Teaching Tonal Discrimination Based on Statistical Properties and Acoustical Characteristics of the Chinese Four Tones: With Regard to the Contrast between Tone-2 and Tone-3

Shizuo HIKI[1], Kazuya IMAIZUMI[2],
[1]*Waseda University,* [2]*Graduate School of Human Sciences, Waseda University,*
2-579-15 Mikajima, Tokorozawa-shi, 359-1192 Japan,
hiki@waseda.jp, imaizumi@ruri.waseda.jp
Kazuko SUNAOKA[3], Liming YANG[4],
[3]*School of Political Science and Economics,* [4]*School of International Liberal Studies,*
Waseda University, 1-7-14, Nishi-waseda, Shinjuku-ku, Tokyo, 169-8050 Japan,
ksunaoka@waseda.jp, youl@waseda.jp

Abstract

Based on the statistical properties of occurrence frequency and transition probability of the Chinese four tones and the acoustical characteristics of their phonatory control and perceptual response, a Computer-Assisted Instruction system for teaching tonal discrimination to beginners was designed, with special regard to the contrast between Tone-2 and Tone-3. In order to emphasize the difference from the Japanese word accent, intermediate tonal stimulus with synthetic speech and visual display of change in voice pitch were utilized in practicing perception, and the efficiency was improved by introducing the CAI algorithm. In practicing phonation, the target voice pitch patterns were adapted to the talker's voice register which is derived from measurement of fundamental frequency in speaking, and indicated on the screen. The possibility of transferring the system to a self-learning program through the internet was also discussed.

Keywords: *Chinese tones, tonal discrimination, CAI algorithm, synthetic speech, visual display*

1 Introduction

A Computer-Assisted Instruction (CAI) system for teaching discrimination of the Chinese tones to beginners was designed, by utilizing intermediate tonal stimulus with synthetic speech and visual display of change in voice pitch. The statistical properties and acoustical characteristics of the Chinese four tones on which designing of the system based are explained in detail in this report.

In Japanese, there are pairs of words which have the same syllable strings but different meanings according to the accent attached to different syllable positions. However, the words with accent are only about 50% of all the words, and in addition, about 15% of them can be pronounced either with or without the accent (Tokuhiro & Hiki, 2004). Although the number of words in which the accents are indispensable is not so large, it is necessary to perceive and phonate the accent correctly especially for the pairs of frequently used words which are differentiated by the accent each other (Table 1).

In the common Chinese, four kinds of tones (the four tones; namely, Tone-1, Tone-2, Tone-3 and Tone-4) are used. Contrary to the Japanese words, every Chinese syllable is furnished with tone, and the same syllables have different meanings according to the kind of tones. So, there are many groups of words whose meanings are differentiated by the tones (Table 2).

2 Statistical Properties of the Chinese Four Tones

Statistical properties of the Chinese four tones were investigated for the 1st, 2nd, 3rd and 4th syllables, about 124,000 syllables in total, involved in 1, 2, 3 and 4 syllable words, about 55,000 words in total. Those words were compiled in the electronic dictionary of contemporary Chinese (Wu, ed. 1998).

2.1 Occurrence Frequency

As for the occurrence frequency of each kind of tone in the noun words, Tone-2 was close to 25% in average regardless of number of syllables composing a word or position of syllable in the word. Tone-3 was 5% less than 25%, but Tone-4 was 5% more, offsetting each other. In the back syllable positions, Tone-4 occurred 10% more, but Tone-3 10% less. Tone-1 occurred 5% more in the front syllable position, but 5% less in the back. But as a whole, the four kinds of tones in Chinese are used fairly evenly, compared with the deviation found in the Japanese word accent (Figure 1).

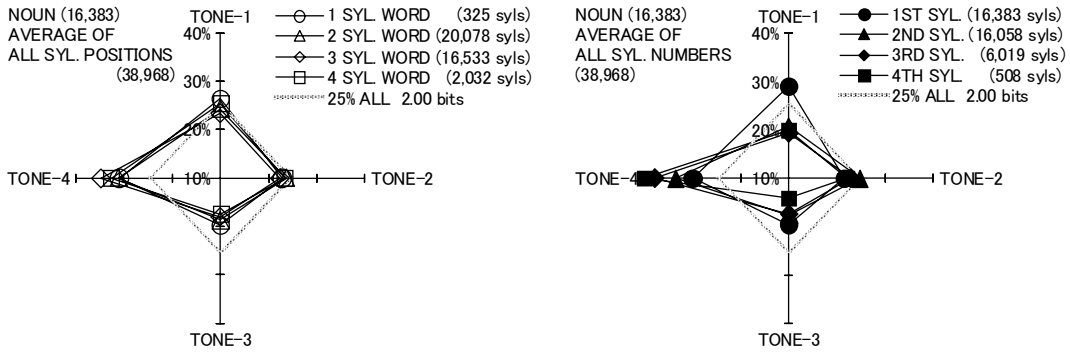

Figure 1. Examples of the occurrence frequency of tones. Classified by number of syllables composing a word (left), and by syllable position in the word (right), for the 1, 2, 3 and 4 syllable noun words. Ratio of occurrence frequency of each tone is plotted on the axis from 10% as the origin up to 40%.

