www.dergipark.gov.tr
ISSN:2148-3736
El-Cezerî Fen ve Mühendislik Dergisi
Cilt: 9, No: 3, 2022 (936-947)

El-Cezerî Journal of Science and Engineering
Vol: 9, No: 3, 2022 (936-947)
DOI : 10.31202/ecjse.1011025

ECJSE

Research Paper / Makale

Determination of Optimum Parameters for Cochlear Implants Speech Processors by Using Objective Measures

Özkan ARSLAN[1a]

[1] Electronics and Communication Engineering Department, Tekirdag Namik Kemal University, Tekirdag, Türkiye
oarslan@nku.edu.tr

Received/Geliş: 17.10.2021 **Accepted/Kabul:** 14.12.2021

Abstract: In a cochlear implant (CI) speech processor, several parameters such as channel numbers, bandwidths, rectification type, and cutoff frequency play an important role in acquiring enhanced speech. The effective and general purpose CI approach has been a research topic for a long time. In this study, it is aimed to determine the optimum parameters for CI users by using different channel numbers (4, 8, 12, 16, and 22), rectification types (half and full) and cutoff frequencies (200, 250, 300, 350, and 400 Hz). The CI approaches have been tested on Turkish sentences which are taken from METU database. The optimum CI structure has been tested with objective quality that weighted spectral slope (WSS) and objective intelligibility measures such as short-term objective intelligibility (STOI) and perceptual evaluation of speech quality (PESQ). Experimental results show that 400 Hz cutoff frequency, full wave rectifier, and 16-channels CI approach give better quality and higher intelligibility scores than other CI approaches according to STOI, PESQ and WSS results. The proposed CI approach provides the ability to percept 91% of output vocoded Turkish speech for CI users.

Keywords: Cochlear implant, vocoder, filter bank, objective intelligibility measures

Koklear İmplant Konuşma İşlemcileri için Optimum Parametrelerin Objektif Ölçütler Kullanılarak Belirlenmesi

Öz: Bir koklear implant (Kİ) konuşma işlemcisinde, kanal sayıları, bant genişlikleri, doğrultma tipi ve kesme frekansı gibi çeşitli parametreler, gelişmiş konuşma elde etmede önemli bir rol oynamaktadır. Etkili ve genel amaçlı Kİ yaklaşımı uzun süredir araştırma konusu olmuştur. Bu çalışmada, farklı kanal sayıları (4, 8, 12, 16 ve 22), doğrultma tipleri (yarım ve tam dalga) ve kesme frekansları (200, 250, 300, 350 ve 400 Hz) kullanılarak Kİ kullanıcıları için en uygun parametrelerin belirlenmesi amaçlanmıştır. Kİ yaklaşımları ODTÜ konuşma veri tabanından alınan Türkçe cümleler ile test edilmiştir. En uygun Kİ yapısı, ağırlıklı spektral eğim (WSS) gibi nesnel kalite, kısa-süreli nesnel anlaşılabilirlik (STOI) ve konuşma kalitesinin algısal değerlendirmesi (PESQ) gibi nesnel anlaşılabilirlik ölçütleri ile belirlenmiştir. Deneysel sonuçlar, 400 Hz kesme frekansı, tam dalga doğrultucu ve 16-kanallı Kİ yaklaşımının STOI, PESQ ve WSS sonuçlarına göre daha kaliteli ve daha yüksek anlaşılabilirlik skorlarına sahip olduğunu göstermektedir. Önerilen Kİ yaklaşımı, implant kullanıcıları için çıkış kodlu konuşmanın %91'ini algılama yeteneği sağlamaktadır.

Anahtar Kelimeler: Koklear implant, ses kodlayıcı, filtre bankası, nesnel anlaşılabilirlik ölçütleri

1. Introduction

The cochlear implants (CIs), a prosthetic device, is implanted in the inner ear and is widely used to provide more comfortable hearing for people with hearing impairment. The CIs allow users with hearing problems to communicate without lip reading or hand signals. The CIs are the product of an interdisciplinary study such as physiology, otolaryngology, speech and signal processing [1]. These

How to cite this article
Arslan O., "Determination of Optimum Parameters for Cochlear Implants Speech Processors by Using Objective Measures", El-Cezerî Journal of Science and Engineering, 2022, 9 (3); 936-947.

Bu makaleye atıf yapmak için
Arslan O., "Koklear İmplant Konuşma İşlemcileri için Optimum Parametrelerin Objektif Ölçütler Kullanılarak Belirlenmesi", El-Cezerî Fen ve Mühendislik Dergisi, 2022, 9 (3); 936-947.
ORCID : [a]0000-0003-1949-3688

disciplines have different roles in the design of CIs. The use of different techniques in the electrical transmission of sounds is very important in the development of the CI, especially in signal processing [2-3]. Recent studies have proved that the general model of listening and intelligibility performance for CI users in audio/speech coder simulations can be predicted by a lot of parameters. These parameters are listed as background noise, speech masker type, number of electrodes, rectification type used in envelope extraction and cutoff frequency of the low-pass filter [4-6]. Speech encoder simulations cannot predict absolute performance for CI users. These encoders are expected to impact performance when a certain parameter is changed [7].

