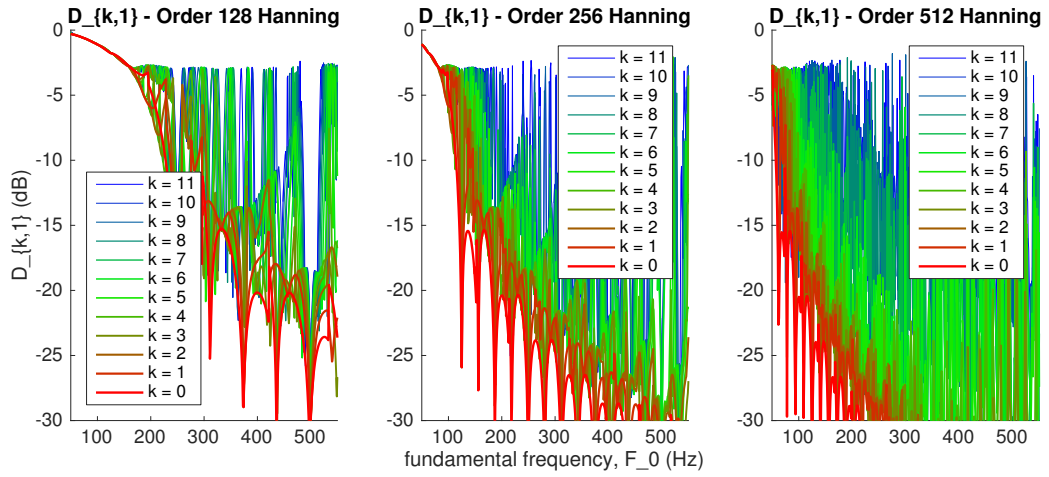


Figure 4.30:  $D_{k,1}$ , hanning filter, pitch estimate quantization and  $\pm 5\text{Hz}$  estimation error

Regardless of extraction method there is a fundamental limit on performance with poor pitch estimates. From all of the above examples, performance degrades proportional to  $k \times \text{error}/F_0$ .

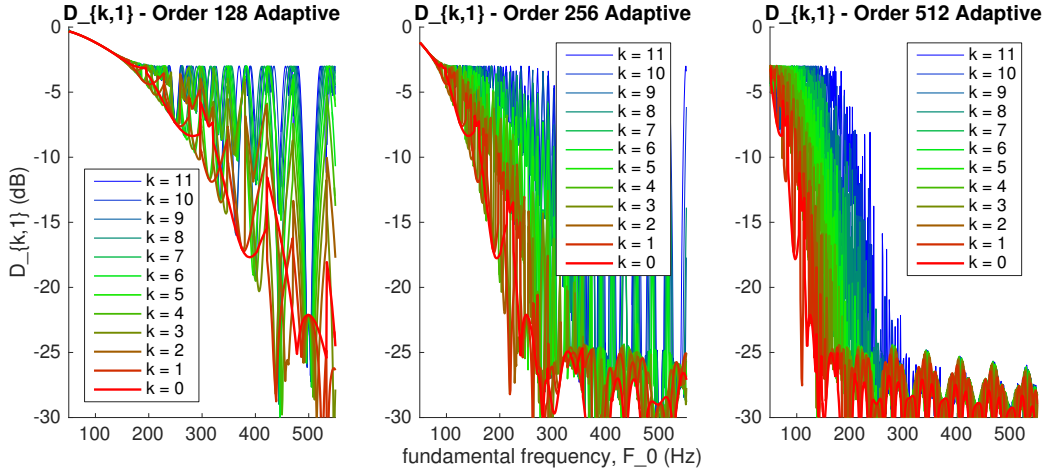


Figure 4.31:  $D_{k,1}$ , hanning filter, pitch estimate quantization and  $\pm 5\text{Hz}$  estimation error

#### 4.8 Mapping and Selection

coherent harmonic processing is different from incoherent in two fundamental ways:

- 1) envelopes are extracted coherently to minimize distortions
- 2) only envelopes relating to harmonics are calculated, this frees

Fixed Greenwood bands are determined offline, corresponding each electrode with a bandwidth. The  $N$  envelopes are then mapped to electrodes by finding the greenwood bands each harmonic falls within.

