

A Low-Power Asynchronous Interleaved Sampling Algorithm for Cochlear Implants That Encodes Envelope and Phase Information

Ji-Jon Sit, Andrea M. Simonson, Andrew J. Oxenham, Michael A. Faltys, and Rahul Sarpeshkar*

Abstract—Cochlear implants currently fail to convey phase information, which is important for perceiving music, tonal languages, and for hearing in noisy environments. We propose a bio-inspired asynchronous interleaved sampling (AIS) algorithm that encodes both envelope and phase information, in a manner that may be suitable for delivery to cochlear implant users. Like standard continuous interleaved sampling (CIS) strategies, AIS naturally meets the interleaved-firing requirement, which is to stimulate only one electrode at a time, minimizing electrode interactions. The majority of interspike intervals are distributed over 1–4 ms, thus staying within the absolute refractory limit of neurons, and form a more natural, pseudostochastic pattern of firing due to complex channel interactions. Stronger channels are selected to fire more often but the strategy ensures that weaker channels are selected to fire in proportion to their signal strength as well. The resulting stimulation rates are considerably lower than those of most modern implants, saving power yet delivering higher potential performance. Correlations with original sounds were found to be significantly higher in AIS reconstructions than in signal reconstructions using only envelope information. Two perceptual tests on normal-hearing listeners verified that the reconstructed signals enabled better melody and speech recognition in noise than those processed using tone-excited envelope-vocoder simulations of cochlear implant processing. Thus, our strategy could potentially save power and improve hearing performance in cochlear implant users.

Index Terms—Asynchronous stimulation, cochlear implant, neural stimulation, phase information.

I. INTRODUCTION

COCHLEAR implants (CIs) have been implanted in more than 100,000 people worldwide, and use a surgically implanted array of electrodes to stimulate the auditory nerve. While CIs can provide a high degree of speech intelligibility in quiet, their performance rapidly deteriorates in the presence of competing speakers and noise [1], [2]. They are also not very successful at conveying pitch information for music perception [3], [4]. Because the sound fidelity perceived by CI

users is not as high as normal hearing listeners, prelingually deafened children have been found to experience difficulty in producing tones [5]. Much research has therefore been directed at improving their performance, and major successes have been achieved since they were first implemented in the 1970s. For example, the technique of continuous interleaved stimulation (CIS) has mitigated the effects of electrode interaction and current spreading that once limited the spectral resolution to a great degree [6]. The number of implanted electrodes has grown from 1 to as many as 24 in current implants [7]. However, there remains much to improve.

In this paper, we identify a few problems with the current stimulation paradigm that might be responsible for the reduced ability of CI users to perceive music, and to perform well in noisy conditions. The problem may lie in the failure of current implants to deliver phase information, or fine time structure [8], which is important especially in conditions of low spectral resolution, such as those experienced by CI users [9]. An average CI user appears to receive at most 10 usable channels of spectral information, regardless of the number of implanted electrodes [10], [11]. Compared to normal-hearing listeners, who potentially have as many as 35 000 independent auditory nerve fibers that can convey spectral information, this is a severe limitation. However, there are data to suggest that the effects of limited spectral resolution can be mitigated by adding fine-structure or timing cues to the information conveyed in neural stimulation [1], [9], [12]. Consequently, at least two published strategies have attempted to incorporate this additional information into new methods of neural stimulation for cochlear implants, by stimulating at the peaks [4] or zero-crossings [13] of CI filter-band outputs. Unfortunately, these strategies have yet to show significant improvement in hearing performance for CI users.

An important caveat to note in this endeavor is that several studies have shown CI users are unable to detect increases in the rate of electrical pulse trains above 300 pps [14], [15]. This could be a constraint that limits the usefulness of frequency-modulated information (i.e., phase information) in cochlear implants. Furthermore, current CI users may experience a perceptual dissonance due to the lack of consistency of phase cues between neighboring channels that would be present in a normal traveling-wave architecture [16]. Nevertheless, the ability of a CI user to discriminate changes in pulse rate varies between individuals [17], and some CI users may yet benefit from rate-coding of pitch and tonal information.

Observing that synchronous stimulation inherently encodes envelope information only, and neglects the encoding of phase

Manuscript received November 15, 2005; revised June 24, 2006. *Asterisk indicates corresponding author.*

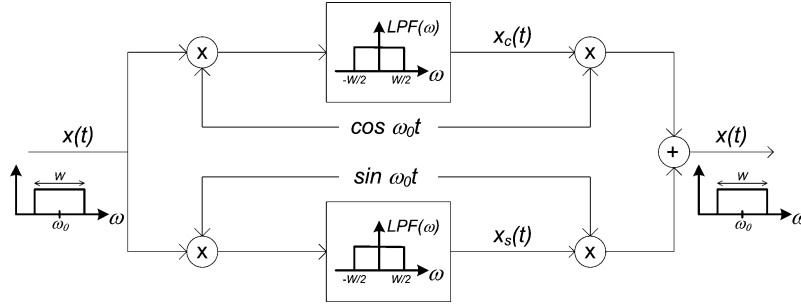
J.-J. Sit is with the Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA.

A. M. Simonson and A. J. Oxenham were with the Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA. They are now the University of Minnesota, Minneapolis, MN 55455 USA.

M. A. Faltys is with the Advanced Bionics Corporation, Sylmar, CA 91342 USA.

*R. Sarpeshkar is with the Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: rahuls@mit.edu).

Digital Object Identifier 10.1109/TBME.2006.883819

Fig. 1. Quadrature decomposition of a bandpass signal $x(t)$.

information, we propose an asynchronous interleaved sampling (AIS) strategy for neural stimulation that may deliver phase information while maintaining minimal electrode interaction in the same manner as CIS, namely, by avoiding the stimulation of more than one electrode at a time.

This paper is organized as follows: Section II gives the motivation for incorporating phase information into an asynchronous stimulation paradigm. Section III describes the AIS strategy in detail. Section IV demonstrates the strategy using MATLAB simulations. Section V presents a spike reconstruction technique that allows us to compare an AIS reconstruction against other CI reconstructions. Section VI presents two psychoacoustic tests performed on normal-hearing listeners as an initial evaluation of the strategy's potential merits. Section VII concludes with a discussion and summary of the AIS strategy.

II. THE IMPORTANCE OF PHASE INFORMATION AND ASYNCHRONOUS STIMULATION

A. What is Phase Information?

It has been shown that any bandpass signal $x(t)$ with center frequency ω_0 and bandwidth W can be completely analyzed into quadrature components $x_c(t) \cdot \cos \omega_0 t + x_s(t) \cdot \sin \omega_0 t$ [18] and hence also into envelope and phase components $E(t)$ and $\phi(t)$ respectively, where $x(t) = E(t) \cdot \cos[\omega_0 t + \phi(t)]$ as shown in Fig. 1. As W can be as large as $2\omega_0$, any band-limited signal can be expressed in the general form of $E(t) \cdot \cos \theta(t)$ where the angle component $\theta(t) = \omega_0 t + \phi(t)$ when ω_0 is specified. The angle component $\theta(t)$ or, equivalently (when ω_0 is defined), the phase component $\phi(t)$ is what we will refer to as the *phase information* in a signal. In the psychoacoustic literature, phase information is commonly referred to as “fine time structure” or “FM (frequency modulation) information,” as it can be thought of as how rapidly the instantaneous frequency is being modulated, as opposed to “AM (amplitude modulation) information” in $E(t)$.

Defining $x_H(t)$ as the Hilbert transform (Fig. 2) of $x(t)$

$$\begin{aligned} x(t) &= x_c(t) \cos \omega_0 t + x_s(t) \sin \omega_0 t \\ x_H(t) &\triangleq x(t) * \frac{1}{\pi t} = \mathcal{F}^{-1} [X(f) \cdot \text{HILBERT}(f)] \\ &= x_c(t) \sin \omega_0 t - x_s(t) \cos \omega_0 t \end{aligned} \quad (1)$$

(since the Hilbert transform serves to exchange sine and cosine components)

$$\begin{aligned} x(t) &= x_c(t) \cos \omega_0 t + x_s(t) \sin \omega_0 t \\ &= E(t) \cos[\omega_0 t + \phi(t)] \\ E(t) &= \sqrt{x_c^2(t) + x_s^2(t)} \\ \phi(t) &= \arctan\left(\frac{-x_s(t)}{x_c(t)}\right) \end{aligned}$$

$$\begin{aligned} \text{Hilbert transform} &= \frac{1}{\pi t} \xleftrightarrow{\mathcal{F}} \begin{array}{c} j \\ \hline f \\ \hline -j \end{array} \\ \text{impulse response} & \end{aligned}$$

Fig. 2. The Hilbert transform in time and frequency domain representation.

