A REVIEW OF COCHLEAR IMPLANT AND SIGNAL PROCESSING STRATEGIES

Maryam Parhizkar, Student ID: 300215734

Review Based Project

Abstract: Hearing is particularly an important factor for communication in human. In the ear, the acoustic signal moves from middle ear to the inner ear, where the acoustic pressure waves create vibration in hair cells [13]. The location and intensity of these vibrations are transmitted to the brain. In case the large number of hair cells are damaged, the acoustic waves cannot be converted to neural impulses by the auditory sensory system. Such condition, shapes the profound hearing impairment [1], [2]. From nearly forty years, Cochlear Implants (CI) have been designed to overcome this hearing problem and they have successfully restored hearing sensation to patients with severe to profound hearing loss. Generally, CIs consist of external parts (microphone, speech processor and a transmitter) and the internal part which is an electrode array implanted in the cochlea [1], [2], [3], [9]. The speech processing strategy which is applied by means of increasing speech perception and intelligibility, plays an important role in cochlear implant platform. There are various prosthesis models available in the market each of which benefits from different signal processing methods.

The intention of this study is to present a review of Cochlear Implant and different signal processing techniques which are being used in CIs. Some of the main techniques that mostly have taken into considerations, are Continuous Interleaved Sampling (CIS), N-of-M and feature extraction by means of speech perception enhancement approaches. On the other hand, noisy listening conditions have remained challenging for most users. During this review, we are looking at some articles proposing solutions to tackle this issue by estimating signal-to-noise ratio or extracting special features of the speech signal. The implementation of CIS Strategy, as a relevant approach for speech processing, will be presented in MATLAB with the input of a recorded audio speech and then the output signal is compared to the input.

1. Introduction

Hearing plays a key role in communication of human. Hearing function is based on transforming sound vibrations into nerve impulses that can be interpreted as sounds in the brain.

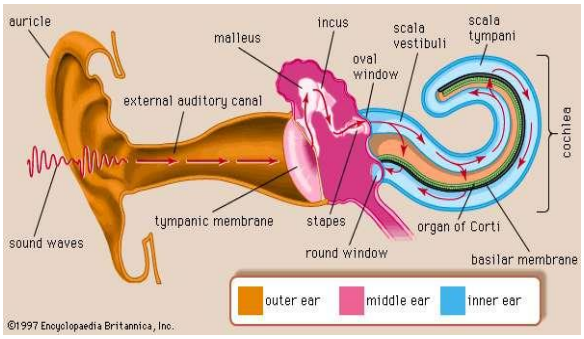

Fig.1: Physiology of Hearing, Joseph E. Hawkins Emeritus

Passing from outer ear which is auricle, the amount of sound waves is enhanced and they reach to tympanic membrane. This canal works as a resonant filter which enhances short wavelength sounds and their transmission efficiency in the frequency range of 2-7 KHz. This range is the one to which the ear is most sensitive and are important for distinguishing the sounds of consonants [2], [19], [13].

The middle ear is an air-filled space that consists of three small bones: malleus, Incus and Stapes that are responsible to conduct the sound vibrations from tympanic membrane to the inner ear [19].

The inner ear is responsible for both balance; by the vestibule or labyrinth contribution, and hearing by snail-shaped structure of cochlea. The stiff structural element within the cochlea is basilar membrane that play an important role in distributing sound energy by frequency along the cochlea. The specific areas in the basilar membrane moves variably in response to the different frequencies of signal and due to cochlea's decreased stiffness, the higher frequencies are distributed in the base and lower frequencies are mapped to apex [2], [5], [19], [13].

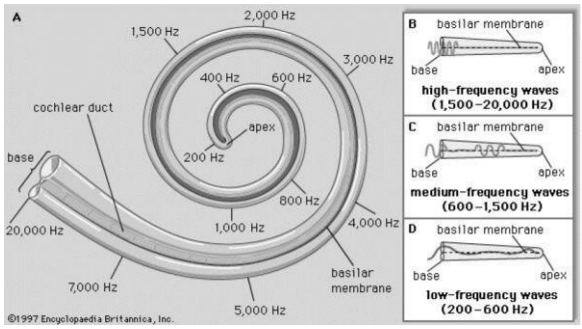

Fig.2: Tuning of basilar membrane, [13]

The process of producing electric spikes being sent to CNS, is the responsibility of the sensory cells named hair cells that are adjacent to the basilar membrane.

Since several mechanisms contribute in hearing process, there is a potential for damage in several points in the ear. While the large number of hair cells are damaged, the inner ear is not able to process the sound as coming in. In fact, although hearing nerves are not damaged, there is no hair cell to communicate with the nerve and therefore, the acoustic waves cannot be converted to neural impulses by the auditory sensory system. This means that frequency resolution cannot be restored. Indeed, damages to cochlear results in lower sensitivity. Generally, the damages to either cochlea or hair cells are termed sensorineural loss which cannot be corrected by hearing aids and direct electrical stimulation of auditory nerves is needed and it is the basis of cochlear implants.

The history of Cochlear Implant goes back to the seventeenth century when the Italian scientist Alessandro Volta (1745-1827) placed the two ends of battery to his ears and felt the sound like crackling and boiling. This was the first time that human figured out that electric stimulation can induce auditory and visual sensation [3].

The first efforts to create cochlear implant was conducted by a French physician named Djourno reported hearing sensation for partially deafened patients by placing electrodes on the trunk of auditory nerve.

In 1970, William House introduced single electrode cochlear implant with low frequency square wave (40-200 Hz) as the carrier of stimulation and the amplitude was modulated by the sound. From this time, cochlear implants have been introduced in different platforms by means of enhancing speech perception by the patients.

In general, cochlear implants are made up of the external and internal units. The external unit or Behind-The-Ear (BTE) consists of a microphone worn at ear to capture the incoming sound and the signal is sampled at 16 KHz so they can be processed by sound processor and a wireless transmitter to transfer the processed signal to the internal unit which has a wireless receiver and electrode array implanted in the cochlea [2], [3], [5]. Stimuli delivered to the electrode by sound processor, preferentially excite the nerve fibers nearby [5]. The speech processor plays an important role in cochlear implant platform and development of techniques to derive electrical stimulation [5]. In fact, the speech processing strategies are classified according to the way they convert sound input to electrode stimulation pattern.

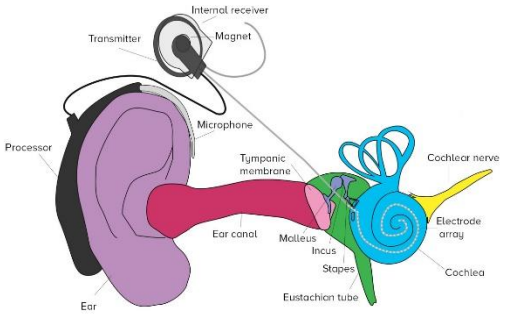

Fig.3: Cochlear Implant Structure

In the following chapters, various prosthesis models have been developed through the years, are introduced each of which applies different signal processing methods.

2. Single Channel Implants

The model that was firstly introduced by William House in the early 1970s, has been modified by 3M company over the years. In figure 3, the block diagram of the House/3M device is shown. According to this figure, the speech signal is amplified after collecting by the microphone and then it is processed through a 340-2700 Hz bandpass filter. The filtered signal goes through an amplitude modulator with 16 KHz carrier and finally, after being amplified, it is applied to the external transmitter. On the other side, the internal receiver transmits the stimulation to the active electrode.