2.2 Transition Probability of Tones

The transition probabilities of tones from the preceding to following syllables in the noun words showed significant deviations. Transition between Tone-2 and Tone-3 was less compared with other combinations of tones. Transition of Tone-2 to/from Tone-1 or Tone-4, and that between Tone-3 and Tone-4 were more frequent (Figure 2).

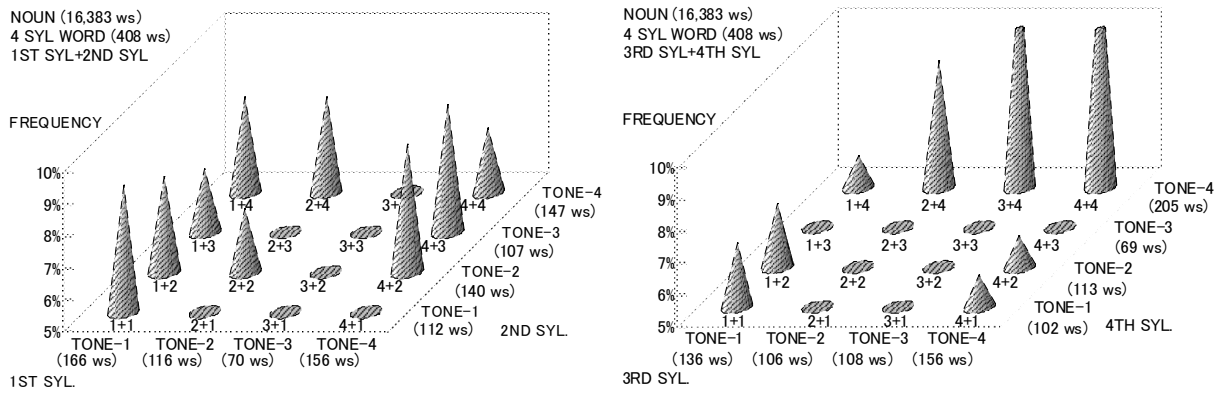

Figure 2. Examples of the transition probability of tones. From the 1st to 2nd syllables (left), and from the 3rd to 4th syllables (right) in 4 syllable noun words. Columns correspond to the preceding tones and rows the following tones. The peaks of cones represent the amount of transition probabilities.

3 Acoustical Characteristics of the Chinese Four Tones

3.1 Phonatory Control

3.1.1 Change in Voice Pitch

In the clear utterance of isolated syllables, the change in fundamental frequency of speech sound (which corresponds to perceptual voice pitch) shows the features of each kind of tone. Tone-1; high flat,

Tone-2; rising, Tone-3; low flat followed by rising, and Tone-4; falling. Features can be seen also in the length of duration and change in amplitude (which corresponds to perceptual voice intensity) (Figure 3).

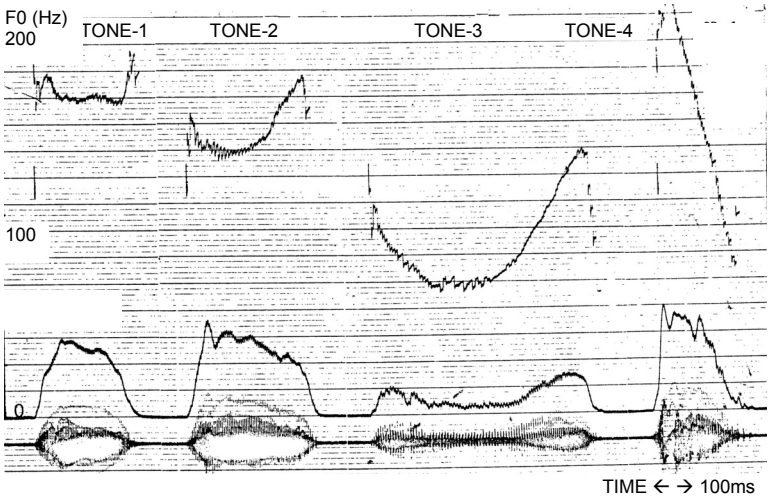

Figure 3. Changes in fundamental frequency (F0, top), amplitude (middle) and waveform (bottom) in the Chinese four tones (Tone-1, Tone-2, Tone-3 and Tone-4, from the left to right). A male speaker.

3.1.2 Concatenation in the Continuous Utterance

When uttered in 2 syllable words with combinations of tones, the changes in voice pitch in each of the two syllables were concatenated. But, the substantial portion of the features for the tones observed in the 1 syllable words were remained unchanged (Figure 4).