Two general solutions

- 1) Adaptive (select loudest)

similar to ACE, we can choose the loudest channels. This suffers from stability issues.

We can apply another heuristic to stabilize the decision based on consistency of signal energy and fundamental frequency

- 2) Fixed

stable, each option suffers from missing key harmonics to the signal

lowest channels will imply no high frequency energy, which could be bad for unvoiced signals



other relationships such as odd harmonics or prime numbered harmonics could miss harmonics critical to timbre perception.

fixed is more computationally efficient!

how to allocate harmonics to channels: combine, select 1st, select largest

hint at hybrid via failure to represent of selection options to represent the high channels well.

how to choose channels, largest vs fixed

heuristics on both!

Because we have isolated individual harmonic envelopes there is no issue of signal energy falling in between channels.

Another bonus to HSSE is that we may add a simple heuristic to maintain channel mapping stability. For example, if F0 has not varied significantly since the previous frame, we can allocate to the same channels to avoid unnecessary switching between channels induced by vibrato or inaccuracies in pitch estimation.

## **4.9 Implementation Considerations**

### *4.9.1 Carrier Synthesis*

what waveform?

raised vs rectified, (raised seems better)

cite that 4-waveform paper

Swanson thesis: “A high-rate pulse train, modulated on and off at frequency F0, had a higher pitch than a train of pulses at the rate of F0. If amplitude modulation of high-rate pulse trains is to be used to convey pitch, then the shape of the modulating waveform is important: a half-wave shape is better than a square-wave (on-off) shape.”

### *4.9.2 Pitch Estimator*

importance of a good pitch estimator

improving pitch estimator

haven't yet mentioned octave errors

#### 4.9.3 *Adaptive Filters using FFT*

...

improved filter quantization using interpolation

#### 4.9.4 *Efficient FFT Interpolation Algorithm*

FFT with changeable window, and interpolate

Can this be done with different filter as function of F0? We probably need to design the filters such that they pass reconstruction requirements

Is the actual equation just a sinc function times a phase shift?!

READ THIS: [An Intelligent FFT-Analyzer with Harmonic Interference Effect Correction and Uncertainty Evaluation]

#### 4.9.5 *Stimulation Rate*

“ it was important not to change other 12 HWR strategy take-home study 224 aspects of the strategy, in particular, stimulation rate. It would not be a fair comparison to trial HWR at 1800 pps against ACE at 900 pps, as the increased stimulation rate in itself could affect performance. A higher rate could potentially represent amplitude modulation cues more faithfully (McKay et al. 1994). Conversely, there is evidence that sensitivity to temporal modulation is worse at higher rates (Galvin and Fu 2005).” [swanson thesis]

#### 4.9.6 *encoding transients*

explicit transient encoding using transient detectors [find a vocoder ref]

#### 4.9.7 Hybrid Methods

Most everything so far has assumed the signal has an  $F_0$ , what if it doesn't? What if it is well outside the boundaries of  $F_0$ ? What about polyphonic music? What about SNRs below what is needed for accurate  $F_0$  estimation.

hybrid considerations

1) to achieve harmonic and inharmonic at same time

2) to better model the critical bands in the cochlea

maybe narrower filters could improve SiN, this was not investigated

effective information bandwidth: should I really get into this? From the theoretical standpoint, envelope extraction is exactly the same in ACE and F0mod. In implementation ACE typically uses a lower order FFT. In [laneau] the authors consider 128-point for ACE and 512-point for F0mod and both will be considered here. with respect to bandwidth we actually have to different things, filter bandwidth and effective information bandwidth. The former is obvious, the later refers to what frequencies are encoded on a electrode channel. If multiple narrowband filters are somehow combined on the same channel, they may have the same information bandwidth as one wideband filter.