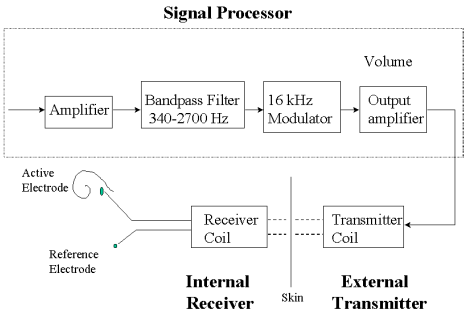

Fig.4: Block diagram of the House/3M device, [2]

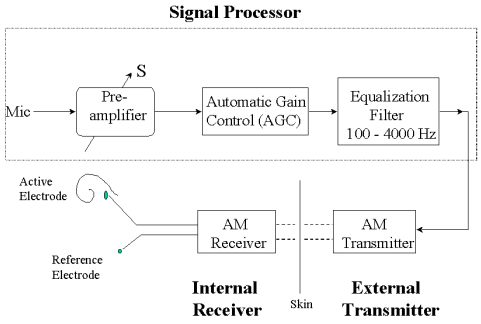

Fig.5: Block diagram of the Vienna/3M implant, [2]

According to this method, what is being applied to the electrode is modulated signal rather than the speech signal. Since it is Amplitude Modulated, so any fluctuation relating to the signal, is preserved. However, since the processor does not limit the input dynamic range, the shape of the signal is affected by the input signal. This means that for sound pressures that are more than a specific range like 80 dB, the envelope output saturates at a level just below the patient's level of discomfort. This issue results in the output be clipped and the temporal information of the signal is discarded and speech recognition is remarkably low.

To overcome this shortage, another method was proposed at the Technical University of Vienna, Austria, in the early 1980s. According to the block diagram of the Vienna/3M implant, the signal is first pre-amplified and then it goes through an Automatic Gain Control which is adjusted by the patient's dynamic range, so the temporal information in the signal are preserved. In the next stage, the compressed signal goes through a frequency-equalization filter which is a band pass filter with the range of 100-4000 Hz. This range contains the frequencies which are important for speech understanding. Unlike the previous method, the amplitude modulated signal will be demodulated by the implanted receiver and stimulate the electrode. This strategy has shown a better speech perception and more accurate word identification in a sentence.

2.1 Single Channel Implant shortages

Although single channel implants are able to convey temporal information and even some frequency information, few patients were capable of understanding speech with the limited spectral information.

Since there is only one electrode implanted in cochlea, the frequency encoding cannot be achieved like cochlea frequency distribution. On the other hand, considering the refractory period of the nerve fiber, the temporal frequency encoding is limited to 1 KHz which is not sufficient for speech perception [1]-[3].

3. Multi-channel Cochlear Implants

The mentioned limitations of single channel implants, made the researchers switch to a new approach in terms of electrode numbers implanted in cochlea. In this modification, an array of electrodes is implanted in the cochlea instead of only one electrode. This approach provides more ability to stimulate different places in cochlea and accordingly, more information can be transmitted to CNS by nerve fibers [1]-[3].

At this stage, various signal processing strategies have been introduced that can be divided into two main categories: waveform strategies and feature-extraction strategies. The main difference between the categories is in the information being extracted from the speech signal and interpreted by brain.

4. Waveform Strategies

4.1 Compressed-Analog (CA) approach

The Compressed-Analog method was firstly used by Symbion, Inc., Utah and now it in no more applied to cochlear implants. In this platform, the

electrodes have 4mm distance one to another. The algorithm consists of an automatic gain control to compress the signal and then the signal goes through a filter bank and divides to four channels with center frequencies at 0.5, 1, 2, and 3.4 kHz. Each channel is assigned to an adjustable gain control and then is sent directly to the relevant electrode [1], [2], [9].

Compared to the single-channel implants, this device was more successful in providing patients with speech understanding.

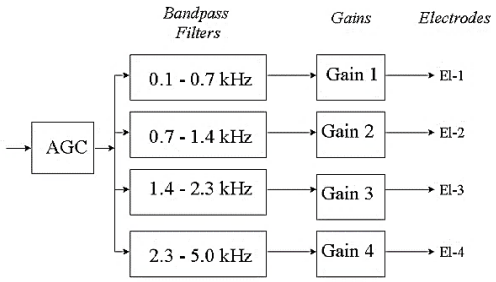

Fig.6: Block diagram of the Compressed-Analog method [1]

4.2 Continuous Interleaved Sampling (CIS) strategy

Compressed-Analog approach was suffering from interaction between channels due to simultaneous stimulation of four electrodes. The main concern is that the interactions can distort speech spectrum information and decrease speech recognition by the users.

CIS technique was the next generation of waveform strategies, firstly proposed by Wilson B.S in1991 and mainly introduced to tackle the channel interactions in CA method by applying asynchronous and alternate pulses [1], [2], [3]. In fact, the CIS strategy preserves the energy of the signal and trains of non-overlapping biphasic pulses are delivered to the electrodes in a way that only one electrode is activated at a time [3], [9], [11].

The signal; which is picked up by the microphone, is firstly sampled at 16 kHz (for speech intelligibility, frequency range of 2-6 kHz in sufficient. So, according to Nyquist, 16 kHz is well enough for sampling rate). After that, the signal is pre-emphasized using a first order Butterworth high-pass filter with cut-off frequency of 1200 Hz [4]. However, there are many different filters have been tried for this strategy including Chebyshev I or Chebyshev II and so on. This pre-emphasis filter is applied to reduce low-frequency energy before the signal is encoded and of course, the signal to noise ratio can become better. In the next stage, the signal needs to be divided into some channels. The number of channels; which is equivalent with the number of implanted electrodes, changes from 8 to 16. For this purpose, a group of bandpass filters are applied to the signal. Different filter banks have been used in articles such as Chebyshev I, II or Butterworth filter. To extract the temporal envelope, the signal goes through a full wave rectification and a low pass filter with cut-off frequency between 200 to 500 Hz. The result would be a line that traces over the positive peaks of the waveform. The envelope is then modulated using a biphasic pulse train generator to produce the desired output for each electrode [3], [4], [11].

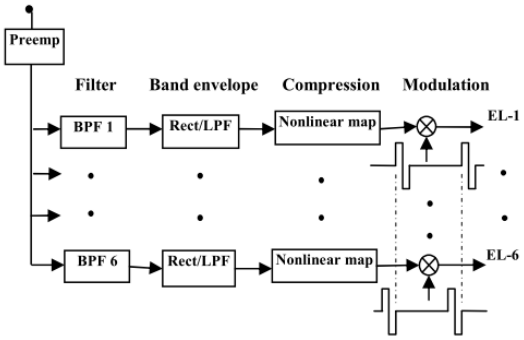

Fig.7: Block diagram of CIS strategy, [11]

4.2.1 Stimulation Parameters

a. Pulse rate

Pulse rate is the number of pulses per second which is delivered to each electrode. This rate changes for patients depending on their sensitivity and the range varies between 100 to 2500 [2].

b. Stimulation Sequence

The stimulation sequence is defined as the order of electrode activations in a way that the minimum interaction happens between electrodes. This issue has a direct effect on speech recognition by the user. In CIS approach, this order makes the maximum distance between consecutive electrodes. However,

the order of electrode stimulation can be varied by the patients [2].

c. Compression Function

Since the acoustic signal amplitude is remarkably higher than the dynamic range of the patient, there must be a compression function for acoustic-to-electric mapping of speech sounds. There have been many studies on optimal speech recognition based on different methods of mapping input dynamic range. Fan-Gang Zeng, propose a logarithmic map for low-frequency channels and a more compressive map for high-frequency channels to improve overall speech recognition for cochlear-implant users [3]. Generally, in CIS strategy the logarithmic compression function of the form $Y = A \log(X) + B$ is applied to the rectified output of X, where A and B are constants.