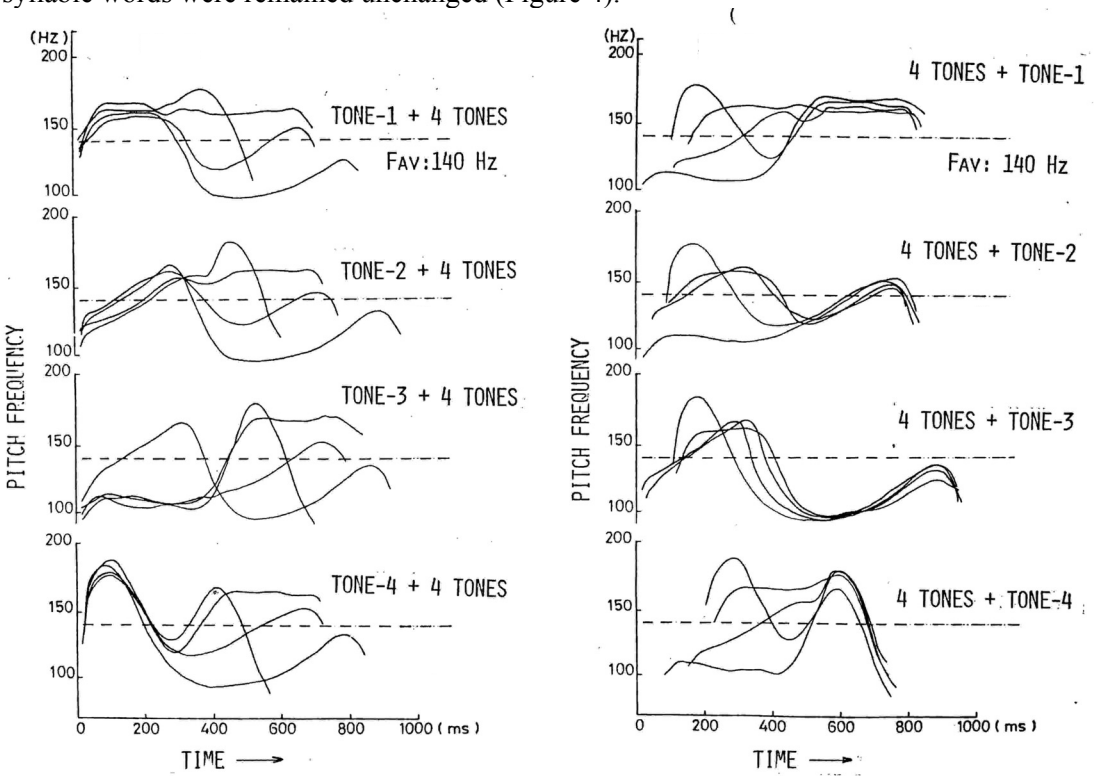

Figure 4. Change in voice pitch of the utterances of all possible combinations of tones in the 1st and 2nd syllables in 2 syllable words. A male speaker. (Tone-3 + Tone-3 is modified to Tone-2 + Tone-3.)

Also in the continuous utterance of sentence speech, those features of change in voice pitch are superposed on the slower rising and falling, which convey syntactic information over phrase, clause and sentence. In the perceptual process, they are decomposed and recognized in parallel.

3.2 Perceptual Response

3.2.1 Boundary of Discrimination

As a result of perceptual discrimination tests utilizing synthetic speech stimulus in which the changes in voice pitch were interpolated stepwise between Tone-2 and Tone-3, the probabilities of identification were smoothly overlapped each other. On the other hand, the range of acceptable voice pitch pattern of Tone-1 was narrow and the boundary of its identification was steep, compared with smooth boundary of Tone-2 (Figure 5).

