Smart Things and Cochlear Implants

Tyler Ganter

A dissertation†
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

1985-2014

Program Authorized to Offer Degree: UW Information Technology

†an egocentric imitation, actually

University of Washington

Abstract

Smart Things and Cochlear Implants

Tyler Ganter

Chair of the Supervisory Committee:
Title of Chair Name of Chairperson
Department of Chair

This document is about extracting harmonic envelopes, what matters, what doesn't and how to design your system accordingly. It is broken into three parts:

- envelope extraction techniques and their relationships
- phase preservation
- system design (filter and downshift)

Many strategies consider the effects of leaving modulations in the signal, but nothing really talks about what the envelope should be, independent of the modulations. If we do this first, we can than think about how the modulations affect this envelope as a separate modulation component.

If explicitly inducing modulations it is important to remove any other modulations, and this is how.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

ARGUMENT: replacement

BACK-UP: a copy of a fi

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to University of Washington, where he has had the opportunity to work with the TeX formatting system, and to the author of TeX, Donald Knuth, *il miglior fabbro*.

DEDICATION

to my dear wife, Joanna

Chapter 1
INTRODUCTION

this is the introduction

Why harmonic encoding? Help differentiate signals, (timbre), improve SIN performance, free up channels for other information

1.1 Overview

we are considering what is the ideal matched filter, and how close of an approximation do we need?

1.2 Survey of Literature

Equivalent noise bandwidth (ENBW) considers BW of noise if squished into a box of gain 1 around the downshift frequency. [windows for harmonic analysis] This isn't entirely applicable since our harmonic has BW ¿ epsilon, and since for any window most of the energy is close to 0, most of the so-called noise is actually desired harmonic signal. If this were not the case, (I think) rectangular window would be the best, but since it distributes the noise more heavily to higher frequencies away from zero, it is actually worse (higher sidelobes)

"some windows have a high rate of sidelobe decay that allows minimizing the error due to interference. However the steeper the sidelobe decay the wider the main lobe width and then the worse the minimum resolution bandwidth." [An Intelligent FFT-Analyzer with Harmonic Interference Effect Correction and Uncertainty Evaluation]

"For NH listeners, the timbre space was best represented in three dimensions, one correlated with the temporal envelope (log-attack time) of the stimuli, one correlated with the spectral envelope (spectral centroid), and one correlated with the spectral fine structure (spectral irregularity) of the stimuli. The timbre space from CI listeners, however, was best represented by two dimensions, one correlated with temporal envelope features and the other weakly correlated with spectral envelope features of the stimuli. “temporal envelope is dominant cue for timbre perception in CI listeners” [Temporal and Spectral Cues for Musical Timbre Perception in Electric Hearing]

Hypothesis: –temporal envelope (log-attack time) this is in some cases smeared in time (F0mod) and in other cases mixed across harmonics –one correlated with the spectral envelope (spectral centroid) this is not as clearly represented as it could be (are we talking about resonance or per-harmonic details such as clarinet?) –one correlated with the spectral fine structure (spectral irregularity) this manifests in the envelope for CI processing, the problem though is that it is blurred across harmonics so the noise-like characteristics will be smoothed.

1.3 Contents of Thesis

Chapter 2

BACKGROUND

"By definition, timbre is the perceptual attribute that distinguishes two sounds that have the same pitch, loudness, and duration (American National Standards Institute, 1973)."

Chapter 3

HARMONIC ENVELOPES

We model our harmonic signal with a sum-of-products model as:

$$x[n] = \sum_k x_k[n] = \sum_k m_k[n]c_k[n] \tag{3.1}$$

our extracted envelope can be generally defined as:

$$\tilde{m}_k[n] = \left| \widehat{x}[n]e^{-j\omega_k[x]n} * h_k[n,x] \right| \tag{3.2}$$

This is a good generalization of any envelope extraction (harmonic or not). The design can be summarized by two things:

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

If $w_k[\cdot]$ and $h_k[\cdot]$ are functions of $x[n]$ we have coherent envelope extraction. If they are time-invariant, we have incoherent extraction.