Another function which is being used, is the power-law compression function as $Y = AX^p + B$ ($p < 1$), $A=(MCL-THR)/(X^p_{max} - X^p_{min})$ and $B = THR – AX^p_{min}$. the constants A and B are chosen in a way that $[X^p_{min} \quad X^p_{max}]$, as the input acoustic range falls into the electrical dynamic range [THR, MCL], where THR is the threshold level at which the patient can just hear the stimulus and MCL is the maximum comfortable level, that is, the level which produces a loud but comfortable sensation [1], [2].

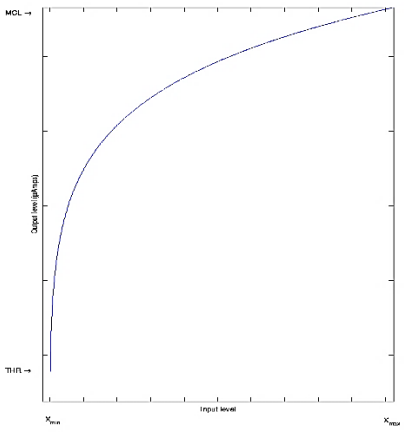

Fig.8: Example of logarithmic compression map for CIS strategy, [2]

5. Strategies based on feature extraction

Previous strategies are based on waveform information obtained from the original signal. However, there are other strategies which are focusing on spectral features such as formant frequencies, spectral maxima and so on. In the following sections, we are providing an overview of some feature extraction strategies.

5.1 F0/F2 and F0/F1/F2 Strategies

The resonant frequencies of the vocal tract are introduced as the formants. Their energy around a particular frequency can be seen in speech spectrogram. There are several formants which mostly occur at 1000 Hz intervals. the lowest frequency of a periodic waveform is known as fundamental frequency or F0. F0, is crucial for sound discrimination, music perception and detecting pitch; which shows how deep or shrill the sound is according to the frequencies [1]-[3], [9].

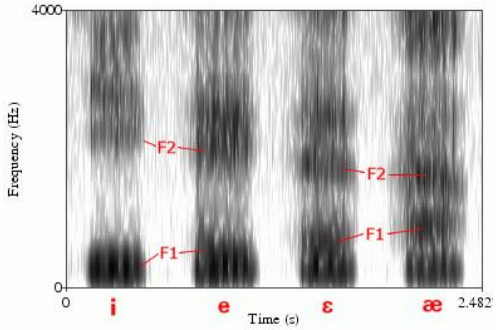

Fig.9: Formants change frequency for vowels in sound spectrogram

The F0/F2 strategy was first developed by Nucleus device in 1980. In this strategy the zero-crossing detector is used to extract F0, as the fundamental frequency and F2, as the second formant from speech signal. For F0, a lowpass filter with cut-off frequency of 270 Hz is applied to the signal and F0 is estimated using zero-crossing detector. By the same method and another zero-crossing detector, F2 will be estimated from the output of a bandpass filter with cut-off frequencies of 1000 and 4000 Hz. The output is then rectified and goes through a lowpass filter to estimate F2 amplitude. The number of electrodes implanted in cochlea is 22 and F0/F2 processor, activates the relevant electrodes for F2 frequency information and information in terms of voicing is transmitted according to a rate of F0 pulses per second [1], [2].

Later, this strategy was modified to include F1 information in the output signal by adding one more zero-crossing detector to estimate F1 from a bandpass filter output with cut-off frequencies of 280-1000 Hz [1]. In this approach, signal is represented by F1 and F2. Since F1 is up to 1000 Hz; among 22 electrodes, first 5 electrodes from apex are dedicated to F1. Meanwhile, F2 represent frequencies above 1000 Hz and accordingly the remained 15 electrodes are stimulated with F2. 2

electrical pulses are produced for voiced segments according to F1 and F2. This modification, improved speech recognition of patients using Nucleus CI. [20]

5.2 MPEAK Strategy

The next generation of cochlear implants was modified by adding three additional frequency bands to F0/F1/F2 strategy. This approach could improve the F2 participation in output signal as well as including higher frequencies which play key role in consonant perception. It should be noted that vowel frequency band ranges from 250 to 2000KHz and voiced consonants can be presented by 250 to 4000 KHz. Unvoiced consonants lie between 2000 to 8000 KHz. On this base, four electrodes are allocated to the rate of F0 pulses per second for voiced sounds. Also, F1 and F2 electrodes and high-frequency electrodes for 4 and 7 are assigned to voiced sounds. As in the spectrum above 4KHz, there is a little energy for voiced sound, the electrode 1 is not active. However, for unvoiced sounds, higher frequencies dedicated to electrodes 1,4 and 7 as well as F2 electrode, are activated. So, by adding three bands of higher frequencies, better understanding of consonants is expected. Wallenberger and Battmer showed that there is an increase of 28% on open-set sentence recognition. [1], [2], [9].

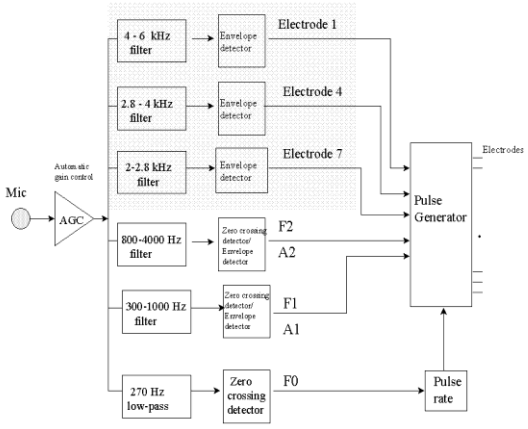

Fig.10: Block diagram of MPEAK strategy [9]

5.3 N-of-M Strategies

This coding strategy have been developed the 1990s by the purpose of increasing temporal resolution. Such strategies divide the original signal into M bands and after extracting the envelope information for every frequency bands, the N bands with largest amplitudes are selected for electrode stimulation. This approach neglects the less important spectral components which results in higher user performance [1], [2], [3], [5].

5.4 Advanced Combination Encoders (ACE)

The ACE strategy is based on the principle of N-of-M which is the speech processor designed for Nucleus 22 by Cochlear Corporation. This device is using an array of 22 electrodes [5], [8]. The first stage of digitalizing with sampling rate of 16 KHz and pre-emphasizing, is similar to CIS strategy. In ACE, CI processing is based on frequency analysis and envelope extraction. Signal is buffered using blackman or hann windowing methods into 8 ms frames. The 128-point-FFT of each frame is calculated which creates 128 samples. As we are applying the FFT to a real signal, the result would be a symmetric complex signal with absolute and imaginary parts. Therefore, we discard half of the FFT data and we keep 64 bins with frequency resolution of 125 Hz. The number of bins exceed the number of electrodes. Accordingly, we need to pass the bins through 22 weighting filters. So, we will have 22 frequency bands. These bands are spaced linearly for frequencies up to 1000 Hz and logarithmically spaced for above 1000 Hz.

Band number z	1	2	3	4	5	6	7	8	9	10	11
Number of bins	1	1	1	1	1	1	1	1	1	2	2
Center freqs. (Hz)	250	375	500	625	750	875	1000	1125	1250	1437	1687
Gains g_z	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.68	0.68

Band number z	12	13	14	15	16	17	18	19	20	21	22
Number of bins	2	2	3	3	4	4	5	5	6	7	8
Center freqs. (Hz)	1937	2187	2500	2875	3312	3812	4375	5000	5687	6500	7437
Gains g_z	0.68	0.68	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65

Fig.11: Number of FFT bins, Centre frequencies and gains per filter in ACE with 22 channels

So, for each channel the envelope is computed by taking the absolute value of each channel and 8-12 channels with highest amplitudes are mapped to the patient's electrical dynamic range and stimulates the corresponding electrodes. For each frame of audio signal, 8-12 electrodes are stimulated sequentially which creates one cycle [3], [8].

