COCHLEAR IMPLANT ACOUSTIC SIMULATION MODEL BASED ON CRITICAL BAND FILTERS

P Mahalakshmi¹, M R Reddy²
Bio-medical Engineering Group, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai-600 036, Tamil Nadu, India.
¹E-mail: maha_50@yahoo.com, ²E-mail: rsreddy@iitm.ac.in

ABSTRACT

Cochlear Implants (CI) are surgically implanted biomedical devices that can provide hearing to profoundly deaf people by direct electrical stimulation of the auditory nerve. Choosing the optimal number of electrodes for the improved perception of speech is an especially difficult problem for users of CI, as the prostheses system depends on various parameters and processing strategies. It has widely been recognized that using the auditory models in cochlear prostheses might significantly improve the speech perception. In this paper, we propose a new CI acoustic simulation model based on a 16-channel Critical Band Finite-Impulse Response (CB-FIR) filter bank and its performance is discussed. The place of stimulation in the cochlear electrode array is identified through the computation and selection of significant spectral band energies. We also investigate the temporal information through the simulation model based algorithm. The amplitude levels of the pulses are identified by envelope detection method. The simulation model is tested with TIMIT database for 15 phonemes. The experimental results indicated that, a reduced number of channels are only required for stimulation of cochlear electrodes, one at a time, thereby minimizing channel interaction. The proposed model should be helpful for the development of novel speech processing strategies to improve the speech perception of cochlear implantees.

Keywords

Auditory model, cochlear implant, speech processing, speech perception, spectral analysis, acoustic simulation.

1. INTRODUCTION

A cochlear implant (CI) is an auditory neural prosthesis for restoring hearing function in patients with sensori-neural hearing loss. Hearing restoration is achieved by electrically
stimulating the auditory nerve [1]. The electrical stimulation pulse parameters are derived from incoming speech by speech processors contained within the CI devices [2]. Several CI devices developed over the years have the following features in common. It is comprised of external/ and internal units. The external unit includes microphone, speech processor and transmitter. The internal unit consists of receiver-stimulator and an electrode array embedded in cochlea [3]. The sound is picked up by the microphone and processed by the signal/speech processor that analyzes sound and encodes the results into instructions. By RF (radio frequency) signal, the transmission system transmits the instructions to the implanted receiver. The internal circuit decodes the data, then generates electrical stimulus and drives the electrode array to stimulate auditory nerve fibers at different places [4]. The speech processor is responsible for decomposing the input audio signals into different frequency bands and formulating the most appropriate stimulation pattern to the electrodes [5]. The design of speech processors must be such that it extracts the speech parameters that are essential for intelligibility and then encode them for electrical stimulation [6]. Speech understanding requires the analysis of various speech components such as vowels and consonants. The estimate of perceptual importance of these component features will lead to intelligible speech stimuli [7]. This speech stimuli depends on various parameters such as speech material, filter cutoff frequencies, order of band pass and low pass filters, number of channels, number of electrodes, type and parameters of the compression function, pulse width and pulse duration of modulating pulses, sequence of channel stimulation [8].

Various speech processing strategies have been developed and reported in literature over time for cochlear prostheses which include Continuous Interleaved Sampling (CIS), Spectral Peak (SPEAK), Advanced Combination Encoder (ACE), Spectral Maxima Sound Processor (SMSP), Simultaneous Analog Strategy (SAS) and various formant based strategies [9,10]. Most implant manufacturers provide research speech processors for use on human subjects that allow researchers to develop and test new signal processing algorithms [11]. Much research has therefore been directed at improving the performance and major successes have been achieved since they were first implemented in the 1970s. The number of implanted electrodes has grown from 1 to as many as 24 in current implants [12]. However, current cochlear implants are not very successful in delivering phase information that might be responsible for the reduced ability of CI users to perceive music.

Spectral enhancement is another major concern that has benefits of improving performance in noise. It is possible that this is one of the reasons for why the human auditory system with a biological cochlea front end is known to have good performance in noise[13]. Spectral enhancement is a natural consequence of our trying to mimic the operation of the biological cochlea. Also, the effect of electrode interactions and current spreading limits the spectral resolution to a great degree. A common speech processing strategy used in implants and in speech-recognition systems, employs a mel cepstrum filter bank. In implants, the number of functioning electrodes used for stimulation is between 8-24 and having more electrodes is often not useful due to spatial interactions amongst the electrodes [14].

