Title	Simulation of Cochlear Implant Using the Mel Scale
Author(s)	Sakakihara, Junji; Takeuchi, Mariko; Kojima, Hisayoshi; Ito, Juichi; Honjo, Iwao
Citation	音声科学研究 = Studia phonologica (1990), 24: 55-60
Issue Date	1990
URL	http://hdl.handle.net/2433/52483
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

STUDIA PHONOLOGICA XXIV (1990)

Simulation of Cochlear Implant Using the Mel Scale

Junji SAKAKIHARA, Mariko TAKEUCHI, Hisayoshi KOJIMA, Juichi ITO
and Iwao HONJO

I. INTRODUCTION

Nucleus' multiple-electrode cochlear implant (CI) selects and activates 2 electrodes from 22 scala tympani electrodes in the cochlea according to the first and second formants extracted by the speech processor[1)].

Vowel recongition by this device is not complete even though its stimulation strategy is aimed at vowels[2)]. The major reason may be that the stimulation site is restricted to the basal turn which elicits high pitch perception and causes formants to shift to higher frequency region.

In 1940, Stevens et al.[3)] proposed the mel scale (the unit "mel" originates from the word melody) as a psychophysical scale of pitch. They defined the pitch of 1 kHz pure tone as 1000 mels, a pure tone which has twice as high as a pitch of 1 kHz as 2000 mels, a pure tone which has half as high as a pitch of 1 kHz as 500 mels, and so on.

The mel scale is thought to reflect the way in which a change in the stimulus frequency causes a change in the point of maximum vibration along the basilar membrane[4)]. Thus, we tried to improve vowel discrimination in CI patients by formant shift using the mel scale.

Shoji et al.[5,6)] simulated vowels and consonants perceived by CI patients, and pointed out the possibility of improving speech discrimination using a CI. Through the simulation using the mel scale, we evaluated the efficacy of formant shift, and then applied this strategy in CI patients.

II. SUBJECTS AND METHODS

① Simulation of vowels perceived by CI patients.

The position of the electrode array in the cochlea was investigated using the postoperative X-ray film of one CI patient with good speech perception (Case 1). The characteristic frequency (CF) at each electrode was determined according to the place theory. Next, output from the electrodes of one receiver-stimulator was digitally recorded when a Japanese vowel sample /u,o,a,e,i/ was inputted to his speech

---

Junji SAKAKIHARA (榊原淳二), Mariko TAKEUCHI (竹内真理子), Hisayoshi KOJIMA (児嶋久剛), Juichi ITO (伊藤壽一), Iwao HONJO (本庄 巖): Department of Otolaryngology, Faculty of Medicine, Kyoto University
Shuji FUJITA (藤田修治): Osaka Red Cross Hospital

processor and coded signals were sent from it to the receiver-stimulator. Accurate values of the number of electrodes and the timing of the discharge were measured with NEC's 16 bit computer and a commercial software program for analysing sound waves (Wave Master^{TM}).

Thereafter, we simulated vowels which may be "heard" by CI patients. Tone bursts were synthesized with a 16 bit home-made software program written in C with a sampling frequency of 20 kHz. Each tone burst was comprised of two pure tones. Whose frequencies approximately corresponded to CFs at two positions which were stimulated quasi-simultaneously by intracochlear electrodes to represent the first and second formants. The duration of tone bursts was 5 ms, and their frequencies were integer multiples of 200 Hz in order to synchronize their phases at the beginning and end of tone bursts. In addition, the frequencies approximated the CFs at the stimulated position in the cochlea as much as possible. The tone bursts were arranged on the time-axis according to the temporal information obtained from the stimulation. Manipulation of the tone bursts was accomplished using Wave Master^{TM}.

A discrimination test of simulated vowels was performed by 5 normal untrained subjects. The simulated vowels were presented randomly 10 times through a digital audio taperecorder. Sound spectrograms of the simulated vowels were obtained using an Apple Macintosh II^{TM} computer and the MacSpeech Lab II^{TM} software program.

