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 is my abstract

- here's an item [1]

- item number 2

[1]here's a footnote

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

1.1 Overview

1.2 Survey of Literature

1.3 Contents of Thesis

Chapter 2
BACKGROUND

2.1 Acoustic Hearing

Before discussing how cochlear implants are able to restore hearing in people with profound hearing impairment, it is useful to talk about how acoustic hearing works in normal listeners.

2.1.1 Anatomy of the Ear

tonotopic, critical bands, ...

2.1.2 speech

2.1.3 pitch

2.1.4 other characteristics of sound

2.2 cochlear implants

CI basics (CIS) vague envelope concepts without math

”In spite of the fact that this analog signal itself preserves most of the original temporospectral in- formation, the signal transfers to the auditory nerve is handicapped by the limited maximal firing fre- quency of the auditory nerve in response to the electrical stimulation. High synchronization of nerve fibers and the neural refractory period only allow for frequency transmission up to 1 kHz via tempo- ral coding alone. For frequencies above 1 kHz, the spectral information cannot be sufficiently trans- ferred by temporal coding alone. Multichannel im- plants have been developed to make use of the tonotopic organization in the cochlea and thus transmit more spectral information to the auditory nerve.” [1]

”The HiRes120 strategy, used in the Advanced Bionics implant, is the first commercial stimulating strategy that uses the virtual channel technique. Virtual channels are created by adjusting the current level ratio of two neighboring electrodes.”

2.3 signal models

A classic way of analyzing audio signals is the sum-of-products model. In this model a signal is represented as a sum of narrowband signals $x_k[n]$ at distributed center frequencies. Each of the narrowband components is modeled as a product of a slow-time-varying envelope $m_k[n]$ and a fast-time-varying carrier $c_k[n]$.

$$x[n] = \sum_k x_k[n] = Re\Big\{ \sum_k m_k[n] c_k[n] \Big\}$$

This model works especially well for harmonic signals, where each individual harmonic is represented by a narrowband component. In the case of speech, one could think of the collection of carriers as the harmonic signal generated by the vocal tract, while the modulators together represent the resonant structure that distinguishes separate vowels. ¡–figure?! Continue...

In this model information such as gender of talker, intonation or musical pitch would be dominantly characterized by the carrier. The resonant information that distinguishes different vowels or the unique formant structure of a particular instrument would be encoded in the modulator.

The sum-of-products model is particularly convenient for CIs for a few different reasons. For starters, this model is naturally similar to the way our ears work. ”mechanical fourier transform” [chimeara] The initial goal of CI researchers was to achieve speech recognition. Given the limitations of CIs, encoding only the modulation information acts as a form of lossy data compression.

More reasons this model is good, musically? Why is it good for CIs?

Provided the above motivation for sum-of-products signal modeling we come to the challenge of how exactly to decompose a signal. There are many different ways of doing this but they broadly fall into one of two categories: incoherent methods and coherent methods.

2.3.1 Incoherent Methods

With incoherent methods, an envelope is extracted independent of the carrier; no information about the carrier is taken into account prior to envelope extraction.

One such method is the short-time Fourier transform (STFT), which has two classic interpretations: a series windowed Fourier transforms or a collection of uniform bandpass filters. For our purposes we will be using the later.

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

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

Defining a new variable $r' = r - n$ and defining our window such that $w[n] \neq 0$ for $0 \leq n \leq N-1$,

$$X[n,k] = \sum_{r'=0}^{N-1} x[n+r']w[r']e^{-j\frac{2\pi}{N}k(n+r')}$$

$$= e^{-j\frac{2\pi}{N}kn} \sum_{r'=0}^{N-1} x[n+r']w[r']e^{-j\frac{2\pi}{N}kr'}$$

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

$$X[n,k] = |X[n,k]|e^{j\angle X[n,k]}$$

If we assume that the window $w[n] \neq 0$ for $0 \leq n \leq N-1$ then we have the inverse

$$x[n+r'] = \frac{1}{Nw[r]} \sum_{k=0}^{N-1} X[n,k]e^{j\frac{2\pi}{N}k(n+r')}$$

$$= \frac{1}{Nw[r']} \sum_{k=0}^{N-1} |X[n,k]|e^{j(\frac{2\pi}{N}k(n+r')+\angle X[n,k])}$$

