STUDIA PHONOLOGICA XXIII (1989)

Hardware Simulation of Cochlear Implant System

Shigeyoshi KITAZAWA, Masatake DANTSUJI and Shuji DOSHITA

ABSTRACT

Cochlear implant systems have been developed and used for many patients, however, the processing system is still imperfect in transferring speech information. Usually, processing is evaluated with the cooperation of implanted patients, but this implies difficulties and limits the application. Obviously the speech processor needs to be improved because it extracts only limited speech features. In order to accelerate improvements, we developed acoustic simulator hardware which is intended to imitate the stimulation received by implanted patients. It consists of a set of active filters with operational amplifier circuits, like the original speech processor. Hard-wired analog circuits are used to achieve real time examination on various sources of speech material. The acoustic process and electro-neurophysiological process in action around the electrode implanted into the cochlea and haircell nerve were with band pass filters and stimulation pulse trains which are amplitude modulated. This system has been evaluated under several conditions, such as monosyllables, words, and sentences. Comparison with other reports, our results have shown that this hardware simulation approximately matches the articulation score of implanted patients. This simulator developed here provides a useful tool for evaluation of acoustic processing and parameter coding schemes for cochlear implant systems.

1. INTRODUCTION

Cochlear implant systems have been developed and used for patients over the world, including Japan. The number of implanted patients is increasing. Present state of the art requires months of postsurgery patient training to adapt to the artificial characteristics of the stimulus fed by the implanted cochlear electrode. Sound is coded by a special speech processor so called the Wearable Speech Processor (WSP in short). Although the effects of implantation system are remarkable, the level of speech communication achievable is still limited, even after

Shigeyoshi KITAZAWA(北澤茂良): Associate professor, Department of Computer Science, Faculty of Engineering, Shizuoka University
Masatake DANTSUJI(壇辻正剛): Instructor, Department of Linguistics, Faculty of Letters, Kyoto University
Shuji DOSHITA (堂下修司): Professor, Department of Information Science, Faculty of Engineering, Kyoto University

training. In order to improve the quality of information transferred to the artificial cochlea, repeated evaluation of the implant system, the speech processor in particular, is necessary. If every experiment require participation of real implanted patients, the number of experiments is necessarily limited. Therefore a sort of simulator needs to be developed to obtain performance similar to real patients, to improve processing algorithms. This paper describes our attenpt to develop an acoustic simulator.

2. TYPES OF COCHLEAR IMPLANT SYSTEM

Cochlear implant system currently used are classified in several categories : inside the cochlea/outside, analog/digital stimulation, single/multiple channel [8]. "Nucleus system", manufactured by Cochlear Inc. is an digital 22-channel inner type using stimulation by alternating current. This system is the most widely used, and has shown the most successful results. It has been applied several times in Japan [7].

The processing used by "Nucleus" has the following characteristics : (1) The features of a vowel can be represented by formants F1 and F2. Interpretation of F2 is loosely defined as the most dominant frequency component between 800 through 4000 Hz. F2 corresponds to the second formant of vowels, but can reach 10 kHz for some consonants. (2) There is no explicit way to represent features of consonants : Fundamental frequency varies randomly, and F2 takes on values in