② Simulation of formant shifts using the mel scale.

The most prominent peaks in the FFT power spectrum of the speech sounds obtained by Wave Master^{TM} were taken as formants, and used in the calculations below.

A function which converts frequency F (Hz) to M (mels) within a limited frequency range is described as follows.

$$M = (1000/\log 2) \times \log (F/1000+1)^{7)}$$

By this equation, all the formant frequencies of the speech sample were converted to mels. The lowest formant (first formants of /u/ and /i/, 254 Hz, 327 mels) was shifted to 1000 Hz (1000 mels) which corresponded to the lowest CF of the most apical electrode. The other formants were also shifted the same mel width (673 mels). With these data, the frequencies of shifted formants were obtained by back-calculation and vowels were simulated using the same method as described in ①. Using these simulations, discrimination tests were performed by the same 5 normal subjects.

③ Application of the mel scale to two CI patients.

To evaluate the efficacy of the mel scale, we conducted discrimination tests on two successful CI patients, case 1 and 2. The appropriate synthetic vowels were inputted to the speech processor which activates electrodes in the cochlea where

CFs approximate formant frequencies shifted by the mel scale. Table 2 shows the frequencies of the synthetic vowels and the number of the electrodes activated.

III. RESULTS

The position of the electrode array in one patient's cochlea is shown in Fig. 1, and the electrode discharge pattern of the vowel /a/ is shown in Fig. 2. An example of the simulation is shown in Fig. 3 (from our previous report). Fig. 4 shows a sound spectrogram of this simulated /a/ sound, which is quite similar to Fig. 2, indicating the correct or proper simulation.

The synthesized vowels /u,o,a,e,i/ sounded like "Donald Duck speech" and greatly differed from the original speech sounds.

Table 1. Data of five Japanese vowels pronounced by an adult male (fundamental frequency was 127Hz).

Vowel		Original vowel sound	Numbers and CFs of the intracochlear electrode	F1 and F2 shifted by the mel scale
/u/	F1	250 Hz	20 (1000 Hz)	1000 Hz
	F2	1260	10 (2800)	2200
/o/	F1	500	18 (1200)	1400
	F2	740	12 (2200)	1600
/a/	F1	740	15 (1600)	1800
	F2	1120	13 (2000)	2200
/e/	F1	510	18 (1200)	1400
	F2	2010	9 (3200)	3800
/i/	F1	250	20 (1000)	1000
	F2	2290	6 (4600)	4400

Table 2. Electrodes activated by live speech and synthetic vowels whose formant frequencies were shifted according to the mel scale.

Case Vowel (F0=130Hz)		case 1		case 2	
		live	synthetic	live	synthetic
/u/	F1	20	20 (260 Hz)	18	19 (260 Hz)
	F2	10	11 (1560)	14	12 (1560)
/o/	F1	18	16 (520)	17	18 (390)
	F2	12	14 (910)	15	15 (1040)
/a/	F1	15	14 (910)	16	15 (780)
	F2	13	11 (1560)	14	13 (1560)
/e/	F1	18	17 (520)	18	16 (520)
	F2	9	9 (2080)	10	9 (2730)
/i/	F1	20	20 (260)	19	19 (260)
	F2	6	8 (2470)	9	8 (2990)

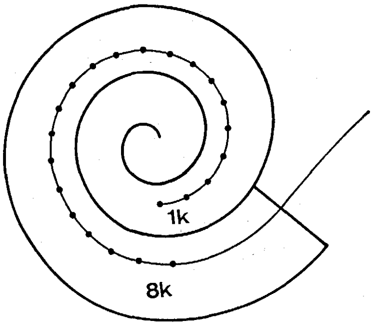

Fig. 1. The positions of 22 electrodes in the cochlea of case 1.