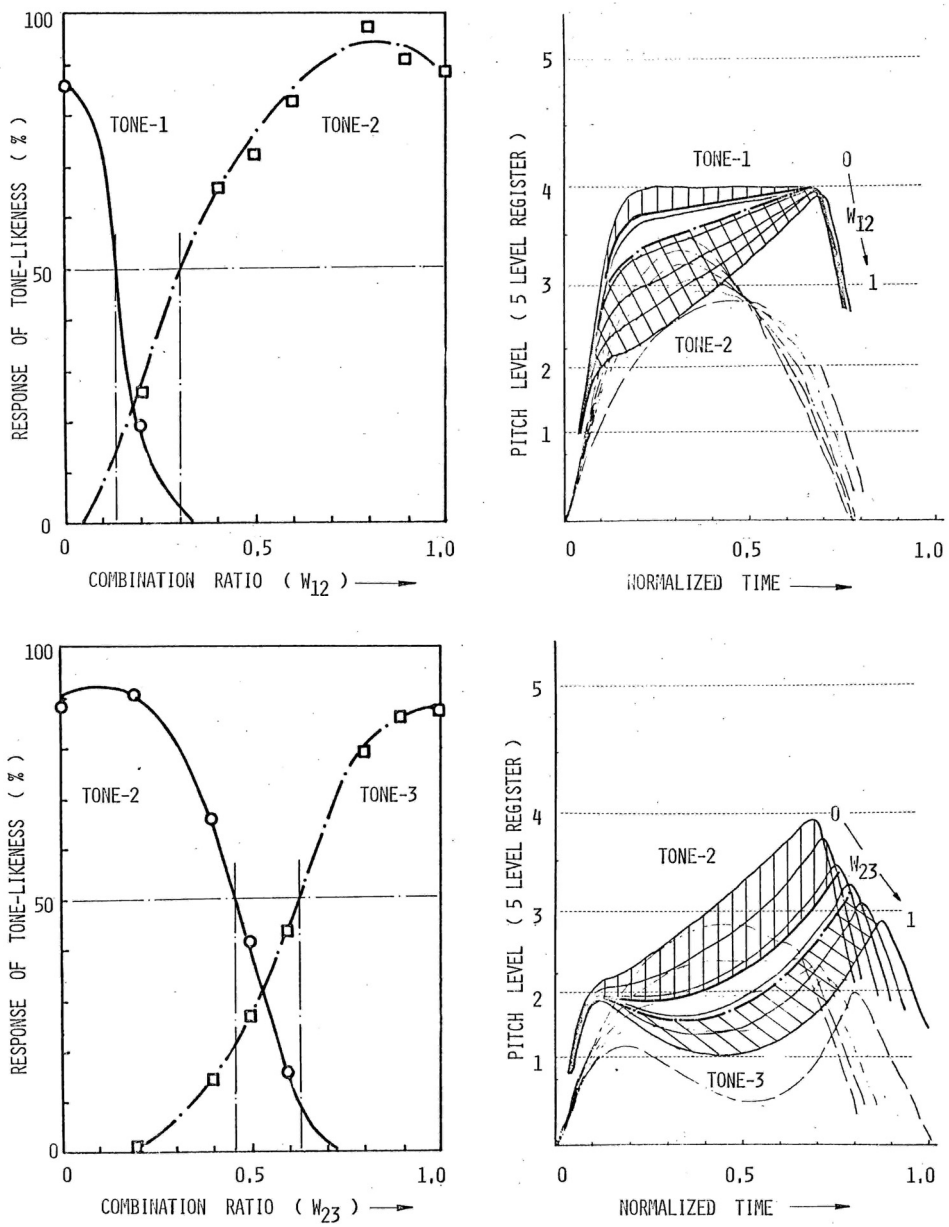

Figure 5. Probability (tone-likeness) of identifying Tone-1 or Tone-2 for the intermediate voice pitch patterns (upper left), and Tone-2 and Tone-3 (lower left). The hatched ranges of the change in voice pitch (upper and lower right) were 50% acceptable for each tone. Average of the four listeners.

3.2.2 Perceptually Acceptable Range

Perceptually acceptable range of voice pitch patterns for each of the four tones were summarized from the results of discrimination test utilizing synthetic speech stimulus (Figure 6).

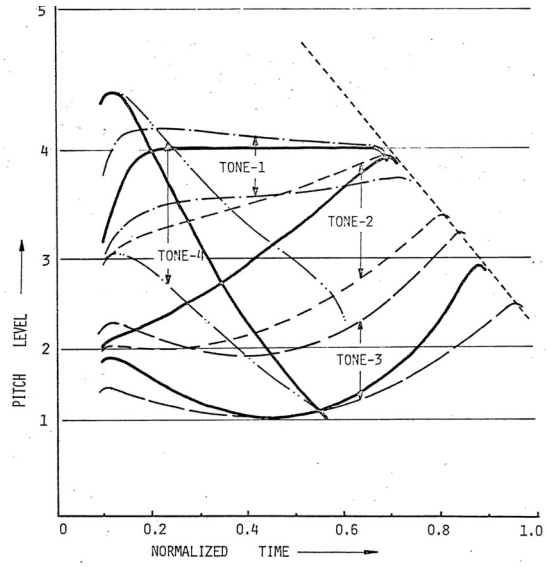

Figure 6. Perceptually acceptable ranges of change in voice pitch for the Chinese four tones. Solid lines show the typical change, and dotted lines the upper and lower limits of the range.

4 Distinction Between Tone-2 and Tone-3

4.1 Rate of Change in Voice Pitch

The results of acoustical analysis and perceptual test were schematized on the diatonic scale. The Japanese word accent is conveyed by the fall of voice pitch at the end of the accented syllable. The rate of fall is about two whole steps per 0.2 sec, at normal speaking speed.

Among the changes in voice pitch in the Chinese four tones, contrast between high flat of Tone-1 and low flat portion of Tone-3 is as large as three whole steps. The contrast between rising of Tone-2 and falling of Tone-4 can be discriminated clearly, as the falling of Tone-4 is as steep as three whole steps per 0.1 sec. But, the rising of Tone-2 and the rising portion at the end of Tone-3 are both around two whole steps per 0.4 sec, so their distinction is not clear (Figure 7).

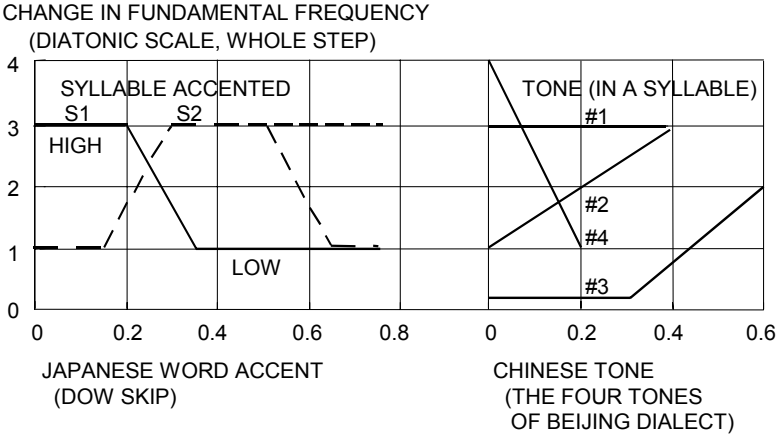

Figure 7. Schematic patterns of changes in voice pitch for Japanese word accent and the Chinese four tones on the diatonic scale.