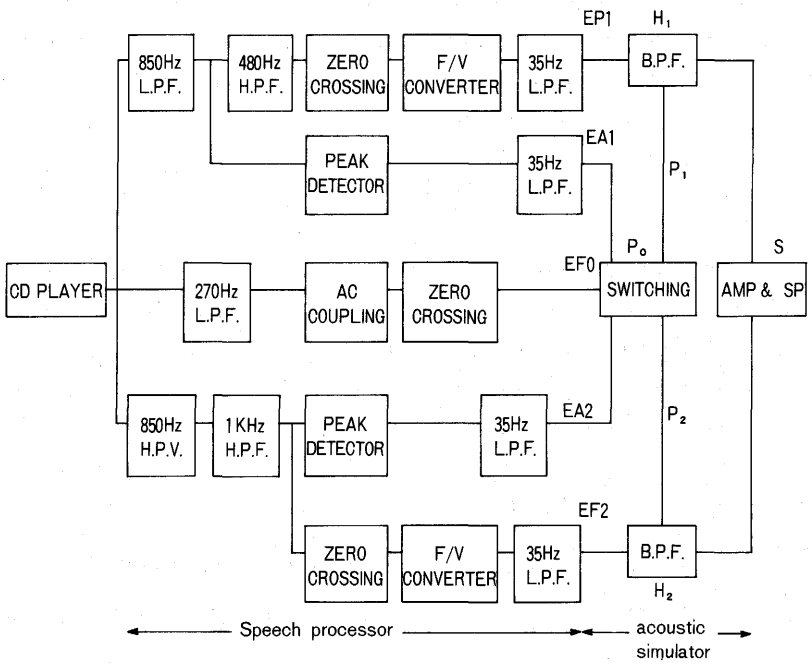

Figure 1. Block diagram of hardware simulator.

high and wide range.

2. Bis Our System

The speech processor (WSP) was rebuilt based on the reference article [4]. The block diagram is shown in Figure 1. The WSP extracts EF0, EF1, EF2, EA1, EA2 in real time with analog circuits. Where EF0 is the estimated fundamental frequency, EF1 is the estimated first formant frequency, EF2 is the estimated second formant frequency, EA1 is the estimated amplitude envelope of the first formant, and EA2 is the estimated amplitude envelope of the second formant frequency. Analog filters designed were using a general purpose operational amplifier (type 4558) to achieve -12 dB/oct slope characteristics of the second order Butterworth type. The time constants of the peakdetector circuits were set to 22-10 ms which corresponds to the lowest fundamental periods. The device used for frequeney-to-voltage conversion is the DATEL VFQ-3 chip which is inexpensive and has sufficient performance.

3. Acoustic Simulation

The parameters extracted by the WSP are pulse coded and transmitted to the receiver stimulator unit implanted under the skin using a high frequency modulated pulse which stimulates the auditory nerve with a biphasic pulse, to wake the perception of a speech-like sound. The "Nucleus" cochlear implant system functions as follows:

The cochlea performs a frequency analysis along the basilar membrane. A frequency component activates the characteristic auditory nerve adjacent to the frequency specific location.

An exponentially decaying sinusoidal periodic signal, like a formantic resonance, produces auditory nerve firing strongly synchronized with the driving fundamental frequency at the location around the central resonance frequency.

A pair of electrodes inserted into the cochlea produces auditory nerve firing that results in the perception of a noisy *"sound"*.

In order to simulate this process acoustically, an impulse train at the EF0 frequency is fed to a pair of band pass filters with center frequencies controlled by EF1 and EF2. This analysis and synthesis system is similar to an F1, F2-based two pole parallel formant synthesizer. The driving pulse is made from a zero crossing waveform of the EF0 and is reshaped into a 100 ns width impulse which is shown as P0 in Figure 1. The pulse frequency approximately corresponds to the vowel fundamental frequency but sometime rises to double pitch. During voiceless consonants it can rise to several thousand hertz.

This pulse train is amplitude modulated so that the peak value of the pulse

equals EA1 or EA2. Two pulse trains, H1 modulated by EA1 and H2 modulated by EA2, are thus composed. The center frequencies of each band pass filter are controlled by voltages directly derived from EF1 and EF2 respectively. The voltage tuned filters we used (DATEL FLJ-VB) are cascaded second order filters, to form a fourth order tuned Butterworth type filter (Q=5). The tuning characteristics approximately simulate critical bandwidth (Q=6). The sum of the two signals from the bandpass filters, results in the synthetic speech signal.

4. ANALYSIS OF SYNTHETIC SPEECH BY SONAGRAM

The synthetic speech signal produced by our hardware simulator was analyzed by a digital sonagram program running on a MACHINTOSH, and compared with the original speech. The original is shown in Figure 2, and the processed speech Figure 3, for five Japanese vowels /a, i, u, e, o/. Basically, the first and second formant are well represented in the processed speech and the envelope of the waveform is very similar to the orignial. We can observe some vertical lines around the middle part of each vowel which implies that the parameters EF0, EF1, EF2, EA1, EA2 are approximately correct. If we look in greater detail, we can see, for example, in the case of /i/ and /u/, a tendency for the third formant to be extracted instead of the second one. Owing to this fact, sonagram patterns of /i/ and /u/ become very similar. This is also true for some /a/'s where extracted second formant is somewhat higher than the true F2, or nearly equal to F3. This is because higher frequency components are enhanced by the cascaded high pass filters of cutoff frequencies of 850 Hz and 1000 Hz. The range between these two filters is emphasised with a slope of 12 dB/oct. This high boost is needed for the