3.0.1 harmonic signals

Since harmonic signals have a specific structure, we can define our carriers from equation 7.1 as 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.

$$\theta_k[n] = 2\pi(k+1)\frac{F_0}{F_s}n + \phi_k[n] \quad (3.3)$$

$$x[n] = \sum_{k=0}^{K} m_k[n]cos(\theta_k[n]) \quad (3.4)$$

$$\widehat{x}[n] = \sum_{k=0}^{K} m_k[n]e^{j\theta_k[n]} \quad (3.5)$$

3.1 Steady-State Analysis

We start with the simplest scenario, where $x[n]$ is a steady-state signal. The conditions we require for this are:

- constant pitch: $F_0[n] = F_0$
- narrowband modulator: $m_k[n] \approx constant$ over short periods of time
- constant phase term: $\phi_k[n] = \phi_k$, we choose $\phi_k[n] = 0$ for cleaner equations however this is not necessary

3.1.1 3 Harmonic Example: Desired Envelope

We visualize the frequency domain for a signal with three harmonics ($K = 2$) in figure **??**. For this example we consider the 1st harmonic ($k = 1$), centered at $2F_0$.

Figure **??**(d) is the spectrum of the squared envelope, $|\mathcal{F}\{m_1^2[n]\}|$. We see this relationship in equation 3.9

$$(a) \quad \widehat{x}[n] \Longleftrightarrow \widehat{X}[n, f] \quad (3.6)$$

$$(b) \quad \widehat{x}_1[n] \Longleftrightarrow \widehat{X}_1[n, f] \quad (3.7)$$

$$(c) \quad \widehat{x}_1^*[n] \Longleftrightarrow \widehat{X}_1^*[n, -f] \quad (3.8)$$

$$(d) \quad m_1^2[n] = \widehat{x}_1[n]\widehat{x}_1^*[n] \Longleftrightarrow \widehat{X}_1[n, f] * \widehat{X}_1^*[n, -f] \quad (3.9)$$

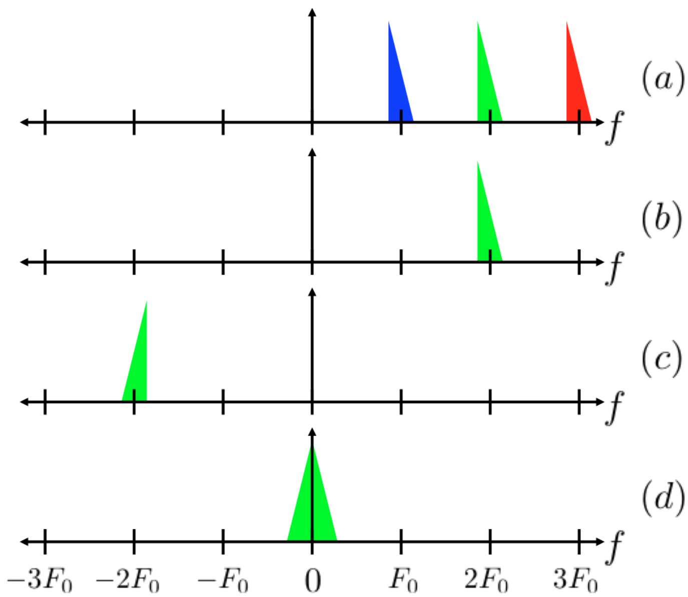

Figure 3.1: Magnitude of spectrum for equations 3.6 through 3.9

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 our analysis we can consider the squared envelope. This final square root operation will remain constant across all examples which allows us to not consider it.

$$m_1[n] = \left|\widehat{x}_1[n]\right| = \left[\widehat{x}_1[n]\widehat{x}_1^*[n]\right]^{\frac{1}{2}} \tag{3.10}$$

3.1.2 Estimated Envelope

Let's now evaluate our estimate, using equation 3.2. As stated above, we consider the squared envelope.

$$\begin{aligned}
\tilde{m}_k^2[n] =& \left| \widehat{x}[n] e^{-j\omega_k n} * h_k[n]] \right|^2 \\
=& \left| \sum_{l=0}^{K} m_l[n] e^{j(\theta_l[n] - \omega_k[n])} * h_k[n] \right|^2 \\
\approx& \left| \sum_{l=0}^{K} m_l[n] \Big(e^{j(\theta_l[n] - \omega_k[n])} * h_k[n] \Big) \right|^2 \\
=& \left| \sum_{l=0}^{K} m_l[n] e^{j(\omega_{k,l} n + \phi_k)} H_k\big(e^{j\omega_{k,l}}\big) \right|^2 && (3.11) \\
\omega_{k,l} =& 2\pi \frac{(l+1)F_0 - F_{ds,k}}{F_s} && (3.12) \\
h_k[n] \Longleftrightarrow& H_k\big(e^{j\omega}\big) && (3.13)
\end{aligned}$$

$\omega_{k,l}$ is the downshift of the l'th harmonic for the estimate of the k'th envelope. $H_k\big(e^{j\omega}\big)$ is the discrete Fourier transform (DFT) of $h_k[n]$.