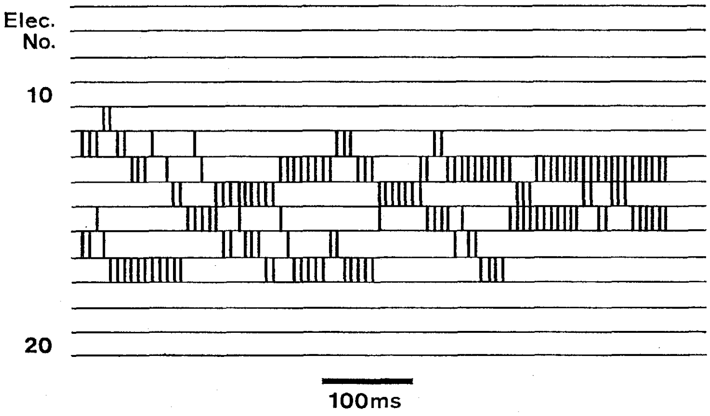

Fig. 2. Electrical discharge from each electrode in vowel /a/.

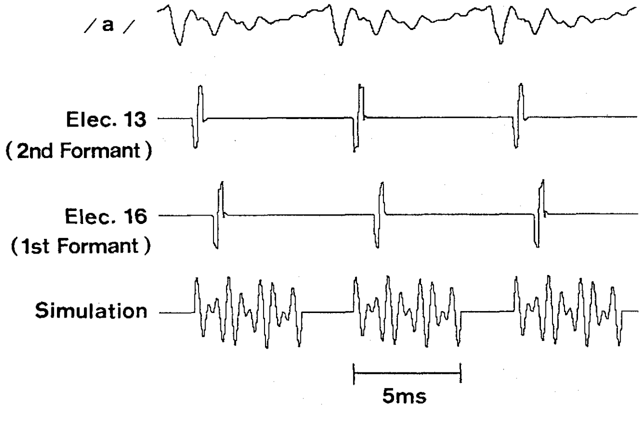

Fig. 3. An example of the simulation.
The original speech sample /a/ (top), output from the two electrodes corresponding to the second and first formants, and the simulated sound wave (bottom). (redrawn from Shoji et al).

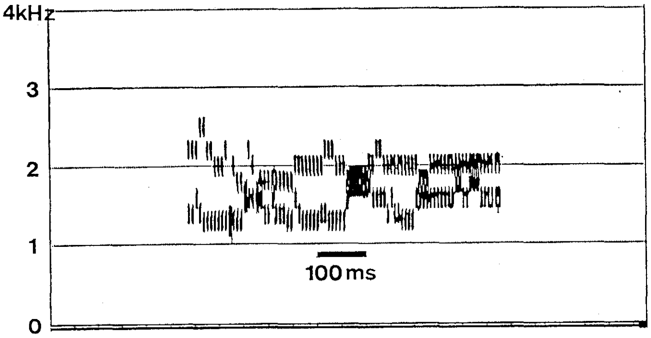

Fig. 4. A sound spectrogram of the simulated /a/ sound.

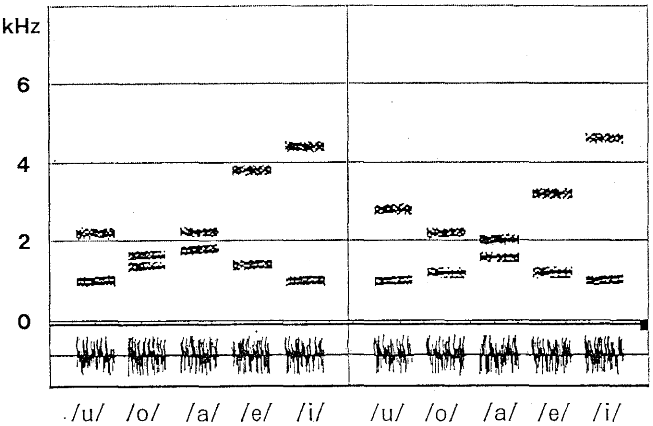

Fig. 5. Sound spectrograms of simulated vowels, the ones CI patient perceive (left), and the ones whose formants were shifted by the mel scale (right).

