HARMONIC COHERENT DEMODULATION FOR IMPROVING SOUND CODING IN COCHLEAR IMPLANTS

Xing Li¹, Kaibao Nie², Les Atlas¹ and Jay Rubinstein²³⁴

Departments of ¹Electrical Engineering, ²Otolaryngology and ³Bioengineering
⁴Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, USA

ABSTRACT

The paper presents a promising application of the coherent demodulation technique to improving hearing with cochlear implants. It has been a challenge to encode temporal fine structure (TFS) in cochlear implants for better speech and music perception. We propose a pitch-synchronized coherent demodulation strategy—Harmonic Single Sideband Encoder (HSSE)—to efficiently encode TFS cues. A dynamic carrier estimation approach based on harmonic detection was chosen to enhance temporal pitch coding in cochlear implants. Acoustic analysis results showed that the HSSE strategy preserves temporal fine structure cues including fundamental frequency (F0) information as well. The coding of temporal fine structure cues in cochlear implants could be potentially improved with the F0-synchronized coherent demodulation.

Index Terms—coherent demodulation, cochlear implants, speech analysis, modulation, single-sideband demodulation

1. BACKGROUND

In the late 1930’s, Homer Dudley proposed that speech and other audio signals, such as music, are actually low bandwidth envelope processes that modulate higher bandwidth carriers [1]. A pioneering approach to signal decomposition of this form was implicit in the modulation filtering studies of Drullman et al. [2]. In their method, the speech signal was passed through a filterbank. For each filterbank band, the amplitude envelope was computed as the absolute value of the analytic signal of that subband (the “Hilbert envelope”). This envelope was filtered, and a new subband signal was constructed by multiplying the original subband signal’s temporal fine structure (TFS) carrier.

Contrary to common expectation, the rectified or Hilbert envelope of a speech signal is not a low-pass replica of the subband’s magnitude spectrum, but is typically much broader in frequency than the subband itself. The envelope, as estimated incoherently by the magnitude operation, is much broader in frequency than the original signal. Consequently, the TFS or carrier is much broader in frequency than the original signal. These excess distortion frequencies will usually fall outside a filterbank subband.

As we have shown in previous work, it is possible to avoid this envelope distortion by using a coherent approach to find modulation envelopes [3,4]. These coherent approaches then mathematically require that the envelope (and associated carrier) be complex. But in this paper, we describe a cochlear implant (CI) application which allows a coherent approach which can retain a real envelope and carrier. Thus, simpler concepts and processing are possible.

Recent advances in cochlear implants (CIs) have made it successful to treat profound hearing loss with electrical current stimulation. A majority of CI users can understand speech to a high degree in quiet, however their performance deteriorates rapidly in the presence of background noise or competing talkers. Also, meaningful music perception is still a challenge for CI users. One major cause is the limitation of current speech processing strategies in delivering temporal fine structure (TFS) cues [5-10]. TFS is traditionally defined as the unimodular part of an analytic signal. In the widely used continuous interleaved sampling (CIS) or similar CI strategy [6], the slowly-varying magnitude envelope is typically extracted from each subband and then encoded as channel stimuli, whereas the carrier-like TFS is discarded due to lack of appropriate CI coding schemes.

The encoding of TFS in CIs is ultimately restricted by the perceivable range of temporal pitch in electric hearing. Zeng [11] and other studies have shown that CI users can not discern differences in temporal pitch more than 300-500 Hz. Also, speech perception studies with hearing impaired people indicate a greatly reduced ability to use TFS [7]. However, the temporal fine structure is typically a rapidly varying signal with frequency content extending up to 20 kHz, making it impossible to directly code TFS in cochlear implants without any signal transformation.

Nie et al. [8] proposed to encode TFS in the form of frequency modulation. Sit et al. [9] suggested an
asynchronous interleaved sampling strategy, in which the phase information is encoded by a race-to-spike algorithm. Rubinstein et al. [12] also proposed to use high-rate conditioning pulse trains to enhance the coding of temporal information by stochastic resonance. In most of the existing or these newly proposed strategies, the envelope detection is still implemented by the traditional incoherent demodulation technique such as the Hilbert transform or half-wave rectification/filtering. But the phase or TFS information is independently extracted and encoded in the pulse train stimuli.