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

We can now clearly see our sum-of-products model

$$m_{k,STFT}[n] = \frac{1}{Nw[0]}|X[n,k]|$$

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

Alternatively, a Hilbert transform uses the analytic signal to separate envelope from carrier.

$$x_k[n] = x[n] * h_k[n]$$

Where h_k is a bandpass filter and k has arbitrary limits

$$\widehat{x_k}[n] = x_k[n] + jH\{x_k[n]\}$$

$$m_{k,hilbert}[n] = |\widehat{x_k}[n]|$$

$$c_{k,hilbert}[n] = cos(\angle\widehat{x_k}[n])$$

OR

$$c_{k,hilbert}[n] = e^{j\angle\widehat{x_k}[n]}$$

Since "the Hilbert transform of a convolution is the convolution of the Hilbert transform on either factor" [wikipedia] we have

$$\widehat{x_k}[n] = x_k[n] + jH\{x_k[n]\}$$

$$= x[n] * h_k[n] + jH\{x[n] * h_k[n]\}$$

$$= x[n] * h_k[n] + x[n] * jH\{h_k[n]\}$$

$$= x[n] * [h_k[n] + jH\{h_k[n]\}]$$

Now let us define our filter specifically as

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

If we assume the sidelobes of $w[n]$ roll-off sufficiently fast in relation to the center-frequency $\frac{2\pi k}{N}$, we may approximate

$$H\{h_i[n]\} \approx \frac{1}{Nw[0]} w[-n] H\{cos(\frac{2\pi}{N} in)\}$$

$$= \frac{1}{Nw[0]} w[-n] sin(\frac{2\pi}{N} in)$$

To verify the previous equation, consider the extremes: 1) $w[n] = 1$ 2) $w[n] = \delta[n]$ Now I'll go into details! (TODO)

$$\widehat{x}_i[n] \approx x[n] * \frac{1}{Nw[0]} w[-n] e^{j\frac{2\pi}{N}in}$$

$$= \frac{1}{Nw[0]} \sum_{r=-\infty}^{\infty} x[n-r] w[-r] e^{j\frac{2\pi}{N}ir}$$

Let $r' = -r$

$$= \frac{1}{Nw[0]} \sum_{r'=0}^{N-1} x[n+r'] w[r'] e^{-j\frac{2\pi}{N}kr'}$$

$$= \frac{1}{Nw[0]} \left[e^{-j\frac{2\pi}{N}kn} \sum_{r'=0}^{N-1} x[n+r'] w[r'] e^{-j\frac{2\pi}{N}kr'} \right] e^{j\frac{2\pi}{N}kn}$$

$$= \frac{1}{Nw[0]} X[n,i] e^{j\frac{2\pi}{N}kn}$$

What this tells us is that under the assumption of fast sidelobe rolloff we may define a filter bank of N filters

$$h_k[n] = w[-n] cos(\frac{2\pi}{N} kn), 0 \leq k \leq N-1$$

We have

$$m_{k,hilbert}[n] \approx m_{k,STFT}[n]$$

$$c_{k,hilbert}[n] \approx Re\{c_{k,STFT}[n]\}$$

Hilbert Envelope - where is this done in practice?? sounds like hires120 does this

ACE - STFT as a bank of filters pitch modulation in ACE

CIS - BPF -¿ rectification -¿ LPF typically 200400Hz cutoff frequency ”Unlike ACE, all 16 frequency bands are then stimulated in sequence”

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. [2]

One 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 frequency 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 doesn't escape the limitation of fixed and pre-determined band limits that each filter operates within.

To escape this, Atlas and Others 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 get a pitch estimate $F_0[n]$ of the signal. We then define k complex carriers where there is a hard limit as a function of Nyquist sampling rate, $k \leq \lfloor \frac{F_s}{2F_0} \rfloor$

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

where

$$\phi_0[n] = \phi_0[n-1] + 2\pi \frac{F_0[n]}{F_s}$$

[modulation toolbox]

We define our bandpass signal as

$$x_k[n] = x[n] * h_k[n]$$

where the key difference between this and the incoherent methods is that $h_k[n]$ is centered at F_0, that is, it changes adaptively as a function of F_0.