Our work proposes an acoustic simulation model for cochlear implants that uses a 16-channel CB-FIR filter bank. Spectral enhancement and audibility could be improved by using a set of independent filters with independent and easily programmable parameters. Input speech signals are filtered using this filter bank, then rectified, low-pass filtered and the amplitude levels are estimated. Spectral energies are computed for various speech components such as phonemes from each channel. The most significant channels, which are few in number, constituting maximum signal energy are selected thus aiding to minimize channel interactions. These channel energies are used in estimating the stimulation pulse level to be presented to the selected cochlear electrodes.
The outline of the paper is as follows. In Section 2, we describe the essential ideas behind the choice of filter bank strategy, the improved design aspects and how various speech components are assessed for its speech intelligibility. In Section 3, we present Matlab simulation data for the designed filter bank system with speech input that illustrate the working and benefits of the system. In Section 4, we illustrate and discuss the results of our proposed cochlear implant acoustic simulation model through the spectral and temporal analysis performed for various speech components and paving a way for the electrical stimulation procedure for cochlear implants. In Section 5, we conclude by summarizing our contributions.

2. MATERIALS AND METHODS

2.1 Digital speech processing

The greatest common block of all recognition systems is the signal-processing front end, which converts the speech waveform to some type of parametric representation for further analysis and processing. Parametrical representation of speech signal uses short-time energy and zero crossing rates [15]. Speech signals are more consistently analyzed spectrally than in the time domain and human hearing pays more attention to spectral aspects of speech [16]. Filter bank representation is one such spectral analysis method that uses a set of band pass filters, each analyzing a different range of frequencies of the input speech. There are two main motivations for the filter bank representation. First, the position of maximum displacement along the basilar membrane for stimuli such as pure tones is proportional to the logarithm of the frequency of the tone. This hypothesis is part of a theory of hearing called the ‘place’ theory. Second, the experiments in human perception have shown that frequencies of a complex sound within a certain bandwidth of some nominal frequency cannot be individually identified. When one of the components of this sound falls outside this bandwidth, it can be individually distinguished. This bandwidth is referred to as the critical bandwidth [17].

The digital speech processing proposed for ‘cochlear implant - acoustic simulation model’ is built around a critical band filter bank followed by rectifiers and low pass filters. The cascade of a rectifier and a low pass filter yields an envelope detector. Input sound signals are filtered using a digital band-pass FIR (Finite Impulse Response) filter bank, with one filter for each electrode of the implant. The number of filter bands depends on the number of stimulation channels to be considered in the implant of cochlear prostheses system [18]. The filter bank was designed to have linear frequency spacing at low center frequencies and logarithmic spacing at high center frequencies in correspondence to the properties of cochlear filters [19]. In accordance with Rekha et al., [20], the human speech frequency, approximately 100 to 3500 Hz, is allocated to filter bands using critical band scale. This frequency range is covered in 16 bands, which is shown in Table 1. In the signal analysis application, filter bank model with non-uniform spacing has the potential to offer better performance [21].

The general specifications of the filter bank include the stop band attenuation (e.g., 40-50dB), number of channels (e.g., 16) and the pass band spacing. For the advanced cochlear devices, we specify the need for a linear phase finite-impulse response (FIR) filter bank as opposed to the nonlinear phase infinite-impulse response (IIR) filter bank for the following reasons [22].

- FIR filters are intrinsically stable
- They can easily realize various features that are not possible or are difficult to achieve with IIR filters
- To eliminate channel interactions.
To achieve more stop-band attenuation, a higher filter order is required with consequent cost in CPU usage. We impose higher stop band attenuation (≤60dB) compared to −50dB achieved by the current devices.

Table 1. Critical bandwidth as a function of center frequency: Non-uniform, non-overlapping band-pass filter cutoff frequencies that correspond to cochlea filter properties are listed here.