Incoherent demodulation, such as inherent in the common CIS Hilbert envelope approach, can transform a signal into a non-negative, real envelope within the perceivable frequency range of electric hearing, yet it is not capable of encoding the fast-changing temporal fine structure. In contrast, coherent demodulation spectrally shifts the entire signal down to a base band with regard to a single carrier signal. TFS information can be accurately represented in the coherently demodulated envelope, potentially providing a better solution for TFS encoding in cochlear implants.

Using the coherent demodulation technique, Nie, Atlas and Rubinstein [10] proposed the single sideband encoder (SSE) in which the phase and envelope information can both be inherently preserved in the demodulated signal. In the SSE strategy, the carrier frequency is fixed at the lower corner of each band-passed signal, analogous to single sideband demodulation. Although the proposed SSE strategy converts the fast-varying TFS into perceivable cues for CI users, it introduces asynchronous pitch cues when multiple stimulation channels are enabled [13]. The amount of frequency shift in the SSE strategy is independently controlled from channel to channel. Fixing the carrier frequency at the lower edge could result in the loss of harmonic structure after demodulation for voiced sounds or melodies.

In the normal auditory system, synchronized neural firing (phase locking) is an important aspect of neural coding [14]. To mimic the neural firing pattern observed in neurophysiologic experiments and also to avoid inconsistent pitch rate across channels in SSE, this paper proposes a pitch-synchronized coherent demodulation technique. Similar to the SSE strategy, the new method can preserve both the envelope and the phase information in the demodulated signals. Meanwhile, the fundamental periodicity information (F0) can also be enhanced, which is a critical cue for music perception, tonal language perception, and speaker identification.

The organization of the paper is as follows. In Section 2, the proposed pitch synchronized coherent demodulation technique is described. To demonstrate the efficacy of the proposed method, Section 3 presents the demodulation results from a speech sample. Discussion about the results and a summary are given in Section 4.

2. PITCH SYNCHRONIZED COHERENT DEMODULATION

For coherent demodulation, Schimmel and Atlas [3] proposed a time-varying carrier estimate based on the center of spectral gravity. In this study, we propose a time-varying carrier estimate based on pitch detection. A fundamental frequency tracker [15] was used to estimate the timing-varying fundamental frequency (pitch) \( F_0(t) \). For aperiodic parts of a signal, interpolation of \( F_0(t) \) was performed to generate a smoothed function. We separately estimated this time-varying pitch.

The proposed pitch synchronized coherent envelope detection scheme is shown in Fig. 1.

![Block diagram of the proposed pitch synchronized coherent demodulation for cochlear implants.](image)

For each band-pass filtered signal \( x(t) \), mathematically it can be modeled as:

\[
x(t) = E(t)\cos(2\pi f_c t + \phi(t))
\]

where \( E(t) \) is the envelope, \( f_c \) is the center frequency, and \( \phi(t) \) is the phase information. Since the Hilbert transform serves to exchange sine and cosine components, the Hilbert transform of \( x(t) \) can be obtained as:

\[
x_H(t) = E(t)\sin(2\pi f_c t + \phi(t))
\]

Consequently, the analytic signal \( \hat{x}(t) \) can be derived as:

\[
\hat{x}(t) = x(t) + jx_H(t) = E(t)\exp[j(2\pi f_c t + \phi(t))]
\]

To demodulate the subband signal incoherently, we can take the magnitude of the analytic signal and derive the Hilbert envelope. To demodulate the subband signal coherently, we need to estimate the carrier frequency. Then coherent demodulation can be performed by multiplying the analytic signal with the conjugate of a complex carrier function.

Assume there is only one harmonic component contained in the subband. If we denote the index of the harmonic as \( m \), then the harmonic carrier frequency will be equal to \( mF_0(t) \).

Define the complex carrier function as:
Essentially, the carrier phase is computed as the integration of the instantaneous carrier frequency over time. Physically, this means the frequency shift amount in the demodulation process is always an integer multiple of the instantaneous fundamental frequency. Note that the carrier frequency is set to \((m-1)F_0(t)\) in order to demodulate each subband signal to the location of the fundamental frequency \(F_0(t)\).

Multiply \(\hat{x}(t)\) with the conjugate of the defined carrier function:

\[
\hat{y}(t) = \hat{x}(t) \cdot e^{-j\theta(t)} = E(t) \exp\{j(2\pi f_0 t + \phi(t)) - 2\pi \int_0^t (m-1)F_0(\tau)d\tau\} \tag{5}
\]

Then, taking the real part of this analytic signal, we obtain the demodulated signal:

\[
y(t) = E(t) \cos(2\pi f_0 t + \phi(t) - 2\pi \int_0^t (m-1)F_0(\tau)d\tau) \tag{6}
\]

which carries both the original envelope \(E(t)\) and the phase information \(\phi(t)\) in a slowly-varying manner. Also, the periodicity information is preserved after \(F_0\)-synchronized demodulation. Fig 2 illustrates the demodulation process in frequency domain. Here we assume \(F_0(t)\) is constant at a given time. Note that the demodulation is actually implemented by one-sided frequency shift (Fig. 1).