4.2 Resolution Necessary for Discrimination

Acceptable areas and resolution necessary for discriminating the Japanese word accent and the Chinese four tones were schematized on a two-dimensional plane representing the change in fundamental frequency vertically and duration horizontally. The Japanese word accent can be recognized by detecting about two whole steps falling at the end of accented syllable from slower falling over a span of utterance in Japanese.

In the Chinese tones, much finer fundamental frequency and time resolution is required, especially for identifying the areas of Tone-2 and Tone-3 separately (Figure 8).

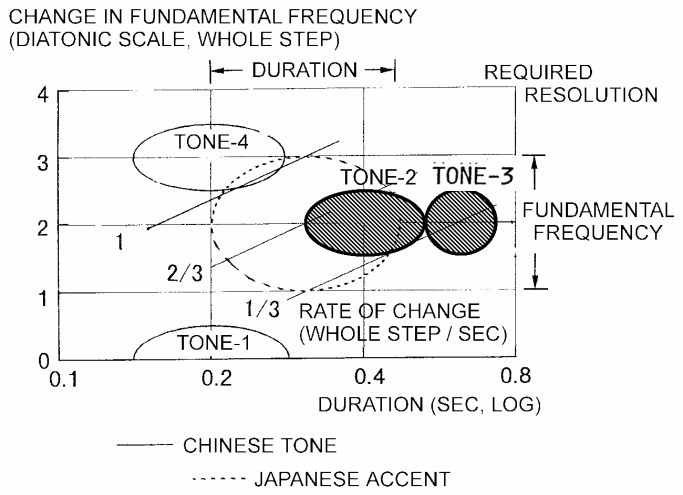

Figure 8. Resolution necessary for discriminating Japanese word accent and the Chinese four tones.

4.3 Clue of Voice Quality

In Tone-1, a very steady phonatory control of voice pitch within 1/4 of whole step produces a sharp voice quality. In Tone-3, the electro-myographic observation showed that the sterno-hyoid muscle is active for extreme phonatory control for lowering voice pitch. This results in almost rough voice quality (Chuang & Hiki, 1976) (Figure 9, left). A heavy lowering of the larynx for Tone-3 was also observed by thyrometer (Kakita & Hiki, 1972) (Figure 9, right). Those voice qualities associated with the kind of tone also serve as clues to identifying the tones.

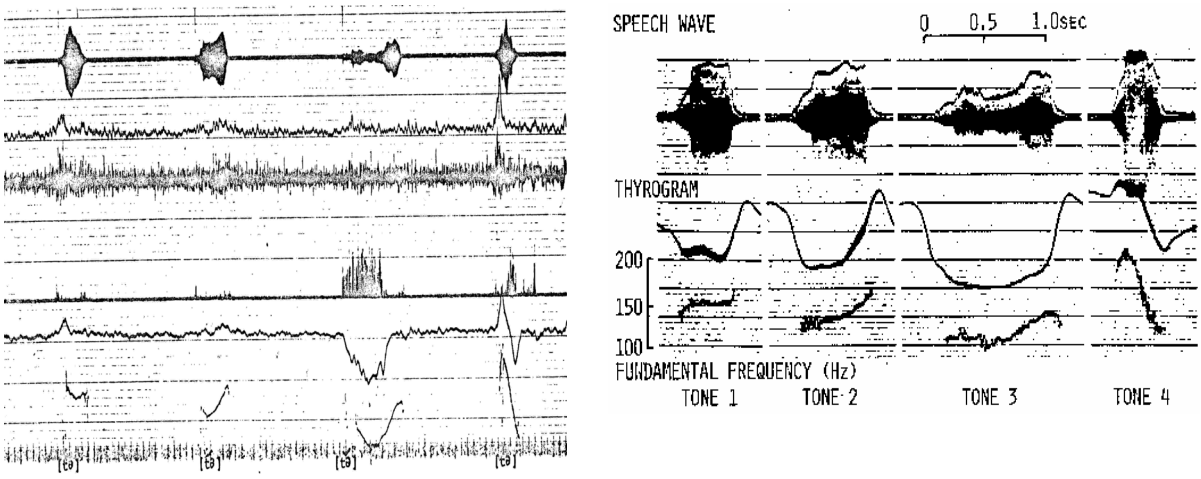

Figure 9. Electro-myographic observation of sterno-hyoid muscle activity (third from the bottom) particularly for Tone-3 (left). Recording of lowering of the larynx for Tone-3 by the thyrometer (right).

5 CAI System for Teaching Tonal Discrimination

For the practicing discrimination especially between Tone-2 and Tone-3, which is most difficult for beginners, a CAI (Computer-Assisted Instruction) system was designed by utilizing synthetic speech stimulus and visual display of voice pitch pattern.

5.1 Use of Synthetic Speech

In the practice of perceptual discrimination, real speech samples by teachers' utterances were acoustically analyzed and re-synthesized, in order to interpolate or extrapolate changes among the typical voice pitch patterns of the tones (Figure 10). Those re-synthesized speech, in which the contrast of voice pitch patterns were moderated or exaggerated, were used as stimulus in the CAI system.