If we then define the analytic signal $s(t) = x(t) + j \cdot x_H(t)$

$$\begin{aligned} s(t) &\triangleq x(t) + jx_H(t) \\ &= x_c(t) [\cos \omega_0 t + j \sin \omega_0 t] - jx_s(t) [\cos \omega_0 t + j \sin \omega_0 t] \\ &= [x_c(t) - jx_s(t)] e^{j\omega_0 t}. \end{aligned} \quad (2)$$

We should note $E(t)$ and $\theta(t)$ can then be obtained *identically* from $s(t)$ as $|s(t)|$ and $\angle s(t)$, respectively

$$\begin{aligned} |s(t)| &= |x_c(t) - jx_s(t)| \cdot |e^{j\omega_0 t}| \\ &= \sqrt{x_c^2(t) + x_s^2(t)} = E(t) \\ \angle s(t) &= \angle [x_c(t) - jx_s(t)] + \angle e^{j\omega_0 t} \\ &= \arctan\left(\frac{-x_s(t)}{x_c(t)}\right) + \omega_0 t = \phi(t) + \omega_0 t = \theta(t). \end{aligned} \quad (3)$$

When analyzing a signal, it is useful to use the Hilbert transform to look at the $E(t)$ and $\theta(t)$ components separately, especially when comparing two signals, as they may be similar in one component but not the other.

B. How Important is Phase Information for CI Users?

Experiments by Smith *et al.* have shown that the melody in music will be carried primarily in the fine structure, or phase component $\phi(t)$, when the number of bandpass analysis channels is smaller than 32 [19]. The current generation of cochlear implants is limited to 16–24 analysis channels and delivers only $E(t)$ at a fixed rate, discarding $\phi(t)$ altogether [6]. Thus it is no surprise that CI users have trouble perceiving tones, much less the melody in a piece of music. In an attempt to restore tonal perception, which is important in many Asian languages like Mandarin Chinese and Cantonese, researchers have attempted to modulate the bandpass center frequencies ω_0 by the fundamental frequency F_0 in each channel, with significant success in normal-hearing listeners [20]. Lan *et al.*'s results can be interpreted as modulating $\phi(t)$ in each bandpass channel, to better approximate the true phase of the signal.

Furthermore, recent work by Nie *et al.* using acoustic simulations on normal-hearing listeners shows that speech recognition

in noise is significantly improved by the addition of FM information, which is essentially what we refer to as $\phi(t)$, band-limited to 400 Hz. They hypothesize that additional FM cues allow better segregation of a target sound into a perceptual stream that is more distinct from competing noise [1]. If CI users are able to utilize FM cues similar to normal-hearing listeners, they should then improve in speech and music perception when provided with phase information in a noisy environment.

Overall, these studies and others [21], [22] suggest that the introduction of phase information, if delivered in ways that were usable, could provide substantial benefit for CI users.

C. How Can Phase Information be Delivered?

While there are many reasons to believe that phase information is important for higher fidelity listening, it is not at all clear how phase adjustments or frequency modulation should be conveyed to a CI user. The variation of $\phi(t)$ is usually very wide and rapid, which means it cannot be delivered directly to neurons, which have an absolute refractory period and hence a bandwidth usually no larger than 1 kHz. With regard to CI users in particular, perceptual data show that they cannot detect FM rate changes at base rates of more than a few hundred hertz [23], [24]. Hence any scheme proposing to convey phase information to CI users must practically impose both a rate limitation and a bandwidth limitation on the delivery of $\phi(t)$ [1].

One method is to utilize a zero-crossing detector: we know $\theta(t)$ is exactly 0 or π at the zero-crossings, and so delivering this information (e.g., by stimulating at the zero-crossing times) should provide the brain with fairly rich knowledge of what the phase should be, even with the band-limiting requirement. However, to stimulate at zero-crossing times, we need some quick and intelligent way of choosing a zero-crossing in one channel over another, because zero-crossings may arrive simultaneously on different channels, and also very rapidly in the high-frequency channels.

Even if zero-crossing stimulation is not used, channel selection is still required to enforce firing in only one channel at a time, which we will term the “interleaved stimulation requirement” as taken from CIS, to avoid simultaneous interactions between electrodes. The channel selection process must also prevent one or two strong (intense) channels from dominating the firing and not giving other channels a chance to fire.

However, zero-crossing stimulation may not be the ideal method of delivering phase information. In the presence of noise, the reliability of zero-crossing times as an indicator of $\phi(t)$ rapidly deteriorates. Furthermore, stimulation at a deterministically precise phase in the signal is very different from what actually happens in biology, as nerves tend to fire at times which are smoothly distributed over all phases, but concentrated near the peak in signal energy [22].

To introduce FM information into the stimulation paradigm, Nie *et al.* have proposed that either a carrier be frequency modulated with $\phi(t)$, or that the carrier itself be replaced by the band-limited $\phi(t)$ signal [1]. In this paper, and in the spirit of Nie *et al.*'s latter idea, our objective is to convey $\phi(t)$ but depart entirely from the concept of either an AM or FM carrier, for the reasons that we outline below.

D. The Problems With Synchronous Stimulation

While multi-channel sound coding strategies such as CIS and the Advanced Combinational Encoder (ACE) [25] achieve good performance among CI users to date [26], many problems have been identified with the current paradigm of synchronous (i.e., fixed-rate) pulsatile stimulation. In particular, there is unnecessary synchronization of the neural response to the fixed-rate electrical carrier, despite the carrier containing no information. Physiological studies have shown that many neural artifacts are produced, which are never seen under normal acoustic stimulation. To name a few, at low-rate stimulation, there is deterministic entrainment to the carrier, while at high-rate stimulation, there are severe distortions in the temporal discharge patterns caused by neural refractoriness [22], [27]. Finally, synchronous stimulation inherently excludes the possibility of delivering phase information to CI users. These problems can be avoided if fixed-rate stimulation is replaced by some form of asynchronous stimulation that also takes the limits of neurobiology into account. For example, the stimulation should refrain from driving the nerves heavily into refractory-rate limited firing, where fiber synchrony becomes pronounced.

E. Power Savings Available Through Asynchronous Stimulation

In order to deliver the temporal information in $\phi(t)$ within a pulsatile stimulation paradigm, the ability to stimulate at precise times is essential, and therefore high temporal resolution is required to take full advantage of asynchronous stimulation. If conventional discrete-time signal processing is used to generate asynchronous stimulation times, a high-rate sampling clock is required to achieve high temporal resolution, which can be costly in power. However, an event-based asynchronous digital or continuous-time analog signal processing scheme can provide high temporal resolution without the need for a fast sampling clock that is constantly running whether there are events or not.

Second, during asynchronous stimulation, the average firing rate (AFR) is able to fall significantly below the worst case maximum firing rate. In contrast, synchronous paradigms need a clock that is fast enough to handle the fastest event rate. Asynchronous stimulation allows the input firing rate to be averaged across time and across channels such that no power is spent during quiet periods or on quiet channels to continuously sample the signal. Sensory codes that adapt to signal statistics are efficient [28] and asynchronous stimulation allows us to adapt our sampling rate in time and spectral space to the needs of the signal.

III. AN AIS “RACE-TO-SPIKE” STRATEGY

If we do not use zero-crossings, how might we generate good asynchronous firing times? In this section, we propose a bio-inspired way to provide stimulation pulses at asynchronous times which still have a definite correlation with $\phi(t)$. However, if we make the firing times a sole function of $\phi(t)$, this decouples the local decision to fire from global knowledge of activity in other channels, which may be important. In addition, we also do not want to decouple the decision to fire from the envelope strength,

as it should be probabilistically more important to provide correct FM information when the AM signal is strong. Therefore, our strategy attempts to incorporate global information across all channels about $E(t)$, and favors the firing of a pulse when $E(t)$ in that channel is larger than others. The resulting times will naturally be asynchronous, and should even appear pseudorandom for natural, arbitrary sounds, given complex interactions between channels during the channel-selection process. Further biological justification for this technique is presented in Section V.

The system is comprised of coupled electronic neurons that incorporate information about $E(t)$ across all channels, and in competition with each other, generate the asynchronous firing times. The strategy is also termed a “race-to-spike” algorithm, as the neurons are set up in a race where the winning channel gets to deliver a stimulation spike on its electrode. The algorithm is described in the following steps:

- 1) The system receives as inputs half-wave rectified currents from a bank of bandpass analysis filters, which could be actual currents like those generated by an analog processor [29], [30], or a digital version as produced by a digital signal processor.
- 2) There is one integrate-and-fire neuron [31] associated with each channel, receiving the above-mentioned current input from that channel, to charge up its neuronal capacitance from the ground state. This begins the “race-to-spike.”
- 3) The first neuron to reach a fixed voltage threshold “wins” and resets all capacitors back to zero. This ensures that the interleaved stimulation requirement is satisfied, as there can be only one winner.
- 4) The winning neuron then fires a current spike (which is an asynchronous timing event) on its electrode that is scaled by the channel envelope’s energy.
- 5) Once a neuron wins, its input current is inhibited (i.e., weakened) for a period determined by a relaxation time constant, to prevent it from winning repeatedly.
- 6) After the winning neuron has fired its spike, we start the neuronal “race-to-spike” (Step 2) again.