$$m_{k,harmonic}[n] = x_k[n] c^*_{k,harmonic}[n]$$

We may choose to design our filters such that

$$h_k[n] = \frac{2}{Nw[0]} w[-n] cos(k\phi_0[n])$$

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

$$w[n] \neq 0, 0 \leq n < N$$

$$= 0, else$$

In this case each bandpass filter has the same response and is simply shifted to a different center frequency as a function of F_0 and k. This is a natural design since the spacing of harmonics is uniform in frequency.

$$m_{k,harmonic}[n] = \Big[x[n] * h_k[n]\Big] e^{-jk\phi_0[n]}$$

$$= \Big[x[n] * \frac{2}{Nw[0]} w[-n] cos(k\phi_0[n])\Big] e^{-jk\phi_0[n]}$$

$$= \sum_{r=-\infty}^{\infty} x[n-r] \frac{2}{Nw[0]} w[-r] \Big[\frac{1}{2} e^{jk\phi_0[r]} + \frac{1}{2} e^{-jk\phi_0[r]}\Big] e^{-jk\phi_0[n]}$$

let $r' = -r$

$$= \frac{1}{Nw[0]} \Big[e^{-jk\phi_0[n]} \sum_{r'=0}^{N-1} x[n+r'] w[r'] e^{-jk\phi_0[r']}\Big]$$

$$= \frac{1}{Nw[0]} X[n, k\phi_0[n] \frac{N}{2\pi n})$$

$$= \frac{1}{Nw[0]} X[n, k\lambda[n])$$

where we define $\lambda = \phi_0[n] \frac{N}{2\pi n}$. The ")" is a reminder that $k\lambda[n]$ is not restricted to integer values. We can see the relationship to the STFT envelope as

$$\lambda[n] = 1 \Rightarrow |m_{k,harmonic}[n]| = m_{k,STFT}[n]$$

$$\lambda[n] = 1 \Rightarrow c_{k,harmonic}[n] e^{j\angle m_{k,harmonic}[n]} = c_{k,STFT}[n]$$

We can now see that the STFT envelope and magnitude of the harmonic envelope are related with two key differences:

1) $\lambda[n]$ is not necessarily an integer

2) $\lambda[n]$ is time-varying

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. That being said, there is no guarantee that this quantization will lead to integer values of $\lambda[n]$.

Moving on to the carriers, note that in the coherent harmonic method the phase information split into two components, one of which is part of the envelope. We can think of this component as the fast variations unaccounted for by the smooth F_0 estimate as well as any drifting from kF_0 because the relationship is often not a perfect integer multiple or because the estimate of F_0 is off.

Talk about filter bandwidth, what is optimal? F0/2? F0/4?

With harmonic decomposition we are faced with a decision. Our envelope is now complex, so if we need to use the envelope independent of carrier we must somehow convert it to a real signal. The two natural options are two either take the real component or the magnitude. Let me describe the advantages of each...

2.3.3 Phase Vocoder

phase vocoder or other?

2.3.4 Modulator Bandwidth Tradeoff

incorporating multiple harmonic components VS eliminating transient characteristics

Chapter 3

CI DSP STRATEGIES

3.1 A General Framework

3.2 F0 Estimation

3.3 Envelope Extraction

tie the DSP theory into the pros and cons and limitations of CIs

3.4 Electrode Subset Selection

Chapter 4
HHE

come up with a better name!!!

4.1 Algorithm

1) Filter Center Frequency 2) Filter BW 3) Effective Channel Information BW

4.2 Freedom details

Chapter 5

SUBJECT TESTS

initial results are...

Chapter 6

CONCLUSION

6.1 Summary

6.2 Future Work

Chapter 7

OOOOOOOOOOOOOOOO

Chapter 8
THE GOODS

8.1 the tools

8.1.1 types of audio

speech, music, harmonic, inharmonic, voiced, unvoiced tonal, non-tonal, consonant disonant..., transient, steady state

8.1.2 audio signal models

SUM [A(t) * cos(wt + phi(t))] etc... TFS and envelope

8.1.3 cochlear implants

what is important? hearing for any general reason...safety, functionality speech recognition what is important and lacking? music appreciation tonal language SiN quality

8.2 A General Framework for CI Processing

8.3 F0 Estimation

our (shared) technique e-tone? harmonic sieve, etc. latency, accuracy, octave errors and range restrictions, quantization

8.4 Envelope Extraction

8.4.1 incoherent methods

CIS... fft hilbert

8.4.2 coherent methods

spectral center of gravity harmonic

8.4.3 the relationships

how different are they? talk about interpolation, fft padding what does this mean when considering hilbert envelope or other method without identical filters what does it mean to have a filter that changes between frames? does it cause artifacts?