Table 1. shows the first and second formant frequencies of the original vowel sounds, the numbers and CFs of the intracochlear electrodes which were mainly activated, and the final formant frequencies, after being shifted by the mel scale. Sound spectrograms (Fig. 5) show the difference between the two groups of simulated vowels, the ones CI patients perceive (left), and the ones whose formants were shifted by the mel scale (right). A score of 140/250 (56%) on the discrimination test performed by normal subjects was obtained for the left group, whereas the score improved significantly to 175/250 (70%) for the right group.

A vowel discrimination test was performed by two CI patients. During this test, the appropriate electrodes, according to the mel shift were activated. Case 1 and 2 scored 96% and 80% respectively. These scores were better than their best (88% and 63%, respectively) during their ordinary rehabilitation.

IV. Discussion

At first, we evaluated the efficacy of the formant shift using the mel scale in normal subjects. The discrimination scores improved by the use of the mel scale. It was therefore suggested that the psychophysical scale of pitch should be taken into account during a formant shift. Secondly, we evaluated this formant shift in two CI patients, and in both cases the discrimination score improved. We supposed that the mel scale may be effective in electrical stimulation as well as in acoustical stimulation.

It is generally accepted that the multiple-electrode CI is able to utilize the place coding of pitch in the auditory nerve innervating to the cochlea, e.g. electrical stimulation of different places along the cochlea elicits different pitch sensations[8]. Because the mel scale represents a correlation between the basilar membrane position and the pitch sensation, it is thought that the relative position of formants on the mel scale must be preserved for better vowel discrimination even after the formants are shifted to a higher frequency region.

At present, we use the same electrode frequency assignment table for all patients regardless of the degree of electrode insertion into the cochlea. In view of the auditory psychophysics, it may be appropriate to modify this table for each patient, especially in a case with shallow insertion, where the formant shifts become large and the number of available electrodes is limited. In this case, vowel discrimination becomes more difficult. Formant shift using the mel scale may be helpful to some extent in such a case.

This device send the fundamental frequency of the vowels to the auditory nerve by activating intracochlear electrodes in the repetition rate corresponding to the voice pitch. The fundamental frequency therefore takes no shift regardless of the formant shifts. This dissociation between the fundamental frequency and the formants may cause the vowel discrimination more difficult. It will be appropriate to rise the stimulation rate according to the mel scale. In order to confirm this in CI patients, some alteration of the speech processing strategy is necessary.

REFERENCES

1) Seligman, PM, Patrick, JF, Tong, YC, et al: A signal processor for a multiple-electrode hearing prosthesis. Acta Otolaryngol Suppl 411: 135–139, 1984.
2) Ito, J, Tsuji, J, Sakakihara, J, et al: Studies on speech perception by cochlear implant. Studia Phonologica 22: 50–55, 1988.
3) Stevens, SS, Volkman J: The relation of pitch to frequency: a revised scale. Am J Psychol 53: 329–353, 1940.
4) Zwislocki, J: Analysis of some auditory characteristics. in Handbook of Mathematical Psychology Vol 3 (ed by Ruce RD, Bush RR, and Galanter EH). Wiley, New York, 1965.
5) Shoji, K, Fujita, S, Omori, K, et al: Simulation of vowels through multi-electrode cochlear implant. Pract Otol 81: 1709–1713, 1988. (in Japanese)
6) Shoji, K, Sakakihara, J, Fujita, S, et al: Simulation of consonants with a multi-electrode cochlear implant. Pract Otol 82: 1569–1573, 1989. (in Japanese)
7) Miller, JD: Auditory-perceptual interpretation of the vowel. J Acoust Soc Am 85: 2114–2134, 1989.
8) Evans, EF: How to provide speech through an implant device. Dimensions of the problem: An overview. Cochlear Implants (ed by Schindler RA and Merzenich MM). pp 167–183, Raven Press, New York, 1985.