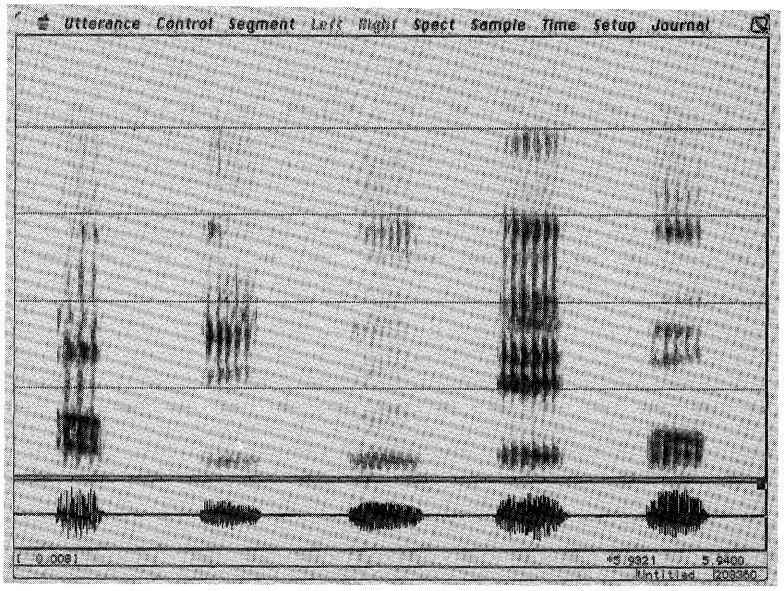

Figure 2. Sonagram of original speech of Japanese five vowels (/a i u e o/).

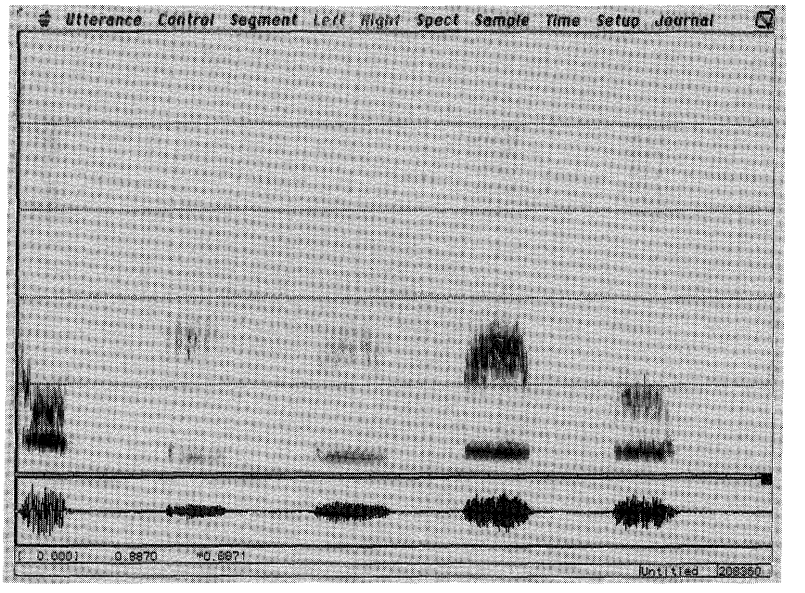

Figure 3. Sonagram of processed speech with the hardware simulator. Syllables are Japanese five vowels corresponding to Figure 2.

separation of F2 from a high F1.

For the sake of perceptual vowel separation in the F1-F2 domain, the F2′, introduced by Carlson et al. [12] is preferable to the actual F2. The F2′ requires to be lower in case of high F1 and lower F1 is combined with higher F2 or approximately F3 is chosen. This is somewhat contradictory situation because it is usually difficult to separate two formants of high F1 and low F2 and it is also difficult to extract high F2 and F3 separately. The optimal F2′ value is not easily extracted with a simple hardware : some higher level processing is necessary.