![Figure 2 Illustration of the pitch synchronized coherent demodulation in the frequency domain. Vertical lines indicate the harmonics contained in each subband.](image)

In practice, the harmonic index \(m\) in a given subband is typically time-varying. We can estimate the instantaneous harmonic index as the integer function:

\[
m(t) = \left\lfloor \frac{f_L}{F_0(t)} \right\rfloor + 1 \tag{7}
\]

where \(f_L\) is the lower cutoff frequency of the bandpass filter. The lowest harmonic that can be contained in the bandpass signal is determined from (7). Consequently, the complex carrier function will be determined as follows:

\[
e^{i\theta(t)} = \exp\{j2\pi \int_0^t [f_L / F_0(\tau)]F_0(\tau)d\tau\} \tag{8}
\]

The physical explanation for (8) is the same as that for (4). Instead we estimate the harmonic index dynamically.

If there are multiple harmonics contained in one bandpass signal (e.g., the third subband displayed in Fig. 2), the demodulated signal will contain higher harmonics in addition to the fundamental frequency. The new coding strategy based on harmonic demodulation is termed harmonic single sideband encoding (HSSE).

### 3. ACOUSTIC ANALYSIS

To verify the proposed HSSE strategy, we processed some speech samples according to the demodulation procedure described above. A sound signal was filtered into 20 bands according to a frequency allocation map in the Nucleus cochlear implants by Cochlear Ltd., Australia. The cutoff frequencies of each bandpass filter presented here are 188, 313, 438, 563, 688, 813, 938, and 1063 Hz which are related to the 7 lowest frequency apical electrodes.

![Figure 3.](image)

The speech sample is a female voice saying the word ‘hard’. The bandpass signals from the 2nd subband (313 to 438 Hz), 4th subband (563 to 688 Hz), and 6th subband (813 to 938 Hz) are plotted separately on the top panels of Fig. 3. For the selected sound segment, the estimated pitch \((F_0(t))\) is changing from 215 to 220 Hz. The demodulated signals from the Hilbert transform, SSE and HSSE are plotted respectively in Fig. 3. The Hilbert envelops only tracks the smoothed envelope of each band-passed signal. In contrast, the demodulated signals from both SSE and HSSE carry not only the envelope cues but also the temporal fine structure cues (low to high frequency change in the 4th subband) in the original signals. Comparing SSE with HSSE, we found that the demodulated signals from SSE vibrate independently at different frequencies, whereas the demodulated signals from HSSE are completely synchronized to the same fundamental frequency \(F_0(t)\). To generate pulse stimuli, these demodulated signals need to be half-wave rectified and then modulated with a relatively high-rate biphasic pulse train. TFS can be then coded in the gaps between pulse bursts.

### 4. DISCUSSION AND SUMMARY

The proposed harmonic coherent demodulation technique provides an innovative but promising way of encoding temporal fine structure in cochlear implants. Our acoustic analysis results demonstrate that the HSSE strategy utilizing the coherent demodulation technique is effective in preserving TFS in a highly synchronous manner across channels. The demodulated signals from different subbands oscillate at the same fundamental frequency; therefore, a salient temporal pitch cue might be perceivable to cochlear
implant users. Our preliminary results from 2 Nucleus implant subjects showed that our proposed coherent HSSE strategy produced significantly better melody and musical instrument recognition with a mean improvement of 15.4 ± 5.9% (standard deviation) over the subjects conventionally and daily used incoherent strategy. This paper provides key new insights about the advantageous acoustic signal processing and representation potential of our coherent strategy. Details on experimental methods for these subject test experiments were presented in [16]. Further investigations are being conducted with more cochlear implant patients to evaluate the performance of the proposed coding strategy in speech perception.

This study was supported by the AFOSR Grant FA9550061019, NIH R01 DC007525 and P30 DC004661.

5. REFERENCES


Figure 3 Bandpass signals from 2nd, 4th, 6th subbands and their demodulated signals from the Hilbert detection, SSE and proposed HSSE.