Natural variations of this algorithm could be implemented, and we will list just a few. The inputs to the integrate-and-fire neurons could be generated by any signal analysis front-end, for example a modulated filter bank or a cochlear cascade of low-pass filters. The voltage threshold could be different for different channels, to create pre-emphasis or to accommodate different sensitivity to stimulation. The level of the electrode current spike could be an arbitrarily complex function of its input, such as an average past envelope stored dynamically on a capacitor in that channel. The level of inhibition following a spike, and also its duration could be arbitrarily complex functions of the past. For the purposes of this paper, we will implement in simulation only a simple version of this strategy, to be elaborated on below.

We wanted to build on existing work on an ultra-low-power analog bionic ear processor (ABEP) to eventually implement a very low power strategy in hardware [29], [30]. Thus, to examine the effectiveness of this strategy in conjunction with low-power analog signal processing components already built, we simulated as our front-end the bandpass filters and envelope de-

tectors exactly as they were implemented in the ABEP: The bandpass filters were conventional two-stage fourth-order filters with $Q = 4$ and 12 dB/octave high-pass and low-pass rolloffs; 16 channels were implemented, with center frequencies scaled logarithmically from 116 Hz to 5024 Hz. The envelope detectors were implemented with asymmetric attack and release first-order low-pass time constants, with $\tau_A = 1$ ms and $\tau_R = 3$ ms respectively, acting on full-wave rectified filter outputs in order to compute an approximation for $E(t)$. The envelope detector may therefore be thought of as a low-pass filter with cutoff frequencies of 1 kHz and 333 Hz for rising and falling transients respectively. To simulate the finite sampling rate in digitizing $E(t)$, $E(t)$ was passed through a sample and hold running at a rate of 1.8 kHz. Half-wave rectified versions of the bandpass filter outputs were used as inputs to the integrate-and-fire neurons.

Neuronal capacitances are simply modeled as capacitive state variables, and their voltage threshold was set to be the same for all channels, where a low value of around 35 mV (in simulation units) turned out to give good results. Using this threshold and input speech tokens at conversational sound pressure level, all channels reach their threshold and fire within ~ 400 μ s of receiving an input pulse, with a time-to-spike of 50–200 μ s being most common. The size of the spike fired was simply set to the sampled-and-held value of $E(t)$, which could be easily implemented in hardware by D/A converters.

The last important detail is the time course of the inhibition described in Step (5) above. We wanted to ensure that firing was absolutely prohibited for a minimum amount of time that is determined by the absolute refractory period of biological neurons. This prohibition avoids wasting stimulation power when biological neurons are unable to respond. However, after the absolute refractory period, we would like to softly turn off the inhibition current, thus enabling a very strong input to overcome the imposed inhibition.

To accomplish these objectives, we designed the time course of the inhibition current to be modeled by a Fermi-Dirac exponential roll-off, given by the equation $h_{inh}(t) = (1/1 + e^{k(t-\tau_{inh})})$ where τ_{inh} sets the time where the inhibition falls to half its maximum value, fixed at 0.8 ms in our simulations and k sets the steepness of the rolloff, fixed at 10^4 . The value of τ_{inh} was chosen to be 0.8 ms to enforce a minimum interspike interval near the refractory period of auditory neurons [32]. The shape of this time course is shown in Fig. 3. It also happens to match closely with the decrease in current output of a subthreshold current source if the gate voltage on a pass transistor is linearly decreased, which can be easily implemented in electronics. The inhibition current $I_{inh}(t)$ is then defined as $A_{inh} \cdot h_{inh}(t)$, where A_{inh} sets the maximum inhibition current, fixed at -0.5 in our simulation units.

The algorithm is described in the following pseudocode.

```
Initialize: capacitor voltages  $V_{cap}(t) = 0 \forall$  channels.
Initialize: time-of-last-spike  $t_{lastspike} = -\infty \forall$  channels.
Initialize: spiking output  $spike(t) = 0 \forall$  channels.
At each timestep  $t$ ,  $\forall$  channels do the following.
```

```
    Compute:  $I_{inh}(t) = A_{inh} \cdot h_{inh}(t - t_{lastspike})$  (Step 5).
```

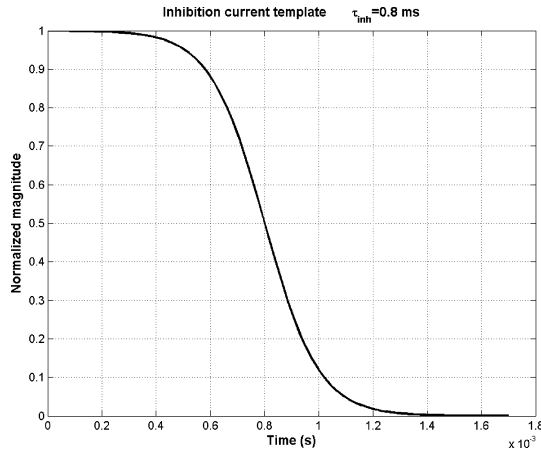


Fig. 3. Time profile of inhibition current, with exponential rolloff modeled by the Fermi-Dirac equation. Note the current falls to half its value when $t = \tau_{inh}$.

Increment: $V_{cap}(t) += \max[HWR(t) - I_{inh}(t), 0]$
 (Steps 1 and 2) [where $HWR(t)$ is the half-wave-rectified bandpass filter output for that channel].

If $\max_{\forall channels} [V_{cap}(t)] > (V_{thresh} = 35 \text{ mV})$

Find: $\max_ch = \text{channel with } \max_{\forall channels} [V_{cap}(t)].$

Set: $V_{cap}(t) = 0 \forall \text{ channels}$ (Step 3).

Set: $spike(t) = E(t)$ in channel \max_ch (Step 4).

Set: $t_{lastspike} = t$ in channel \max_ch (Step 5).

IV. RESULTS FROM MATLAB SIMULATIONS OF AIS

Various sound files in the .wav format were loaded into MATLAB and input to the simulation. We tested speech utterances of the words “die” and “buy,” a Handel chorus containing a word vocalization of “hallelujah,” and a piece of music (jazz) that does not contain words.

Being a digital simulation, we had to choose some level of timing resolution with which to discretize the time steps, thus a “sampling rate” of $F_s = 44100 \text{ Hz}$ was used, indicating a timing resolution of about $23 \mu\text{s}$. Preliminary results (not shown) suggest that a timing resolution much worse than that degrades the accuracy of our experiments. We present results from “die,” which is one of the two speech utterances tested, to illustrate the performance of the system.

A. Capacitor Voltage and Inhibition Current Waveforms

Fig. 4 is a zoomed-in figure of typical capacitor voltage and inhibition current waveforms. We note that spikes are fired when the capacitor voltage reaches a threshold, turning on a negative inhibition current that has the profile as described in Fig. 3. Immediately after a spike, all capacitors are then reset to zero as given by the algorithm. As the level of inhibition current A_{inh} at -0.5 was set to be higher than the highest input level, no channel fires again until its inhibition current has almost returned to baseline.

B. Half-Wave Rectified Inputs Against Spiking Outputs

Fig. 5 shows the half-wave rectified outputs of the bandpass filters, used as inputs to the asynchronous race-to-spike system,

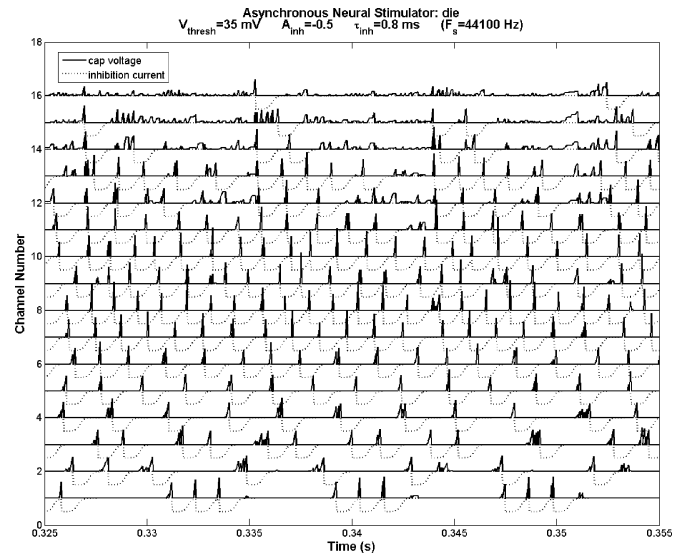


Fig. 4. Capacitor voltage (solid lines) and inhibition current (dotted lines) waveforms from race-to-spike simulation of the speech utterance “die.”