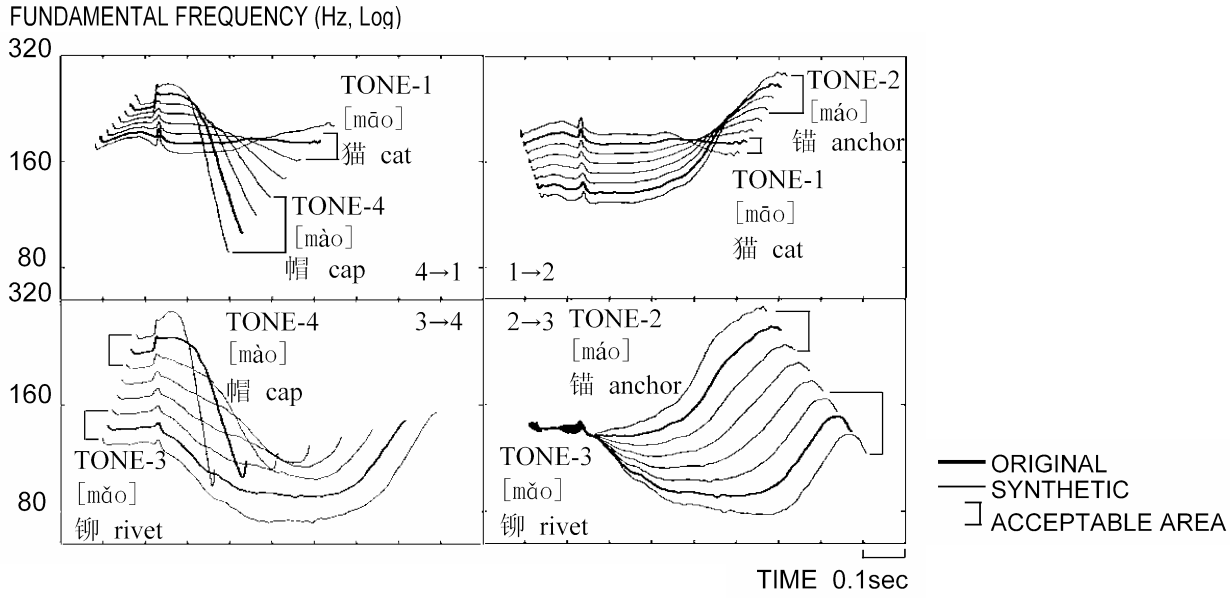

Figure 10. Voice pitch patterns of the re-synthesized speech interpolated or extrapolated from Tone-1 to Tone-2, from Tone-2 to Tone-3, from Tone-3 to Tone-4, and from Tone-4 to Tone-1. A female speaker.

In the interpolated and extrapolated voice pitch patterns, the contrast between Japanese 2 syllable word with and without accent at the first syllable seems to be close to that between Tone-4 and Tone-1, whose distinction is most clear among the Chinese four tones (Figure 11).

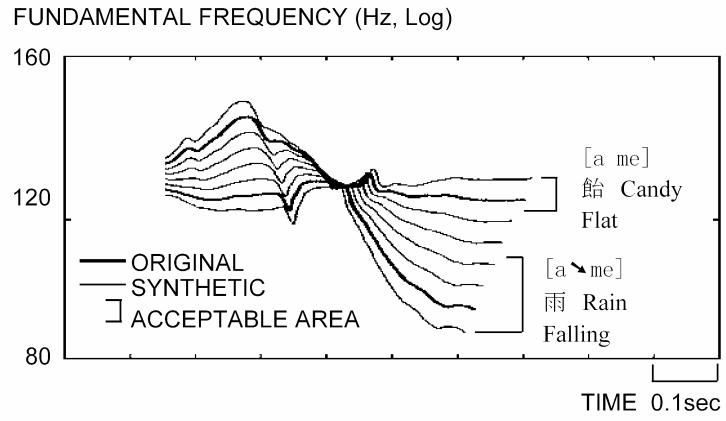

Figure 11. Interpolation and extrapolation between Japanese 2 syllable words with and without accent at the 1st syllable (upper). A male speaker.

5.2 CAI Algorithm

It was attempted to minimize the steps before achieving the goal of perceptual discrimination, by adjusting the next stimulus easier after incorrect response or harder after correct response (Figure 12).