<table>
<thead>
<tr>
<th>Channel Number</th>
<th>Center Frequency (Hz)</th>
<th>Frequency Band (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150</td>
<td>125-175</td>
</tr>
<tr>
<td>2</td>
<td>250</td>
<td>225-275</td>
</tr>
<tr>
<td>3</td>
<td>350</td>
<td>325-375</td>
</tr>
<tr>
<td>4</td>
<td>450</td>
<td>420-480</td>
</tr>
<tr>
<td>5</td>
<td>570</td>
<td>530-605</td>
</tr>
<tr>
<td>6</td>
<td>700</td>
<td>655-745</td>
</tr>
<tr>
<td>7</td>
<td>840</td>
<td>790-890</td>
</tr>
<tr>
<td>8</td>
<td>1000</td>
<td>940-1060</td>
</tr>
<tr>
<td>9</td>
<td>1170</td>
<td>1105-1235</td>
</tr>
<tr>
<td>10</td>
<td>1370</td>
<td>1285-1455</td>
</tr>
<tr>
<td>11</td>
<td>1600</td>
<td>1505-1695</td>
</tr>
<tr>
<td>12</td>
<td>1850</td>
<td>1745-1955</td>
</tr>
<tr>
<td>13</td>
<td>2150</td>
<td>2005-2295</td>
</tr>
<tr>
<td>14</td>
<td>2500</td>
<td>2345-2655</td>
</tr>
<tr>
<td>15</td>
<td>2900</td>
<td>2705-3095</td>
</tr>
<tr>
<td>16</td>
<td>3400</td>
<td>3145-3655</td>
</tr>
</tbody>
</table>

2.2 FIR filter design using kaiser window

The problem of designing an FIR filter consists of determining the impulse response sequence, h[n], or its system function H(z), so that given requirements on the filter response are satisfied. The filter requirements are usually specified in the frequency domain, and only this case is considered here [23]. The frequency response \( H_d(e^{j\omega}) \) corresponding to the impulse response h[n] is expressed in a fourier series as

\[
H_d(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_d(n) e^{-j\omega n}
\]

and known as fourier coefficients having infinite length. FIR filter is obtained by truncating the infinite fourier series at \( n = \pm \left( \frac{N-1}{2} \right) \), where N is the length of the desired sequence.

Abrupt truncation of the fourier series results in oscillations in pass-band and stop-band regions of the filter. An optimal way to reduce these oscillations is to use an appropriate finite-length window \( w(n) \), which controls the overall filter to yield a smooth frequency response. Therefore, a window should possess some of the following spectral characteristics [24,25].

(i) The main lobe width in the frequency response of the window should be narrow [17]
(ii) The ripple ratio should be small
(iii) The side-lobes should decrease in energy rapidly as \( \omega \) tends to \( \pi \).
(iv) To adequately block the stop-band frequencies, it is necessary to have good stop-
band attenuation.

The symmetrical frequency spectrum of an odd length window \( w(nT) \) can be written as [25],

\[
W(e^{j\omega T}) = w(0) + 2 \sum_{n=1}^{(N-1)/2} w(nT) \cos(wnT)
\]

(2)

Where \( T = 2\pi/\omega_s \) is the sampling period.

Standard windows such as Rectangular, Hamming and Hanning have only one independent parameter namely, the ‘window length’ \( N \), which controls the ‘transition width’ of the filter. An increase in window length decreases the transition width of the filter. But the minimum stop band attenuation of the filter is independent of \( N \). In order to overcome this, a Kaiser window is used which has an adjustable parameter \( \beta \) that controls the stop-band attenuation and an independent parameter \( N \), that controls the transition width [26]. Kaiser window is widely used for the spectral analysis and FIR filter design applications.

The design of band pass filter using Kaiser Window has been carried out based on the following steps [24,26,27].

**Step 1:**

The length ‘\( N \)’ of the band pass filter is computed using Eq (3).

\[
N \equiv -\frac{20\log_{10}\left(\sqrt{\delta_p \delta_s}\right)-13}{14.6(\omega_s - \omega_p)/2\pi}
\]

(3)

where, \( \delta_p \) and \( \delta_s \) are the pass band and stop band ripples respectively and \( \omega_p \) and \( \omega_s \) are the pass band and stop band edge frequencies respectively.