soft decisions (e-Tone)

“bowed string tones are inharmonic during both their attack and decay (Beauchamp, 1974)”

mention transients can't be reliably represented with a pitch tracker

#### 4.10 Summary

## Chapter 5

# CONCLUSION

### **5.1** *Summary*

### **5.2** *Future Work*

## BIBLIOGRAPHY

- [1] Oded Ghitza. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *The Journal of the Acoustical Society of America*, 110(3):1628–1640, 2001.
- [2] Fredric J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.
- [3] Ying-Yee Kong, Ala Mullangi, Jeremy Marozeau, and Michael Epstein. Temporal and spectral cues for musical timbre perception in electric hearing. *Journal of Speech, Language, and Hearing Research*, 54(3):981–994, 2011.
- [4] Waikong Lai and Norbert Dillier. Investigating the mp3000 coding strategy for music perception. *ACE*, 10(70.8):85–4, 2008.
- [5] Johan Laneau, Jan Wouters, and Marc Moonen. Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neurotology*, 11(1):38–52, 2006.
- [6] Qin Li and Les Atlas. Time-variant least squares harmonic modeling. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 2, pages II–41. IEEE, 2003.
- [7] Xing Li, Kaibao Nie, Les Atlas, and Jay Rubinstein. Harmonic coherent demodulation for improving sound coding in cochlear implants. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5462–5465. IEEE, 2010.
- [8] Xing Li, Kaibao Nie, Nikita S Imennov, Jay T Rubinstein, and Les E Atlas. Improved perception of music with a harmonic based algorithm for cochlear implants. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 21(4):684–694, 2013.
- [9] Consolatina Liguori, Alfredo Paolillo, and Alfonso Pignotti. An intelligent fft analyzer with harmonic interference effect correction and uncertainty evaluation. *Instrumentation and Measurement, IEEE Transactions on*, 53(4):1125–1131, 2004.
- [10] Jeremy Marozeau, Alain de Cheveigné, Stephen McAdams, and Suzanne Winsberg. The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, 114(5):2946–2957, 2003.

- [11] Waldo Nogueira, Andreas Büchner, Thomas Lenarz, and Bernd Edler. A psychoacoustic nofm-type speech coding strategy for cochlear implants. *EURASIP Journal on Applied Signal Processing*, 2005:3044–3059, 2005.
- [12] Margaret W Skinner, Laura K Holden, Timothy A Holden, Richard C Dowell, Peter M Seligman, Judith A Brimacombe, and Anne L Beiter. Performance of postlinguistically deaf adults with the wearable speech processor (wsp iii) and mini speech processor (msp) of the nucleus multi-electrode cochlear implant\*. *Ear and Hearing*, 12(1):3–22, 1991.
- [13] Branko Somek, Siniša Fajt, Ana Dembitz, Mladen Ivković, and Jasmina Ostojić. Coding strategies for cochlear implants. *AUTOMATIKA: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 47(1-2):69–74, 2006.
- [14] Nancy Tye-Murray, Mary Lowder, and Richard S Tyler. Comparison of the fof2 and fof1f2 processing strategies for the cochlear corporation cochlear implant. *Ear and hearing*, 11(3):195–200, 1990.
- [15] Andrew E Vandali, Catherine Sucher, David J Tsang, Colette M McKay, Jason WD Chew, and Hugh J McDermott. Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. *The Journal of the Acoustical Society of America*, 117(5):3126–3138, 2005.
- [16] Andrew E Vandali and Richard JM van Hoesel. Development of a temporal fundamental frequency coding strategy for cochlear implants. *The Journal of the Acoustical Society of America*, 129(6):4023–4036, 2011.
- [17] Blake S Wilson, Charles C Finley, Dewey T Lawson, Robert D Wolford, and Mariangeli Zerbi. Design and evaluation of a continuous interleaved sampling (cis) processing strategy for multichannel cochlear implants. *Journal of rehabilitation research and development*, 30:110–110, 1993.
- [18] Fan-Gang Zeng. Trends in cochlear implants. *Trends in amplification*, 8(1):1–34, 2004.