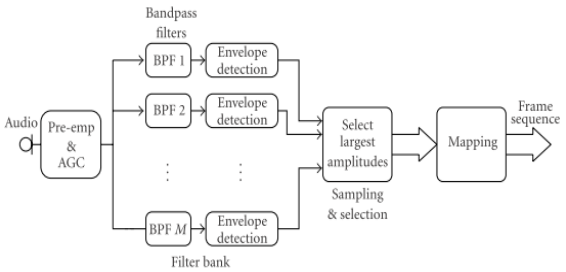

Fig.12: ACE strategy block diagram

6. Factors Effecting Cochlear Implant Performance

6.1 Stimulation Rate

The rate of stimulation is defined on each electrode as the number of cycles per second. In fact, this factor can determine the temporal resolution of the simulated signal. Temporal resolution is the ability of the system to detect amplitude changes over time which means that the more stimulation rate can provide us with more information being extracted from the original signal [10], [17]. In other word, the ability to detect the change in stimulus over time can show the temporal resolution of the signal. As an example, the figure below, compares the temporal resolution in processing the syllable /ti/ in channel 5 and shows the ability of the processor to build the original signal by higher stimulation rate [6], [17].

In view of signal processing, it is expected that increasing the stimulation rate ends up with a better speech perception, however in practice, it is not always the best idea to choose the maximum possible stimulation rate in that significant improve in speech recognition cannot be necessarily achieved by increasing number of pulses per second [6].

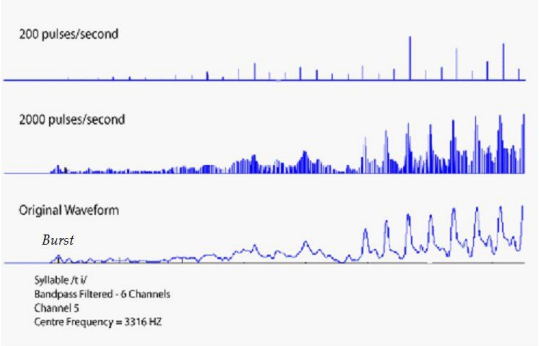

Fig.13: Simulated waveform for the syllable /ti/ in channel 5 of a cochlear implant with an array of 6 electrodes with two stimulation rates of 200 pps and 2000 pps, [17]

Robert V. Shannon 2011, [6] has shown that except for a small difference for vowel recognition in quiet, there were no significant differences in performance among the experimental stimulation rates for any of the speech measures.

Meanwhile, some animal studies to examine neural responses to electrical stimulation concluded that the rates of stimulation higher than 800 pps per channel lead to poor phase locking [3]. Phase locking is when neurons only fire preferred phases of the sound in each cycle which is usually in the peak amplitudes of low frequencies vowel. It is actually related to the coding of the temporal pattern of the sound which is important to discriminate two sounds or frequencies. At low number of pulses per second, the action potential can occur for every cycle, however if we increase the number of pulses remarkably; like 1 to 4 kHz, neurons cannot fire for every cycle because this firing rate is limited by the refractory period of action potentials [17]. Refractory period is the time duration which a cell is not capable of repeating an action potential. Accordingly, increasing the number of pulses per second can not bring extra information to the patient.

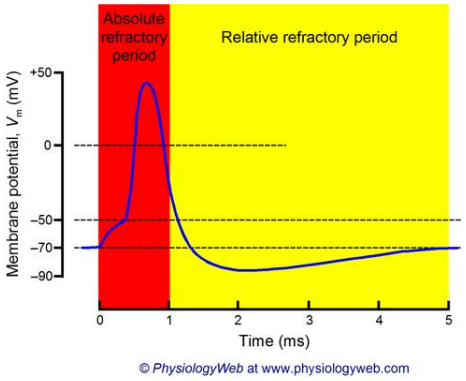

Fig.14: Phases of action-potential through a nerve fiber

6.2 Number of Electrodes and Spectral Resolution

Another effective term to achieve high speech understanding is determined by the number of channels or electrodes implanted through the cochlea. So, by this definition, we can say that the number of electrodes determines the spectral resolution of the whole processor and shows how close we can put the electrodes to simulate the basilar membrane performance [10].

In sound coding, the pitch discrimination is highly dependent on the distribution of electrodes in the cochlea. In cochlear implant, we are limited by the electrode array and we know that with really few bands, comparing to the frequency bands in human cochlea, we are unable to achieve the original signal with the whole context [5]. Just like temporal resolution, experiments suggest that no matter how many electrodes are implanted in the cochlea, the patients do not use the full spectral information and only four to ten channels are being used even by the implant listeners with the best speech recognition for each cycle [10]. One hypothesis justifies this issue due to the electrode interactions [6], [10] which may have negative impact on both temporal and spectral resolution [18].

Anyhow, in some certain strategies; like CIS, there is pulse train that are interleaved in that only

one electrode is active in time which tries to prevent the decrease in spectral resolution. One the other hand, CIS strategy has considered a constant interval between pulses while natural hearing does not follow this pattern for temporal resolution.

6.3 Location of the Electrodes

In general, electrode arrays are inserted into the cochlea in a way that there is a typical distance of 22-30 mm between the electrodes. In fact, the electrode array is not fully inserted into the cochlea which obviously creates a mismatch between the analysis and stimulation frequencies [3]. According to the anatomy of the cochlea, having access to the lower frequencies would be possible only if we can insert the electrodes in a high enough depth of the cochlea. The frequency components of the low frequencies between 200 to 1200 Hz are located in the depth of 540° in the cochlea. However, the current available cochlear implants are not able to insert the electrodes deeper than 400° due to the possible intracochlear damages [3]. This damage occurs when the force from the electrode gets so high that the tissue cannot resist and may result in tearing of the basilar membrane.

From this view point, the cochlear implants are still under development to achieve better design with higher functional channels and deeper insertion with safety improvements.

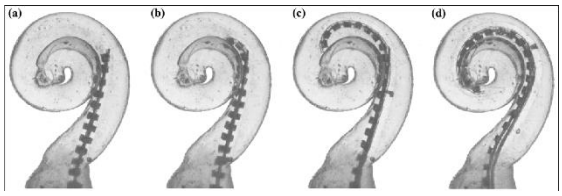

Fig.15: Due to possible intracochlear damages, there is a mismatch between the analysis and stimulation frequencies

6.4 Availability of Noise

Although there have been significant improvements in cochlear implant signal processing strategies, and while the users have quite good speech recognition in quiet listening conditions, they complain of difficulty to understand speech in noise. In fact, speech processors are mostly tested in quiet and the results are evaluated [3, [8]]. However, daily speech perception involves various situations that there are multiple sounds at the same time like restaurants or a simple party where the noise is dynamically changing. Such conditions seem so stressful for the users and demands lip reading to communicate. This issue has become a major problem for all introduced algorithms so far.

There are many reasons for the loss of original signal and speech perception in noisy environments. One reason is described as the lack of spectro-temporal details in the processed speech. For example, strategies based on N-of-M principle, basically focus on the energy of the signal and they only select the highest amplitudes of the spectrum which is not always the main components of the speech signal. When the user is in a noisy condition; such as a party or even when more than one person is talking, the processor fails to have a good performance, because it may select some high peaks of the noise as the target [8].

To tackle this problem, there have been different studies through the years. One of these studies tries to increase the performance of the users by modifying the way the processor selects the channels in ACE strategy and in the next section, this algorithm is described.