The design is based on the narrower of the two transition bandwidths (\( B_t \)), that is

\[
B_t = \min\left\{ (\omega_{p_2} - \omega_{s_2}),(\omega_{s_1} - \omega_{p_1})\right\}, \quad \text{where} \quad (\omega_{s_1},\omega_{p_1}) \quad \text{and} \quad (\omega_{s_2},\omega_{p_2}) \quad \text{are the initial and later transition edge frequencies of the band pass filter.}
\]

The cut-off frequencies are given by

\[
\omega_{c_1} = \omega_{p_1} - B_t/2 \quad \text{and} \quad \omega_{c_2} = \omega_{p_2} + B_t/2.
\]

**Step 2:**

The Kaiser window sequence is given by

\[
w[n] = I_0 \left( \frac{\beta \sqrt{1 - \left[2n/(N-1)\right]^2}}{I_0(\beta)} \right), \quad -(N-1)/2 \leq n \leq (N-1)/2
\]

(4)

where \( \beta \) is an adjustable parameter, and \( I_0(x) \) is the modified zeroth order Bessel function, which can be expanded into the following series.

\[
I_0(x) = 1 + \sum_{r=1}^{\infty} \left[ \left( \frac{x}{2} \right)^r \right] \frac{2}{r!}
\]

(5)

which is seen to be positive for all real values of \( x \).
Step 3:
The desired minimum stop band attenuation is calculated by \( \alpha_s = -20 \log_{10} \delta_s \). \( \beta \) is computed from Eq.(6).

\[
\beta = \begin{cases} 
0.1102 (\alpha_s - 8.7), & \text{for } \alpha_s > 50, \\
0.5842 (\alpha_s - 21)^{0.4} + 0.007886 (\alpha_s - 21), & \text{for } 21 \leq \alpha_s \leq 50, \\
0, & \text{for } \alpha_s < 21. 
\end{cases}
\]  

(6)

Step 4:
The coefficients of linear phase filter with delay \( \gamma = \left( \frac{N-1}{2} \right) \) are computed using

\[
h_d(n) = \begin{cases} 
\left( \frac{\omega_{c_2} - \omega_{c_1}}{\pi} \right), & \text{for } n = 0 \\
\left( \frac{\sin \omega_{c_2}(n-\gamma) - \sin \omega_{c_1}(n-\gamma)}{\pi(n-\gamma)} \right), & \text{for } n \neq 0 
\end{cases}
\]  

(7)

Step 5:
The impulse response \( h(n) \) is obtained by multiplying the desired impulse response \( h_d(n) \) of the ideal filter and the window coefficients generated in Step 2 to yield the coefficients of FIR filter.

\[
h(n) = h_d(n) \cdot w(n)
\]  

(8)

Step 6:
Filter specifications are usually given in terms of the magnitude response \( |H(e^{j\omega})| \). The magnitude function of FIR filter when the impulse response is symmetric and when \( N \) is odd is generated by

\[
|H(\omega)| = h \left( \frac{N-1}{2} \right) + \sum_{n=1}^{\frac{N-1}{2}} 2h \left( \frac{N-1}{2} - n \right) \cos(\omega n)
\]  

(9)

Filters designed with this window had the level of stop-band approximately equal to \(-60\text{dB} \). 

2.3 CI-Acoustic Simulation Model

The schematic of the proposed CI acoustic simulation model is shown in Figure 1. The input speech signal is passed through a bank of 16 band pass filters, whose coverage spans the human speech frequency range. The spectral output is obtained by the summation of independent band pass filters responses. Spectral energies are estimated for each of 16 channels. Temporal analysis is done by full-wave rectification and low pass filtering the decomposed signal in each band. The stimulus amplitude levels are estimated through envelope detection procedure.
Fig. 1. Schematic of the proposed CI-acoustic simulation model: Spectral energies are estimated by a bank of band pass filters, followed by temporal estimation by rectification and envelope detection, determining the place and stimulus level information.