## Appendix A

### DERIVATIONS

$$\phi_0[n+r] = \phi_0[n] + 2\pi \frac{F_0[n]}{F_s} r, \quad 0 \leq r < N \quad (\text{A.1})$$

$$\begin{aligned} m_{k,harmonic}[n] &= \left| x[n] e^{-jk\phi_0[n]} * \frac{1}{Nw[0]} w[-n] \right| \\ &= \frac{1}{Nw[0]} \left| \sum_{r=-\infty}^{\infty} x[n-r] e^{-jk\phi_0[n-r]} w[-r] \right| \end{aligned}$$

Let  $r' = -r$

$$\begin{aligned} &= \frac{1}{Nw[0]} \left| \sum_{r'=0}^{N-1} x[n+r'] e^{-jk\phi_0[n+r']} w[r'] \right| \\ &= \frac{1}{Nw[0]} \left| e^{-jk\phi_0[n]} \sum_{r'=0}^{N-1} x[n+r'] e^{-j \frac{2\pi F_0[n]}{F_s} k r'} w[r'] \right| \\ &= \frac{1}{Nw[0]} \left| e^{-jk \left( \phi_0[n] - \frac{2\pi F_0[n]}{F_s} n \right)} \left[ e^{-j \frac{2\pi F_0[n]}{F_s} k n} \sum_{r'=0}^{N-1} x[n+r'] w[r'] e^{-j \frac{2\pi F_0[n]}{F_s} k r'} \right] \right| \\ &= \frac{1}{Nw[0]} \left| X \left[ n, \frac{N}{1} \frac{F_0[n]}{F_s} k \right] \right| \\ &= \frac{1}{Nw[0]} \left| X \left[ n, \lambda[n] k \right] \right| \quad (\text{A.2}) \end{aligned}$$

#### A.0.1 CIS VS Hilbert

Provided the envelope definitions

$$m_{k,CIS}[n] = |x_k[n]| * h_{lp}[n] = |x[n] * h_k[n]| * h_{lp}[n]$$

$$m_{k,Hilbert}[n] = |x_k^+[n]| = |x[n] * h_k[n] + j\mathcal{H}\{x[n] * h_k[n]\}|$$

and an ideal brick-wall filter defined as

$$H_k(f) = \mathcal{F}\{h_k[n]\} \quad (\text{A.3})$$

$$H_k(f) = \begin{cases} 1, & f_k - \frac{1}{2}f_{bw} < |f| < f_k + \frac{1}{2}f_{bw} \\ 0, & \text{else} \end{cases} \quad (\text{A.4})$$

let:

$$Y_{k,Hilbert}(f) = \mathcal{F}\{|x_k^+[n]|^2\} \quad (\text{A.5})$$

$$Y_{k,CIS}(f) = \mathcal{F}\{|x_k[n]|^2\} \quad (\text{A.6})$$

$$(\text{A.7})$$

$$Y_{k,Hilbert}(f) = \begin{cases} X_k^+(f) * X_k^{*+}(-f), & |f| < f_{bw} \\ 0, & |f| \geq f_{bw} \end{cases} \quad (\text{A.8})$$

$$Y_{k,CIS}(f) = \begin{cases} 2Y_{k,Hilbert}(f), & |f| < f_{bw} \\ 0, & f_{bw} \leq |f| \leq 2f_k - f_{bw} \\ X_k(f) * X_k^*(-f), & 2f_k - f_{bw} < |f| < 2f_k + f_{bw} \\ 0, & |f| \geq 2f_k + f_{bw} \end{cases} \quad (\text{A.9})$$

