DESIGN OF COCHLEAR IMPLANT DEVICE FOR TRANSMITTING VOICE PITCH INFORMATION IN SPEECH SOUND OF ASIAN LANGUAGES

*HIKI, Shizuo, IMAIZUMI, Kazuya and FUKUDA, Yumiko**

Graduate School of Human Sciences, Waseda University, Tokorozawa, Japan
*Research Institute, National Rehabilitation Center for the Disabled, Tokorozawa, Japan

ABSTRACT

Resolution of the fundamental frequency of speech sound required for the design of a speech processor of a cochlear implant device is investigated, with special regard to transmitting voice pitch information in Asian languages.

Clinical application of the cochlear implant has spread rapidly in recent years to Asian countries where a variety of languages having different voice pitch information from English and other European languages are spoken.

The perceptually acceptable area and required resolution of duration and fundamental frequency is estimated on a two-dimensional chart consisting of logarithmic time and frequency scales, based on the typical voice pitch contours of Japanese word accent and Chinese syllabic tone.

As a result, it is shown that much finer quantizing and time sampling for the change in fundamental frequency is required compared with sentence intonation and emphasis common to other languages. It is also shown that the amount of information conveyed by combined use of lipreading with a cochlear implant is not sufficient for supplementing the voice pitch information.

A possible way of transmitting such voice pitch information by transmission of the waveform of speech sound directly to the auditory area of cortex, where the waveform is reconstructed and voice pitch is extracted, is discussed.

1. COCHLEAR IMPLANT

1.1. Processes involved in the device

In a cochlear implant device (or an artificial inner ear device), speech sound is converted into an electrical signal and is transmitted to the electrode implanted in the inner ear, in order to directly stimulate the auditory nerve of a deaf person.

The processes involved in a cochlear implant device are schematized in Figure 1 (References 1 and 2). Speech sound through a microphone and a pre-processor is input to a bank of band-pass filters, which are assigned from low to high frequency components of speech sound waveform. The output, after being rectified and smoothed by low pass filters, is fed to the speech signal processor, where the peak channels are enhanced, and formants and voice pitch are extracted.

The information from the filters is encoded to map to the electrodes, transmitted transcutaneously from the radio frequency sender to receiver, and decoded and fed to the electrode array.

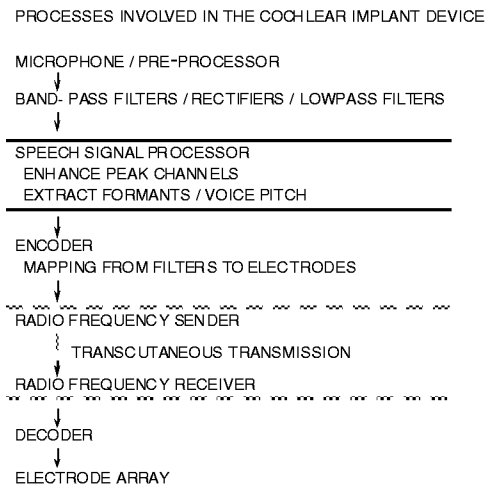

Figure 1: Processes involved in a cochlear implant device

1.2. Spread to Asian countries

Various types of cochlear implant devices have been developed in the United States, Australia and European countries, and clinically applied to several thousand deaf persons around the world effectively (Reference 3).

In recent years, clinical applications of the cochlear implant are rapidly spreading to Asian countries such as Korea, China, Taiwan, and Japan as well (References 4, 5, 6, 7 and 8). Information conveyed by the voice pitch of the speech sound of the languages spoken in these Asian countries, however, is quite different from that of English and European languages.

2. SPEECH INFORMATION

2.1. Phonemic and prosodic features

Speech information can be divided into two categories, namely; phonemic features and prosodic features. Phonemic features are mostly conveyed through the sensation of timbre, which is caused by a frequency spectrum or formant frequency of speech sound, while, prosodic features are conveyed through the sensation of loudness and voice pitch, which are caused by amplitude and fundamental frequency.

2.2. Syllabic pitch change

The relationship between acoustical property, auditory sensation and speech information, and change in fundamental frequency is schematized in Figure 2.

Among the speech information conveyed by the time change in fundamental frequency, there is change within segments of the syllable length for word accent in Japanese and syllabic tone in Chinese, in addition to change over segments of the phrase, clause and sentence for sentence intonation and emphasis found common to many other languages. Therefore, those Asian languages require special consideration concerning the design of a cochlear implant device regarding fundamental frequency transmission.

