MULTI-CHANNEL PULSATION STRATEGY FOR ELECTRIC STIMULATION OF COCHLEA

Shigeyoshi KITAZAWA, Hiroyuki KIRIHATA, and Tatsuya KITAMURA

Department of Computer Science, Faculty of Information, Shizuoka University

ABSTRACT

This paper describes a speech coding strategy for a cochlear implant system assuming a Nucleus Cochlear Implant receiver stimulator. Speech processor converts input speech into a series of stimulation electrode position and stimulation current intensities. This process can be optimized with a decomposing process of an acoustic signal into a given set of impulse responses corresponding to a set of electrode channels. An error minimization algorithm can find a optimal stimulation sequence that minimizes distortion of transferred speech and maximize transferred phonological information as well as sound qualities. Re-synthesized sound quality was qualitatively evaluated. Environmental sound can also be recognizable with this method.

1. INTRODUCTION

Cochlear implant (C.I.) technology has made progress in transferred speech quality, however, a number of implantees remain at poor response under hearing only condition. And environmental and musical sounds are yet unsatisfactory heard. Although an increased pulsation rate improved recipient's speech quality, efficient way of coding is still required. We employed the multi-channel pulsation algorithm based on the principle of the mullet-pulse voice coding to get an optimized pulse train for the implanted electrodes.

2. ELECTRODE STIMULATION STRATEGY

The early cochlea implant (C.I.) systems such as WSP stimulated cochlea hence auditory nerves at every fundamental frequency interval around the places corresponding to the first and the second formant frequencies. This is as it were an old speech analysis and synthesis system based on speech production theory of which reproduced speech quality is intelligible but is never good. Recently speech coding technology has achieved much better speech quality than before, apart from speech production theory therefore applicable to any kind of acoustic signals. An efficient conversion was designed by decomposition of speech wave into some of stimulation pulses estimated with LPC coding that may elicit an acoustic image similar to the spectral pattern and a sense of pitch within a frame of speech. Here the cochlear implant (C.I.) means the Nucleus 22 Channel Cochlear Implant System. Parameters of the implanted receiver are electrode position, stimulation time, and electrode current.

2.1. Electrode Position

Electrode positioning is different in every implantee, but we assumed one typical positioning and hence corresponding center frequency of the auditory nerve to be stimulated.

Eletr#	CF	Eletr#	CF	Eletr#	CF	Eletr#	CF
1	3796	6	2235	11	1309	16	634
2	3411	7	2007	12	1176	17	525
3	3066	8	1803	13	1058	18	431
4	2752	9	1623	14	917	19	360
5	2478	10	1458	15	760	20	164

Table 1: Electrode number(Eletr#) and corresponding center frequency(CF) in Hz.

2.2. Stimulation Timing

Stimulation interval conveys a sense of periodicity which is important property of speech as well as small perturbation in fundamental frequency is essential for naturalness for speech sound. At any time electrode can be stimulated, that is, asynchronously, but there are restrictions of hardware. Each stimulation must be separated with an interval more than 1 ms.

2.3. Electrode Current

Allowable current magnitude is small in real implantee. Therefore rather small number of discrete current level have to be used. This causes degradation of speech quality with C.I. system.

3. ERROR MINIMIZATION IN SPEECH CODING

In wave form coding, squared errors in wave form are minimized. In our application of speech coding, errors are minimized for the re-synthesized signal. We assume that each electrode corresponds to an impulse response of a broadened critical band pass filter of which center frequency approximated the characteristic frequency of the auditory nerve surrounding the electrode. We expect that an error minimized re-synthesized wave represent the spectrum and pitch of the original speech. What we do here is to re-synthesize the input speech wave by combining impulse responses of different electrode superimposing together with more than 1 ms interval which is a restriction of the receiver unit. We minimized error between input and re-synthesized speech wave. The re-

synthesized speech was perceptually evaluated as a simulated cochlear implanted speech. As a result we get a series of a pair of electrode number and activation time and magnitude of activation.

3.1. Decomposition into C.I. Parameters

Speech processing extracts C.I. parameters stimulation timing and electrode position and current intensities. In order to optimize these parameters, we regarded this process as decomposition into impulse responses using the multi-pulse coding algorithm applied to multi-channel coding. In the original form, each analysis frame is LPC analyzed then the frame wave is decomposed into number of pulses placed at somewhere in the frame and impulse response corresponding to the LPC parameters is used to decompose in to multi-pulses.

In C.I. system, LPC parameters and corresponding impulse response is related to each electrode channel. Under these given impulse response set, we can select one of impulse response switching one by one in a frame at a different time point and magnitude of current, since electrode can be stimulated one at a time.

As a result in a frame, number of pulses are placed at an appropriate time on an appropriate channel of electrode with appropriate magnitude of currents. These combinations of channels are supposed to approximate spectral pattern in a frame.