Over the decades, CI algorithms have been the subject of many researches [8-9]. The developments and advances in chip technology are also effective in CI devices. Currently, two basic audio coding schemes that produced by four main manufacturers are used [10]. Med-El® and Advanced Bionics® manufacturers apply the “CIS” (Continuous Interleaved Sampling) [11] voice coding strategy in sound processors, while Cochlear® and Oticon Medical® manufacturers apply the “NofM” (for N out of M) strategy [12] in sound processors. The most obvious difference between the CIS and NofM strategies relates to the number of electrodes activated. NofM strategies activate channels with the highest energy, while CIS strategies activate all electrodes [13-14]. All coding strategies for CIs are currently implemented by vocoder (voice encoder) technique. [15-16]. Vocoder is an audio operating system that can analyze and synthesize audio and speech signals. Vocoder is used extensively in applications such as compressing, encoding, transmitting and modifying voices. In research and development of CIs, the vocoder has made very important and profound contributions. Also, it has been used frequently to create the general model of speech recognition and perception for CI users [17-18].

Subjective quality and intelligibility tests are used to ensure that CI devices reach acceptable hearing levels [17, 19-20]. Subjective tests for CI devices are performed in two ways. First, encoded sounds received from the CI output are presented and evaluated to listeners with no hearing problems. Secondly, the sound and speech obtained by the algorithm are directly tested with CI users with hearing impairment. The second subjective test approach is mostly used in hearing aids for perception of effect noise suppression and dynamic range. However, subjective tests are time consuming and expensive as the evaluation needs to be done in several different sessions. It also extends the development and evaluation process of CI algorithms. Therefore, the testing process is accelerated by using fast and repeatable objective quality and intelligibility measures instead of subjective tests. In CI system performance evaluation, computational objective measures based on human auditory model are used instead of listener tests. Objective quality and intelligibility measures such as normalized covariance metric (NCM) [21-22], short-time objective intelligibility (STOI) [23], perceptual evaluation of speech quality (PESQ) [24], hearing aid speech quality (HASQI) [25], hearing aid speech perception indices (HASPI) [26], and weighted spectral slope (WSS) [27] are widely used in speech processing applications [28-29].

Kates et al. [30] determined the relationship between objective and subjective intelligibility scores with the Pearson correlation for English sentences database. They reported that the Pearson correlation results had values of 0.89, 0.92, 0.93, 0.81 and 0.79 for the NCM, STOI, HASQI, HASPI and PESQ criteria, respectively. In another similar study [19], PESQ and WSS objective quality and intelligibility scores were correlated with subjective scores from listeners without hearing problems for English and Mandarin Chinese speech database. The results of the study showed that there was 0.91 and -0.87 correlation between the objective and subjective scores according to the PESQ and WSS values, respectively. In a study conducted with different noise reduction algorithms in a cochlear implant speech processor, the effect on objective intelligibility of speech was investigated. The objective assessment and listening tests were performed using the Mandarin Chinese version of Hearing in Noise Test (MHINT) database. The objective intelligibility scores of vocoded speech output from an 8-channel CI were determined using STOI and NCM. It was stated that there is a high correlation between objective intelligibility and listening tests [17, 30]. Research and studies prove that the quality and intelligibility scores of the designed CIs algorithms differ in terms of the language of the CI user. In general, it can be said that CIs algorithms developed using English and Chinese databases pose a disadvantage for CI users using other languages. Therefore, it is very important to develop CIs algorithms considering the phonetic features of the language.

The aim of this study is to determine the effect of cochlear implants on the intelligibility levels of vocoded speech by channel and electrode numbers, frequency bandwidths, wave rectification type, filter degrees and modulation type. The quality and intelligibility scores of the speech signals electrically transmitted to the auditory nervous system are determined by STOI, PESQ and WSS measures for Turkish speech database. It is ensured that optimum parameters are obtained for CI users to understand voice and speech at high level.

The remainder of this work is organized as follows. In section 2, speech corpus, cochlear implant algorithm and objective measures are explained. The results and discussions of the proposed approach are presented in section 3. In section 4, the conclusions are presented.

2. Material and Methods

Figure 1 shows the block diagram of the CI vocoder simulation system designed with different channel numbers. As can be seen in Fig. 1, speech signals received with the microphone are first pre-emphasis processed with an adaptive gain control (AGC).

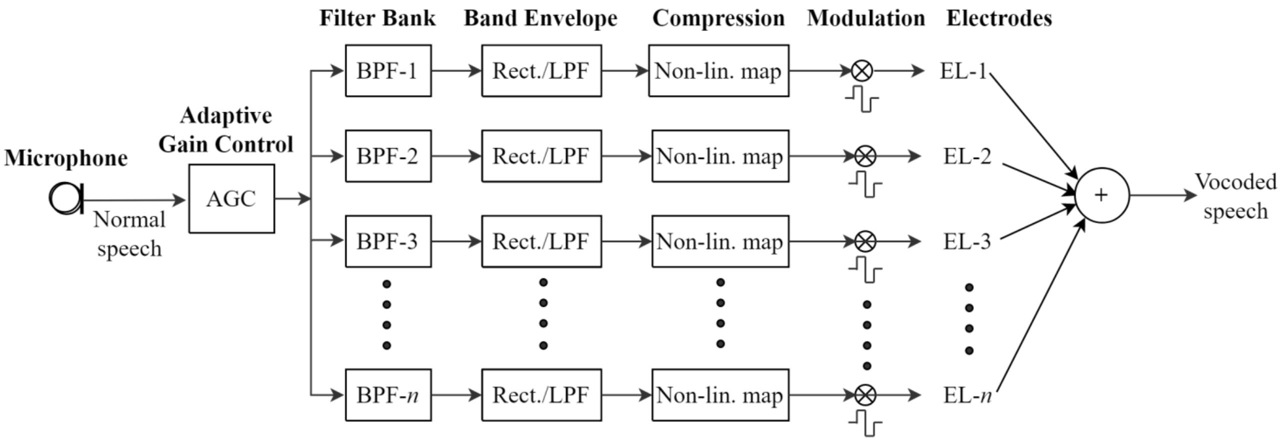

Figure 1. Signal processing stages for a cochlear implant (CI) vocoder speech processing strategy. AGC, adaptive gain control; BPF, band-pass filter; Rect, rectification; LPF, low-pass filter; Non-lin. Map, non-linear mapping; EL, electrode.