8.4.4 other factors from below

phase preservation, etc.

8.5 channel mapping

8.5.1 background

why only 8 at a time? ACE vs CIS and benefits of each

8.5.2 harmonic to channel allocation

8.5.3 channel selection

8.6 alternative coherent envelope computation

8.7 less theoretical stuff

8.8 test results

8.8.1 simulated real-time

8.8.2 mandarin tones pitch tilt

8.8.3 freedom processor

speech recognition... timbre recognition... other...

Chapter 9

GETTING STUFF INTO THIS DOCUMENT ASAP

These are things that I have done so far that I would like to start putting into written form. The figures will probably need to be edited, so matlab scripts that can do just that may be found in appropriate folders.

9.1 A General Framework for CI Processing Strategies

There are numerous stages to processing in cochlear implants. The main components are visualized in Fig. ?? below. 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 section I will talk about the general differences between ACE, F0mod, HSSE

Figure 9.1: Signal Flow in CI

9.1.1 ACE

The simplest of the considered strategies is the Advanced Combination Encoder (ACE). ACE has become a clinical standard for CI processing and is used in a vast number of users. In essence, hilbert envelopes are extracted from the signal using an FFT. The magnitudes of the FFT bins are then combined to form a single envelope per channel. During each processing frame, a subset of the channel envelopes is selected for stimulation on the internal implant.

Figure 9.2: ACE Flow Diagram

While ACE does a sufficient job for many CI users in speech recognition tasks, a large gap remains between NH and CI in pitch discrimination. ACE uses place cues as the primary source of encoding a sound's characteristics. To this day it is still unclear as to what implications this has. This is due to a combination of factors including the subjective nature of pitch and absence of a ground truth baseline in many CI users. For example, high-pass filtering a sound may cause it to sound brighter. In contrast low-pass filtering would cause a warm quality. As stimulus change electrodes a CI user could claim to experience changes in the high-low quality of pitch when really they are experiencing changes in the bright-warm quality of spectral distribution, or more likely an ambiguous combination of both.

There is general consensus that place cues are not sufficient for encoding pitch. Alternatively, temporal cues encoded as time-domain carrier modulations have shown to be promising.

ACE does provide limited temporal modulations through the bandwidth of the STFT filters. Further discussion of STFT filters may be found later in this document.

9.1.2 F0mod

To get at the problem of pitch discrimination, (Laneau et al 2006) developed a new research strategy, F0mod. F0mod 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 (F_0) is acquired, all output envelopes are modulated by a raised sinusoid at a rate of F_0. This raised sinusoid is constant modulation depth, (full dynamic range), and same across channels, (phase aligned). *maybe show a figure? The details of modulator type are discussed later in section ?.?. The important point here is that modulations are applied at a rate of F_0 and full depth.

Figure 9.3: F0mod Flow Diagram

F0mod has shown promising results in acute tests for pitch discrimination. It has also inspired other processing strategies such as eTone, which uses a more sophisticated harmonic sieve pitch estimator as well as soft decisions to overcome the problem of encoding both harmonic and inharmonic sounds as well as those that fall somewhere in between.

9.1.3 HSSE

Looking for a novel approach to improved pitch perception and more broadly music perception, (Li, Atlas, Nie) came up with Harmonic Single Sideband Encoder (HSSE). HSSE uses coherent demodulation to extract harmonic envelopes. These envelopes are then combined with carrier modulators, just as in F0mod. These combined carrier-envelope modulators are then assigned to channels based on the harmonic index and F_0.

Sparing details which we will soon investigate deeper, the differences between F0mod and HSSE can be summarized quite simply; Every stage of typical ACE processing is now done coherently using F0 information.