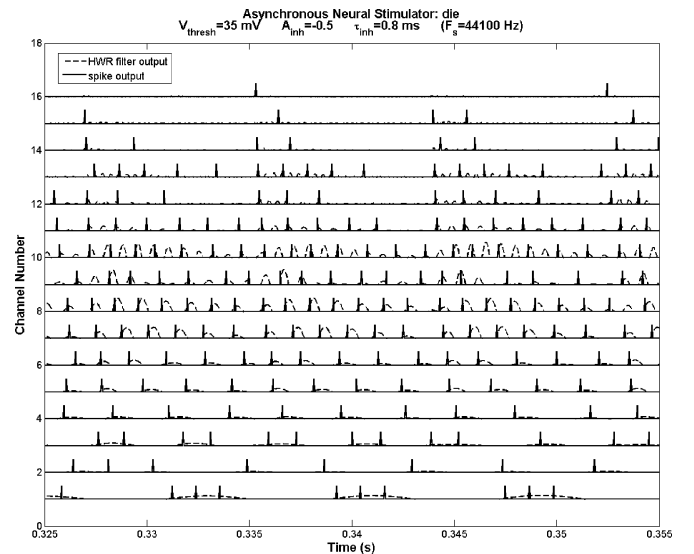


Fig. 5. Half-wave rectified bandpass filter outputs (dashed lines) used as inputs to the race-to-spike simulation, plot against the spiking outputs (solid lines). The spikes in this figure are shown before scaling by $E(t)$, for clarity.

and plotted against its spiking outputs for the same time window as in Fig. 4. We should note that spikes tend to fire near the beginning of each positive excursion in the filter output, but the phase at which a spike is fired is not deterministic. Instead, the time of firing exhibits some pseudostochastic variation due to the competition between channels. We should point out that our strategy does not explicitly model stochastic firing that arises normally in the auditory system, but may, however, introduce stochastic responses that are similar to those encountered normally. By firing only when enough charge has accumulated within a fraction of a half-wave rectified cycle, and when the intensity of a channel is high enough to be the first to spike, a channel will generate spike times that are correlated with but not precisely determined by important features in the phase of the signal $\phi(t)$. The resulting spike trains, as we shall show later,

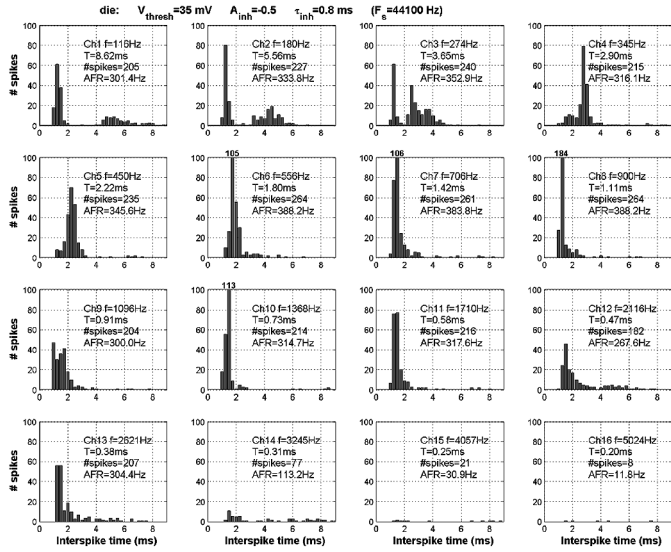


Fig. 6. Interspike-interval histograms for the 16 simulation channels. Also shown for each channel are f , the center frequency of each channel, $T = 1/f$, the number of spikes in that channel, and the AFR. Spike counts > 100 are clipped (but reported numerically above the figure), and spike intervals > 9 ms are not shown in the histogram.

are sufficient for high-fidelity signal reconstruction, and encode phase information with a high degree of correlation.

Of note are also the low-frequency channels, where we note multiple spikes may be fired over a single pulse of energy. This allows pulses of long duration to be well represented, which would not occur in zero-crossing-based stimulation.

C. Interspike-Interval Histograms

The interspike-interval histogram of each channel is presented in Fig. 6. An “absolute refractory period” of about 1 ms is shown to be enforced, keeping the instantaneous firing rate in each channel below a maximum of 1 kHz. The distributions also look fairly natural, where a perfectly natural distribution due to spontaneous firing would be a smooth gamma distribution indicating Poisson arrival of inputs to a neuron. The histograms do not show exact gamma distributions, but the distribution is nevertheless more natural than any distribution produced by synchronous firing.

Finally, we note that the AFR in each channel is about 300–388 Hz for most channels, and when averaged over all 16 channels comes down to only 279 Hz per channel. This AFR is lower than the firing rate in conventional synchronous stimulation where firing rates are not adapted to the input stimuli. Table I shows the power savings possible with our technique by stating the worst-case channel’s AFR and also contrasting it with the mean AFR of all the channels. We see that averaging in time reduces the worst-case AFR below 1 kHz, thus saving power, and that averaging across channels reduces the mean AFR, also saving power.

V. AIS SPIKE RECONSTRUCTION AND COMPARISONS

In this section, we present a method of comparing the AIS strategy against other acoustic simulations of cochlear implants. Many acoustic simulations reconstruct the sound input to a CI

TABLE I
MAXIMUM AND MEAN (ACROSS CHANNELS) OF EACH CHANNEL’S AFR

Sound file	Maximum AFR (Hz)	Channel with max AFR	Mean AFR (Hz)
“die”	388.2	6	279.4
“buy”	466.5	7	288.9
Handel chorus	772.0	9	451.9
Jazz music	556.9	7	275.3

from its output of channel envelopes by using white-noise or tonal carriers for each channel [33], [34], and are known as noise/tone vocoding reconstructions. To similarly reconstruct a sound from the train of spikes generated by AIS, we use a spike-based reconstruction technique that has its foundation in prior neurophysiology work, showing that analog waveforms can be accurately reconstructed from spiking waveforms. For example, optimal low-pass filters can be designed, that when applied to the recorded spike trains from the *Eigenmannia* electric fish, produce as their output a very well-correlated reconstruction of the input voltage variations in the fish’s sensed aquatic environment [35]. Such experiments show that frequency modulations in neuronal spike trains can encode an analog input so well that only conventional low-pass filters are needed for stimulus reconstruction.

Recent work has also shown that a spike-based auditory code can very efficiently encode speech, outperforming gamma-tone, wavelet and Fourier decompositions, achieving the highest signal-to-noise ratio (SNR) in representing speech at coding rates < 60 kbits/s [28]. In this paper, sound signals are also reconstructed from spikes using tuned kernel filters that have impulse responses closely resembling those of high-order resonant low-pass filters. Interestingly, these impulse responses independently matched physiological data; the dictionary of 32 kernel filters that were adapted to give good reconstructions turned out to match the tonotopic population of auditory neuron impulse responses in cats, as derived from reverse correlation analysis [36]. These results lend further support to the validity and usefulness of impulse response reconstructions for neuronal spiking codes.

A. Channel-by-Channel Spike Reconstruction

Motivated by prior work on spike-based reconstruction techniques like that in the electric fish, we applied two-stage fourth-order low-pass resonant filters (with the same center frequency and Q of 4 as the bandpass filters in each channel and with 24 dB/octave low-pass rolloffs) to the spiking outputs from each channel. The spikes effectively behave like envelope-scaled impulses and the filters sum the impulse responses from various spikes to recreate the analog information in each channel

$$\text{reconstruction}_{\text{ch}}(t) = \frac{\int_0^\infty \text{spike}_{\text{ch}}(\tau) \cdot h_{\text{ch}}(t - \tau) d\tau}{\max_{0 \leq t < \infty} h_{\text{ch}}(t)} \quad (4)$$

where $\text{spike}_{\text{ch}}(t)$ contains the $E(t)$ -scaled spikes from channel ch , $h_{\text{ch}}(t)$ is the low-pass filter impulse response for channel ch , and $\text{reconstruction}_{\text{ch}}(t)$ is the reconstructed signal on channel ch .

As the peak in the impulse response $h_{\text{ch}}(t)$ for each channel increases linearly with the center frequency, we needed to

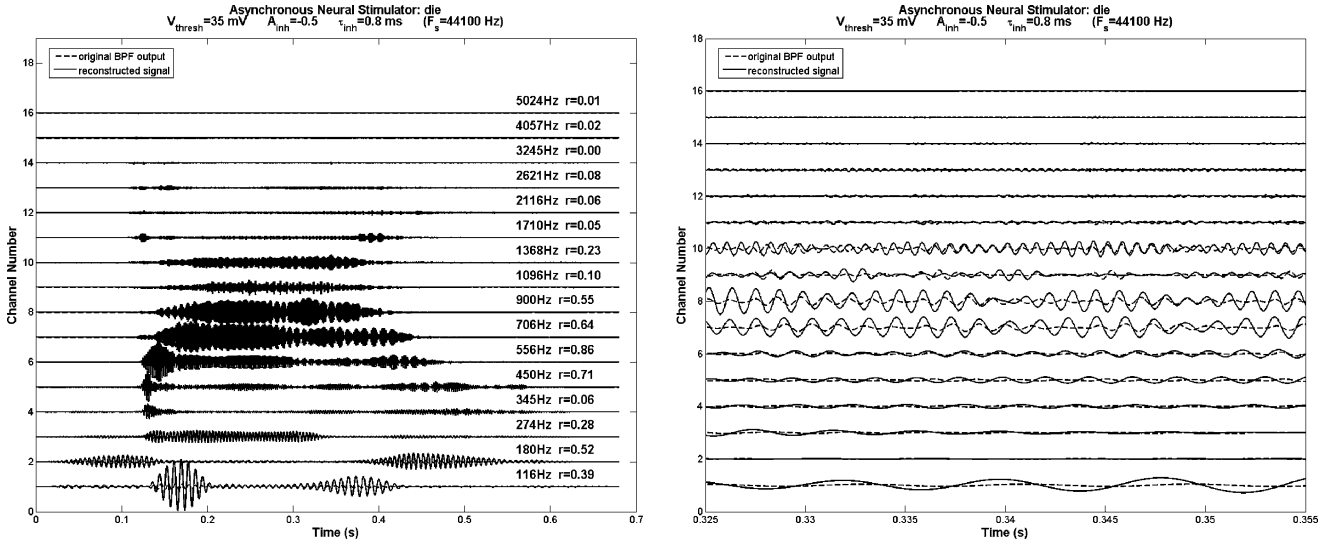


Fig. 7. Sixteen-channel spiking reconstruction (solid lines) of the bandpass filter outputs (dashed lines). A magnified version of the same time window as in earlier figures is shown on the right.