ACOUSTICAL PROPERTY	FREQUENCY SPECTRUM (FORMANT FREQUENCY)	FUNDAMENTAL FREQUENCY
		AMPLITUDE
AUDITORY SENSATION	TIMBRE	VOICE PITCH
		LOUDNESS
SPEECH INFORMTION	PHONEMIC FEATURES (SEGMENTAL FEATURES)	PROSODIC FEATURES (SUPRA-SEGMENTAL FEATURES)
	VOWELS	SENTENCE INTONATION
	CONSONANTS	EMPHASIS
CHANGE IN FUNDAMENTAL @ FREQUENCY*		*OVER SEGMENT OF PHRASE / CLAUSE / SENTENCE
		WORD ACCENT
		SYLLABIC TONE
		*WITHIN SEGMENT OF SYLLABLE LENGTH

Figure 2: Speech information and change in fundamental frequency

3. PITCH CONTOURS

3.1. Sentence intonation

Typical voice pitch contours of sentence intonation in Japanese for affirmative and interrogative modes are shown in Figure 3, with duration in seconds on the horizontal axis, and changes in the fundamental frequency in whole steps over the complete diatonic scale (musical scale) on the vertical axis.

3.2. Japanese word accent

Typical voice pitch contours of the down skip of Japanese word accent, when the first syllable (Japanese mora) is accented, and when the second syllable is accented, in Figure 4 (Reference 9).

3.3. Chinese tones

Typical voice pitch contours of the four kinds of Chinese syllabic tones in the Beijing dialect (or Standard Colloquial Chinese), namely, tone 1, tone 2, tone 3 and tone 4, are also shown in Figure 4.

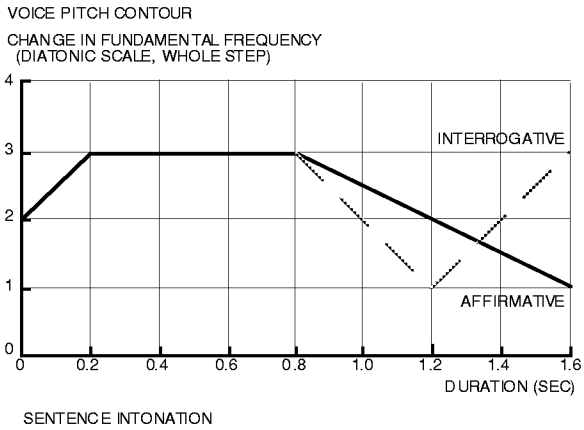

Figure 3: Typical pitch contours of sentence intonation of Japanese for affirmative and interrogative modes

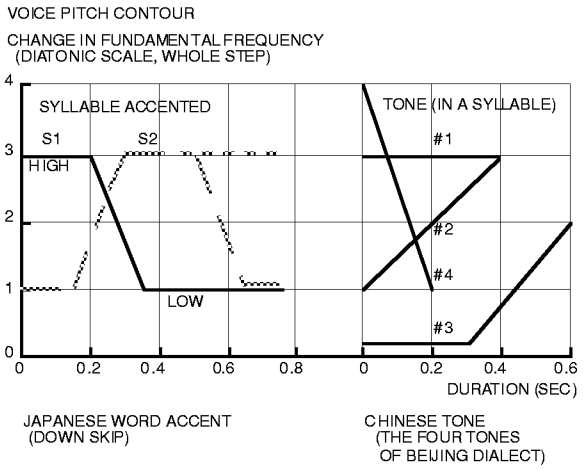

Figure 4: Typical pitch contours of down skip of Japanese word accent and four kinds of Chinese syllabic tone

The contours for both the Japanese word accent and the Chinese tone are normalized to fit the utterances with normal tempo. (Reference 10). The rate of change in fundamental frequency for sentence intonation is as slow as one whole steps per 0.2 seconds, while for word accent and syllabic tone it is up to 3 whole steps per 0.2 seconds.

4. RESOLUTION

4.1. Perceptually acceptable area

Perceptually acceptable areas of Japanese word accents and the Chinese tones are shown in Figure 5, on a two dimensional plane of duration expressed in a logarithmic second scale on the horizontal axis, and change in fundamental frequency in whole steps of the diatonic scale on the vertical axis.

These areas are derived from the typical voice pitch contours in Figure 4, considering the ability to discriminate among the accented syllables and among the four tones. The rate of

change in fundamental frequency is indicated by a set of diagonal lines.