2.3.1 Rectification

Speech signals are complex patterns that do not permit direct analysis of their wave shape. The envelope of the overall signal carries important information. The envelope of a signal can be derived by obtaining the absolute value of the signal at each time instant, i.e., performing full-wave rectification [28]. A FIR low pass filter of order 100 and a cut-off frequency of 400 Hz are used to obtain smooth envelopes of speech signals [29].

2.3.2 Envelope detection

Envelope detection is done using a mathematical method called ‘Linear Interpolation’. This technique is used to detect the envelope of the low pass filter output. In order to obtain smooth envelope of the low pass filter output, the estimation of values for intermediate time instances are carried out. A point in the low pass filter output is interpolated using Eq (9).

\[ y = y_1 + \frac{(y_2 - y_1)(x - x_1)}{(x_2 - x_1)} \]  

where \(x_1, x_2\) are the two time instances and \(y_1, y_2\) are their corresponding amplitude values estimated from the output of the low pass filter. If these quantities are known, then for a desired intermediate time instant \(x\), a new value \(y\) can be obtained. Matlab code is developed for interpolation and stimulus amplitude estimation.

2.4 Speech sounds and features

The perceptually important information of a speech signal needs to be preserved in order to assess its intelligibility. Speech sounds called phonemes are produced depending on the state of input excitation (vocal cords). Voiced speech (e.g., vowels) is produced if the vocal cords are tensed and vibrate periodically. Unvoiced speech (e.g., consonants) is produced if the vocal cords are not vibrating and the resulting waveform is aperiodic or random [15]. Based on the position of the tongue, vowel articulatory configurations are produced and classified as front vowel /i/, mid vowel /a/ and back vowel /u/. A ‘diphthong’ is a complex speech sound that begins with one vowel and gradually changes to another vowel within the same syllable (e.g.s., /ay/, /oy/, /aw/, /ey/). Voiced (/b/, /d/, /g/) and unvoiced stop consonants (/p/, /t/, /k/) are produced by the complete stoppage and sudden release of the breath. Nasal consonants (/m/, /n/) are produced with glottal excitation and the vocal tract totally constricted. In the design of speech analysis system, how the spectral and temporal...
properties of a speech signal affect its perception need to be studied [30]. In this paper, we have estimated the spectral energy to separate the voiced and unvoiced parts of speech and also to determine the stimulus pulse level to be given to the cochlear electrodes, using the developed ‘CI acoustic simulation model’ discussed in the earlier section. Table II lists the classification of phonemes into sound classes taken for the study.

Table 2. Classification of the standard phonemes into sound classes: Three vowels: front, mid, back corresponding to the position of the tongue hump; four diphthongs, two nasal consonants; three voiced and unvoiced stop consonants are listed here.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Type of sound</th>
<th>Phonetic symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vowels</td>
<td>/a/, /i/, /u/</td>
</tr>
<tr>
<td>2</td>
<td>Diphthongs</td>
<td>/aw/, /oy/, /ay/, /ey/</td>
</tr>
<tr>
<td>3</td>
<td>Nasal consonants</td>
<td>/m/, /n/</td>
</tr>
<tr>
<td>4</td>
<td>Voiced consonants</td>
<td>/b/, /d/, /g/</td>
</tr>
<tr>
<td>5</td>
<td>Unvoiced consonants</td>
<td>/p/, /t/, /k/</td>
</tr>
</tbody>
</table>

2.5 Speech data

The speech data used for the current study is obtained from TIMIT database which consists of phonemes, sampled at 16 KHz. In a total of 48 standard phonemes, 15 phonemes as listed in Table 2, are used for the analysis.

3. EXPERIMENT

The function of signal processor in CI devices primarily consists of filter bank model that divides an input speech signal into a number of frequency bands in order to extract signal strength in each band for exciting the implanted electrodes accordingly. As per the frequency bands shown in Table 1, the input speech signals are digitally filtered using critical band FIR filter based on the specifications discussed in section 2.2. MATLAB code is written to generate filter coefficients by using Kaiser Window. We adopt the linear-phase FIR filter and use its symmetrical property to reduce the number of multiplications by half [24]. Figure 2 depicts the magnitude response of 16 channels of the filter bank; each channel is an 877th-order FIR filter sampled at 16KHz. The stop-band attenuation obtained is approximately –60dB for a $\beta$ value of 6.