Expanding this equation we get:

$$\begin{aligned}
\tilde{m}_k^2[n] =& \sum_{l=0}^{K} \sum_{i=0}^{K} m_l[n] m_i^*[n] e^{j(l-i)F_0} H_k\big(e^{j\omega_{k,l}}\big) H_k^*\big(e^{j\omega_{k,i}}\big) && (3.14) \\
=& \sum_{l=0}^{K} \Big| m_l[n] \Big|^2 \Big| H_k\big(e^{j\omega_{k,l}}\big) \Big|^2 \\
&+ e^{-j2\pi \frac{F_0}{Fs} n} \sum_{l=0}^{K-1} m_l[n] m_{l+1}^*[n] H_k\big(e^{j\omega_{k,l}}\big) H_k^*\big(e^{j\omega_{k,l+1}}\big) \\
&+ e^{j2\pi \frac{F_0}{Fs} n} \sum_{l=1}^{K} m_l[n] m_{l-1}^*[n] H_k\big(e^{j\omega_{k,l}}\big) H_k^*\big(e^{j\omega_{k,l-1}}\big) \\
&+ e^{-j2\pi \frac{2F_0}{Fs} n} \sum_{l=0}^{K-2} m_l[n] m_{l+2}^*[n] H_k\big(e^{j\omega_{k,l}}\big) H_k^*\big(e^{j\omega_{k,l+2}}\big) \\
&+ e^{j2\pi \frac{2F_0}{Fs} n} \sum_{l=2}^{K} m_l[n] m_{l-2}^*[n] H_k\big(e^{j\omega_{k,l}}\big) H_k^*\big(e^{j\omega_{k,l-2}}\big) \\
&+ \ldots \\
&+ e^{-j2\pi \frac{KF_0}{Fs} n} m_0[n] m_K^*[n] H_k\big(e^{j\omega_{k,0}}\big) H_k^*\big(e^{j\omega_{k,K}}\big) \\
&+ e^{j2\pi \frac{KF_0}{Fs} n} m_K[n] m_0^*[n] H_k\big(e^{j\omega_{k,K}}\big) H_k^*\big(e^{j\omega_{k,0}}\big) && (3.15)
\end{aligned}$$

We can now think of $\tilde{m}_k[n]$ as a combination of terms each centered at iF_0 where the magnitude of each term is:

$$\left|\tilde{m}_{k,iF_0}[n]\right| = \left[\sum_{l=0}^{K-|i|} \left|m_l[n]\right|\left|m_{l+i}[n]\right|\left|H_k\left(e^{j\omega_{k,i}}\right)\right|\left|H_k\left(e^{j\omega_{k,l+i}}\right)\right|\right]^{\frac{1}{2}}, \quad -K \leq i \leq K \tag{3.16}$$

$$\left|\tilde{m}_{k,0F_0}[n]\right| = \left[\sum_{l=0}^{K} \left|m_l[n]\right|^2 \left|H_k\left(e^{j\omega_{k,l}}\right)\right|^2\right]^{\frac{1}{2}} \tag{3.17}$$

3.1.3 3 Harmonic Example: Estimated Envelope

Let's go back to our three harmonic example. We are again trying to acquire the 1st harmonic, $m_1[n]$ (green). We define $\omega_1 = 2F_0$.

We can see the relationships

$$\widehat{x}[n] \Longleftrightarrow \widehat{X}[n,f] \tag{3.18}$$

$$\widehat{x}[n]e^{-j2\pi\frac{2F_0}{F_s}n} \Longleftrightarrow \widehat{X}[n,f-2F_0) \tag{3.19}$$

$$\widehat{x}[n]e^{-j2\pi\frac{2F_0}{F_s}n} * h_2[n] \Longleftrightarrow \widehat{X}[n,f-2F_0)H_1(f) \tag{3.20}$$

$$\tilde{m}_1^2[n] \Longleftrightarrow \widehat{X}[n,f-2F_0)H_1(f) * \widehat{X}^*[n,-f+2F_0)H_1^*(-f) \tag{3.21}$$

The interesting part of figure **??** is (f). We see our green component that we were looking for, however there are a whole lot of other things that we didn't want.