4.2. Required resolution

The required resolution for the duration and fundamental frequency for is shown in Figure 5, taking the smallest area of tone 3 as an example. In this way, the required resolution can be derived on this chart common to any language. Radius of the area is 0.1 second in duration and 1/2 whole step in fundamental frequency in the smallest case. The rate of change in fundamental frequency ranges from 3 to 15 whole steps per second.

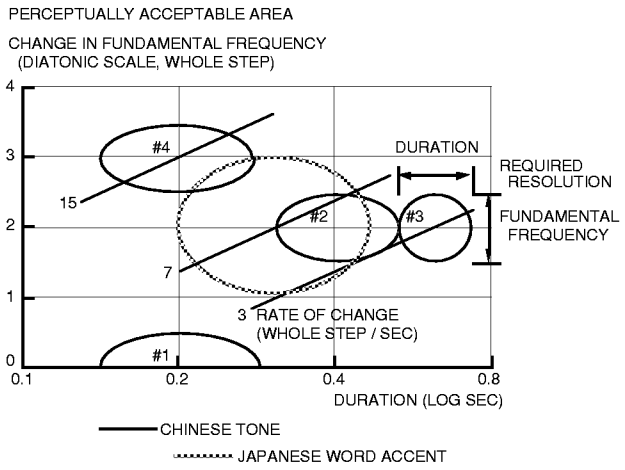

Figure 5 : Perceptually acceptable area and required resolution for duration and fundamental frequency

5. USE OF LIPREADING

The relationship between auditory and visual reception of phonemic and prosodic features of speech is schematized in Figure 6.

About one fourth of the information obtainable through normal hearing for the phonemic features for vowels and consonants can be received visually by lipreading (or speech reading) (Reference 11). In a cochlear implant, amount of information through auditory reception decreases by 25% to 75% from that through the normal hearing. But, they are effectively supplemented by the combined use of lipreading, as the phonemic features obtained by auditory and visual receptions are mutually complementary .

Change in fundamental frequency for prosodic features is always accompanied by changes in amplitude and duration. In the accented syllable if the Japanese word, for example, the amplitude and duration increase. But, these visual cues are not distinct, so that the amount of information obtained visually is about half of that of phonemic features. So, lipreading is not sufficiently useful for supplementing voice pitch information.

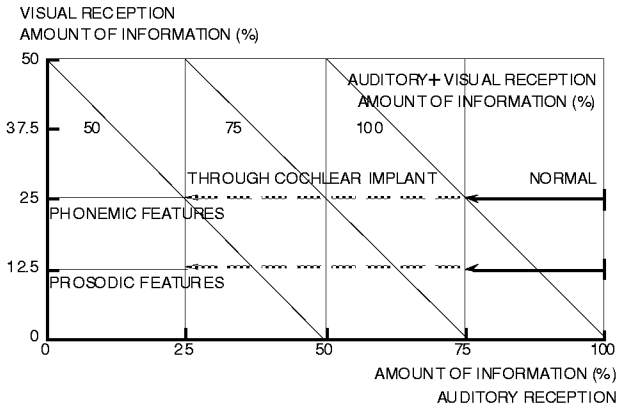

Figure 6: Auditory and visual reception of phonemic and prosodic features of speech

6. PROCESSING SOUND

6.1. Spectral analysis

The processing of sound waveforms in the cochlear and auditory area of the cortex is schematized in Figure 7.

Frequency components of the sound waveform are distributed in the cochlear from the low frequency component at the apical end to the high component in the basal end. They are transmitted to the auditory area of the cortex, through each of the auditory nerve fibers in the form of a neural pulse train.

In the cortex, timbre is perceived based on formant frequency, and voice pitch is perceived based on the fundamental frequency or pitch period of the speech waveform.

The number of sensors (or units of haircells) in the cochlear is as many as 20,000, but, they are mapped through the band-pass filter banks to the electrodes of only around 20, in the current cochlear implant device.

The number of the electrodes is not sufficient even for perception of up to several thousand Hz formant. For voice pitch perception, resolution of 1/2 whole step in the voice pitch range up to several hundred Hz for male and female talkers requires more than 20 channels.

6.2. Pitch extraction

To overcome this difficulty, the most promising way is to transmit sound waveforms as accurately as possible, directly to the auditory area of the cortex.

The neural pulse train transmitted through each nerve fiber conserves information of the sound waveform in the form of pulse density modulation, if superimposed over the whole of the nerve fibers. So, the sound waveform can be reconstructed by means of pulse density demodulation in the cortex. Then, the voice pitch can be extracted from the reconstructed waveform.