The AGC adjusts the imbalances that may occur in the amplitude of the high frequency components of the normal speech signals received from the microphone. Speech signals are then processed in the band-pass filters (BPFs) stage, which are created with filter banks, envelope extraction, compression and modulation stages. After the pre-emphasis stage, the signal is divided into channels 4, 8, 12, 16, and 22 with BPFs in the range of 80 Hz to 6800 Hz frequency band interval. The bandwidths of each channel created with BPFs are calculated as follows:

$$f_{range} = log_{10}(f_u/f_l) \tag{1}$$

$$f_{int} = f_{range}/N \tag{2}$$

$$B_U = f_l * 10^{i*f_{int}} \quad (3)$$

$$B_L = f_l * 10^{(i-1)*f_{int}} \quad (4)$$

$$BW = B_U - B_L \quad (5)$$

where N, f_l and f_u represent the number of channels, lower and upper frequencies, respectively. BW denotes the bandwidth of channels and i represents each of channels.

After BPF bank stage, half and full rectifiers are applied to the filtered signals for envelope extraction. Then, low-pass filters (LPFs) with different cutoff frequencies (200, 250, 300, 350 and 400 Hz) are applied to the rectified speech signals for each channel. The envelopes are compressed logarithmically to prevent an abnormal increase in speech amplitude over a small dynamic range. The non-linear amplitude mapping function is used to compress the envelopes. The amplitude of the compressed envelope signals for each channel is modulated with a pulse sequence. At this stage, where amplitude modulation is used, the envelopes of the simulation modulate sinusoid carriers, whose frequencies corresponded to the center frequencies of the BPFs. Finally, the output of the vocoded speech signal is the sum of the signals received from the envelope-modulated sinusoids of each channel.

2.1. Speech Corpus

In this study, a Turkish speech database [31] containing 40 sentences from 193 subjects (104 males and 89 females) was used. The sentences were recorded at sampling rate of 16 kHz. The age distribution of subjects is between 19 and 50 (the average age is 23.9), and subjects do not have any hearing impairment. The database contains 2462 different sentences and 30 sentences which are phonetically balanced and with equal difficulty levels are selected for CIs algorithms evaluation.

2.2. Objective Evaluation Measures

Performance evaluation of the designed CIs algorithms is obtained by objective quality and intelligibility measures. These measurements, which proved to be highly correlated with subjective tests, can provide significant information about the performance of the CI approach [19, 32]. Speech quality measure which is WSS and speech intelligibility measures which are STOI and PESQ are widely used in speech processing applications. The WSS calculates the weighted spectral slope differences between speech signals which are divided into frequency bands. The spectral slope is a difference value in decibels between adjacent spectral amplitudes. The WSS quality measure is defined as [27]:

$$d_{WSS} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W(j,m) \left(S_c(j,m) - S_p(j,m)\right)^2}{\sum_{j=1}^{K} W(j,m)} \quad (6)$$

where $W(j,m)$, K and M represent the weight in the jth frequency band, the number of bands and the total number of frames in the signal, respectively. $S_c(j,m)$ and $S_p(j,m)$ denote the spectral slope in the jth frequency band of the mth frame of the normal and vocoded speech signals, respectively. The WSS quality value decreases when the vocoded speech approaches the original speech.

STOI is a correlation-based measure that provides subjective intelligibility evaluations of speech signals [23]. In the calculation of STOI, the vocoded and normal speech signals are divided into short time frames and grouped into 15 one-third (1/3) octave bands. The intermediate intelligibility measure $d_j(m)$ for each frame and band is defined as the coefficient of correlation between temporal envelope vectors derived from vocoded and normal speech signals. Finally, the STOI scale is determined by averaging intermediate values over the octave band and all speech frames. The STOI model is designed for a sampling rate of 10 kHz to cover the relevant frequency range for speech intelligibility. The average intelligibility measure $d_j(m)$, which is correlation between vocoded and normal speech signals, is expressed as:

$$d_j(m) = \frac{\sum_n \left(X_j(n) - \frac{1}{N}\sum_l X_j(l)\right)\left(Y'_j(n) - \frac{1}{N}\sum_l Y'_j(l)\right)}{\sqrt{\sum_n \left(X_j(n) - \frac{1}{N}\sum_l X_j(l)\right)^2 \sum_n \left(Y'_j(n) - \frac{1}{N}\sum_l Y'_j(l)\right)^2}} \tag{7}$$

where $X_j(m)$ and $Y_j(m)$ represent the time-frequency representation of the normal and vocoded speech signal, respectively. The STOI measure is calculated by mean the intermediate intelligibility values for all frames and bands. Finally, STOI is expressed as:

$$STOI = \frac{1}{JM}\sqrt{\sum_{j,m} d_j(m)} \tag{8}$$

where M and J represent the total number of frames and one-third octave bands, respectively. The STOI score is calculated in the range of [0 - 1], and higher STOI scores indicate better intelligibility.