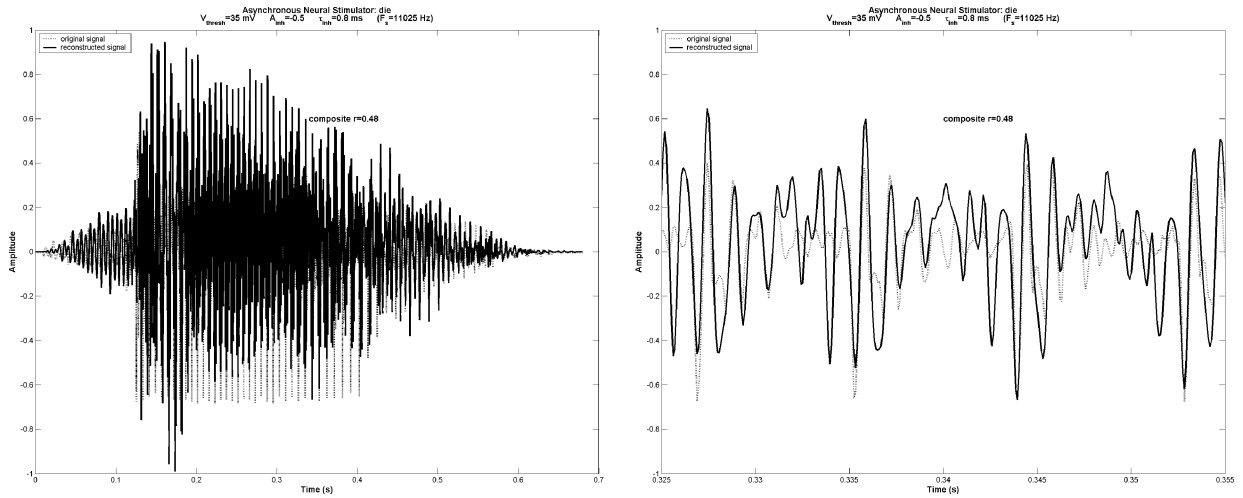


Fig. 8. Composite spiking AIS reconstruction (solid line) of the waveform "die," by summing the channel reconstructions together. Note the reconstruction is downsampled to match the sampling rate of the original sound signal (dotted line). The composite correlation coefficient is 0.48.

normalize the reconstructions across channels by dividing by the peak value of $h_{\text{ch}}(t)$. The reconstructed waveforms on a channel-by-channel basis are shown in Fig. 7, with the correlation coefficients (r) for each channel, computed from the following equation:

$$r = \frac{\int_0^{\infty} \text{BPFoutput}_{\text{ch}}(t) \cdot \text{reconstruction}_{\text{ch}}(t) dt}{\sqrt{\int_0^{\infty} \text{BPFoutput}_{\text{ch}}^2(t) dt \cdot \int_0^{\infty} \text{reconstruction}_{\text{ch}}^2(t) dt}} \quad (5)$$

where $\text{BPFoutput}_{\text{ch}}(t)$ is the bandpass output for channel ch , and both $\text{BPFoutput}_{\text{ch}}(t)$ and $\text{reconstruction}_{\text{ch}}(t)$ have had their means removed, i.e., are zero-mean signals.

The correlation coefficient is on a scale from 0 to 1, with 1 indicating a perfect correlation between the bandpass filtered output and the spike reconstruction. The correlations between each channel are fairly good, with a low-frequency channel

having a correlation coefficient as high as 0.86. In performing these correlations, it was also important to account for group delay introduced by the bandpass and low-pass filters, which causes the composite reconstruction(t) (as defined in the next section) to lag the original signal. Thus, to compensate for the group delay (which does not affect the sound fidelity), the cross-correlation between reconstruction(t) and the original signal was performed, and the lag corresponding to the peak in the cross-correlation was then used to time-shift reconstruction(t) and align it with the original signal. The lags for the 4 sounds ranged from 4.8 – 9.5 ms.

B. Composite Signal Spike Reconstruction

If we sum all the reconstructed channels together, we can generate a composite reconstruction of the original signal just as in

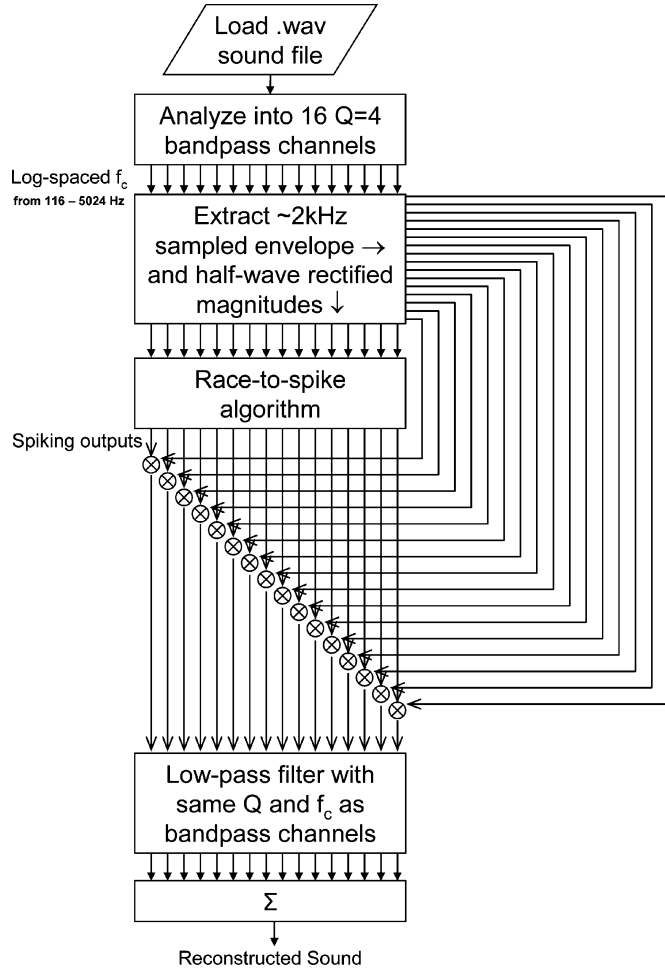


Fig. 9. Flowchart of the entire AIS reconstruction process.

CI acoustic simulations using vocoding reconstruction, defined as follows:

$$\text{reconstruction}(t) = \sum_{\text{ch}=1}^{16} \text{reconstruction}_{\text{ch}}(t). \quad (6)$$

The summated signal is shown in Fig. 8, with a zoomed-in version again on the right. The phase relationships are clearly preserved, and the envelope is also fairly well tracked. The composite correlation coefficient is 0.48 in this example, the word *die*, and is calculated as follows:

$$r = \frac{\int_0^{\infty} \text{original}(t) \cdot \text{reconstruction}(t) dt}{\sqrt{\int_0^{\infty} \text{original}^2(t) dt \cdot \int_0^{\infty} \text{reconstruction}^2(t) dt}} \quad (7)$$

where $\text{original}(t)$ is the original sound signal, and both $\text{original}(t)$ and $\text{reconstruction}(t)$ have had their means removed, i.e., are zero-mean signals

Note that the sampling rate was reduced to match the same sampling rate of the original signal, to make a fair comparison. A flowchart of our entire reconstruction technique is shown in Fig. 9.