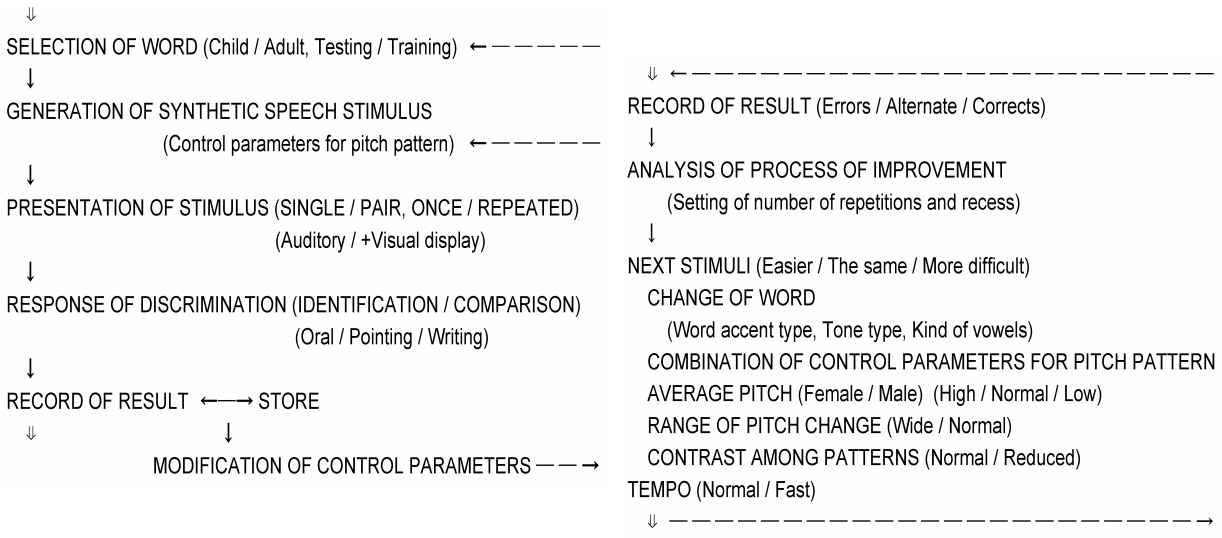

Figure 12. Flow of the CAI algorithm. Scheme of the program for perceptual discrimination practice (left), and modification of control parameters of re-synthesized speech (right).

5.3 Use of Visual Display

Along with the auditory presentation of speech stimulus, pattern of change in fundamental frequency of the stimulus can be presented visually, if requested by the learner, in order to provide a clue to perceptual discrimination of the contrast between tones (Figure 13).

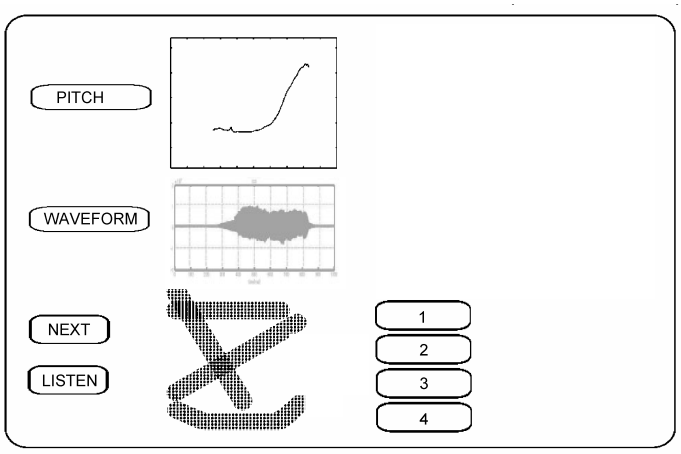

Figure 13. Visual display of speech waveform and change in fundamental frequency on the screen.

5.4 Target Patters for Phonation Practice

In practicing phonation, the target voice pitch patterns are indicated on the screen. They are normalized by the learner's voice register, which is derived from the extraction of fundamental frequency in speaking (Figure 15). Voice register in normal speaking ranges about one octave (Hiki, 1967), and the target voice pitch patterns of Chinese tones is set over 3 whole steps within it. The real time pitch extraction and display program serves as a bio-feedback for practicing phonatory control of the tones.

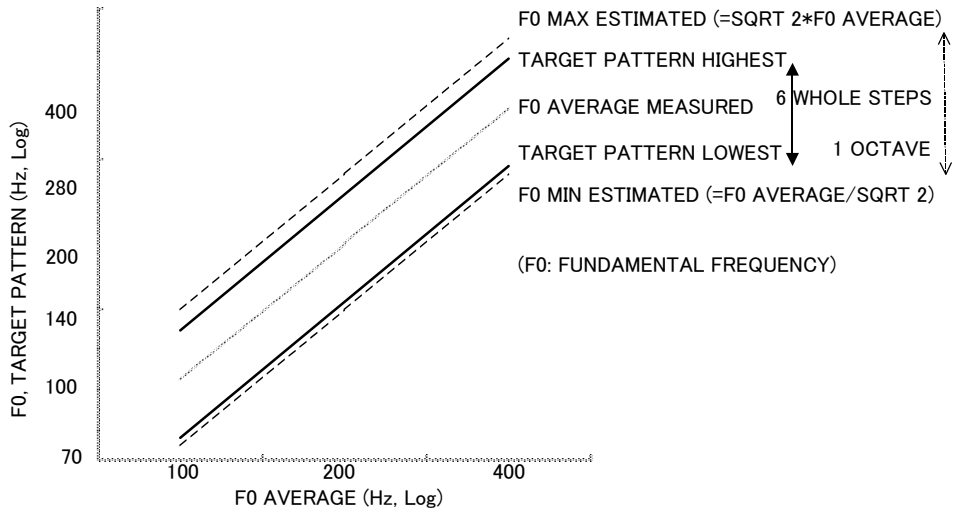

Figure 15. Each learner's speaking voice register and the range of target voice pitch patterns of tones.