PESQ is an international standard that grades speech samples in five quality categories and is used for intelligibility score estimation. These grades are from 1 to 5 and speech samples are scored in sound quality categories such as very bad (1), bad (2), normal (3), good (4) and very good (5) [24, 33]. In listening tests, which are subjective evaluations, the fact that the tests have to be done repeatedly may cause the subjects to memorize the listened speech. In many speech processing applications, PESQ is widely used as a measure of quality and intelligibility as alternative to listening tests [28-29, 32].

3. Results and Discussion

In this study, the output vocoded speech signal was obtained for five different channel numbers (4, 8, 12, 16, and 22), half and full rectifier types and five different cutoff frequencies (200, 250, 300, 350, and 400 Hz) of the LPFs. The performance of the cochlear implant algorithm was evaluated by calculating the objective quality and intelligibility scores of the output vocoded speech signals for Turkish speech database. Simulation results showed that using LPF with 400 Hz cutoff frequency in envelope extraction stage gives the highest intelligibility scores. Therefore, the LPF cutoff frequency was chosen as 400 Hz independent of the number of channels and rectification types.

Figure 2 shows the distribution of STOI scores obtained for five different channel numbers and two rectification types. As seen in Figure 2, the highest average STOI scores for vocoded speech were obtained by 16-channel approach. It is seen that the full wave rectifier has 0.91 ± 0.02 (mean ± std) and the half-wave rectifier has 0.88 ± 0.02 (mean ± std) STOI score when the 16-channel structure is evaluated in which the best STOI scores are observed. In the 22-channel full rectifier structure, which is the closest to the 16-channel structure, the mean STOI score was 0.79 ± 0.022. According to the STOI results, the lowest scores are obtained in 4 and 8 channel CI approaches.

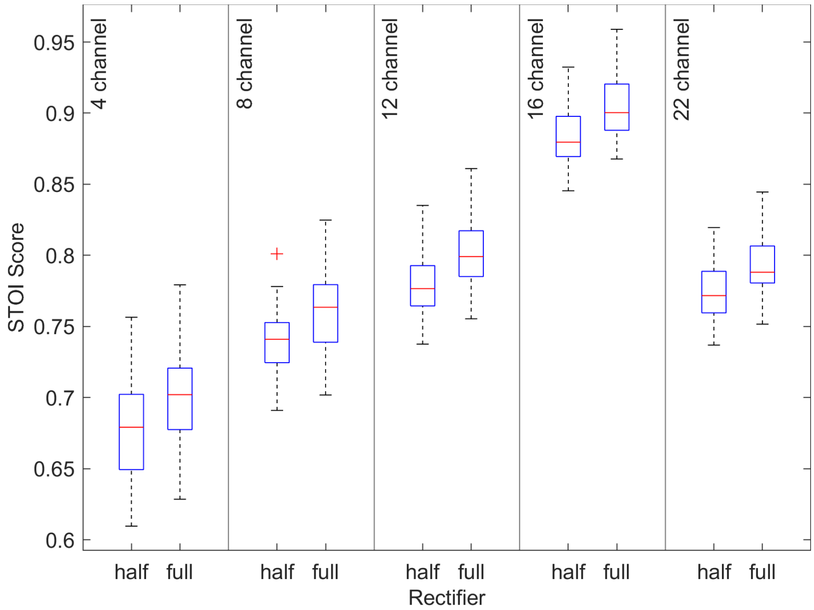

Figure 2. STOI scores for CI algorithms with half and full rectifier type and 4, 8, 12, 16, and 22 channels

The average WSS values of cochlear implant approaches, which are a measure of the distance between vocoded speech and normal speech, are shown in Figure 3. As can be seen from Figure 3, minimum WSS values are obtained in all channel numbers when 400 Hz cutoff frequency is used in both half and full wave rectifiers.

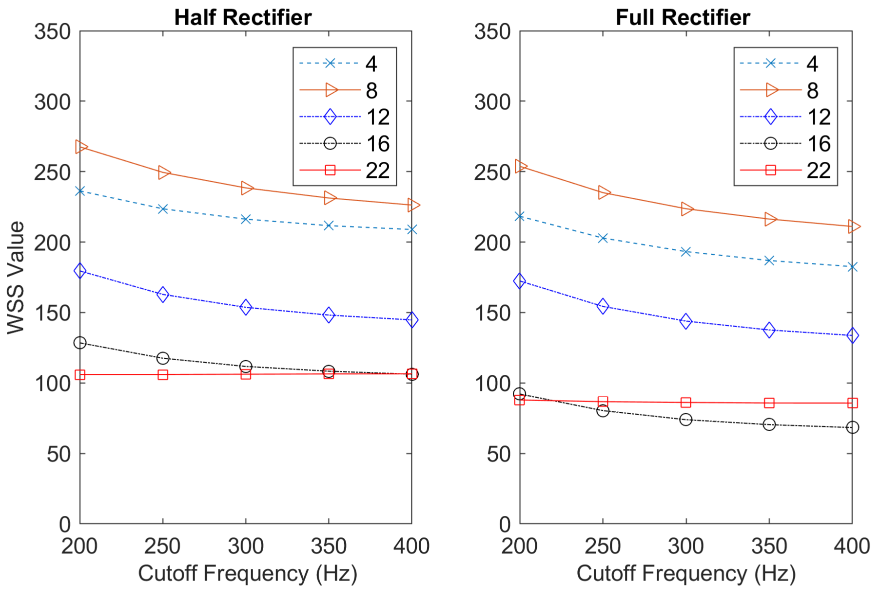

Figure 3. Average WSS values for CI algorithms with half and full rectifier type, {4, 8, 12, 16, and 22} channels, and five different cutoff frequencies

The simulation results show that the average WSS values are {208.93, 226.27, 144.75, 106.21, 106.53} for {4, 8, 12, 16, 22} channel approaches for half-wave rectifier with 400 Hz cutoff frequency, respectively. Similarly, the average WSS values are {182.58, 211.07, 133.72, 68.35, 85.73} for {4, 8, 12, 16, 22} channel approaches for full-wave rectifier with 400 Hz cutoff frequency, respectively.