Because the envelopes are now harmonically related, they do not directly correspond to channels. The number of envelopes will typically be different than the number of channels and more importantly the frequency ranges of the channels and the harmonic envelopes will never be a one-to-one mapping. For these reasons an additional channel allocation stage

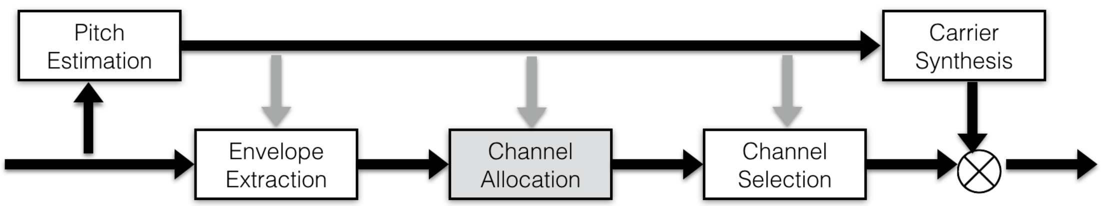

Figure 9.4: HSSE Flow Diagram

must be added, mapping each harmonic envelope to a channel and potentially combining multiple harmonic envelopes into one channel.

9.1.4 Other Strategies

any hybrid considerations? maybe hint at hsse ace hybrid

talk about unmentioned methods (AB, MedEl)

9.1.5 Summary

REWRITE THIS WHOLE SECTION

To summarize, the considered strategies can be described by four general processing blocks, Pitch Estimation, Envelope Extraction, Channel Allocation and Channel Selection.

1) Pitch Estimation

Fundamental Frequency Modulation is a key prospect in temporal encoding, shared by F0mod and HSSE, but not ACE. There are various ways to estimate pitch with trade-offs for each. We are going to assume the pitch estimator is the same when analyzing F0mod and HSSE.

2) Envelope Extraction

This has not been talked about in detail intentionally. The goal of this section was preparation and justification of a deeper analysis of envelope extraction methods.

3) Channel Allocation

Although this processing block was only introduced with HSSE, it may be view somewhat like an identity operation for the other strategies, allocating envelope to channel in a constant one-to-one mapping independent of input signal or time frame.

4) Channel Selection

This is a key stage to N-of-M strategies. There are a few different options considered for HSSE. In the case that only N channels are non-zero after the Channel Allocation stage, this may be viewed as an identity operation.

9.2 F0 Estimation

talk about F0 estimator and alternatives...

9.3 Envelope Extraction Methods

This is about Coherent vs Incoherent Envelopes and the details of each Incoherent method

how much detail are we going into? Compare CIS, ACE, Hilbert. Compare ACE128 to ACE512...

where do the matlab figs fall into play here?? (probably way later after math, etc)

Let's do more detailed figures highlighting the differences between F0mod/ACE and HSSE envelope techniques!

Fig: FFT -¿ Sum into Channels

Fig: Coherent Demod, maybe we need to talk about types of coherent envelopes first?

up until now we haven't needed to be explicit in what envelopes we are talking about, we now need to define these various terms:

incoherent vs coherent

harmonic vs hilbert (FFT bin)

9.3.1 Phase Preservation

previous version of HSSE looked like: (preserve phase)

phase alignment has proven to be important in both HSSE and F0mod, but is that because we're screwing it up???

reduces the posibilites to just rectification, instead of various forms of modulation possible in CIs, (who cares though)

F0mod tries to separate magnitude and phase information. HSSE recognizes this is a bad idea and keeps them together.

Math!

RESULTS:

- do we need harmonic envelopes for this to matter? doesn't matter even when you use them

- what are implications of downshift? what does it mean to apply phase to a different carrier? does this introduce a phase delay? what is it? Don't need to worry about it

- hypothesis: phase needs to be scaled: phik(t) / k when signal is downshifted Totally true! explain this math

- okay, magnitude is not really different, but still is sorta...phase is potentially very important though. What happens to the phase with downshift, how does my hypothesis play into this, what if the pitch estimate is off by a little bit? how well are the signals aligned when voiced or unvoiced?

used "shh" vs "saw" test. at least when listening to the simulations, the processing essentially sounds like narrowband resonant filters. The noise-like sounds are completely dominated by the filter bandwidth and the phase-information is not noticeable at all.

for high frequencies this works out because we would have needed to find a way to combine these non-phase-aligned envelopes

downshift frequencies are quantized to same as FFT (256 frequencies spaced 30Hz apart) doesn't matter though, gain is same... (¡ -1.4dB dip) NOT TRUE!!! Roll-off is not linear in dB, so since signal is not pure tone, components will roll off at faster or slower rates

9.3.2 Magnitude and Bin Alignment

The human ear has much better resolution than the cochlear implant sound processor when decomposing a signal into frequency bands. The artifacts of this can be clearly demonstrated by example. In case1, the energy of the signal falls directly on the center frequency of an FFT bin. In case2 the signal falls in between two bins. In this case, neither bin represents the true energy of the signal.