$$X(f) = \mathcal{F}\{x[n]\} \quad (\text{A.10})$$

$$X_k(f) = \mathcal{F}\{x_k[n]\} \quad (\text{A.11})$$

$$X_k^+(f) = \mathcal{F}\{x_k^+[n]\} \quad (\text{A.12})$$



The ideal brickwall filter sets (??) and (??) to zero outside of its bandwidth:

$$X_k(f) = \begin{cases} X(f), & f_k - \frac{1}{2}f_{bw} < |f| < f_k + \frac{1}{2}f_{bw} \\ 0, & \text{else} \end{cases} \quad (\text{A.13})$$

$$X_k^+(f) = \begin{cases} X(f), & f_k - \frac{1}{2}f_{bw} < f < f_k + \frac{1}{2} \\ 0, & \text{else} \end{cases} \quad (\text{A.14})$$

Stemming from the CIS and Hilbert envelope equations, consider the following two functions.

$$Y_{k,Hilbert}(f) = \mathcal{F} \left\{ \left| x_k^+[n] \right|^2 \right\} \quad (\text{A.15})$$

$$Y_{k,CIS}(f) = \mathcal{F} \left\{ \left| x_k[n] \right|^2 \right\} \quad (\text{A.16})$$

$$(\text{A.17})$$

$Y_{k,CIS}(f)$  is equivalent to the squared CIS envelope if the final lowpass filter is not applied.  $Y_{k,Hilbert}(f)$  is the squared Hilbert envelope.

$$Y_{k,CIS}(f) = 2Y_{k,Hilbert}(f) \quad (\text{A.18})$$

These functions can be computed by convolution in the frequency domain. For the Hilbert function:

$$Y_{k,Hilbert}(f) = \widehat{X}_k(f) * \widehat{X}_k^*(-f) \quad (\text{A.19})$$

$$= \int_{-\infty}^{\infty} \widehat{X}_k(f-r) \widehat{X}_k^*(-r) dr \quad (\text{A.20})$$

$$= \int_{-\infty}^{\infty} \widehat{X}_k(r+f) \widehat{X}_k^*(r) dr \quad (\text{A.21})$$

We can narrow the integration bounds provided the restrictions

$$\widehat{X}_k^*(r) \neq 0 \Rightarrow f_k - \frac{1}{2}f_{bw} < r < f_k + \frac{1}{2}f_{bw} \quad (\text{A.22})$$

$$\widehat{X}_k(r + f) \neq 0 \Rightarrow f_k - \frac{1}{2}f_{bw} - f < r < f_k + \frac{1}{2}f_{bw} - f \quad (\text{A.23})$$

$$a = \max\left(f_k - \frac{1}{2}f_{bw}, f_k - \frac{1}{2}f_{bw} - f\right) \quad (\text{A.24})$$

$$b = \min\left(f_k + \frac{1}{2}f_{bw}, f_k + \frac{1}{2}f_{bw} - f\right) \quad (\text{A.25})$$

$$Y_k^1(f) = \int_a^b \widehat{X}_k(r + f) \widehat{X}_k^*(r) dr, \quad -f_{bw} < f < f_{bw} \quad (\text{A.26})$$

$$= 0, \quad \text{else} \quad (\text{A.27})$$

For  $Y_k^2(f)$  there are actually three non-zero bands.

$$Y_k^2(f) = X_k(f) * X_k^*(-f) \quad (\text{A.28})$$

$$= \int_{-\infty}^{\infty} X_k(r + f) X_k^*(r) dr \quad (\text{A.29})$$

Case 1:  $-2f_k - f_{bw} < f < -2f_k + f_{bw}$

$$Y_k^2(f) = \int_a^b X_k(r + f) X_k^*(r) dr \quad (\text{A.30})$$

$$a = \max\left(f_k - \frac{1}{2}f_{bw}, f_k - \frac{1}{2}f_{bw} - f\right) \quad (\text{A.31})$$

$$b = \min\left(f_k + \frac{1}{2}f_{bw}, f_k + \frac{1}{2}f_{bw} - f\right) \quad (\text{A.32})$$