Figure 4 shows the average PESQ intelligibility scores of vocoded speech for all channel numbers, rectification types and cutoff frequencies. As can be seen from Figure 4, the highest PESQ scores for all channel numbers and rectification types are observed at 400 Hz cutoff frequency. The simulation results show that the average PESQ scores are {2.75, 2.74, 2.67, 2.88, 2.79} for {4, 8, 12, 16, 22} channel approaches for half-wave rectifier with 400 Hz cutoff frequency, respectively. Similarly, the average PESQ scores are {2.93, 2.94, 2.87, 3.13, 2.98} for {4, 8, 12, 16, 22} channel approaches for full-wave rectifier, respectively. Considering the PESQ results, the 16-channel cochlear implant approach with full wave rectifier gives the highest PESQ score.

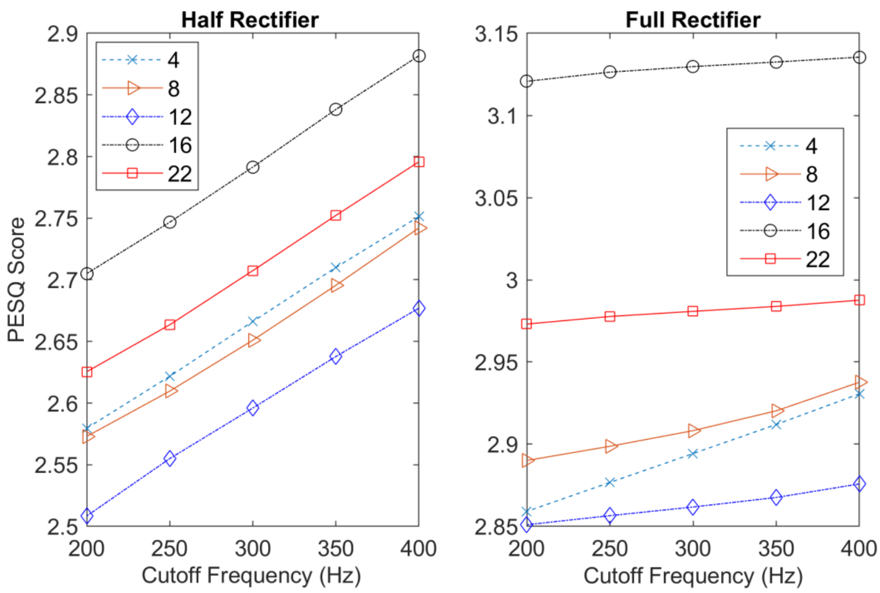

Figure 4. Average PESQ scores for CI algorithms with half and full rectifier type, {4, 8, 12, 16, and 22} channels, and five different cutoff frequencies

Table 1. Bandwidths and center frequencies of band-pass filters.

Channel number	Bandwidth (Hz)	Center frequency (Hz)
1	80-105	91.91
2	105-139	121.33
3	139-184	160.16
4	184-242	211.42
5	242-320	279.08
6	320-423	368.40
7	423-558	486.31
8	558-737	641.95
9	737-973	847.41
10	973-1285	1118.61
11	1285-1696	1476.62
12	1696-2239	1949.21
13	2239-2956	2573.05
14	2856-3902	3396.54
15	3902-5151	4483.58
16	5151-6800	5918.53

The STOI, PESQ and WSS values show that the best vocoded speech for Turkish speech signals is obtained by 16-channel, full-wave rectifier and 400 Hz cutoff frequency. Therefore, this CI approach has been proposed as it has optimal parameters for the Turkish language. In this proposed approach, normal speech signals are divided into 16 channels using 6th order Butterworth filters. The bandwidths of each channel and center frequencies, which are the geometric mean of the bandwidth, are listed in Table 1.

Figure 5 shows the frequency responses of the BPFs for the proposed 16-channel CI algorithm and the frequency responses of filters are shown on logarithmic and linear frequency scale. In this filter structure, the last BPF is used as an anti-aliasing filter and it filtered out the frequency components above 6800 Hz.

The next step of the implant algorithm is the process of enveloping the speech signal received from each channel. A full-wave rectifier was used for each channel output at this stage, which included a rectifier and LPF. The 2nd order Butterworth LPF with a cut-off frequency of 400 Hz was used, as it gave the best scores in providing speech quality and intelligibility. After the envelope extraction step, the envelope amplitudes from each channel modulate a pulse sequence.