Envelope detection is also a difficult task for analog hardware. We can compare waveform envelopes displayed at the bottom of the two sonagrams. They are approximately similar each other, but detailed comparison shows that the processed speech has less dynamic range than original. The shape of the envelope depends on the precision of the peak detector and the relative timing of the driving pulse with respect to the peak of the waveform. The peak detector is made fast so it can follow the sharp rise of amplitude peak during a stop burst or fricative noise, or the impulse resonance of higher formants of front vowels. In particular, a higher speed is required to detect the peak value for the second formant band than for the first formant. The detected phase of the driving pulse does not necessarily correspond to the phase of the amplitude peak, because the EF0 is obtained from the zero-crossing waveform of the low-pass filtered speech component, and the zero-crossing is then differentiated to get a thinner pulse, like a delta function, which means the rising edge and the falling edge of the zero-crossing, is detected, therefore the resulting driving pulse shifts some amount of time forward or backward. Another error in both amplitude and timing of the driving pulse is

due to the low-pass filtering (35 Hz) after detection. In order to follow the envelope change, the peak detector has a rather long discharging time constant, so the envelope signal is an amplitude modulated sawtooth pulse train. Low-pass filtering of this sawtooth pulse train introduces a delay, that is, the maximum value of the amplitude of the low-pass filtered envelope does not correspond to the maximum of the original signal. Consequently, the dynamic range is considerably reduced.

These considerations explain the differences in envelope. This probably affects intelligibility of consonants more seriously than that of vowels.

5. PERCEPTUAL TESTS OF RESYNTHESIZED SPEECH

The perceptual tests were carried out as follows.

(1) Source speech: The original speech source is part of the speech data base constructed by Denshi-kyo (The Electronics Manufactures Association) and titled "The standard data of Japanese: place names". The data base consists of Japanese 100 city names selected such that occurrences of various syllable are balanced. The data of four male and four female speakers were used. (2) Stimulus speech: The stimulus speech prepared for perceptual tests consists of the original speech plus the speech processed in 6 ways as shown in Table 1., using the speech processor described in the previous sections. In the first condition, EF0 was extracted to the original description, from the 270 Hz low-pass filtered signal of the fullwave rectified full-band speech. In the second condition, EF0 was synchronized with the gating signal of the peak detector for the F1 band. In the

Table1. Results of the Identification Test

Stimulus Number		Total errors	Percentage of errors	Percentage of Correct Responses
1	F0=original, F1, F2	149	18.6	81.4
2	F0=F1, (F1), F2	104	13.0	87.0
3	F0=original, F1=fixed(500 Hz) F2=fixed(1500 Hz)	141	17.6	82.4
4	F0=fixed(120 Hz), F1, F2	210	26.3	73.7
5	F1 only	241	30.1	69.9
6	F2 only	385	48.1	51.9

third condition, the center frequencys for EF1 and EF2 band-pass filter were fixed, but amplitude modulated with EA1 and EA2. In the fourth condition, F0 is fixed to 120 Hz, therefore no prosodic information was conveyed by the fundamental frequency. The fifth and sixth condition use only one formant. (3) Subjects: The subject is a trained phonetician, one of co-authors, who carefully tried to discriminate the stimuli.

Table2. Perception test results of the place names.

(a) Confusion Matrix for Vowels.

C\E	i	e	a	o	u	%
i	1334	3			7	99.3
e	32	852	3	1		95.9
a		2	1506	1	3	99.6
o	1	6	9	1254	2	98.6
u	81	19		10	1042	90.5
average						97.1