TABLE II
CORRELATION IN $E(t)$ ENVELOPE COMPONENT
FOR DIFFERENT PROCESSING METHODS

Sound file	AIS spike-based reconstruction	CIS noise-based vocoding (100 trials)	CIS tone-based vocoding	CIS spike-based reconstruction (1.4 kHz firing rate)
“die”	0.60	0.50 ± 0.04	0.57	0.48
“buy”	0.60	0.46 ± 0.08	0.46	0.37
Handel chorus	0.54	0.27 ± 0.06	0.33	0.27
Jazz music	0.49	0.18 ± 0.02	0.29	0.52

TABLE III
CORRELATION IN $\theta(t)$ PHASE COMPONENT
FOR DIFFERENT PROCESSING METHODS

Sound file	AIS spike-based reconstruction	CIS noise-based vocoding (100 trials)	CIS tone-based vocoding	CIS spike-based reconstruction (1.4 kHz firing rate)
“die”	0.19	0.03 ± 0.02	0.07	0.16
“buy”	0.13	0.03 ± 0.02	0.01	0.07
Handel chorus	0.42	0.03 ± 0.01	0.02	0.01
Jazz music	0.40	0.01 ± 0.01	0.01	0.01

C. Hilbert Decomposition and Correlation

It is not immediately obvious how much of the phase information $\phi(t)$ is retained in the spike output from the AIS strategy. In order to better quantify the transmission of $\phi(t)$, we performed a Hilbert decomposition of the reconstructed signal into $E(t)$ and $\theta(t)$ components as described in Section II-A. These components were then correlated separately with the original signal’s envelope and phase. In order to see whether the correlations for $E(t)$ and $\theta(t)$ were significant for AIS, we then compared them against CIS noise and tone vocoding reconstructions, and also a CIS spike-based reconstruction which employs the same reconstruction filters as described in (4), except that the nonoverlapping spike input is now sequentially and synchronously firing at a fixed rate of 1.4 kHz. Other firing rates were tested as well but found not to make a significant difference to the results. The sound samples used were the words *die* and *buy*, to be representative of speech, and snippets from Handel’s *Hallelujah chorus* and a jazz piece, *Highway Blues*, to be representative of music. As the results for CIS noise vocoding could vary significantly between trials due to randomness in the noise input, we conducted 100 trials and present the mean and standard deviation of those trials. Envelope correlations are shown in Table II, phase correlations are shown in Table III, and composite correlations are shown in Table IV.

Correlation coefficients for $\theta(t)$ are in general lower than that for $E(t)$ because $\theta(t)$ by nature varies much more rapidly than $E(t)$. However, $\theta(t)$ correlation coefficients are clearly higher for AIS than other reconstruction techniques, especially for the music pieces. As the $E(t)$ correlation coefficients are not that much different between reconstruction strategies, the improved

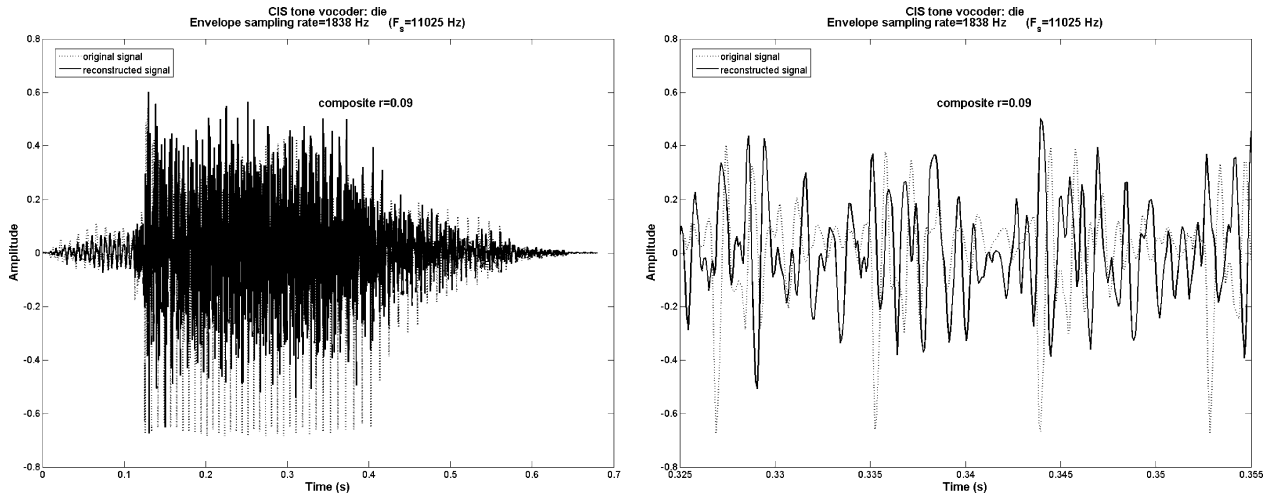


Fig. 10. Composite tone vocoder reconstruction (solid line) of the original speech waveform “die” (dotted line). The composite correlation coefficient is 0.09.

TABLE IV
CORRELATION IN COMPOSITE RECONSTRUCTION
FOR DIFFERENT PROCESSING METHODS

Sound file	AIS spike-based reconstruction	CIS noise-based vocoding (100 trials)	CIS tone-based vocoding	CIS spike-based reconstruction (1.4 kHz firing rate)
“die”	0.48	0.10 ± 0.02	0.09	0.07
“buy”	0.43	0.10 ± 0.02	0.08	0.04
Handel chorus	0.67	0.08 ± 0.02	0.04	0.01
Jazz music	0.58	0.04 ± 0.01	0.02	0.01

match in the composite reconstruction for AIS may likely be due to the improved transmission of $\phi(t)$ matching up with $E(t)$ in the spike output. A tone vocoder reconstruction analogous to the AIS reconstruction in Fig. 8 is shown in Fig. 10, and the effect of poor phase correlation can be clearly seen.

D. What do They Sound Like?

Correlation coefficients only give us a very general indication of the signal fidelity. The added value of performing a spike reconstruction is that the level of information encoded can be assessed by listening. If the sound fidelity is poor, it does not necessarily mean that a high level of information is absent, but if the sound fidelity is good, it demonstrates that that a high level of information must be present. Some sample reconstructions can be listened to at <http://avns1.mit.edu/AIS>.

The sound quality in AIS reconstructions is noticeably improved in that they sound more natural, and AIS channels should contain sufficient information for tonal languages to be correctly represented. In the case of music, while other reconstructions retain only the rhythm, a clear melody and even different musical instruments are perceptible in the AIS reconstructions.

VI. PERCEPTUAL TESTS IN NOISE

In order to determine whether AIS can provide any real advantage in cochlear implants, testing our strategy with CI users is absolutely necessary. However, such testing is costly in time

and resources, and hence often prohibitive unless we are convinced of the new strategy’s potential merits. One faster and less costly way of evaluating a new CI strategy is to perform perceptual tests on normal-hearing listeners, using acoustic reconstructions of cochlear implant outputs as described above. While previous results from CI simulations have been found to correlate with CI performance on some perceptual measures [11], it should be emphasized that they can only gauge the best possible outcomes for CI users, as many differences between acoustic and electric stimulation are not accounted for in these reconstruction techniques, such as the channel interactions and poor spatio-temporal coding in real implants. Furthermore, the tests we perform to evaluate AIS are tasks relying on the perception of fine time structure, and unlike tasks that rely only on the perception of envelope cues, there is only a small body of evidence to suggest that the results will correlate with actual CI performance [37]. Nevertheless, these perceptual tests should provide an indication of whether a new CI strategy is worth testing on real CI users. In this section, we present two psychoacoustic experiments that were designed to verify whether AIS provides any advantage in coding speech and music, particularly in the presence of noise, as perceived by normal-hearing listeners.

A. Methods

Eight normal-hearing listeners were recruited from a local on-line bulletin board to participate in this study. Five subjects were female and three were male, ranging in age from 23–48 with a mean age of 29.5 years. Their hearing thresholds were screened before the test to be at 20 HL or better. Signals were presented at 70 dB SPL over Sennheiser 580 headphones in a sound-attenuating booth. Speech-spectrum shaped noise was used in conditions where noise was added to the stimulus. Each experiment began with practice trials, during which feedback was provided.

The target sounds and noise maskers were mixed before processing, using either AIS spike reconstruction as described in the previous section, or envelope vocoding with tonal carriers (as in [33] and [38]). In both cases, 16 contiguous frequency channels were used, with center frequencies spaced equally on

a logarithmic scale, and an overall passband extending from 100 to 5000 Hz. Pilot data from CIS spike-based reconstructions and CIS noise-based vocoding both resulted in poorer performance than CIS tone-based vocoding for speech recognition in noise and were therefore not used, as we wanted to compare the best performing CIS acoustic simulation against our AIS acoustic simulation.

In the first experiment, subjects were told that they would be listening to distorted speech sounds in a noisy background. They were told that some of the utterances would be very hard to understand, and that they should type all the words that they think they hear. In the practice trials, they were presented with 8 lists of 10 HINT sentences [39], counterbalanced across subjects for the 2 processing conditions tested, namely AIS spike reconstruction and CIS tone vocoding. Half of the practice trials were presented in quiet ($\text{SNR} = \infty$) and the other half at +6 dB SNR, in alternating sequence. During this stage, subjects were given the chance to hear sentences again after typing their response, and were shown on the screen what they had just heard. Actual trials used 16 different lists of 10 HINT sentences at four different SNRs (6, 3, 0, and -3 dB). Lists were randomly selected, and SNR and processing condition were randomized for each list. This resulted in two lists (or 20 sentences) for each condition. Repetition of the stimuli was not allowed and no feedback was provided.

