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

Chapter 2
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.

2.1 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 2.1: Signal Flow in CI

2.1.1 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

While ACE does a sufficient job for most 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

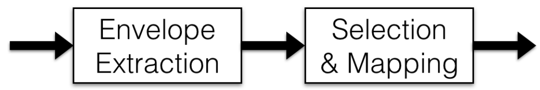

Figure 2.2: ACE Flow Diagram

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 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 rates are not exactly at F_0 and even worse, they aren't aligned with one another (what are the implications of this?). Furthermore, the cutoff is fixed and decided by parameters of the FFT and sampling rate which have nothing to do with the signal itself. (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)

2.1.2 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.

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.?.?

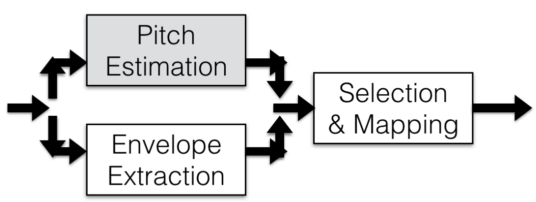

Figure 2.3: F0mod Flow Diagram

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.

2.1.3 HSSE

Looking for a novel approach to improved pitch perception and more broadly music perception, (Li, Atlas, Nie) came up with 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.

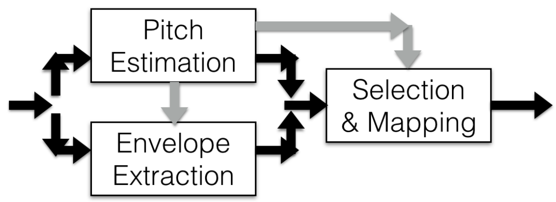

Figure 2.4: HSSE Flow Diagram

Sparing details which we will soon investigate deeper, the differences between F0mod and HSSE can be summarized in two simple points:

1) Envelope extraction is done coherently using F_0 from the pitch estimator.

2) Selection and Mapping are now on a different type of envelope and new considerations must be taken into account.

I will argue that all differences can actually be isolated to the envelope extraction stage, however, an important background topic must first be discussed.

2.1.4 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.

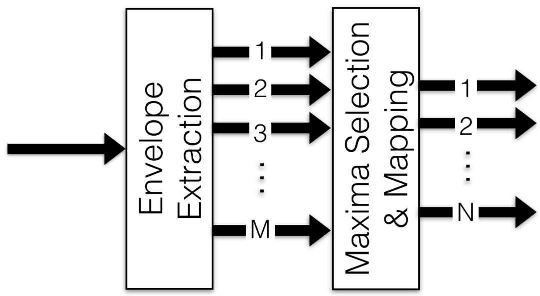

Figure 2.5: N-of-M Maxima Selection

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.

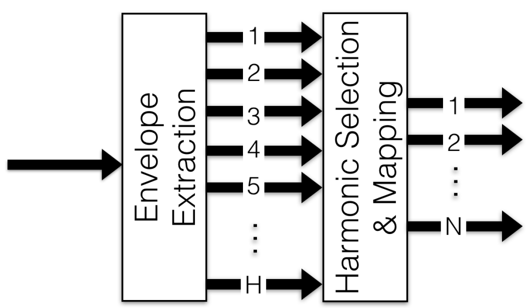

Figure 2.6: N-of-H Harmonic Selection

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.

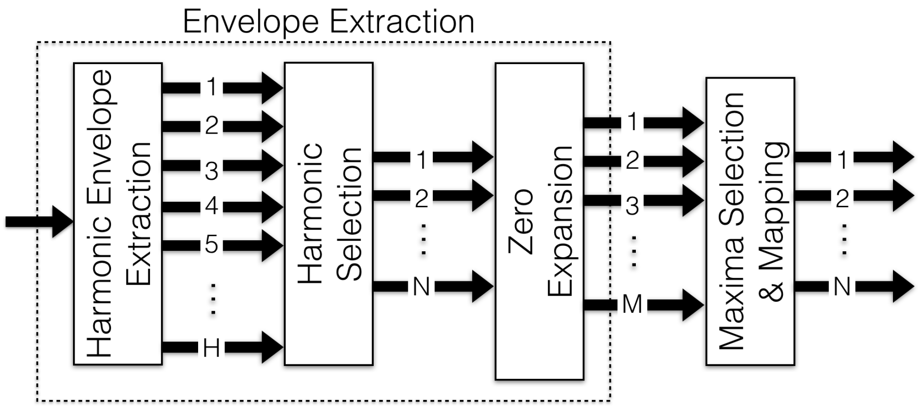

Figure 2.7: Reinterpreted N-of-H as N-of-M

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.

2.1.5 Other Strategies

any hybrid considerations? maybe hint at hsse ace hybrid

talk about unmentioned methods (AB, MedEl)

2.1.6 Summary

To summarize, the considered strategies can be described by three general processing blocks, Pitch Estimation, Envelope Extraction and Selection & Mapping.

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) Selection and Mapping

By appending harmonic selection to the envelope extraction block analysis becomes simpler in that all strategies are the same within the selection and mapping block.

2.2 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)

2.2.1 Why We Can't Take Real

I think that this stems from the complex filter. If our F_0 estimate is off by just the slightest bit, then the energy of the signal shifts into the imaginary component. As a result the energy seems to be lost when taking the real.

furthermore, phase alignment has proven to be imprtant in both HSSE and F0mod

reduces the posibilites to just rectification, instead of various forms of modulation possible in CIs

phase alignment has been found to be important

2.2.2 Coherent is the Same (mathematically) as Hilbert, as ACE, as CIS except...

...for the downshift frequency

2.2.3 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.

2.3 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:

2.4 Critical Bands

2.4.1 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...

2.4.2 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

2.4.3 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.

2.5 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 3

NEW ATTEMPT: HSSE VS F0MOD

3.1 The Components

1) Envelope Extraction

2) Channel Combination

3) Channel Selection

3.2 Envelope Extraction

3.2.1 Magnitude

I have looked into a few different ways to get magnitude envelopes. As it turns out...they're all pretty much the same.

3.2.2 Real

Not sure where this is going to lead me

I think that this stems from the complex filter. If our F_0 estimate is off by just the slightest bit, then the energy of the signal shifts into the imaginary component. As a result the energy seems to be lost when taking the real.

furthermore, phase alignment has proven to be imprtant in both HSSE and F0mod

reduces the posibilites to just rectification, instead of various forms of modulation possible in CIs

phase alignment has been found to be important. Can we do real version without phase misalignment problem?

3.3 Channel Combination

does remanian/geodesic distance have anything to do with sum of squares? I think it's just euclidean...

What happens when bins are summed? Does gain change or do the bins just become averaged? In that sense, how important is the gain stage? What are those gains?

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

3.4 Analysis Directions

0 - F0mod 1 - HSSE

EE — CC — CS

0 — 0 — 0

0 — 0 — 1

0 — 0 — 0

0 — 0 — 0

Chapter 4

LESS THEORETICAL STUFF

About this chappy

4.1 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!

4.2 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

4.3 Modulation Types

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`.