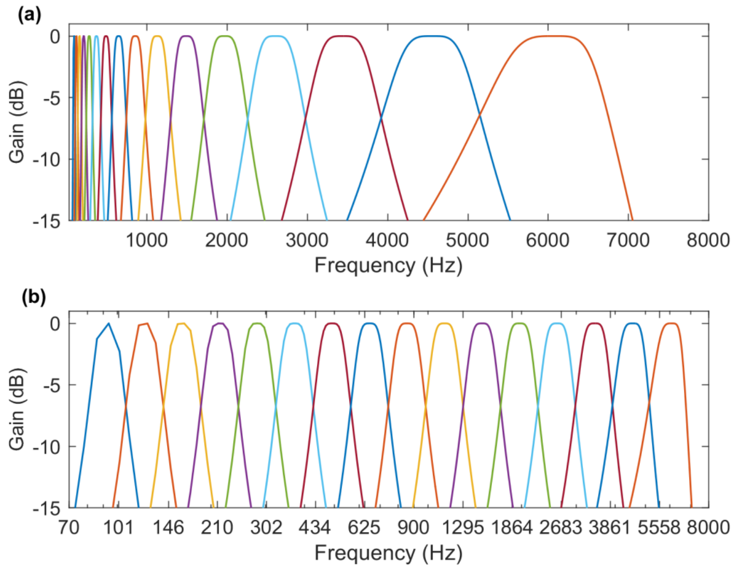

Figure 5. Frequency response of band-pass filters for 16-channels cochlear implant algorithm (a) logarithmic (b) linear frequency scales

The envelopes modulate sinusoid carriers and the modulated sinusoids of each channel are summed to generate the output vocoded signal. Figure 6 shows time and spectrogram representation of normal and output vocoded speech. As can be seen in Fig. 6, especially low frequency components of normal speech applied to the CI algorithm input are successfully transferred to the CI output.

Experimental results show that the optimum parameters of a CI algorithm can be determined by the objective quality and intelligibility measures. An optimum CI structure is presented according to the characteristics of the Turkish language when the objective STOI, WSS and PESQ scores are evaluated for each algorithm. Therefore, as shown in Fig. 2, the average STOI intelligibility score of 0.91 shows that the 16-channel and full-wave rectifier structure is the most suitable approach for CIs.

In Fig. 3, the WSS quality results show that the 16-channel full-wave rectifier has the lowest WSS value with an average of 68.35 and that structure has optimum parameters according to WSS values. According to the average PESQ scores, which is the perceptual evaluation of speech quality shown in Fig. 4, a 16-channel and full-wave rectifier system is preferred with a PESQ score of 3.13 which is the highest score. This 16-channel approach is superior to other approaches in providing intelligibility of vocoded speech. Thus, a 16-channel with 400 Hz cutoff frequency full wave rectifier system is recommended for CI users.

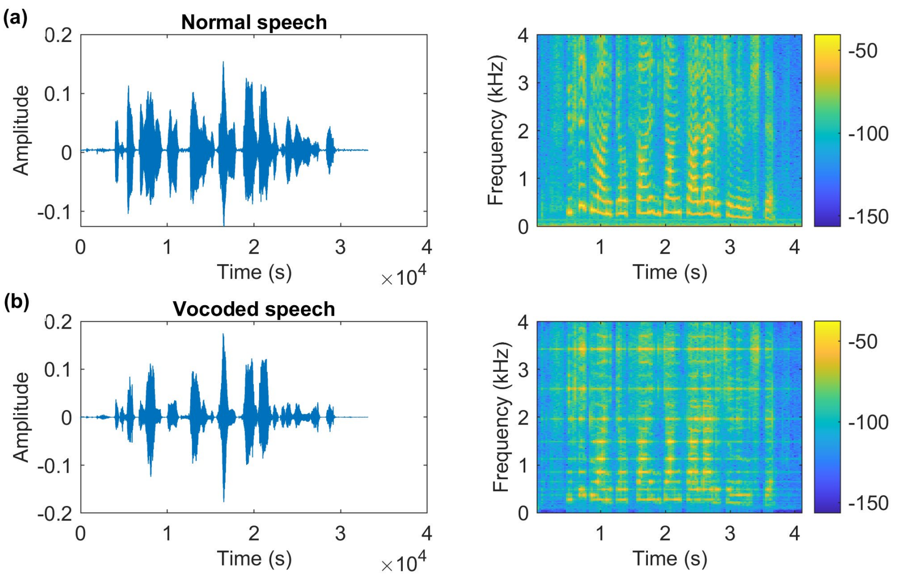

Figure 6. Time and spectrogram representation for (a) normal (b) vocoded speech

The channel numbers, n-of-m strategies, rectification type, LPF cutoff frequency, and electrode placements have long been the subject of research in CI sound processors. In Table 2, some studies in the literature on this subject and this study are summarized. CI algorithm evaluations are usually performed with subjective criteria. However, this method is quite time consuming.

Table 2. Comparison of proposed method and related works (CNs: channel numbers, RT: rectifier type, LPF: low-pass filter, LR: lower rate, EMD: Electrode-to-modiolus distances).

Author	Dataset	Channel numbers	Measures	Score	Optimum Parameters
Berg et al. [4]	English	4, 8, 10, 16, and 22	Vowel (V) and consonant (C) recognition	V: %60 C: %70	CNs: 4 of 10 (n-of-m) EMD: 0.4 to 1.5 mm
Jain and Ghosh [34]	English	1, 2, 4, 8, and 16	PESQ, WSS and Composite	Var: 0.796 R^2=0.9	CNs: 2 and 4 LR =500 pps/ch
Chen [35]	English	8	Correlation Coef. (r) between Percent Correct and PESQ	r=0.91	CNs: 5-of- 8 (n-of-m) RT: Full-wave rect. LPF:400 Hz cutoff freq.
Chen [35]	Mandarin Chinese	8	Correlation Coef. (r) between Percent Correct and PESQ	r=0.48	CNs: 5-of- 8 (n-of-m) RT: Full-wave rect. LPF:400 Hz cutoff freq.
This study	Turkish	4, 8, 12, 16, and 22	WSS, PESQ and STOI	WSS: 68.35 PESQ: 3.13 STOI: 0.91	CNs: 16 RT: Full-wave rect. LPF: 400 Hz cutoff freq.