(b) Confusion Matrix for Consonants

C\E	p	t	k	b	d	g	s	h	z	m	n	r	j	w	N	Q	R	Ø	%
p	48																		100
t		573	28				1	3	1									18	94.6
k		48	584			11											1	4	90.7
b				400				4		3	1								98.0
d					240														100
g						383			2	5	12		1					5	95.0
s							576	2	2										99.3
h				22		3	2	426		2	18							55	90.0
z				2		9	1		417		1							2	97.0
m				8		8				493	6		1	9				3	93.9
n				17	1	10				24	420		5					3	88.1
r				5		8						389		2				4	96.3
j										25			468					11	94.9
w											2		21	97					80.8
N															479			1	100
Q																72			100
R															1		667	4	99.9
Ø		2		3			1	16		4			30		1	1	8	4	—
E	p	t	k	b	d	g	s	h	z	m	n	r	j	w	N	Q	R	Ø	
N→i	1														average				95.2
Ø→i	1																		
N→u	1																		

The obtained results are summarized in Table 2. as a confusion matrix. The first three conditions were pooled because there was no significant difference observed between them. The major results are as follows.

The highest discrimination score was observed with the pooled with condition 1 and 3 where EF0 is extracted from F1. The next was the first one, the original form of processing.

Analyzing misperceptions of vowels, we find that a number of /u/ are heard as /i/. Phonologically this phenomena can be interpreted as follows (see Figure 4. to follow) ; phoneme /u/ is usually non-labialized as [ɯ] by "Kanto" dialect and furthermore it is sometimes centralized and advanced as [ɯ̈+], then among distinctive features which reside in original contrasting phonemes between [i] and [u] such as the feature [+/−back] and [+/−labial], the one is neutralized to [ɯ]

	i	u	ɯ	ü+
high	+	+	+	+
low	−	−	−	−
back	−	+	+	−
labial	−	+	−	−

Figure 4. Distinctive feature matrix of Japanese vowel /i/ and /u/.

and the other is also neutralized to [ü+]. As a result, the distinctive contrasts between /u/ and /i/ are neutralized.

Spectrographic observation has revealed that F3 rather than F2 tends to be extracted from /i/ and /u/. This is one of direct factor for the confusion.

6. DISCUSSION

Comparisons with previous researche is necessary to validate our results. There are no comparable reports in Japan, but we could find several in Australia, although for a different language.

The acoustic simulation of the cochlear implant system has been described by the Australian group of University of Melbourne as an acoustic model of a multiple-channel cochlear implant. They prepared eight bandpass filters to represent electrodes 0 to 7 of an implant patient. The filters corresponded to 1140 Hz to 10880 Hz covering only the F2 region, and each electrical pulse was represented by a pulse of acoustic noise. [1], [2] In the acoustic model reported in 1985[3], they employed 12 channel filters corresponding to 320 to 10880 Hz to cover 200 to 4000 Hz of speech analyzed into F1 and F2 parameters. They achieved 75% correct vowel recognition and 58% consonant recognition.

Acoustic simulation has also been tried in Japan. Shoji et al. simulated the 'sounds' implanted patients hear through multi-electrode cochlear implants[9]. In order to simulate vowels, they synthesized waves which have translated formant frequencies by reverse Fourier transformation. Through a digital-analog converter, they listened to the synthesized sounds. Synthetic sounds are very different from real vowels, but they can easily be distinguished each other. The characteristic frequency estimated based on the place theory was used to synthesize each formant.

Sakakibara et al. simulated the 'sounds' by a waveform summation method[10]. From the measurement of electrode pulses, they composed a sinusoidal burst of 5ms duration. Pitch perturbation drawn from the speech processor sounded more natural than a fixed pitch. Frequency transformation by the Mel-scale emphasized the phonetic characteristics of the vowel.

The multi-electrode cochlear implant has been evaluated by implanted patients in both Australia and Japan. The evaluation is mainly in three parts;

phoneme or syllable recognition rate, word recognition rate and sentence recognition rate. Evaluation conditions are hearing without lip reading, and composite recognition with hearing and lip reading. Here we are interested in the speech processing strategies but not the overall perceptual capability achievable by the implantation, it is reasonable to compare phoneme level recognition rate because word and sentence recognition depends on lexical and higher level intelligence and lip reading induces more complexity to evalute speech processing mechanism. Results are summarized in Table 3. including acoustic simulation.