3.2. Acoustic Simulation

Recomposition process is nessary in order to evaluate information losses during C.I. parameterization. That is how much distortions have been incurred during decomposition. Electrode stimulation to a channel is regarded as a delta function or as a impulse and then acoustically simulated as an impulse response to that channel. A stimulation pulse sequence of each channel was convolved with the impulse response and finally summed for all channel to reproduce a simulated input acoustic signal that is regarded as simulated auditory perception of cochlear implantee.

4. SPEECH CODING ALGORITHM

The minimization process progresses as follows.
(1) Take a 20 ms frame of 8 k Hz sampled speech wave.
(2) Using multi-pulse coding algorithm(Ozawa, 1986) for each of 22 channel, find a most dominant magnitude pulse.
(3) Compare the sum square of error between original speech wave and coded wave for each one of 22 channel. Select the channel which minimizes the sum square error and then the location and magnitude and channel of the first pulse is determined.
(4) From the original speech wave, the impulse response at the first pulse position is subtracted. This residue signal is processed as a target, 2nd pulse is searched for in the same way as in (2).
(5) The process continues until 20 pulses have been determined. However, each pulse have more than 1 ms gap to keep the constriction of the receiver stimulator unit.

The following descriptions are processings in an analysis frame.

4.1. Representation of C.I. Parameters

We have 20 channels of C.I. electrode and one of them is selected and switched one by one. The predicted signal after k-th selection of stimulation pulse is

$$\bar{x}_k(n) = \sum_{l=0}^{k} g_l^{(i)} h^{(i)}(n - m_l)$$

where, k means the k-th pulse in a frame, i means the i-th channel, $h^{(i)}$ means the impulse response of the i-th channel, m_l means l-th pulse position in a frame and $g_l^{(i)}$ is the pulse amplitude.

4.2. Residue Error

The residue error between the input signal and the predicted signal is

$$e_{w,k}^{(i)}(n) = \left\{ x(n) - \bar{x}_k^{(i)}(n) \right\} * w(n)$$

where the error signal is weighted with the weighting function, x is a input signal, and w is the weighting function. This weighting is called as an auditory weighting in speech coding so as to reduce perceptual noise. In the coding algorithm, we intend different channel is likely to be selected once a channel was selected.

4.3. Error Minimization Criteria by Selection of Channel

Supposing i-th channel to be selected to find the next driving pulse, and supposing the series $\{m_k\}$ and $\{g_k\}$ have been found, then next channel i is selected so as to minimize the following mean square error and the find the sequence $\{i\}$.

$$\min_i \left[\sum_{n-1}^{N} \left\{ e_{w,k}^{(i)}(n) \right\}^2 \right]$$

Pulse positions under C.I. restrictions, i.e. 1 ms interval between pulses is also considered then minimization is under condition.

4.4. Solution with Multi-Pulse Coding

The k-th pulse in a frame is determined as follows: $h^{(i)}$ is the impulse response of the i-th channel, then the amplitude of a candidate k-th pulse is determined as follows;

$$g_k^{(i)}(m_k^{(i)}) = \frac{\varphi_{h_{h(i)}x}(m_k^{(i)}) - \sum_{l=1}^{k-1} g_l^{(C_l)} \varphi_{h^{(C_l)}h^{(C_k)}}(|m_l - m_k|)}{R_{h^{(i)}h^{(i)}}(0)}$$

where

$$\varphi_{h_{h(i)}x}(m) = \sum_{n=1}^{N} s_w^{(i)} h_w^{(i)}(n-m) \qquad (1 \le m \le N)$$

$$R_{h(i)h(i)}(0) = \Sigma h_w^{(i)}(n) h_w^{(i)}(n)$$

$$\varphi_{h(C_l)h(C_k)}(m) = \sum_{n=1}^{N} h_w^{(C_l)}(n) h_w^{(C_k)}(n-m)$$
$$(1 \le m \le L)$$

where $s_w^{(i)}$ is an auditory weighted (with respect to the i-th channel) input signal $s(n)$, $h_w^{(i)}(n)$ is an auditory weighted impulse response, and L is the sample length of the auditory weighted impulse response which is usually less than N.

4.5. Re-synthesized Sound

The analysis-synthesis sound $\bar{s}_k(n)$ was regenerated in the following way to evaluate the transferred speech information.

$$\bar{s}_k(n) = \sum_{l-1}^{20} \sum_{k-0}^{N-1} d^{(l)}(k) h(n-k)$$

where, $d^{(l)}(k)$ is the driving source sound for channel l. And,

$$d^{(l)}(k) = \sum_{j=1}^{K^{(l)}} g_j^{(l)} \delta_{k,m_j^{(l)}} \delta(C_j - l)$$

where, δ is a Kronecker's delta.

Reproduced sound is not auditory weighted that used during decomposition.

5. RESULTS AND DISCUSSIONS

The above algorithm has been implemented and tested as a software simulation.