6.4.1 Improving channel selection in ACE strategy in presence of noise background

In an article by Ali et al, 2014, [8], a new method is proposed, to modify ACE strategy by giving weights to each time-frequency unit based on the formant location of speech and signal to noise ratio and also by giving attenuation factor to low signal-to-noise ratios.

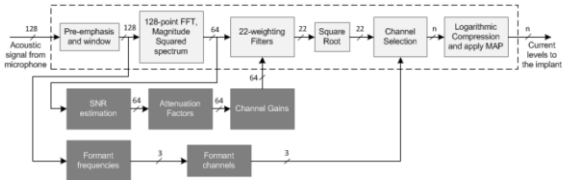

Fig.16: Block diagram of the signal processing proposed in [8]. Dotted block: Standard ACE strategy processing steps. Darker tone: Modifications proposed by the article [8]. Numbers on connecting arrows show the frame size in number of samples in each step

As mentioned before, the formant with the lowest frequency is called F1, the second is F2, and the third is F3. Most often the first two formants, F1 and F2, are sufficient to identify the vowels. In this study, the first three formants; which are F1, F2 and F3, are calculated for every cycle and the priority in channel selection, is assigned to the channels corresponding to these three formants. The way the formant extraction is being done, is by applying Linear Predictive coefficient (LPC) model which is generally used to estimate formants in the speech signal. Actually, this model uses the points of the signal to calculate the next sample. In fact, it considers the next point as a linear combination of past samples which creates the prediction

coefficients. It also calculates the error between the signal and its own prediction. Linear Prediction Coefficients can be determined by minimizing the sum of squared differences between the actual speech samples and the predicted ones [12].

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + e[n]$$

Here, a_k is the Linear Predictive Coefficients, s[n-k] is the linear combination of past samples and e[n] is the error between the signal and predicted value.

Another modification is done by giving attenuation factor according to SNR. For this purpose, firstly, the SNR is estimated for each time-frequency unit: X(i,j) which is magnitude squared spectrum of the ith frame and jth frequency bin. So, we will have 64 SNR values from 64 FFT bins that we previously calculated in ACE coding.

There are two methods to estimate the SNR: priori SNR, which is the ratio between original, pure speech signal and the noise.

The other one, is estimation using Improved Minimum Controlled Recursive Average Algorithm (IMCRA) which is proposed by another article and it estimates the noise in adverse environments by averaging power values of noise in the past, when there is no speech signal and saves it to be multiplied by a constant when speech signal appears [23]. Actually, this averaging technique is being done when there is no speech signal, and then it will be stored until a new input; such as speech, is adding to the spectrum. In some studies, this saved noise will be subtracted from the signal when the speech signal comes. However, in this article, a smoothing parameter is used which is adapted according to the speech presence probability and then noise estimation updates continuously even for weak speech signals.

In both ways, we need to assign priority to the channels with higher SNR, so, we compare the results for binary and soft functions.

In binary masking, we assign zero to any SNR value which is less than 0 dB while in Soft masking, the weighting function is more flexible, since we are using a sigmoidal shape function to be multiplied with the signal. So, for any SNR>15 dB, the signal is multiplied with one or close to one and the attenuation factor of zero or close to zero, is assigned to SNR<-15 dB.

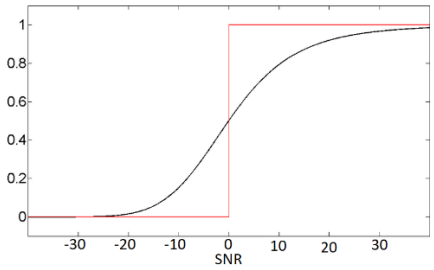

Fig.17: Binary (black) and soft masking (red) functions applied to the signal based on SNR by the proposed method [8]

3 American English speakers who use Nucleus 24 manufactured by Cochlear Corporation whose device is using the ACE strategy, were recruited, to understand the performance for each feature extracted, they have been tested separately.

So, for formant prioritizing method, Patients were asked to score their speech intelligibility and speech perception in 6 different noisy conditions. We can see that when the SNR is getting worse, the difference between standards ACE and Formant prioritizing approach is considerable and for the quality of perception, they stated better to no difference on all tests.

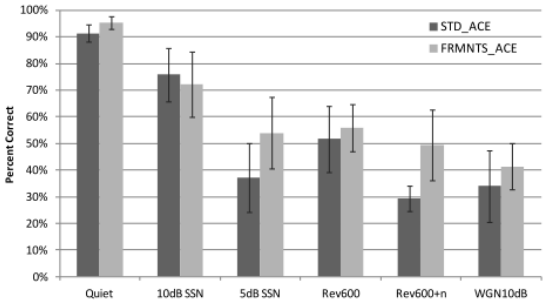

Fig.18: mean speech intelligibility score of 3 CI users of the experiment of formant prioritizing [8]

In the next experiment, binary and soft masking functions, have been compared for both noise estimation methods, again, in terms of speech intelligibility and speech perception. And according to the bar chart, we see that both binary and soft masking are able to restore the original signal even better than standard ACE in low SNR and the Ideal masking had a better performance of all. In terms of quality, the participants asserted much better quality for ideal conditions and indeed, they gave the average score of 2 for estimated SNR conditions which is showing better quality. In general, and according to the scores, soft masking had a better result than binary masking to enhance SNR.

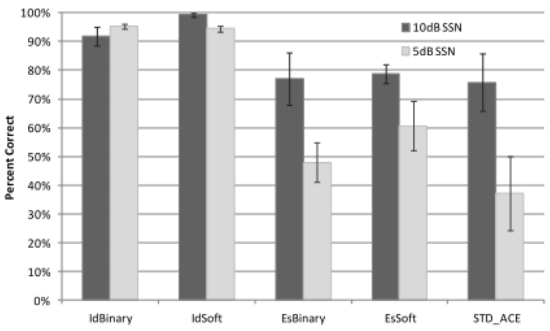

Fig.19: speech intelligibility score of 3 CI users of the experiment of giving attenuation factor according to SNR. Comparison of the proposed technique using IdBinary, IdSoft, EsBinary and EsSoft with STD_ACE. Error bars are standard error of the mean. [8]

According to the study, significant improvement can be acquired by applying this technique when the user is in a modest noisy condition. Such method shows promising result to enhance speech perception of CI users in noisy environments.

7. Comparison between different strategies

As mentioned in this review study, there have been various signal processing strategies introduced into the market over the years. It has been accepted that multi-channel platform provides the patient with better speech perception [3]. According to the Fig.20, to understand the performance of each device and strategy to communicate better in daily conversations and other conditions, the recognition scores in quiet for each device is summarized which shows that due to lack of speech recognition, single-channel approaches have been disappeared just after the multi-channel systems designed and developed. Meanwhile, it appears that recently introduced strategies have provided better speech recognition [16] and ACE strategy has shown a better performance, compared to other multi-channel approaches.

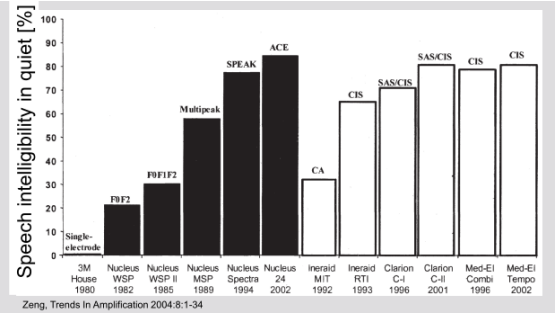

Fig.20: Sentence recognition score for various strategies development over the year [3]

However, since individuals are different in terms of number of healthy hair cells and deafness, they may have different performance for strategies. On this base, it is better to test various processors for each patient. Indeed, due to limitations some of which explained in this review, none of the signal processing strategies, can provide the users with a perfect speech understanding which shows that the best signal-processing strategy must be used after clinical testing for users [16].