Table3. Evaluation of the multi-channel cochlear implants

	parameters	ranges	cochlea	vowel	consonant	
acoustic						
Dantsuji'89	EF0EF1EF2	200-10000Hz	200-10000Hz	97%	95%	cityname
Kitazawa'89	EF0EF1EF2	200-10000	200-10000	84	40	monosyllable
Kitazawa'89	EF0EF1EF2	200-3000	1000-8000	25	18	no traning
Blamey'85	EF0EF1EF2	300-4000	300-10880	75	58	12 traning
implants						
Fukuda'89	EF0EF1EF2	300-4000		75	28	monosyllable
Kyoto'87	EF0EF1EF2	300-4000		90	27	monosyllable
Blamey'87	EF0EF1EF2	300-4000	600-8000	70	71	3Best /hVd//aCa/
Blamey'87	EF0EF1EF2	300-4000	600-8000	49	37	Average
Blamey'84	EF0EF2	800-4000	1140-10880	48	38	(30) 4 choice 7 channel
Blamey'84	EF0EF2	800-4000	1140-10880	39	51	(23) acoustic

Comparison between Australian and Japanese is illustrated in Figure 5 concerning the vowel and consonant recognition rate. In terms of phoneme recognition rate, Japanese achieves much better recognition rate in vowel discrimination compared with Australin, while recognition rate in consonants in Japanese is much inferior to Australian. This result can be explained as follows. Japanese has just 5 vowels while there are 11 vowels in English. If the confusion between vowels is assumed to be independent, the correct recognition rate of each vowel would be the fifth root of the observed in Japanese, while the 11th root of that observed in Australian. Then we obtain almost the same recognition rate on the both language. This may be a reasonable result, since the cochlear implant system extracts basically only formant information, if it is true in any language that vowels are recognized based on formants.

There seems to be diffrences in consonant recognition rate between Japanese and English, i. e., the recognition rate for Japanese is significantly lower than that of English. Differences between Japanese and English are that the former presents each consonant in monosyllables while the latter presents in the context of /aCa/. This does not suffice to explain the observed gap between two languages. A hypothetic interpretation would be that English listener relies on temporal cues more than Japanese do, since the prosodic information is reserved well in the stimulation intervals.

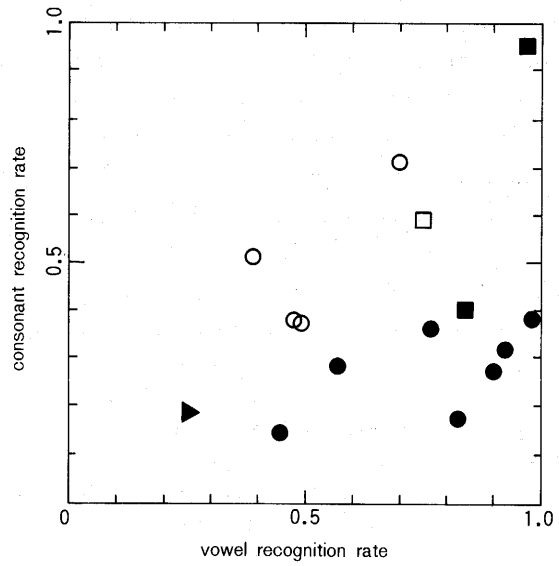

Figure 5. Comparison of recognition rate between English and Japanese along the dimensions of vowel and consonant recognition rate. Black circles indicate Japanese cochlear implant patients, white circles indicate English cochlear implant patients, back squares indicate acoustic simulation in Japanese, a white square indicates acoustic simulation in English, and a black triangle indicates acoustic simulation in Japanese with frequency shifting. All the data are drawn from our experiments and reference papers listed at the end of the text.

Fukuda et al.[7] report articulation scores of implanted patients of 23% for monosyllables, 75% for vowels and 28% for consonants under hearing only condition. This shows significant better recognition of vowels compared to that of consonants.

Similar results were observed in Kyoto University ; 90% of vowels and 27% of consonants. These results observed in two different institutes are consistent.

These findings are in good agreements with our observations on our acoustic simulation. In Table 3. we compared previous reported results of experiments in terms of vowel and consonant recognition rate estimated.