6 Discussion

Possibility of transferring the CAI system to a self-learning program over the internet was examined (Figure 16). As a detailed record of the process of practice by each learner is stored in the CAI algorithm, the advantages of connecting the system to internet are that the learner's achievement can be assessed any time, and that additional stimulus can be offered, and a revised version of the CAI algorithm can be uploaded at any stage in the future.

Collection of data on the common and individual defects in perception and production of Chinese tones by the Japanese beginners is planed for use in remodeling the CAI algorithm.

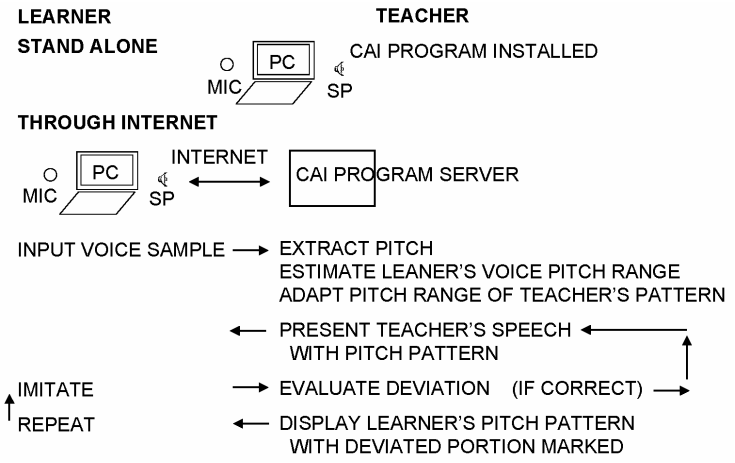

Figure 16. Self-learning program by stand-alone computer or through internet.

Notes

The system utilizes a speech analysis-synthesis program, STRAIGHT, developed in ATR Human Information Processing Research Laboratories by H. Kawahara, Wakayama University, Japan.

References

Tokuhiro, Y. and S. Hiki. 2004. Effects of mora phonemes on Japanese word accent, Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation, December, 2004, Japan, 243-250.

Yu, S-W. ed. 1998. *The Grammatical Knowledge-base of Contemporary Chinese,* Tsinghua University Press, China, (converted into Japanese computer code by M. Yin and K. Sunaoka).

Kakita, Y. and S. Hiki. 1976. Investigation of laryngeal control in speech by use of thyrometer, J. Acoustical Soc. America, 59(3), 669-674.

Chuang, C-K. and S. Hiki. 1972. Acoustical features and perceptual cues of the four tones of standard colloquial Chinese. J. Acoustical Soc. America, 52(1, Part 1), 146.

Hiki, S. 1967. Correlation between increments of voice pitch and glottal sound intensity, J. Acoustical Soc. Japan, Vol. 23(1), 20-22.

Table 1. Examples of the pairs of Japanese bi-syllabic words differentiated by accent. [↘]: downskip.

after 1st syllable	after 2nd syllable	none
[a↘ki]	[a ki↘]	[a ki]
秋autumn	飽き tiresomeness	空き an opening
[o:↘me]	[o: me↘]	[o: me]
青梅name of a town	多め a larger portion	大目 tolerate
[ka↘ki]	[ka ki↘]	[ka ki]
牡蠣 an oyster	垣 a fence	柿 a persimon
[ʦu↘ru]	[ʦu ru↘]	[ʦu ru]
鶴 a crain	弦 a string	釣る fish
[ha↘ʃi]	[ha ʃi↘]	[ha ʃi]
箸 chopticks	橋 bridge	端 edge
[ma↘ku]	[ma ku↘]	[ma ku]
撒く sow seeds	幕 a curtain	巻く wind up
[rjo:↘ba]	[rjo: ba↘]	[rjo: ba]
良馬 a good horse	猟場 a hunting ground	両刃 a double edged sword

Table 2. Examples of the groups of Chinese 1 syllable and 2 syllable words differentiated by tones.

Groups of mono-syllabic words with contrast of tones (in Standard Chinese).

	Tone-1	Tone-2	Tone-3	Tone-4
A-1. Syllable consisted of vowel				
1)	[wā]	[wá]	[wǎ]	[wà]
	蛙 frog	娃 baby	瓦 tile	袜 socks
2)	[yī]	[yí]	[yǐ]	[yì]
	一 one	姨 aunt	椅 chair	亿 hundred million
3)	[wū]	[wú]	[wǔ]	[wù]
	屋 house	无 nothing	五 five	物 matter
4)	[yīn]	[yín]	[yǐn]	[yìn]
	音 sound	银 silver	饮 drink	印 stamp
A-2. Syllable consisted of voiced consonant + vowel				
1)	[lēi]	[léi]	[lěi]	[lèi]
	勒 strangle	雷 thunder	垒 rampart	泪 tear
2)	[liū]	[liú]	[liǔ]	[liù]
	溜 slide	留 remain	柳 willow	六 six
3)	[miāo]	[miáo]	[miǎo]	[miào]
	喵 mew	苗 seedling	秒 second	庙 temple
4)	[māo]	[máo]	[mǎo]	[mào]
	猫 cat	毛 hair	铆 rivet	帽 cap