5.1 Results

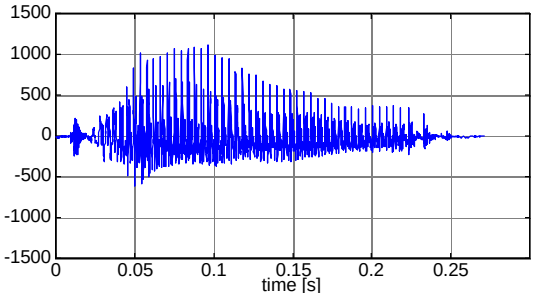
Figure 1 is an example showing how a CV syllable is decomposed into C.I. parameters and then re-synthesized to evaluate how quality of speech is degraded. Figure 1(a) shows the original speech wave. The original multi-pulse coding algorithm can reproduce original speech without degradation of quality if sufficient number of pulses are used. Even if limited number of pulses, such as less than 1 pulse within 1 ms interval, degradation of speech quality is just perceivable and the speech sound natural.

Figure 1(b) shows decomposed and reproduced speech wave applying the algorithm described in section 4. However any channel was selected without restriction, and pulses are placed at any place in a frame. There are obvious deformations in wave form. Degradation of speech quality is much worse than original multi-pulse coded speech with 1 pulse in a 1 ms interval.

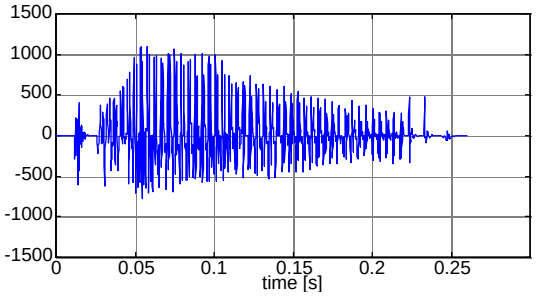
Figure 1(c) shows reproduced speech wave same as above 1(b) and with realistic C.I. restrictions: every pulse must be separated with more than 1 ms interval before and after the pulse, As we can see deformation of wave form, speech quality is degraded more than above examples and sound something unnatural, but still the original characteristics are kept.

We have the following findings:
(1) Selected electrodes corresponded to vowel formants.
(2) Simulated implantee's perception were much better than WSP. Comparing our simulated WSP speech, speech quality is much better and natural.
(3) Environmental sounds were well recognizable. Since this algorithm is based on wave form, this algorithm is applicable any kind of sound such as environmental sound and musics. We have tested some of those kinds of sounds and found results are successful.

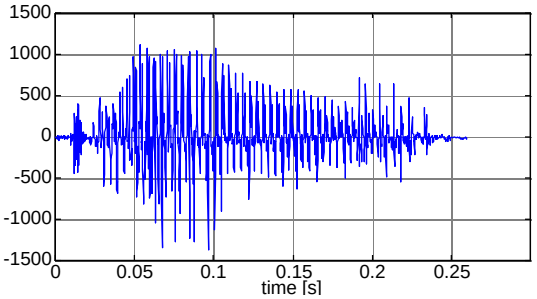
It takes a large computation to get the following results. For example, a CV syllable takes an hour of computation on a SUN 4 work station.



(a) Original speech wave form /ga/.



(b) Re-synthesized speech wave form from above (a).



(c) Re-synthesized speech wave form with restriction in electrode selection and interval between stimulation.

Figure 1: An example of speech coding and re-synthesized

speech wave form to show that the decomposition is successful even if under restrictions incurred into the cochlear implant system.

5.2. Discussion

Impulse response is not sufficient representation of electronic stimulation. Sound heard by implantee is not yet known but quite different from ordinary acoustic stimulation. Since the auditory nerve firing pattern is quite different between acoustic stimulation and electric stimulation. Therefore representation used here such as band-pass filtered acoustic noise is inappropriate for representation of electric stimulation. According to reports from implantees, hearing by electric stimulation is reported sometime as noise and sometime as unusual sound. It is very complicated and difficult to represent as a single impulse response. However we have some experience using this kind of band-pass filtered noise corresponding to electric stimulation and made simulation of WSP speech processor for intelligibility test of phonemes, words, and sentences[2]. The results were comparable with reported results with real implanted persons in such results of vowel recognition accuracy and consonant confusions. Therefore our tentative assumption is that approximation as a set of band-pass filters and their corresponding impulse responses are not too much deviate from real thing but we can draw some insight from this kind of experiments.

Large computation is required to process the above algorithm since pulse searching takes proportional time to number of electrode channels. Progresses of high-speed DSPs and parallel processing will realize real-time computation of the algorithm.

For the further study, we are studying some more realistic representation of electrically stimulated hearings where electrode current is converted into simulated auditory nerve firings and those firings are reverse processed to reproduce the virtual input sound. Then we can really simulate hearings of cochlear implantee's.

6. REFERENCES

1. Ozawa,K.et.al.,"A Study on Pulse Search Algorithms for Multipulse Exited Speech Coder Realization," IEEE Journal on Selected Areas in Communications, vol.SAC-4,No.1,pp.133-141, Jan,1986.

2. Kitazawa, S. "Hardware Simulation of Cochlear Implant System", Studia Phonologica, 23, 57-68, 1989.