8. Simulation of CIS strategy

In this section, the sound coding by CIS strategy is presented, using MATLAB. According to the block diagram and based on what we have explained in the previous sections, different parts of CIS strategy has been introduced. The cochlear implant Lab implementation, consists of four steps: 1) Pre-emphasis, 2) Bandpass filter bank, 3) Envelope extraction and 4) Sine wave modulation.

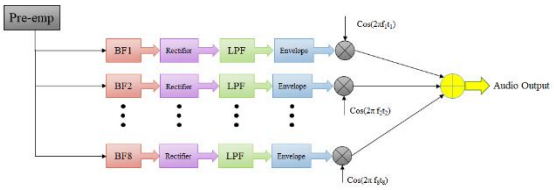

Fig.21: Block diagram of implementation of CIS Strategy in MATLAB

Data consist of spoken digit strings in quiet and noise, downloaded from Columbia University [22].

Before starting the four-step processing, we need to check to see if the signal is mono or stereo. Here, a condition is checked for input data using the following code to load only one channel:

```
[~, c]=size(Data);
if c>1
    MonoSig = sum(Data,2);
else
    MonoSig = Data;
end
```

In this code, in case the input has more than one channel, the second columns are summed and normalizing is done.

Then, the signal is sampled at 16 kHz. As formerly said, for speech intelligibility, frequency range of 2-6 kHz is sufficient. So, according to Nyquist, 16 kHz is well enough for sampling rate.

```
[P,Q] = rat(16000/Srate);
NewSig = resample(MonoSig,P,Q);
NewSig = NewSig(:,1);
fs = P/Q*Srate;
```

In this code, a rational number of P/Q is determined in a way that their multiplication to the

original sampling rate, provides the new frequency of 16kHz. Then these two factors of P and Q are used as inputs to resample the signal at the desired sampling rate [21]. So, the new sampling frequency (fs) would be 16 kHz.

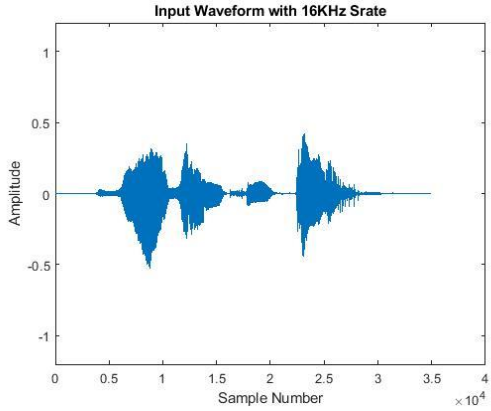

Fig.22: Recorded waveform in quite sampled at 16 kHz

8.1 Pre-emphasis

From previous sections, we figured out that the goal of pre-emphasis filter is to increase the signal to noise ratio just before starting sound coding.

In this implementation, after trying different orders and cut-off frequencies, the Butterworth high pass filter with cut-off frequency of 1200, has been used.

```
fc     = 1200;
w      = 2*fc/fs;
[b,a] = butter(1,w,'high');
SigPreEmp = filter(b,a,NewSig);
```

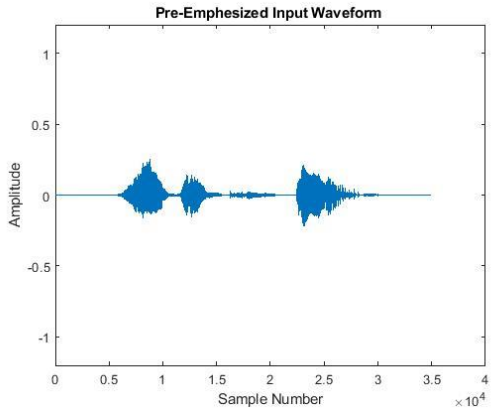

Fig.23: Pre-emphasizing recorded waveform in quite with 16 kHz sampling rate

8.2 Band-Pass Filter Bank

In this step, the signal must be divided into different frequency bands. For this issue, the distance between each band is set to 500 and is equal for all channels. The butterworth bandpass filter with the order of 2 is applied to each frequency band. The number of channels can be varied from 8 to 16. The plotted figures demonstrated in this section, are the outputs of 8-channel implementation.

```
filterOrder = 2;
    BandFiltSig =
butterBandpassFilter(SigPreEmp,
lowerband, higherband, fs,
filterOrder);
    CenterFreq = (lowerband +
higherband) / 2;
```

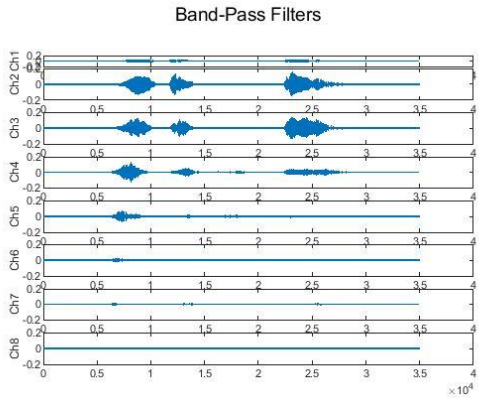

Fig.24: The outputs of 8 channels filtered by Butterworth bandpass filters

The coding details are provided in the Appendix.

8.3 Envelope Detection

In order to extract the envelope of the waveform, the absolute value of the signal is calculated and then the signal goes through a Butterworth filter bank of lowpass with cut-off frequency of 400 Hz. This process would be the full wave rectification.

```
RectSig = abs(BandFiltSig);
SigEnvelope =
butterLowpassFilter(RectSig, 400,
fs, filterOrder);
```

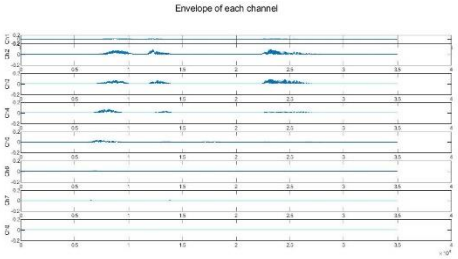

Fig.25: The envelope of 8 channels

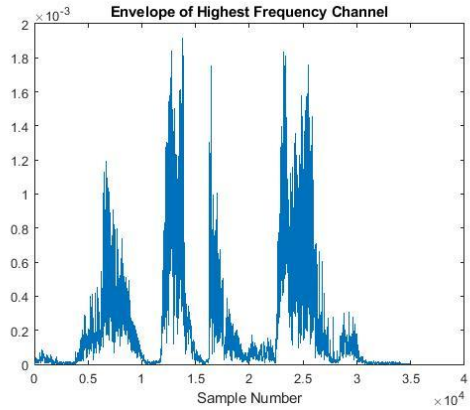

Fig.26: The envelope output of the 8th channel of the waveform

8.4 Amplitude Modulation

The last step, is to generate cosine signal with central frequency of bandpass filters and length of rectified signal. This process is to synthesis the output signal in CIS strategy and can be completed by summation of sine waves with time varying amplitudes and fixed frequencies. This means that envelopes are used to modulate a set of sine signal generators which are summed together to synthesis the original signal.

```
[r, ~] = size(RectSig);
timeDuration = r/fs;
time = linspace(0, timeDuration, r);
CosSig = cos(2*pi*CenterFreq*time);
SigEnvelope = transpose(SigEnvelope);
ModSig = SigEnvelope.*(CosSig);
OutputSig = OutputSig + ModSig;
```

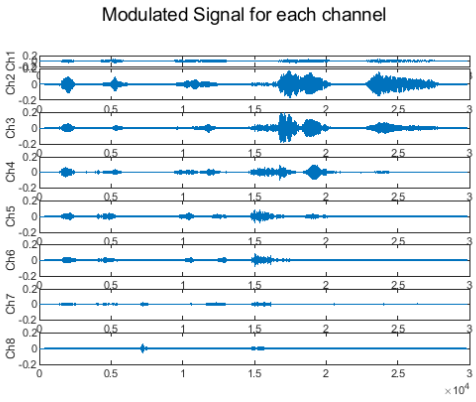

Fig.27: Amplitude-Modulated signal for each channel

After normalizing the signals, the output of the signals has been compared with the input, which was successful to preserve useful information of the original signal.