Case 2:  $2f_k - f_{bw} < f < 2f_k + f_{bw}$

$$Y_k^2(f) = \int_a^b X_k(r+f)X_k^*(r)dr \quad (\text{A.33})$$

$$a = \max\left(-f_k - \frac{1}{2}f_{bw}, -f_k - \frac{1}{2}f_{bw} - f\right) \quad (\text{A.34})$$

$$b = \min\left(-f_k + \frac{1}{2}f_{bw}, -f_k + \frac{1}{2}f_{bw} - f\right) \quad (\text{A.35})$$

Case 3:  $-f_{bw} < f < f_{bw}$

This case is unique because there are two points of intersection. We can break up the integral into a sum. The first integral is exactly the same as in  $Y_k^1(f)$ .

$$Y_k^2(f) = \int_{a_1}^{b_1} X_k(r+f)X_k^*(r)dr + \int_{a_2}^{b_2} X_k(r+f)X_k^*(r)dr \quad (\text{A.36})$$

$$a_1 = \max\left(f_k - \frac{1}{2}f_{bw}, f_k - \frac{1}{2}f_{bw} - f\right) \quad (\text{A.37})$$

$$b_1 = \min\left(f_k + \frac{1}{2}f_{bw}, f_k + \frac{1}{2}f_{bw} - f\right) \quad (\text{A.38})$$

$$a_2 = \max\left(-f_k - \frac{1}{2}f_{bw}, -f_k - \frac{1}{2}f_{bw} - f\right) \quad (\text{A.39})$$

$$b_2 = \min\left(-f_k + \frac{1}{2}f_{bw}, -f_k + \frac{1}{2}f_{bw} - f\right) \quad (\text{A.40})$$

Using the Hermitian symmetry of the real-valued  $x[n]$ ,

$$Y_k^2(f) = \int_{a_1}^{b_1} X_k(r+f)X_k^*(r)dr + \int_{a_2}^{b_2} X_k^*(-r-f)X_k(-r)dr \quad (\text{A.41})$$

$$r' = -r - f$$

$$Y_k^2(f) = \int_{a_1}^{b_1} X_k(r+f)X_k^*(r)dr + \int_{a_1}^{b_1} X_k^*(r')X_k(r'+f)dr' \quad (\text{A.42})$$

$$= 2 \int_{a_1}^{b_1} X_k(r+f)X_k^*(r)dr \quad (\text{A.43})$$

$$= 2Y_k^1(f) \quad (\text{A.44})$$

If we lowpass filter  $Y_k^2(f)$  with a filter defined

$$H_{lp}(f) = \frac{1}{2}, \quad |f| < f_{bw} \quad (\text{A.45})$$

$$= 0, \quad 2f_k - f_{bw} < |f| < 2f_k + f_{bw} \quad (\text{A.46})$$

then

$$Y_k^2(f) = Y_k^1(f) \quad \forall f \quad (\text{A.47})$$

We conclude that

$$\left| x_k[n] \right|^2 * h_{lp}[n] \approx \left| \hat{x}_k[n] \right|^2 \quad (\text{A.48})$$

Things to consider are delay and non-deal filters, however provided the distance between baseband and the  $\pm 2f_k$  terms a sufficient filter is practical in practice.

Now the relationship between  $m_{k,CIS}[n]$  and  $m_{k,Hilbert}[n]$  is muddled by the nonlinear square root operation, however the nonlinearities induced won't be noticeably distorted by  $h_{lp}[n]$ . In practice, there only noticeable difference will be the added delay from the final lowpass filter in the CIS method.

## **VITA**

Tyler Ganter grew up in upstate New York, where the long cold winters inspired him to pick up guitar. While pursuing a BSEE at the University at Buffalo he took an interest in computer programming, which he decided to minor in. Through a continued interest in mathematics, software and music he found himself at home in the field of audio digital signal processing. The desire to delve deeper into this field has led him to study at the University of Washington.