HOWEVER:

Coherent is the Same (mathematically) as Hilbert, as ACE, as CIS except...for the downshift frequency. This leads to a minimal (-1.6dB max) loss of gain for the desired frequency however it may lead to lower SNRs when desired frequency is further from center of filter and noise is closer to center of filter simultaneously.

show plots as well as math :D

9.3.3 Filter Design & Bin Combination

Woah...come back to CIS vs ACE etc for this!

Bin Combo

now that we have considered phase and magnitude, this component of HSSE can essentially be considered as a different combination of FFT bin magnitudes when compared to ACE.

as mentioned above hsse takes F0 into account and avoid bin alignment issues, however, inaccuracies in F0 estimate can lead to loosing high energy harmonics with narrowband filters. likely need to just combine unless F0 estimator can be significantly improved

This is is where the critical band concepts come into play, would this mess up speech in noise goals? probably...but what can be done if we can't get a good pitch estimate? filtering F0 could help this a bit but it introduces further delay

updating only 9 samples of downshift per frame rather than grabbing complete complex exponential could help however once the channels are combined it shouldn't matter

Filter Design

ACE currently uses modulations due to harmonic artifacts and low-order FFT. This is horrible! Let me explain why...it has nothing to do with the harmonic of interest and everything to do with the one harmonic below and one harmonic above the harmonic of interest. Because this demodulation is done incoherently the modulation depths are not directly related to the harmonic of interest. Furthermore, the cutoff is fixed and decided by parameters of the FFT and sampling rate which have nothing to do with the signal itself. This makes the modulation even further unrelated to the signal. (Could this also theoretically be a problem for F0mod? Case: F_0 is very low and the harmonic lands right between two bins. A small modulation could come about, probably not)

An important detail to note is that of low-order-FFT induced modulations mentioned for ACE. Laneau explicitly describes two different methods as ACE128 and ACE512 corresponding to different FFT orders. F0mod uses ACE512 which keeps FFT bin modulations below roughly 60Hz in contrast to ACE128's 240Hz. This sharper cutoff keeps envelope modulations out of the carrier frequency range, isolating this component and leaving the role of carrier modulation to the explicit modulator at F_0.

This segregation allows for easier relation to the modulation model of sounds. Furthermore, F0mod is not prone to the modulation artifacts present in ACE128 and discussed in section 2.?.?

9.3.4 Unvoiced Signals

I really hope!!! This is well handled by two factors.

1) automatically choose high F0 when no good estimate exists. This allows for higher frequencies (more important and more likely to be present in unvoiced) to be acquired.

2) If filters are adaptive bandwidth, the wide-bandwidth filters will preserve more high-frequency noise-like modulations.

- Still no concrete solution for unvoiced signals, best answer so far is to have automatic high-F0 estimate during unvoiced sections (make it more stable than if bouncing between high and low)

9.3.5 Takeaway

- Phase Preservation doesn't matter (shh vs saw)

- Magnitude also doesn't matter (-1.6dB)

- HSSE may be viewed as a different way of combining FFT bin magnitudes. I would argue that we do this using F0 for low frequencies, and fixed for high. (critical bands!!!)

9.4 Channel Allocation

The key to HSSE here, is that we have isolated individual harmonics. Harmonics are mapped to associated fixed channels due to the limitations of a fixed number of channels and fixed locations in the cochlea. Because we have isolated individual harmonic envelopes there is no issue of signal energy falling in between channels.

9.4.1 Regularizer Heuristic

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.

9.4.2 Multiple Harmonics Per Channel

As far as having multiple harmonics in a single channel, there are a few solutions

1) Choose highest energy harmonic.

suffers from stability issues, what about gain?

2) Choose First

suffers from missing important harmonics in channel as well as misrepresenting unvoiced signals

3) Combine

How? via sum of squares?

does a gain factor need to be applied to each channel? how was this determined for ACE?