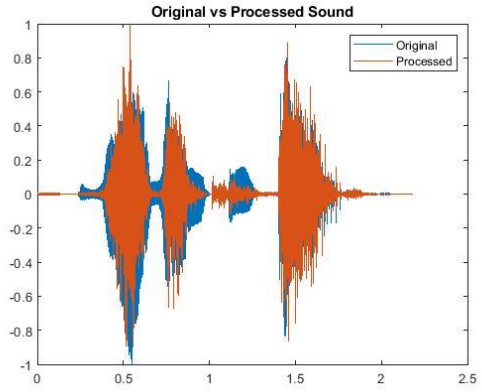

Fig.28: The output of the processed signal with CIS Strategy with 8 channels, recorded in quiet

The process has been tested with higher number of channels and according to the output, we can say that increasing the number of channels does not provide us with more significant information compared to the 8-channel processor.

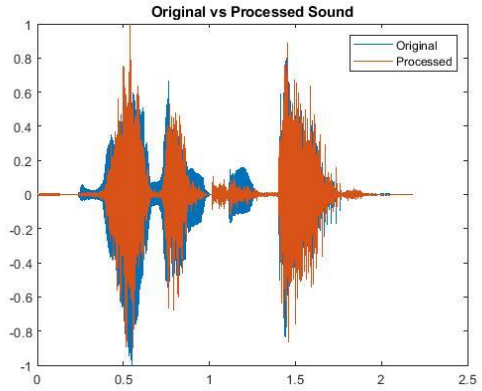

Fig.29: The output of the processed signal with CIS Strategy with 12 channels, recorded in quiet

Also, to understand the efficiency of the proposed code, the process is tried with the same signal, presented in a moderate noisy environment.

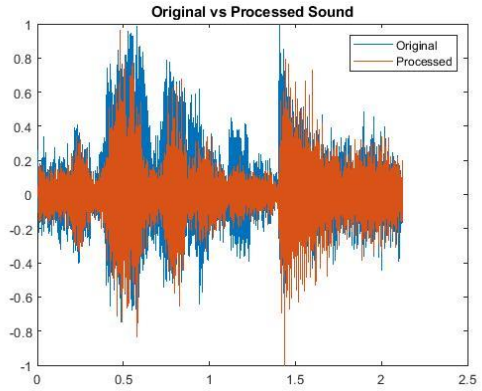

Fig.30: The output of the processed signal with CIS Strategy with 8 channels, recorded in presence of moderate noise

The output of all inputs; presented in quiet or noisy space, have been tested by normal hearing subjects and it can be said that the speech perception was quite good.

Accordingly, we can say that, this code can be used as a lab test for CIS strategy.

9. Conclusion:

This study had a review on the history of cochlear implant design and corresponding signal processing strategies. Thanks to the effort of scientists, there have been remarkable developments over the years that increased the performance of the users in a way that many conversational situations have become easy challenges for the users. However, there are various limitations on the design of a perfect cochlear implant system in terms of hardware and software platforms that have direct impact on clinical results. Refining signal processing strategies by means of improving users' performance in noisy conditions, is one the key challenges in recent developments. In this review, a two-way approach to decrease the effect of noise on cochlear implant performance has been introduced. It seems that having more focus on specific features of the speech signal, can be helpful in practical output. It is fair to conclude that there is a bright future for cochlear implant developments to achieve better results by the time.

References:

1. *Signal-processing techniques for cochlear implants, Laizou, P.C, IEEE Engineering in Medicine and Biology Magazine, 1999-05, Vol.18 (3), p.34-46*
2. *Mimicking the human ear, Loizou, 1998, IEEE signal processing magazine, 1053-5888*
3. *Cochlear Implant: System design, Integration and Evaluation, Zeng et al, IEEE Review in Biomedical Engineering, VOL. 1,2008*
4. *Computer Simulation of Multichannel CIS Strategy for Cochlear Implant, LinJing Wang, 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, 2009-06, p.1-4*
5. *Pschoacoustic "NofM"-type speech coding strategy for cochlear implant, Waldo Nogueira et al, Eurasip Journal, 2005:18, 3044-3059*
6. *Effect of Stimulation Rate on Cochlear Implant Users' Phoneme, Word and Sentence Recognition in Quiet and in Noise, Shannon, Robert V, Audiology & neurotology, 2011, Vol.16 (2), p.113-123*
7. *The coding of sound by a cochlear prosthesis an introductory signal processing lab, Mc Clellan, 2010*
8. *Improving channel selection of sound coding algorithms in cochlear implants, Ali et al, IEEE, ICASSP 2014*
9. *Signal Processing for Cochlear Implant: A Tutorial Review, Laizou, P.C; IEEE MWSCAS'97*
10. *Effect of signal processing strategy and stimulation type on speech and auditory perception in adult cochlear implant users, Int J Audiol, 2019; 58(6):363-372*
11. *Continuous Interleaved Sampled CIS Signal Processing Strategy for Cochlear Implants MATLAB Simulation Program, Y,Srinivas; Int J sci and Eng Research, VOL 3, 2012*
12. *Linear Prediction (LPC) - Columbia EE, https://labrosa.ee.columbia.edu/*
13. *Instrument Identification Through A Simulated Cochlear Implant Processing System, Rebecca Danielle Reich, Mcgill University, Montreal, Quebec, 2000*
14. *Performance of Compressed Analog (CA) and Continuous Interleaved Sampling (CIS) Coding Strategies for Cochlear Implants in Quiet and Noise, Martin Kompis, Mattheus W. Vischer, Rudolf Häusler, Acta oto-laryngologica, 1999, Vol.119 (6), p.659-664*
15. *Cochlear Implant Filterbank Design and Optimization: A Simulation Study, Cosentino, Stefano; Falk, Tiago; McAlpine, David; Marquardt, Torsten, IEEE/ACM transactions on audio, speech, and language processing, 2014-02-01, Vol.22 (2), p.347-353*
16. *Comparison of Mandarin tone and speech perception between advanced combination encoder and continuous interleaved sampling speech-processing strategies in children, Hwang, Chung-Feng, MD; Chen, Hsiao-Chuan, PhD; Yang, Chao-Hui, MD; Peng, Jyh-Ping, MD; Weng, Chia-Hui, American journal of otolaryngology, 2012, Vol.33 (3), p.338-344*
17. *Cochlear Implant Stimulation Rates and Speech Perception, By Komal Arora, 2012, DOI: 10.5772/49992*
18. *Acoustic Models of Consonant Recognition in Cochlear Implant Users, University of Southampton, PHD Thesis, Carl Verschuur, 2007*
19. *Physiology of Hearing, Joseph E. Hawkins Emeritus Professor of Otolaryngology (Physiological Acoustics), Medical School, University of Michigan, Ann Arbor. Editor of Otophysiology*
20. *Evaluation of a Two-Formant Speech-Processing Strategy for a Multichannel Cochlear Prosthesis, Dowell, R. C; Seligman, P. M; Blamey, P. J; Clark, G. M. Annals of Otology, Rhinology & Laryngology, 1987-01, Vol.96 (1_suppl), p.132-134*
21. *Mathworks.com / Help center for MATLAB codes*
22. *https://www.ee.columbia.edu/~dpwe/sounds/digits/*