A CI acoustic simulation model is constructed and simulation experiment is conducted to investigate its contribution for speech perception. All the test materials are analyzed for its dominant energy levels between 125Hz and 3655 Hz. Each utterance of speech signal for the vowels /a/, /i/, /u/, diphthongs /ay/, /oy/, /aw/, /ey/, voiced /b/, /d/, /g/, unvoiced stop consonants /p/, /t/, /k/, and nasal consonants /m/, /n/, are passed through the 16 independent channels. Spectral energies are estimated from each band and the amplitude levels are estimated through envelope detection.
RESULTS AND DISCUSSION

Using the acoustic model discussed in section 2.3, it is proposed that a reduced number of channels is selected for stimulation based on the speech energy obtained in different channels; in order to avoid inter-channel interference among the cochlear electrodes. The experiment is performed for all the phonemes listed in section 2.5. To select the appropriate stimulation electrodes, the spectral energies of the model output are estimated for each of 16 channels, where each phoneme is analyzed for 10 utterances. The following sections discuss the results obtained through spectral and temporal analysis based on the CI acoustic simulation model.

4.1 Vowels

Here an illustration is discussed for selecting the stimulation electrodes based on the channel energies. Typical energy levels of vowel ‘a’ for ten utterances are represented in Table 3. The sixteen rows represent the sixteen channels’ energy and columns U1 through U10 represent utterance of each subject. The energy values are normalized across sixteen channels for each utterance. It is inferred from Table 3 that, nine utterances show maximum energy (energy > 0.9) in five and six channels. This shows that maximum speech information is present in the frequency range of channels 5 and 6 for vowel /a/. If we form a criterion that only channels having normalized energy level > 0.9 are used for stimulation, then for vowel ‘a’, only electrodes 5 and 6 are to be stimulated. This corresponds to a frequency range of 530Hz-605Hz and 655Hz-745Hz. All other channel outputs that possess less than 0.9 of the normalized energy value may not contain much of speech information and these corresponding electrodes need not be stimulated.

The output speech spectra of 16 channel BPF for vowel /a/ for utterance 1 are illustrated in Figure 3a. Two dominant peaks are observed at center frequencies 570Hz and 700Hz, which correspond to channels 5 and 6. The experiment is performed for the remaining vowels /i/ and /u/. The front vowel /i/ showed > 0.9 of the normalized energy in channels 13 and 14, which correspond to the center frequencies 2150Hz and 2500Hz. For back vowel /u/, channels 4 and 11 which correspond to 450Hz and 1600Hz were identified. These significant channels only would be used for the excitation of electrodes.
Table 3. Maximum energy levels corresponding to the utterances of vowel /a/: Distribution of speech energy in 16 channels for ten utterances.