9.4.3 Takeaway

Low Frequencies: stability heuristic keeps from jumping channels when on edge.

High Frequencies: not really relevant if critical bands are used

- gains? maybe just use same as ACE since this should be pretty similar

9.5 Channel Selection

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.

What if we did F0mod with same channel selections as HSSE? What would happen?

9.5.1 Takeaway

- Fixed VS MaximaSelect: this is still up in the air, Fixed is complicated by not necessarily having harmonic envelopes

- for maxima select heuristics can be used to choose same if energy and F0 have not changed significantly

9.6 Selection & Mapping

ACE is Cochlear Ltd's instance of the auditory community's generalized category of N-of-M strategies. M is the number of electrodes (or channels) that may be activated and N is the number of electrodes active during any one stimulation frame, $N \leq M$. In these strategies an incoming audio stream is fed into a filter bank which then computes M modulator envelopes. In each processing frame, N maxima are chosen from the envelopes and are then mapped to their associated electrodes. $M \leq$ the number of electrodes on the device, so envelope to electrode is a simple 1-1 map.

It is important to note that this is the same case for F0mod. The carrier modulation is the same on each envelope and thus does not affect the selection process.

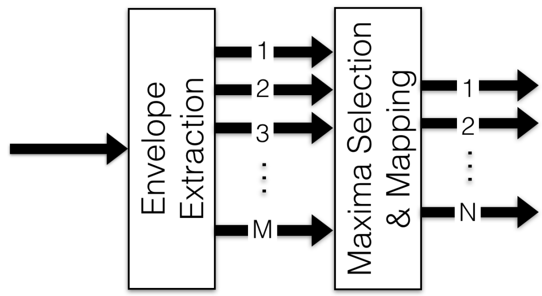

Figure 9.5: N-of-M Maxima Selection

On the other hand, HSSE outputs H harmonic envelopes, (one modulator envelope per harmonic). In this case, it is possible to have $H > N$ but ultimately it must be slimmed down to N envelopes per frame. Various ideas have been proposed including N-largest and lowest-N. 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.

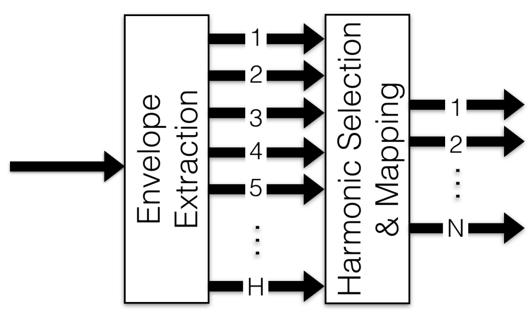

Figure 9.6: N-of-H Harmonic Selection

In HSSE it still holds true that $N \leq M$. While it may seem unnecessary, we could add intermediate steps where first N of M envelopes are set based on some harmonic selection process, then the remaining $M - N$ envelopes are set to zero. The second stage would be N-of-M maxima selection and mapping exactly as is done in ACE.

The added $M - N$ zero expansion give us a new way of interpreting HSSE where an identity-like operation is performed by a selection/mapping stage exactly the same as in ACE.

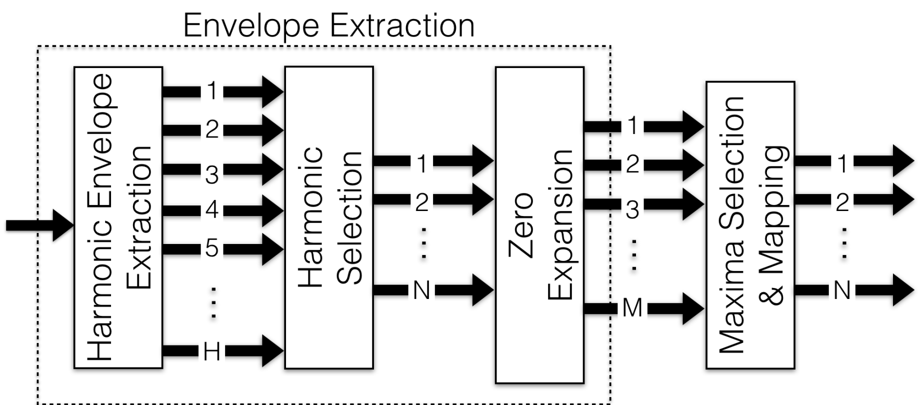

Figure 9.7: Reinterpreted N-of-H as N-of-M

Keeping this in consideration, we can append harmonic selection to the envelope extraction block in HSSE with the benefit to analysis that all differences between F0mod and HSSE are contained in this envelope extraction stage.