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

The other components come from interactions with the unwanted harmonics that we failed to completely filter out. For clarity the convolution is visualized in figures **??**, 3.3, 3.4. Positive and negative components are mirror images so the positive components are not explicitly visualized.

3.2 Steady-State Metrics

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

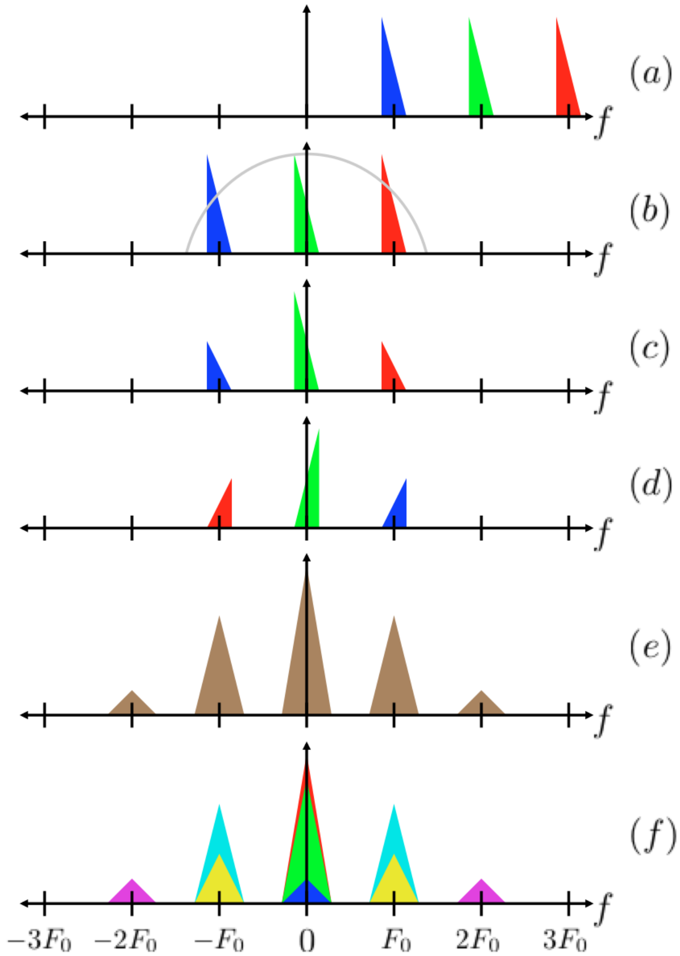

Figure 3.2: $(a)|\widehat{X}[n,f)|$ $(b)|\widehat{X}[n,f-2F_0)|$ $(c)|\widehat{X}[n,f-2F_0)||H_1(f)|$ $(d)|\widehat{X}^*[n,-f+2F_0)||H_1(-f)|$ $(e)|\mathcal{F}\{\tilde{m}_1^2[n]\}|$ (f) contributions of separate components of (e)

3.2.1 Coherent Gain

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

$$G_k = \left| H_k\left(e^{j\omega_{k,k}}\right) \right| \tag{3.22}$$

Recalling equation 3.12, 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.