In the second experiment, subjects were presented with 34 common melodies that had all rhythmic information removed, and were synthesized from 16 equal-duration notes using samples from a grand piano. These melodies were also used in previous studies on melody recognition [19], [40]. Subjects were then asked to select 10 melodies they were most familiar with, which were then played back in random order for them to identify. All subjects were able to find 10 melodies that they could easily identify correctly. Actual trials presented subjects with their 10 melodies which were processed by AIS spike reconstruction and CIS tone vocoding. Melodies were presented at two SNRs (in quiet, $\text{SNR} = \infty$, and in noise, with $\text{SNR} = 0$ dB), counterbalanced across subjects for both SNR and processing condition. All melodies were presented twice in random order for each experimental condition. Subjects were instructed to identify the melody, and were forced to select their response from the closed set of 10 melody names on the screen in front of them. Repetition of the stimulus was not allowed and no feedback was provided.

B. Results

Fig. 11 shows HINT sentence recognition scores as a function of SNR for the two processing conditions, AIS spike reconstruction and CIS tone vocoding. In general, subjects did no worse with tone vocoding than with AIS for the conditions of 6, 3, and 0 dB SNR. However at -3 dB SNR, subjects performed better with AIS by 17 percentage points. Analysis of variance (ANOVA) on arcsine transformed data (to normalize the compression of variance near 100% and 0%) showed a significant main effect for SNR [$F(2, 13) = 145$, $p < 0.001$], processing condition [$F(1, 7) = 5.80$, $p < 0.05$] and an interaction between SNR and processing [$F(2, 15) = 7.53$, $p < 0.005$]. A post-hoc analysis (including Bonferroni correction) using a

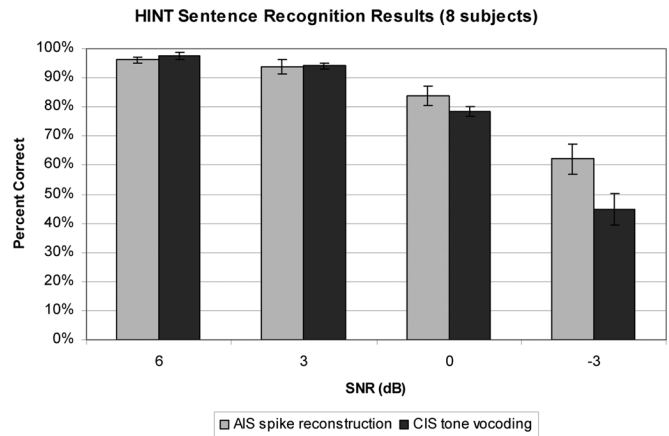


Fig. 11. Sentence recognition scores for AIS versus CIS tone vocoding reconstructions in noise. Error bars show one standard error.

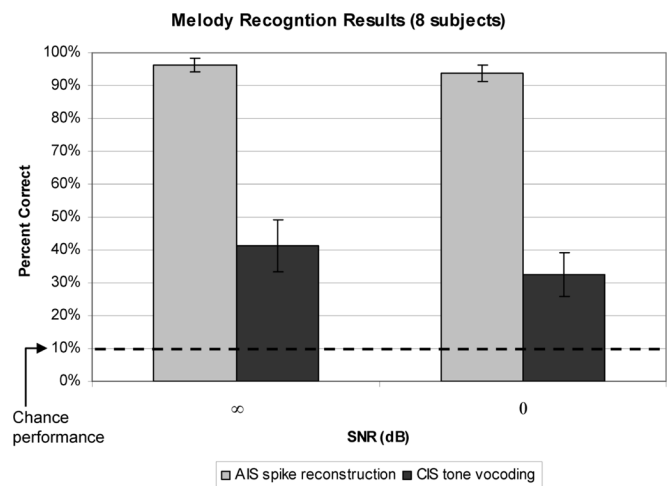


Fig. 12. Melody recognition scores for AIS versus CIS tone vocoding reconstructions in quiet and noise. Error bars show one standard error.

paired samples *t*-test revealed that only the -3 dB SNR condition showed a statistically significant difference between the two processing conditions [$t(7) = 5.61$, $p < 0.005$]. These results confirm the experiments of Nie *et al.* [1], which suggest that additional FM cues improve performance more in noise than in quiet. AIS may therefore improve the hearing of speech in noise, if additional phase information can indeed be delivered to CI users by this strategy.

Fig. 12 shows melody recognition scores for the two SNR conditions and two processing conditions tested. Subjects performed better with AIS by 55 percentage points in quiet (∞ SNR), and by 61 percentage points in noise (0 dB SNR). A repeated-measures ANOVA on arcsine transformed data revealed a significant main effect for processing condition [$F(1, 7) = 107$, $p < 0.001$]. In general, subjects were clearly more able to recognize melodies correctly when listening to AIS spike reconstructions. Thus, tonal perception in CI users may also be improved, if additional phase information is in fact transmitted by AIS. There was no significant main effect for SNR or any interaction. Thus, the addition of noise at an SNR of 0 dB had no significant effect on subjects' ability to recognize melodies in either processing scheme.

VII. DISCUSSION AND CONCLUSION

The AIS technique ensures that the average stimulation rate in each channel is limited by a refractory mechanism, that only one channel is active at one time, minimizing electrode interactions, that there is good timing precision in each channel when it is active, and that the average stimulation rate of a channel is low, saving power. The AIS strategy generates pseudostochastic spike trains that should generate a less artificial neural response than synchronous stimulation.

Earlier stimulation strategies have recognized the importance of providing phase information. For example, the peak derived timing (PDT) strategy presented in [4] and [41] stimulates at times corresponding to positive peaks in the filter-band output, and the spike-based temporal auditory representation (STAR) strategy [13] generates spikes (stimulation pulses) at the zero-crossings. Both also encode phase information in the time of firing, but what truly matters is whether CI users are able to utilize the coded information. CI users in [4] were found to do no better at a pitch ranking task using PDT than other strategies that did not encode phase information. Many factors are likely to limit phase coding with CI users, such as perceptual dissonance [16] and widespread fiber synchrony which is endemic to electrical stimulation, and for these reasons AIS may perform no differently from other strategies. However, in contrast with other strategies, AIS firing times are determined when $E(t)$ in a channel is deemed in a uniquely bio-inspired way to be more pertinent than other channels, using a neuronal "integrate-to-fire" competition. Whether this or other details in AIS make any difference, however, remains to be proven in tests with real CI users.

In conclusion, we have demonstrated a simulation of an AIS strategy for neural stimulation in cochlear implants, that encodes both phase and envelope information known to be important in perceiving tonal languages and music, and for hearing in noise. Stimulus reconstructions with AIS using simple filtering-and-sum techniques show significantly higher correlation coefficients with the input for both speech and music than other stimulus reconstructions which use only envelope information. Perceptual tests in noise show that the improved correlation is reflected in normal-hearing listeners' ability to recognize both sentences and melodies more easily with AIS reconstructions than with more traditional envelope vocoding techniques. Our results confirm that phase information should make a greater difference for perceiving melodies than for speech in noise. However, future tests with CI users will be necessary to verify whether the potential benefits of AIS are borne out, and whether further modifications are necessary.