9.7 Alternative Coherent Envelope Calculation using FFT bins

This could all be achieved by zero padding

Math math math!

Incredibly frustrating...but do we even need this? What about just choosing the nearest FFT bin.

Another consideration:

9.8 Critical Bands

9.8.1 HSSE vs ACE vs Human Ear

In this subsection I will discuss the general differences in critical bandwidth:

1) how HSSE is too fine of a resolution note: HSSE originaly had BW = F0/2, however hard to implement and still not like ear

2) how ACE is overall a poorer resolution

What about doing a hybrid? This would further justify alternative HSUM in it's improved efficiency! If summing together anyway, does it matter if harmonic envelopes are used or incoherent envelopes are used?

How about specifying the bandwidth at each electrode as apposed to the frequency boundaries

Bro, you need to look into Xing's method with multiple harmonics modulated at multiples of F0...

9.8.2 Resolution Simulated by Adaptive Envelopes

The human ear has orders of magnitude more filters than ACE, (roughly 1500/22 I think).

HSSE could simulate this higher resolution by choosing different filter center frequencies based on the input signal

9.8.3 Channel Selection Analysis

ACE is like HSSE but for fixed FoI's. We extract an envelope at the FoI and then transmit it to the associated electrode.

1) this goes back to what are the implications of ACE512 vs ACE128 vs coherent-envelope if we are summing anyway

2) can HSSE be reanalyzed in these terms to better justify wide-bandwidth filters for high frequencies?

Could channel selection concepts in HSSE be important? Reflect on this in hindsight to recent discoveries. By this I mean using memory to not switch channels excessively and other decisions that were brought into account.

9.9 Other Important Components

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. What other flaws do these strategies have? Mention eTone and other possible solutions, or why we justify not considering these problems.

Chapter 10
LESS THEORETICAL STUFF

About this chappy

10.1 Engineering Decisions for Real-time

1) 8 harmonics this assumes we are dealing with musical instruments, speech is going to have characteristics well above the 8th harmonic. A hope is that with inharmonic signals the estimate will automatically bounce to max(F_0 estimate) which will thus hit the highest frequencies. This also goes back to the hybrid idea

2) F_0 estimation downsampling details, ooOOooo, so impressive!

10.2 F_0 tilt, exageration

mention the point that this was already done in Xing's paper, albeit $F_0/2$ without affine shift is more more likely to hit boundaries

10.3 Modulation Types

Chapter 11
TODO

11.1 DONE

11.1.1 ISDL

Brad

LPC (parametric modeling) FDLP

11.1.2 Xing Dissertation

Ch 1-3

11.1.3 HSSE vs F0mod Differences

11.1.4 HSSE vs F0mod More Differences

11.2 TODO

11.2.1 ISDL

Circularity: Asilomar Estimating Latent... Reverb VAD

Homomorphic Modulation Spectra

AAC+

Least-Squares Harmonic Modeling

Fan Chirp sec3-8 Fan Chirp Music

log-envelope (MFCCs?)

11.2.2 Xing Dissertation

Ch 4

Ch 5

Ch 6

11.2.3 HSSE vs F0mod Differences

harmonics are resolved

- how do we deal with should-be-unresolved harmonics?

channel combination

- further considerations are needed

- what does sum of squares mean? is it constant energy within the channel? does it cause a gain or just average the channels? look further into the gain component to ACE it's just a ¡1 gain for multiple bins in one channel

- can harmonics be combined? (higher harmonics) what does it mean to combine channel phase information?

channel selection

- further considerations are needed

- What if we did F0mod with same channel selections as HSSE? What would happen?

11.2.4 HSSE vs F0mod More Differences

recitified modulator (likely not too important)

also, pitch tilts

how can all of this be applied to soft decisions?

how can this all be done in real-time?

how are we accounting for non-linearities: AGC and sensitivity

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`.