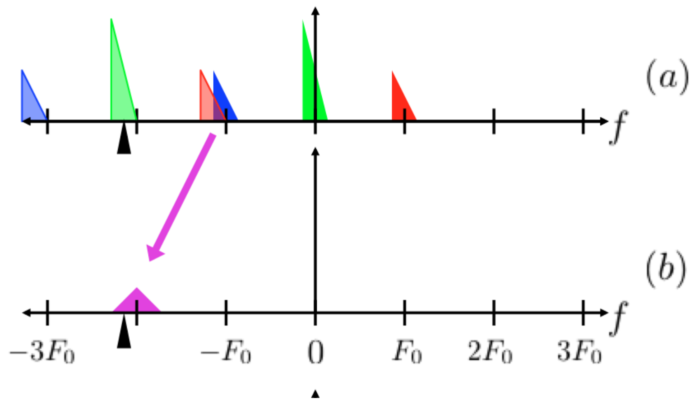

Figure 3.3: Envelope Estimate $-2F_0$ Component

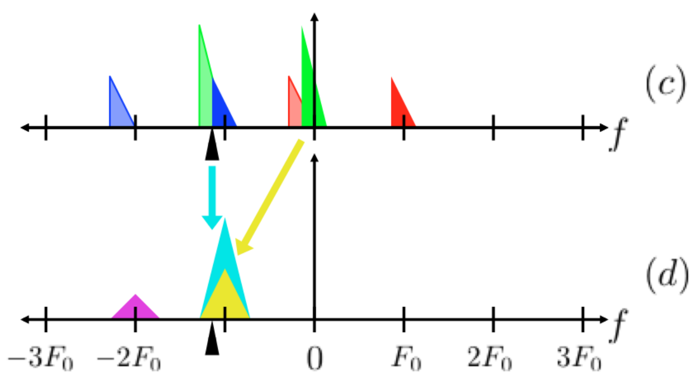

Figure 3.4: Envelope Estimate $-F_0$ Component

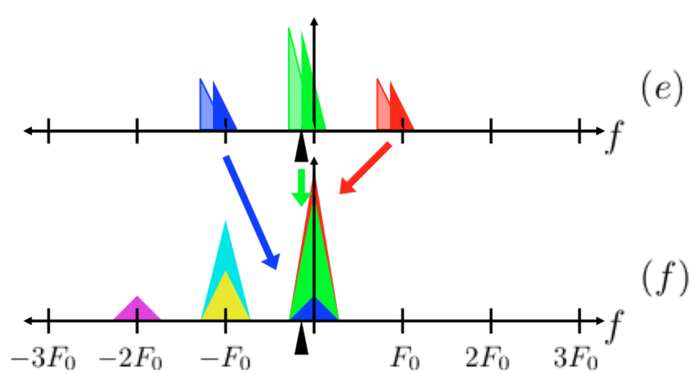

Figure 3.5: Envelope Estimate Baseband Component

$$G_k = \left|H_k(0)\right| = \sum_n h_k[n] \tag{3.23}$$

We may further simplify this by normalizing our filter such that $\left|H_k(0)\right| = 1$. Of course, our 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.

3.2.2 Harmonic SIR

Continuing our focus on the baseband, another question is: what is the contribution of the target harmonic versus 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 **??**(f). The harmonic signal-to-interference-ratio (SIR) quantifies the ratio of target harmonic to spectral leakage.

$$SIR_k = \frac{\left|H_k\left(e^{j\omega_{k,k}}\right)\right|}{\left[\sum_{l=0}^{K}\left|H_k\left(e^{j\omega_{k,l}}\right)\right|^2\right]^{\frac{1}{2}}} \tag{3.24}$$

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. $k-2 \leq l \leq k+2$.

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 our system to arbitrary harmonic inputs.

3.2.3 Modulation Depth

Finally, we consider the magnitude of each bandpass component relative to baseband. These terms appear in our envelope estimate as modulations at rates that are multiples of F_0. Because of the forced symmetry of the real envelope we only need to consider positive frequencies, iF_0.

$$D_{k,i} = \frac{\left[\sum_{l=0}^{K-i} \left|H_k\left(e^{j\omega_{k,l}}\right)\right| \left|H_k\left(e^{j\omega_{k,l+i}}\right)\right|\right]^{\frac{1}{2}}}{\left[\sum_{l=0}^{K} \left|H_k\left(e^{j\omega_{k,l}}\right)\right|^2\right]^{\frac{1}{2}}}, \quad 1 \leq i \leq K \tag{3.25}$$

The largest value and, for that reason, most important value is $D_{k,1}$, the modulation depth at F_0.

3.3 Explicit Temporal Modulation

So our three metrics are coherent gain, harmonic SIR and modulation depth. We aim for a coherent gain of $G_k = 1$, maximum possible SIR and one would think minimum modulation depth.

Interestingly, some current CI processing strategies such as ACE intentionally allow for induced modulations from non-isolated harmonics. This provides a temporal cue to the user which plays into pitch percept.

The alternative is to use narrow enough filter cutoffs to eliminate these modulations, and then explicitly modulate the signal. In this option we need further processing such as a pitch estimator to determine the modulation rate.

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

Let's consider a single note played by two different instruments, clarinet and saxophone. In this example $F_0 = 261Hz$. The clarinet is interesting in that it only has energy at odd harmonics.