In order to distribute the waveform information to the nerve

fibers, difference of individual sensitivity for amplitude can be utilized. By making the individual sensitivity different for the frequency component, the distribution becomes more even and the transmission more accurate. So, the function of frequency analysis in cochlear can be considered as a by-product of transmission of the waveform information (Reference 12).

When the required resolution in Figure 7 is converted into the time domain, it is estimated that the speech processor of a cochlear implant device should be designed to transmit sound waveform with finer time sampling, at less than 2 milli-second, and at finer amplitude quantizing, at less than 3 % of the peak value, in order to make 1/2 whole step of change of voice pitch discriminable at around several hundred Hz.

Such time and amplitude resolutions of the sound waveform can be achieved even with the current technique of cochlear implant devices, if the transmission of the sound waveform rather than the frequency spectrum is taken as its main purpose.

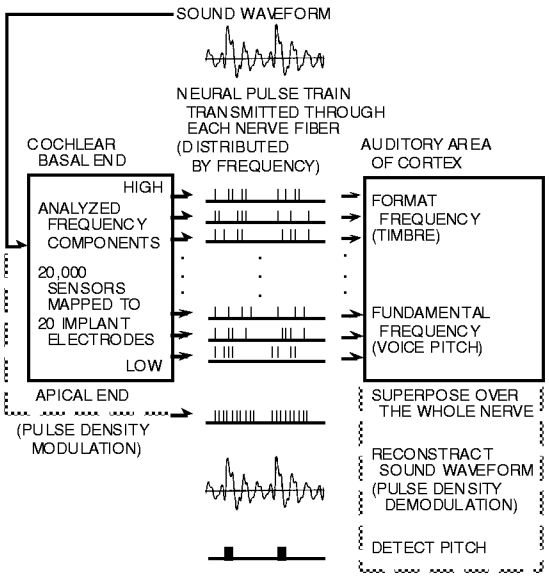

Figure 7: A schematic of the processing of sound waveform in the cochlear and auditory area of the cortex

7. REFERENCES

1. *Book of Abstracts, International Cochlear implant, Speech and Hearing Symposium (hosted by The Australian Bionic Ear and Hearing Research Institute, Australian Hearing Services, University of Melbourn, October 1994, Melbourne.*

2. *Program, International Workshop on Cochlear Implants (organized by Department of Otorhinolaryngology, University of Vienna-Medical School) , October, 1996, Vienna.*

3. *Program and Abstract, 1997 Conference on Implantable Auditory Prostheses (Skinner, M.W., Conference Chair), August, 1997, Pacific Grove, California, U.S.A.*

4. *Standardization of Evaluation Criteria for Speech Perception Abilities Through Artificial Hearing, Record of Waseda Symposium on Artificial Hearing (Hiki, S.,editor),* March, 1996, Waseda University, Tokyo, Japan.

5. Hiki, S. and Hsu, C-J., "Performance of cochlear implants in different languages," *Abstract, The First Asia Pacific Symposium on Cochlear Implant and Related Sciences (Honjo, I. and Takahashi, H., editors), April, 1996, Kyoto, Japan,* p. 66.

6. *Program and Abstracts, Chang Gung International Symposium on Cochlear Implant and Related Sciences November, 1996, Taiwan, Republic of China.*

7. Kim, H-N., Kim, C-S., Kim, L-S, and Lee, S-H., "Sound recognition ability and contrtibuting factors in Korean cochlear implantees - postlingual adults," *same as* 5, p. 69.

8. Cheung, D.M.C., and Lee, K.S., "Speech perception: Assessment for Cantonese (SPAC)," Speech Therapy Clinic, Cochlear Implant Team, Prince of Wales Hospital, The Chinese University of Hong Kong, 1996.

9. Sato, S. and Hiki, S. "Accentual pattern of Japanese with reference to loanwords," *Transactions of Institute of Electrical and Communication Engineers, Vol. 57-D, 1974, pp. 471-478* (in Japanese).

10. Chuang, C-K., Hiki, S., Sone, T. and Nimura, T. "Acoustical features of the four tones in monosyllabic utterances of Standard Chinese," *J. Acoustic. Soc. Japan, Vol. 31, 1975, pp. 369-380* (in Japanese).

11. Fukuda, Y. and Hiki, S., "Characteristics of the mouth shape in the production of Japanese: Stroboscopic observation," *J. Acoustic. Soc. Japan (E),Vol. 3, 1982, pp. 75-91.*

12. Hiki, S., "The function of frequency analysis in cochlear is a by-product of transmission of waveform information," *J. Acoustic..Soc. Japan, Vol. 52, 1996, p. 71* (Reader's forum, in Japanese).