Effects of frequency shift occur because of the location of electrode inserted in the cochlear. Usually the electrode is placed near the entrance far from the apex, therefore the frequency zone below 1 kHz is not stimulated. As far as our results are concerned, an upward frequency shift of 1kHz significantly degrades phonological identification. At present we have no idea to explain the performance gap between implanted patients and acoustic simulation in identification test of vowels, that shows that implanted patients can discriminate vowels as well as acoustic simulation before frequency shifting. In Australian experiments which use acous-

tic stimulus taking the frequency shift into account, the score achieved by subject was better than ours and almost equal to the score by patients, although their subject were trained for one and a half month while we used naive subjects. Another important point is that their frequency range was 600 to 10880 Hz which is widened to cover important information under 1 kHz region.

In Blamey's report [1984], however, they tried frequency shift 800-4000 Hz to 1140-10880 Hz only for F2 range. The recognition score of vowels was a bit lower than implanted patients. But in Blamey's report [1985], they do not warp frequency for some reason.

7. Conclusion

Acoustic simulation of stimulation of cochlear implanted electrode was developed in a hardware equipment and evaluated.

Recognition rate of monosyllabic speech is approximately equal to other previous results.

The amount of training or experience increases the recognition rates.

Among parameters extracted with the speech processor some of 5 can be fixed without significantly degrading recognition rate.

Frequency shift to a higher region considerably degrades recognition rate.

Acknowledgement

We would like to thank Prof. Iwao Honjo, Dr. 's Junji Sakakibara, Michio Kawano and Juichi Ito of faculty of medicine of Kyoto University and Tomoji Haji of Shizuoka Hospital, who suggested direction of our research. We would also like to thank Prof. Tatsuo Nishida who have supported our research in many ways.

References

[1] P. J. Blamey, R. C. Dowell, Y. C. Tong, and G. M. Clark, *An acoustic model of a multiple-channel cochlear implant,* J. Acoust. Soc. Am. 76(1) 97-103 (1984).

[2] P. J. Blamey, R. C. Dowell, Y. C. Tong, A. M. Browns, S. M. Luscombe, and G. M. Clark, *Speech processing studies using an acoustic model of a multiple channel cochlear implant,* J. Acoust. Soc. Am. 76(1) 104-110 (1984).

[3] P. J. Blamey, L. F. A. Martin, and G. M. Clark, *A comparison of three speech coding strategies using an acoustic model of a cochlear implant,* J. Acoust. Soc. Am. 77(1)209-217(1985).

[4] P. J. Blamey, R. C. Dowell, and G. M. Clark, *Acoustic parameters measured by a formant-estimating speech processor for a multi-channel cochlear implant,*J. Acoust. Soc. Am. 82(1) 38-47(1987).

[5] P. J. Blamey, R. C. Dowell, A. M. Brown, and G. M. Clark, *Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors,* J. Acout. Soc. Am. 82(1) 48-57 (1987).

[6] M. Dantsuji and S. Kitazawa, *An auditory study on distinctive features using a speech processor,* The 1989 Autumn meeting of the Acoust, Soc. Jpn. 3-2-9 371-372 (1989).

[7] Y. Fukuda, M. Shiroma and S. Funasaka, *Speech Discrimination by Patients with Cochlear Implants (Nucleus) through Combined Use of Auditory and Visual Stimuli,* "Jpn. J. Logop. Phoniatr. 30 334-339 (1989).

[8] J. Sakakibara and J. Ito, *Hardware of Cochlear Implants —Discussion of Speech Processing—*, Jibirinsho 82 2 181-188(1989).

[9] K. Shoji, K. Omori, J. Tsuji, J. Ito and I. Honjo, *Simulation of Vowels through Multi-electrode Cochlear Implant,* Jibirinsho 81 12 1709-1713(1988).

[10] J. Sakakibara, H. Kojima, I. Honjo and S. Fujita, *Simulation of Vowels heard through Multi-electrode Cochlear Implant*, Audiology(1989).

[11] S. Kitazawa and M. Dantsuji, *Acoustic Simulation of Cochlear Implant System,* The 1989 Autumn meeting of the Acoust, Soc. Jpn. 2-2-20 351-352 (1989).

[12] R. Carlson, G. Fant, and B. Granstrom *Two-formant models, pitch, and vowel perception,* in Fant ed. (1975).