We attempt to estimate the 3rd harmonic, $m_3[n]$. We first downshift by $-3F_0$, then lowpass filter. The spectrum of each signal at this stage is visualized in figure **??**. 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. Here we start to see the problem, that despite the wide bandwidth filter, there is no energy around $\pm F_0$ for the clarinet because of the harmonic structure.

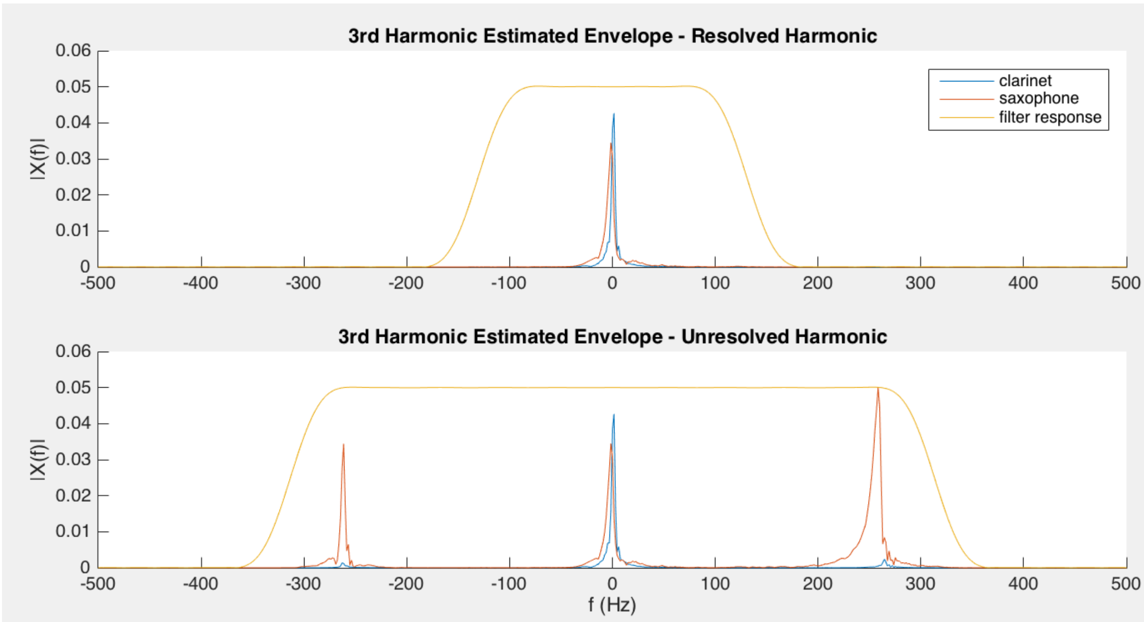

Figure 3.6: Clarinet vs Saxophone Harmonic Components

Figure **??** 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, were we have a very large F_0 modulation in the saxophone envelope but little to no change in the clarinet. The result is that we have 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. Furthermore, if we have uniform bandwidth filters, (as ACE does), the harmonic resolution will not behave as it does in the cochlea.

Beyond this example, explicit modulation decouples F_0 and modulation depth. This way we have much more control over modulation depth while still making optimal design decisions for envelope extraction. We can decide modulation depth as a function of how harmonic the signal is. eTone [REF] uses a harmonic probability metric to do just that.