Therefore, objective evaluation criteria can be used to measure the performance of CI algorithms. Jain and Ghosh [34] have proposed acoustic parameters, including pulse rate, number of channels, 'n of m', number of electrodes, and channel spacing. The authors used 1800 English sentences (3–4 words in length) and PESQ, WSS, and composite criteria to evaluate 2- and 4-channel CI algorithms. They have reported a variance of 0.796 for speech intelligibility. Berg et al. [4] proposed a model based on vowel and consonant recognition scores in CI algorithms with 4, 8, 10, 16, and 22 channels. The authors used English dataset and reported that the vowel recognition rate of 60% and the consonant recognition rate of 70% in this model which 4 of 10 (n-of-m) and 0.4 to 1.5 mm (electrode-to-modiolus distances) parameters were selected. Chen [35] suggested 5-of-8 (n-of-m), the full-wave rectifier and a cutoff frequency of 400 Hz for LPF. The author used English (54 vocoded speech) and Mandarin Chinese (20 vocoded speech) datasets for the proposed CI approach. The performance of the proposed CI approach was evaluated with the correlation coefficient between percent correct and PESQ. The results of the study show that the proposed model has a correlation coefficient of 0.91 for English and 0.48 for Mandarin. In our study for the Turkish database, CI algorithms with 4, 8, 12, 16, and 22 channels were evaluated with objective criteria such as WSS, PESQ and STOI. The results of this study show that the best results are obtained for the 16-channel, full wave rectifier and 400 Hz cutoff frequencies for LPFs.

4. Conclusions

This study provides a CI approach to be able to hear better quality and intelligibility speech for CI users who use Turkish language. In CI approaches, it is seen that the channels obtained by BPFs, the filter type and order used in these channels are very important. The following conclusions can be drawn from the contributions of this study:

(1) As the number of channels is increased from 4 to 16, remarkable improvements are observed in speech quality and intelligibility scores. However, the intelligibility of vocoded speech decreases when the number of channels is increased to 22. Therefore, contrary to popular belief, it can be said that the high number of channels does not show the expected effect on CI users.
(2) In the 16-channel full-wave rectifier approach, the score of intelligibility vocoded speech by CI users is 91% according to STOI evaluation. These scores indicate that comfortable and understandable hearing will be achieved for CI users. Also, an average PESQ score of 3.13 indicates that vocoded speech is perceived even if it is a little annoying.
(3) In the envelope extraction stage, the use of a 400 Hz cutoff frequency LPF with the full wave rectifier positively affects the intelligibility of vocoded speech. The reason for this is that low frequency harmonics cannot be eliminated in the half-wave rectifier. Since the frequencies of harmonics doubles when full wave rectifier is used, they are outside the 400 Hz cutoff frequency of the LPF and thus can be eliminated.
(4) In the development of algorithms for CI users, language structure should be taken into account as each language has its own unique structure and phonetics. Thus, CI users will be able to hear better voice and speech.

Author’s Contributions

All authors have read and agreed to the published version of the manuscript.

Competing Interests

The author(s) declare that they have no competing interests.

References

[1]. Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z., The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels, The Journal of the Acoustical Society of America, 1998, 104(6), 3583-3585.