23. *Noise spectrum estimation in adverse environments: Improved Minimum Controlled Recursive Average algorithm, Israel Cohen, IEEE transactions on speech and audio processing, VOL. 11, 2013*

Appendix

Implementation of CIS Strategy in MATLAB

```
addpath('(folder path)');
[Data,Srate] = audioread('df1_n2H.wav');
%%%%%%%%%%% checking if the signal is mono or stereo %%%%%%%%%%
[~, c]=size(Data);
if c>1
    MonoSig = sum(Data,2);
else
    MonoSig = Data;
end

figure(1), clf
plot(MonoSig);
title('Input Waveform');
set(gca,'ylim',[-1.2 1.2])
xlabel('Sample Number');
ylabel('Amplitude');

%%%%%%%%%%%%%%%%%changing sampling rate to 16KHz %%%%%%%%%%%%%%

[P,Q] = rat(16000/Srate);
NewSig = resample(MonoSig,P,Q);
NewSig = NewSig(:,1);
fs = P/Q*Srate;

figure(2), clf
plot(NewSig);
title('Input Waveform with 16KHz Srate');
set(gca,'ylim',[-1.2 1.2])
xlabel('Sample Number');
ylabel('Amplitude');

%%%%%%%%%%%%%%%%%%%%%%%% Pre-Emphesizing %%%%%%%%%%%%%%%%%%%%

fc    = 1200;
w     = 2*fc/fs; % fc/fs/2 = 2*fc/fs = 1200 Hz
[b,a] = butter(2,w,'high');
SigPreEmp = filter(b,a,NewSig);
%SigPreEmp = NewSig;

figure(3), clf
plot(SigPreEmp);
title('Pre-Emphesized Input Waveform ');
set(gca,'ylim',[-1.2 1.2])
xlabel('Sample Number');
ylabel('Amplitude');

%%%%%%%%%%%%%%%%%%%%%%%%%% Channels %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Channels = 8;
ChBand = 500;
OutputSig = 0;
for i = 1:Channels
    % Set lower and upper corner frequency of bandpass filter
    if i == 1
        lowerband = 100;
        higherband = 500;
    else
        lowerband = 500 + (i-2)*ChBand;
        higherband = lowerband + ChBand;
    end

    filterOrder = 2;
```

```
    BandFiltSig = butterBandpassFilter(SigPreEmp, lowerband, higherband, fs,
filterOrder);
    CenterFreq = (lowerband + higherband) / 2;

  %figure(10)
   %     subplot (8,1,i)
    %    plot(BandFiltSig);
     %   suptitle('Band-Pass Filters');
      %  ylabel(['Ch',num2str(i)]);
       % set(gca,'ylim',[-0.2 0.2])



    %%%%% rectify each band %%%%%%
   RectSig = abs(BandFiltSig);
   %%%%% Envelope Extraction for each band %%%%%
   SigEnvelope = butterLowpassFilter(RectSig, 400, fs, filterOrder);
   %figure(11)
    %    subplot (8,1,i)
     %   plot(SigEnvelope);
      %  suptitle('Envelope of each channel');
       % ylabel(['Ch',num2str(i)]);
        %set(gca,'ylim',[-0.2 0.2])

   % Generate cosine signal with central frequency of bandpass filters and length
of rectified signal
    [r, ~] = size(RectSig);
    timeDuration = r/fs;
    time = linspace(0, timeDuration, r);
    CosSig = cos(2*pi*CenterFreq*time);

    %%%%% AMPLITUDE MODULATION %%%%%
    SigEnvelope = transpose(SigEnvelope);
    ModSig = SigEnvelope.*(CosSig);
%    figure(12)
 %       subplot (8,1,i)
  %      plot(ModSig);
   %     suptitle('Modulated Signal for each channel');
    %    ylabel(['Ch',num2str(i)]);
     %   set(gca,'ylim',[-0.2 0.2])

    % Sum amplitude modulated signals for each channel
    OutputSig = OutputSig + ModSig;

    %%%%% Plotting %%%%%
    if i == 1

        N=length(BandFiltSig);
        sigpow= abs( fft(BandFiltSig)/N ).^2;
        hz = linspace(0,fs/2,floor(N/2)+1);

        figure()
        plot(hz,sigpow(1:length(hz)));
        title('Output Signal of Lowest Frequency Channel');
        xlabel('Sample Number');
        ylabel('Amplitude');
        set(gca,'xlim',[0 1600])

        figure()
        plot(SigEnvelope);
        title('Envelope of Lowest Frequency Channel');
        xlabel('Sample Number');

    elseif i == Channels
```

```
        N=length(BandFiltSig);
        sigpow= abs( fft(BandFiltSig)/N ).^2;
        hz = linspace(0,fs/2,floor(N/2)+1);

        figure()
        plot(hz,sigpow(1:length(hz)));
        title('highest Frequency Channel');
        xlabel('Sample Number');
        ylabel('Amplitude');
        set(gca,'xlim',[0 8000])

        figure()
        plot(BandFiltSig);
        title('Output Signal of Highest Frequency Channel');
        xlabel('Sample Number');
        ylabel('Amplitude');

        figure()
        plot(SigEnvelope);
        title('Envelope of Highest Frequency Channel');
        xlabel('Sample Number');
    end

end
% Normalize the signals by the max of their absolute value
NormOutputSig = OutputSig / max(abs(OutputSig), [], 'all');
%NormInput = SigPreEmp / max(abs(SigPreEmp), [], 'all');
NormInput = NewSig / max(abs(NewSig), [], 'all');

%%%%% Plot the normalized output signal
figure()
plot(NormOutputSig);
title('Synthsized Signal');
xlabel('Sample Number');
ylabel('Amplitude');

% Plot the normalized input signal
figure()
plot(time, NormInput);
hold on;
plot(time, NormOutputSig);
title('Original vs Processed Sound');
legend('Original', 'Processed');
sound(OutputSig, fs);

%for i = 1:Channels
 %       figure(10)
  %      subplot (12,1,i)
   %     plot(BandFiltSig);
    %    suptitle('BP filter');
     %   ylabel(['Ch',num2str(i)]);
 % end

function[y] = butterBandpassFilter(data, lowcut, highcut, fs, order)
    % Nyquist frequency
    nyq = fs/2;

    % Since the cutoff frequency cannot be equal to 1 and nyq = 8000,
    % the upper cutoff frequency must be less than 8000
    if highcut == nyq
        highcut = highcut - 0.00000000001;
    end

    % Normalize the frequencies by dividing by the Nyquist frequency
    lowerband = lowcut/nyq;
    higherband = highcut/nyq;
```

```
    % butter() returns b,a which are transfer function coefficients
    [b, a] = butter(order, [lowerband, higherband], 'bandpass');
    y = filter(b, a, data);
end
function[y] = butterLowpassFilter(data, cutoff, fs, order)
    % Nyquist frequency
    nyq = fs/2;

    % Cutoff frequency cannot equal Nyquist frequency
    % so decrease slightly
    if cutoff == nyq
        cutoff = cuttoff - 0.00000000001;
    end

    % Normalize the cutoff frequency
    cutoffFreq = cutoff / nyq;

    [b, a] = butter(order, cutoffFreq, 'low');
    y = filter(b, a, data);
end
```