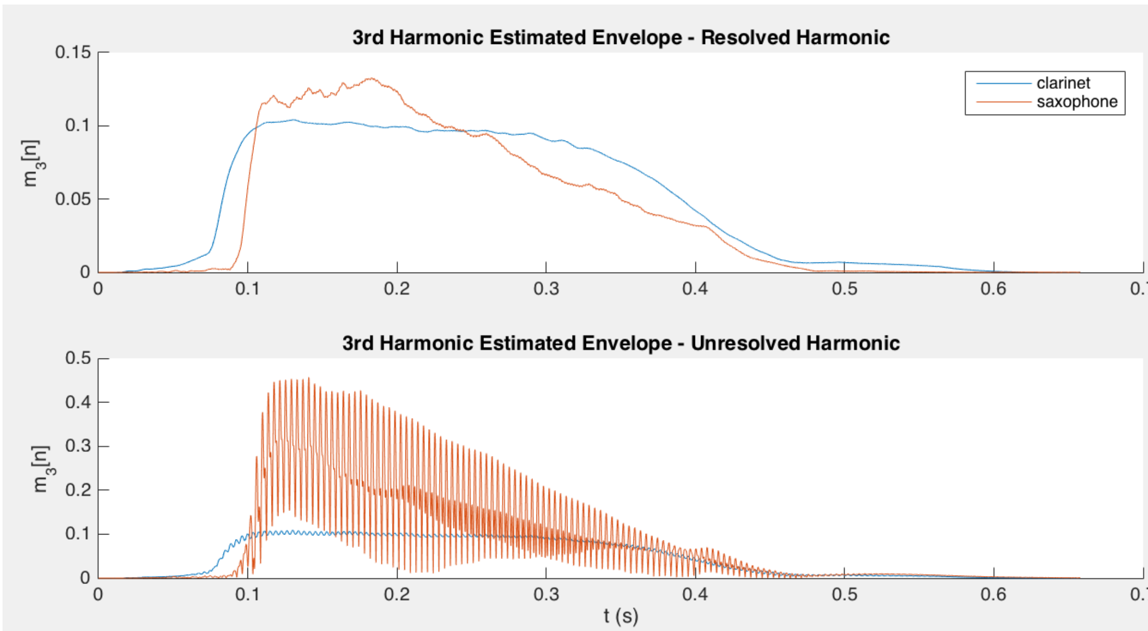

Figure 3.7: Clarinet vs Saxophone Envelope Estimates

3.3.1 Followup Filter

Another thing to note is that regardless of downshift frequency, our harmonic envelope will always have it's energy centered at baseband and 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 [REF?] that have used this additional filter. eTone's envelope follower is an example of this.

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

In terms of our three metrics, the followup filter will provide us with 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 additional processing stage. The additional stage means more memory, clock cycles and processing delay.

3.4 Steady-State Evaluation of Strategies

3.4.1 design parameters

filter and downshift

ideally: BW = F0/2 downshift = exp(kF0)

downshift quantization, bandwidth as function of F, F0?

modulation depth (kind of another SIR) as a function of downshift quantization and filter

3.4.2 figures

3.5 Transients

quantify for best case and worst case where worst case is the fastest transient relevant to music (this should also hinge on CI limitations)

maximum onset dynamic range "90ms - 10ms = 80ms" and ratio of filter smeared range to max range rinse and repeat for CI's

3.6 Changing F0

dips due to quantization

3.6.1 scalloping loss

[windows for harmonic analysis]

scalloping loss or picket-fence effect, ratio of coherent gain for tone located half a bin from DFT sample point to coherent gain for tone located exactly at sample point

$$scallopingloss = \frac{|H(\frac{1}{2}\frac{F_s}{N})|}{H(0)} \tag{3.26}$$

"althought scalloping loss is useful, it's not entirely informative. if the scalloping loss if high, then this relates to a sharp cutoff which is actually good for increasing purity of the harmonic" worst case processing loss = scalloping loss * PL where PL is reduced gain of window (which i have been canceling out) **where does worst case processing loss fit in?**

ALL METRICS:

ENBW (accumulated noise) PL (gain at DC) PG (same as PL?) scalloping loss (down-shift quantization worst case) worst case PL = PL*SL

harmonic SIR harmonic gain (maybe not overly relevant due to AGC, etc) (how is it affected in a relative sense? worst vs best) modulation depth (spectral leakage) transients changing F0 (scalloping loss dip only, and harmonic SIR)

Chapter 4

EFFICIENT 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]

Chapter 5

Chapter 6

CONCLUSION

6.1 Summary

6.2 Future Work

BIBLIOGRAPHY

[1] 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.

[2] 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.

Appendix A

WHERE TO FIND THE FILES

The uwthesis class file, `uwthesis.cls`, contains the parameter settings, macro definitions, and other TeXnical commands which allow LaTeX to format a thesis. The source to the document you are reading, `uwthesis.tex`, contains many formatting examples which you may find useful. The bibliography database, `uwthesis.bib`, contains instructions to BibTeX to create and format the bibliography. You can find the latest of these files on:

- My page.

  `http://staff.washington.edu/fox/tex/uwthesis.html`

- CTAN

  `http://tug.ctan.org/tex-archive/macros/latex/contrib/uwthesis/`

  (not always as up-to-date as my site)

VITA

Jim Fox is a Software Engineer with UW Information Technology at the University of Washington. His duties do not include maintaining this package. That is rather an avocation which he enjoys as time and circumstance allow.

He welcomes your comments to `fox@uw.edu`.