[2]. Guo, R., Ma, X., Liao, M., Liu, Y., Hu, Y., Qian, X., and Tang, M., Development and application of cochlear implant-based electric-acoustic stimulation of spiral ganglion neurons, ACS Biomaterials Science & Engineering, 2019, 5(12), 6735-6741.
[3]. Wouters, J., McDermott, H. J., and Francart, T., Sound coding in cochlear implants: From electric pulses to hearing, IEEE Signal Processing Magazine, 2015, 32(2), 67-80.
[4]. Berg, K. A., Noble, J. H., Dawant, B. M., Dwyer, R. T., Labadie, R. F., and Gifford, R. H., Speech recognition with cochlear implants as a function of the number of channels: Effects of electrode placement, The Journal of the Acoustical Society of America, 2020, 147(5), 3646-3656.
[5]. Bratu, E., Dwyer, R., and Noble, J., A graph-gased method for optimal active electrode selection in cochlear implants, In International Conference on Medical Image Computing and Computer-Assisted Intervention, Peru, pp. 34-43, 2020.
[6]. Berg, K. A., Noble, J., Dawant, B., Dwyer, R., Labadie, R., Richards, V., and Gifford, R., Musical sound quality as a function of the number of channels in modern cochlear implant recipients, Frontiers in neuroscience, 2019, 13, 999.
[7]. Mourão, G. L., Costa, M. H., and Paul, S., Speech intelligibility for cochlear implant users with the MMSE noise-reduction time-frequency mask, Biomedical Signal Processing and Control, 2020, 60, 101982.
[8]. Loizou, P. C., Dorman, M., and Tu, Z., On the number of channels needed to understand speech, The Journal of the Acoustical Society of America, 1999, 106(4), 2097-2103.
[9]. Lee, S., Mendel, L. L., and Bidelman, G. M., Predicting speech recognition using the speech intelligibility index and other variables for cochlear implant users, Journal of Speech, Language, and Hearing Research, 2019, 62(5), 1517-1531.
[10]. Baumann, U., Stöver, T., and Weißgerber, T., Device profile of the MED-EL cochlear implants system for hearing loss: overview of its safety and efficacy, Expert Review of Medical Devices, 2020, 17(7), 599-614.
[11]. Dorman, M. F., and Loizou, P. C., The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels, Ear and Hearing, 1998, 19(2), 162-166.
[12]. Nogueira, W., Büchner, A., Lenarz, T., and Edler, B., A psychoacoustic" NofM"-type speech coding strategy for cochlear implants, EURASIP Journal on Advances in Signal Processing, 2005, 18, 101672.
[13]. Hu, Y., and Loizou, P. C., A new sound coding strategy for suppressing noise in cochlear implants, The Journal of the Acoustical Society of America, 2008, 124(1), 498-509.
[14]. Buechner, A., Frohne-Buechner, C., Boyle, P., Battmer, R. D., and Lenarz, T., A high rate n-of-m speech processing strategy for the first generation Clarion cochlear implant, Taylor & Francis, 2009.
[15]. Churchill, T. H., Kan, A., Goupell, M. J., and Litovsky, R. Y., Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listeners, The Journal of the Acoustical Society of America, 2014, 136(3), 1246-1256.
[16]. Churchill, T. H., Kan, A., Goupell, M. J., Ihlefeld, A., and Litovsky, R. Y., Speech perception in noise with a harmonic complex excited vocoder, Journal of the Association for Research in Otolaryngology, 2014, 15(2), 265-278.
[17]. Lai, Y. H., Chen, F., Wang, S. S., Lu, X., Tsao, Y., and Lee, C. H., A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation, IEEE Transactions on Biomedical Engineering, 2016, 64(7), 1568-1578.
[18]. Stafford, R. C., Stafford, J. W., Wells, J. D., Loizou, P. C., and Keller, M. D., Vocoder simulations of highly focused cochlear stimulation with limited dynamic range and discriminable steps, Ear and Hearing, 2014, 35(2), 262-270.
[19]. Chen, F., Predicting the intelligibility of cochlear-implant vocoded speech from objective quality measure, J. Med. Biol. Eng, 2012, 32(3), 189-194.

[20]. Santos, J. F., Cosentino, S., Hazrati, O., Loizou, P. C., and Falk, T. H., Objective speech intelligibility measurement for cochlear implant users in complex listening environments, Speech Communication, 2013, 55(7-8), 815-824.
[21]. Chen, F., and Loizou, P. C., Predicting the intelligibility of vocoded speech, Ear and Hearing, 2011, 32(3), 331.
[22]. Goldsworthy, R. L., and Greenberg, J. E., Analysis of speech-based speech transmission index methods with implications for nonlinear operations, The Journal of the Acoustical Society of America, 2004, 116(6), 3679-3689.
[23]. Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J., An algorithm for intelligibility prediction of time–frequency weighted noisy speech, IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(7), 2125-2136.
[24]. ITU-T Rec. P.862. International Telecommunication Union. Geneva, Switzerland: Feb. 2001 Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.
[25]. Kates, J. M., and Arehart, K. H., The hearing-aid speech quality index (HASQI), Journal of the Audio Engineering Society, 2010, 58(5), 363-381.
[26]. Kates, J. M., and Arehart, K. H., The hearing-aid speech perception index (HASPI), Speech Communication, 2014, 65, 75-93.
[27]. Hu, Y., and Loizou, P. C., Evaluation of objective quality measures for speech enhancement, IEEE Transactions on audio, speech, and language processing, 2007, 16(1), 229-238.
[28]. Arslan, Ö., and Engin, E. Z., Speech enhancement using adaptive thresholding based on gamma distribution of Teager energy operated intrinsic mode functions, Turkish Journal of Electrical Engineering & Computer Sciences, 2019, 27(2), 1355-1370.
[29]. Möller, S., Chan, W. Y., Cote, N., Falk, T. H., Raake, A., and Wältermann, M., Speech quality estimation: Models and trends, IEEE Signal Processing Magazine, 2011, 28(6), 18-28.
[30]. Kates, J. M., Arehart, K. H., Anderson, M. C., Muralimanohar, R. K., and Harvey Jr, L. O., Using objective metrics to measure hearing-aid performance, Ear and Hearing, 2018, 39(6), 1165.
[31]. Salor, Ö., Pellom, B. L., Ciloglu, T., and Demirekler, M., Turkish speech corpora and recognition tools developed by porting sonic: Towards multilingual speech recognition, Computer Speech & Language, 2007, 21(4), 580-593.
[32]. Rix, A. W., Beerends, J. G., Hollier, M. P., and Hekstra, A. P., Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs, In 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, Salt Lake City, UT, USA, pp. 749-752, 2001.
[33]. Sharma, D., Wang, Y., Naylor, P. A., and Brookes, M., A data-driven non-intrusive measure of speech quality and intelligibility, Speech Communication, 2016, 80, 84-94.
[34]. Jain, S., and Vipin Ghosh, P. G., Acoustic simulation of cochlear implant hearing: Effect of manipulating various acoustic parameters on intelligibility of speech, Cochlear Implants International, 2018, 19(1), 46-53.
[35]. Chen, F., Predicting the intelligibility of cochlear-implant vocoded speech from objective quality measure. J. Med. Biol. Eng, 2012, 32(3), 189-194.