REFERENCES

- [1] K. Nie, G. Stickney, and F.-G. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 64–73, Jan. 2005.
- [2] F.-G. Zeng, K. Nie, G. S. Stickney, Y.-Y. Kong, M. Vongphoe, A. Bhargava, C. Wei, and K. Cao, "Speech recognition with amplitude and frequency modulations," *PNAS*, vol. 102, pp. 2293–2298, 2005.
- [3] H. J. McDermott, "Music perception with cochlear implants: A review," *Trends Amplif.*, vol. 8, pp. 49–82, 2004.
- [4] A. E. Vandali, C. Sucher, D. J. Tsang, C. M. McKay, J. W. D. Chew, and H. J. McDermott, "Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies," *J. Acoust. Soc. Am.*, vol. 117, pp. 3126–3138, 2005.
- [5] L. Xu, Y. Li, J. Hao, X. Chen, S. A. Xue, and D. Han, "Tone production in mandarin-speaking children with cochlear implants: A preliminary study," *Acta Oto-Laryngologica*, vol. 124, pp. 363–367, 2004.
- [6] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236–238, 1991.
- [7] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Process. Mag.*, vol. 15, no. 5, pp. 101–130, Sep. 1998.
- [8] M. K. Qin and A. J. Oxenham, "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.*, vol. 114, pp. 446–454, 2003.
- [9] Q.-J. Fu, R. V. Shannon, and X. Wang, "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," *J. Acoust. Soc. Am.*, vol. 104, pp. 3586–3596, 1998.
- [10] M. F. Dorman, P. C. Loizou, J. Fitzke, and Z. Tu, "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," *J. Acoust. Soc. Am.*, vol. 104, pp. 3583–3585, 1998.
- [11] L. M. Friesen, R. V. Shannon, D. Baskent, and X. Wang, "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.*, vol. 110, pp. 1150–1163, 2001.
- [12] G. S. Stickney, K. Nie, and F.-G. Zeng, "Contribution of frequency modulation to speech recognition in noise," *J. Acoust. Soc. Am.*, vol. 118, pp. 2412–2420, 2005.
- [13] D. B. Grayden, A. N. Burkitt, O. P. Kenny, J. C. Clarey, A. G. Paolini, and G. M. Clark, "A cochlear implant speech processing strategy based on an auditory model," presented at the Intelligent Sensors, Sensor Networks and Information Processing Conf., Melbourne, Australia, 2004.
- [14] R. P. Carlyon and J. M. Deeks, "Limitations on rate discrimination," *J. Acoust. Soc. Am.*, vol. 112, pp. 1009–1025, 2002.
- [15] Y. C. Tong and G. M. Clark, "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients," *J. Acoust. Soc. Am.*, vol. 77, pp. 1881–1888, 1985.
- [16] G. E. Loeb, "Are cochlear implant patients suffering from perceptual dissonance?," *Ear Hear.*, vol. 26, pp. 435–450, 2005.
- [17] B. Townshend, N. Cotter, D. V. Compemolle, and R. L. White, "Pitch perception by cochlear implant subjects," *J. Acoust. Soc. Am.*, vol. 82, pp. 106–115, 1987.
- [18] W. M. Siebert, *Circuits, Signals, and Systems*. Cambridge, Mass.: MIT Press, 1986.
- [19] Z. M. Smith, B. Delgutte, and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, pp. 87–90, 2002.
- [20] N. Lan, K. B. Nie, S. K. Gao, and F. G. Zeng, "A novel speech-processing strategy incorporating tonal information for cochlear implants," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, pp. 752–760, May 2004.
- [21] L. M. Litvak, B. Delgutte, and D. K. Eddington, "Improved neural representation of vowels in electric stimulation using desynchronizing pulse trains," *J. Acoust. Soc. Am.*, vol. 114, pp. 2099–2111, 2003.
- [22] —, "Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains," *J. Acoust. Soc. Am.*, vol. 114, pp. 2079–2098, 2003.
- [23] H. Chen and F.-G. Zeng, "Frequency modulation detection in cochlear implant subjects," *J. Acoust. Soc. Am.*, vol. 116, pp. 2269–2277, 2004.
- [24] F.-G. Zeng, "Temporal pitch in electric hearing," *Hear. Res.*, vol. 174, pp. 101–106, 2002.
- [25] A. E. Vandali, L. A. Whitford, K. L. Plant, and G. M. Clark, "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," *Ear Hear.*, vol. 21, pp. 608–624, 2000.
- [26] A. J. Spahr and M. F. Dorman, "Performance of subjects fit with the advanced bionics CII and nucleus 3G cochlear implant devices," *Arch. Otolaryngol. Head Neck Surg.*, vol. 130, pp. 624–628, 2004.
- [27] L. M. Litvak, Z. M. Smith, B. Delgutte, and D. K. Eddington, "Desynchronization of electrically evoked auditory-nerve activity by high-frequency pulse trains of long duration," *J. Acoust. Soc. Am.*, vol. 114, pp. 2066–2078, 2003.
- [28] E. C. Smith and M. S. Lewicki, "Efficient auditory coding," *Nature*, vol. 439, pp. 978–982, 2006.
- [29] R. Sarpeshkar, M. W. Baker, C. D. Salthouse, J. J. Sit, L. Turicchia, and S. M. Zhak, "An analog bionic ear processor with zero-crossing detection," presented at the IEEE Int. Solid-State Circuits Conf., San Francisco, CA, 2005.
- [30] R. Sarpeshkar, C. Salthouse, J. J. Sit, M. W. Baker, S. M. Zhak, T. K. T. Lu, L. Turicchia, and S. Balster, "An ultra-low-power programmable analog bionic ear processor," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 711–727, Apr. 2005.

- [31] P. Dayan and L. F. Abbott, *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press, 2001.
- [32] C. A. Miller, P. J. Abbas, and B. K. Robinson, "Response properties of the refractory auditory nerve fiber," *J. Assoc. Res. Otolaryngol. (JARO)*, vol. 2, pp. 216–232, 2001.
- [33] M. F. Dorman, P. C. Loizou, and D. Rainey, "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Am.*, vol. 102, pp. 2403–2411, 1997.
- [34] R. V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, pp. 303–304, 1995.
- [35] R. Wessel, C. Koch, and F. Gabbiani, "Coding of time-varying electric field amplitude modulations in a wave-type electric fish," *J. Neurophysiol.*, vol. 75, pp. 2280–2293, 1996.
- [36] E. deBoer and H. R. deJongh, "On cochlear encoding: Potentialities and limitations of the reverse-correlation technique," *J. Acoust. Soc. Am.*, vol. 63, pp. 115–135, 1978.
- [37] L. M. Collins, G. H. Wakefield, and G. R. Feinman, "Temporal pattern discrimination and speech recognition under electrical stimulation," *J. Acoust. Soc. Am.*, vol. 96, pp. 2731–2737, 1994.
- [38] S. F. Poissant, N. A. Whitmal, III, and R. L. Freyman, "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," *J. Acoust. Soc. Am.*, vol. 119, pp. 1606–1615, 2006.
- [39] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, vol. 95, pp. 1085–1099, 1994.
- [40] A. Lobo, F. Toledos, P. C. Loizou, and M. F. Dorman, "The effect of envelope low-pass filtering on melody recognition," presented at the 33rd Neural Prosthesis Workshop, Bethesda, MD, 2002.
- [41] R. J. M. van Hoesel and R. S. Tyler, "Speech perception, localization, and lateralization with bilateral cochlear implants," *J. Acoust. Soc. Am.*, vol. 113, pp. 1617–1630, 2003.



Ji-Jon Sit received the B.Sc. degrees in electrical engineering and computer science from Yale University, New Haven, CT, in 2000, and the Master's degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002. He is currently working towards the Ph.D. degree on neural stimulation for cochlear implants in the Analog VLSI & Biological Systems Lab at MIT.



Andrea M. Simonson received the Ph.D. degree in communication sciences and disorders from Northwestern University, Evanston, IL, in 1991.

She worked as a clinical audiologist for several years before becoming a Research Scientist at the Massachusetts Institute of Technology's Research Laboratory of Electronics in 2002. She is currently a Research Audiologist with the Auditory Perception and Cognition lab in the Psychology Department at the University of Minnesota, Minneapolis.

Dr. Simonson became a clinically certified

audiologist in 1997.



Andrew J. Oxenham received the B.Mus. degree in music and sound recording (Tonmeister) from the University of Surrey, Surrey, U.K., and the Ph.D. degree in experimental psychology from the University of Cambridge, Cambridge, U.K.

Following positions at the Institute for Perception Research (IPO) in the Netherlands, Northeastern University, Evanston, IL, and the Massachusetts Institute of Technology, Cambridge, he is now on the faculty of the Psychology Department at the University of Minnesota, Minneapolis, where

he leads the Auditory Perception and Cognition Laboratory. His interests include auditory perception in normal and impaired hearing, cochlear implants, functional imaging, and music perception.

Dr. Oxenham's awards include an International Prize Fellowship from the Wellcome Trust, the 2001 R. Bruce Lindsay Award from the Acoustical Society of America, and several research grants from the National Institutes of Health. He is a fellow of the Acoustical Society of America, associate editor of the *Journal of the Acoustical Society of America* and the *Journal of the Association for Research in Otolaryngology*, and author of over 40 journal publications.



Michael A. Faltys received the B.S.E. degree in electrical engineering from University of California, Irvine, in 1983.

From 1983 to 1987 he was with TRW working on satellite spread spectrum systems, from 1987 to 1995 he was with Teradata/NCR working as a Computer Architect on a highly parallel computer system for business use, and from 1995 to present he has been architecting and developing cochlear implants for Advanced Bionics, a Boston Scientific Company.

Mr. Faltys is a member of the Acoustical Society

of America.



Rahul Sarpeshkar received B.S. degrees in electrical engineering and physics at the Massachusetts Institute of Technology (MIT). He received the Ph.D. degree from the California Institute of Technology (Caltech), Pasadena, in 1997

After completing the Ph.D. degree he joined Bell Labs as a member of technical staff. Since 1999, he has been on the faculty of MIT's Electrical Engineering and Computer Science Department where he heads a research group on Analog VLSI and Biological Systems, and is currently an Associate

Professor. His research interests include analog VLSI, biomedical and bio-inspired electronics, ultra low-power circuits and systems, and control theory.

Dr. Sarpeshkar has received several awards including the Packard Fellow award given to outstanding young faculty, the ONR Young Investigator Award, and the National Science Foundation (NSF) Career Award. He holds over a dozen patents and has authored several publications including one that was featured on the cover of *Nature*.