<table>
<thead>
<tr>
<th>Channel</th>
<th>U1</th>
<th>U2</th>
<th>U3</th>
<th>U4</th>
<th>U5</th>
<th>U6</th>
<th>U7</th>
<th>U8</th>
<th>U9</th>
<th>U10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>1.3</td>
<td>0.2</td>
<td>0.2</td>
<td>1.7</td>
<td>0.6</td>
<td>1.2</td>
<td>0.6</td>
<td>0.1</td>
<td>4.4</td>
</tr>
<tr>
<td>2</td>
<td>1.1</td>
<td>2.5</td>
<td>0.2</td>
<td>1.1</td>
<td>4.9</td>
<td>2.9</td>
<td>1.4</td>
<td>2.1</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.9</td>
<td>0.4</td>
<td>1.2</td>
<td>1.5</td>
<td>1.9</td>
<td>2.4</td>
<td>0.2</td>
<td>7.2</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.5</td>
<td>0.6</td>
<td>0.2</td>
<td>1.1</td>
<td>4.9</td>
<td>2.9</td>
<td>1.4</td>
<td>2.1</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10.2</td>
<td>3.1</td>
<td>7.6</td>
<td>29</td>
<td>6.9</td>
<td>9.9</td>
<td>24.8</td>
<td>20.6</td>
<td>16.5</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>12.2</td>
<td>5.9</td>
<td>7.5</td>
<td>17.3</td>
<td>9.4</td>
<td>14.2</td>
<td>5.6</td>
<td>42.5</td>
<td>25.1</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>5.1</td>
<td>3.1</td>
<td>2.4</td>
<td>4.1</td>
<td>8.8</td>
<td>13.9</td>
<td>2.6</td>
<td>9.1</td>
<td>26.1</td>
<td>5.5</td>
</tr>
<tr>
<td>8</td>
<td>3.5</td>
<td>3.4</td>
<td>3.2</td>
<td>5.2</td>
<td>13.1</td>
<td>4.5</td>
<td>1.4</td>
<td>9.0</td>
<td>20.9</td>
<td>6.0</td>
</tr>
<tr>
<td>9</td>
<td>6.4</td>
<td>1.4</td>
<td>8.8</td>
<td>18.3</td>
<td>12.9</td>
<td>1.8</td>
<td>2.9</td>
<td>32.8</td>
<td>4.6</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>4.9</td>
<td>0.2</td>
<td>8.3</td>
<td>5.7</td>
<td>1.2</td>
<td>12.9</td>
<td>25.6</td>
<td>2.6</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2.1</td>
<td>0.1</td>
<td>2.5</td>
<td>3.2</td>
<td>4.9</td>
<td>1.1</td>
<td>6.6</td>
<td>3.3</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2.1</td>
<td>0.4</td>
<td>2.7</td>
<td>6.8</td>
<td>7.2</td>
<td>3.6</td>
<td>8.6</td>
<td>3.3</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4.3</td>
<td>1.1</td>
<td>6.7</td>
<td>1.7</td>
<td>5.5</td>
<td>1.8</td>
<td>5.1</td>
<td>1.3</td>
<td>10.2</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2.2</td>
<td>0.6</td>
<td>2.6</td>
<td>1.1</td>
<td>5.2</td>
<td>1.4</td>
<td>2.5</td>
<td>5.8</td>
<td>1.8</td>
<td>3.9</td>
</tr>
<tr>
<td>15</td>
<td>0.9</td>
<td>1.4</td>
<td>4.7</td>
<td>3.4</td>
<td>5.4</td>
<td>3.1</td>
<td>9.2</td>
<td>10.9</td>
<td>1.5</td>
<td>8.1</td>
</tr>
</tbody>
</table>

4.2 Diphthongs

The experiment is performed for four diphthongs and the observations are made as follows. For /aw/ and /oy/, two channels 5 and 7 are found to have normalized energy level > 0.9. These channels only need to be stimulated for diphthong /aw/ and /oy/. The output speech spectra for /aw/ are illustrated in Figure 3b which shows only two peaks at center frequencies 570Hz and 840Hz that correspond to channels 5 and 7 respectively. Similarly, for /ay/, the normalized energy level > 0.9 is observed in channel 6 corresponding to a center frequency of 700Hz. For /ey/, it is channel 5, which correspond to a center frequency of 570Hz.

4.3 Nasal consonants

For nasal consonants /m/ and /n/, channel 2 is found to have normalized energy level > 0.9. Therefore channel 2 needs to be stimulated for nasal consonants /m/ and /n/. The BPF output speech spectra of nasal consonant /m/ are shown in Figure 3c. A peak is observed in channel 2 which correspond to a frequency range of 225Hz-275Hz.

4.4 Voiced and unvoiced stop consonants

The energy levels vary greatly with the frequency content in both voiced and unvoiced stop consonants. The experimental procedure is repeated for voiced (/b/, /d/, /g/) and unvoiced (/k/, /t/, /p/) stop consonants. The results of the significant energy levels are indicated across channels 5, 11 and 16 for /b/, channel 16 for /d/ and channels 10, 11 for /g/. For /k/, it is observed as channels 10 and 11; for /t/, it is channels 15 and 16 and for /p/, it is channels 10 and 14. The BPF output speech spectra of unvoiced stop consonant /t/ are shown in Figure 3d, which shows peaks at channels 15 and 16 that correspond to center frequencies 2900Hz and 3400Hz.
Fig. 3. Spectral analysis outputs showing the energy peaks for various phonemes (a) Two dominant peaks are observed in channels 5 and 6 for vowel /a/, for utterance 1 (b) Diphthong /aw/ exhibiting channels 5 and 7 (c) A peak in channel 2 for nasal consonant /m/ (d) A peak in channel 15 for unvoiced stop consonant /t/.

From the CI acoustic simulation model, it is evident that, low frequencies are significant for vowels and high frequencies are significant for consonants. The events of the consonants are consistent across different subjects, despite that the parameters such as frequency and strength may change to a certain degree. The investigations further demonstrated that, a maximum of three channels are only required for the stimulation of cochlear electrodes for each of the phonemes listed in Table 2. Hence it is inferred that the spectral energies in all 16 channels are not required to be utilized for stimulation of cochlear electrodes. By selecting a few channels for stimulation, our proposed model would help in reducing inter channel interactions among the electrodes. The computation of spectral energies gives a measure of the number of electrodes required for stimulation. The identified channels through spectral analysis are subjected to temporal analysis, in order to get the amplitude information.

4.5 Electrical stimulation procedure

The electrodes in the cochlear array are stimulated by means of biphasic pulses [29]. A biphasic pulse of 15µs width with a repetition rate of 250 µs which produces a stimulation rate of 4000 pulses per second are generated. The ‘place’ of stimulation in the cochlear electrode array is obtained by band energy levels in various channels estimated from the spectral analysis using the CI acoustic simulation model as discussed in section 2.2. The selected channels whose normalized energy >0.9 are subjected to temporal analysis. The amplitudes of the pulses are derived by extracting the envelopes of band passed filter outputs which are obtained by full-wave rectification and low-pass filtering as discussed in section 2.3. The amplitude estimates are useful in fixing the amplitude of the pulses. i.e, the pulse amplitudes are made proportional to the amplitude of the envelope output. Trains of biphasic pulses thus generated are to be delivered to the cochlear electrodes. In our study, the pulses are successfully generated for all the phonemes listed in Table 2, with amplitudes proportional to the outputs of selected channels.

An illustration for vowel /a/ depicting the various stages of temporal outputs in channel 6 (having normalized energy >0.9) is shown in Figure 4(a-d), which correspond to the input
speech signal, band pass filtered output, rectified and low-pass filtered output, and the extracted envelope output. From this extracted envelope output, the stimulus pulse amplitude estimates are derived. Figure 4e shows the stimulation pulse energy levels for vowel /a/ in channel 6.

Fig. 4. Various stages of temporal outputs in channel 6 for vowel /a/.
(a) Input speech signal of vowel /a/ (b) Band pass filtered output in channel 6 (c) Illustration of the rectified and low-pass filtered outputs (d) the envelope tracing of the low pass filter (e) Amplitude level of biphasic stimulation pulses (not expanded).

5. CONCLUSION

In this study, the performance of cochlear implant acoustic simulation model is discussed. We have designed a 16-channel FIR filter bank based on critical band spacing, using Kaiser Window. Higher stop band attenuation and exhibiting linear phase property are the two advantages of the critical band FIR filter design, which is utilized in the proposed CI-acoustic simulation model. This model estimates the spectral output using the designed CB-FIR filter bank and temporal output by envelope detection method through which the ‘place’ and ‘stimulus level’ information are obtained respectively. The spectral band energies of 15 speech components from the TIMIT data base are analyzed and the most significant channels are identified and selected, by computing the maximum energy values for the phonemes taken for the study. The selected channels are subjected to temporal analysis and the amplitudes are extracted by the envelope detection procedure. The simulation results of the selected channels
for different phonemes, demonstrated a good discrimination between vowels, diphthongs, voiced and unvoiced consonants. The other advantage of the proposed model is, it yields a way for minimizing the channel interactions as a reduced number of electrodes only are selected for stimulation. The overall results from the acoustic simulation experiment showed that the CB-FIR filter bank model based speech processor is useful in selecting the channels and estimating the required stimulation pulse level to be generated and presented to the cochlear electrodes.

REFERENCES

Programmable Analog Bionic Ear Processor”, IEEE transactions on biomedical engineering, 52(4): 711-727.


