Acoust. Sci. & Tech. **27**, 2 (2006)

PAPER

Contribution of pitch-accent information to Japanese spoken-word recognition

Ikuyo Masuda-Katsuse*

Department of Information and Computer Sciences, School of Humanity-Oriented Science and Engineering, Kinki University, 11–6, Kayanomori, Iizuka, 820–8555 Japan

(*Received 2 May 2005, Accepted for publication 22 August 2005*)

Abstract: This paper investigates the contribution of pitch-accent information to Japanese spoken-word recognition. Pitch accent of spoken words was manipulated by controlling F_0. First, the present author investigated the relation between word intelligibility in the presence of noise and the adequacy of accent type in those words. In the intelligibility test, participants were presented with speech stimuli and a pink noise together, and were required to identify the word. In the rating test, the same participants were presented with the same speech stimuli and were required to rate the adequacy of the words' accent types. Results indicated that the spoken words with more adequate accent type were more intelligible in the presence of noise. Next, the present author investigated the relation between reaction time in shadowing words and the adequacy of accent type in those words. In the shadowing task, the participants were required to shadow a word whose accent type was manipulated as soon as they identified it. The same participants participated in the rating test. The reaction time in the case of the words with an adequate accent was shorter than in the case of an inadequate one. These results support the hypothesis that pitch-accent information in Japanese spoken words might facilitate word recognition.

Keywords: Pitch accent, Intelligibility, Background noise, Shadowing

PACS number: 43.71.Gv [DOI: 10.1250/ast.27.97]

1. INTRODUCTION

Japanese is one of the pitch-accent languages, in which high or low pitch is determined at every mora, and there is the carrier of accent on the mora just before the mora where the pitch drops to a lower level [1,2]. In a pitch-accent language, pitch height at every mora of spoken words has little importance in comparison to tonal languages such as Chinese. That is to say, in Japanese "where" the pitch falls is important, whereas in Chinese "how" pitch rises and falls is important. In stress-accent languages represented by English, a spoken word can appear with various pitch patterns, although it does of course have a regular pitch pattern [1]. "Clearly, in Japanese accents are realized by pitch as they are in English; but the Japanese accents, involving high pitch and following low pitch, cannot be reversed by intonation as English accents can [3]."

In Japanese, the meanings of homonyms are often distinguished by their accent-position, which is determined by a pitch-fall. Examples of such homonyms are Ha-shi, Ka-ki, and A-ki [1]. When a postpositional particle "ga" trails these words, Japanese standard high-low patterns of pitch in those words are as follows:

Ha-shi	(Ha-shi-gā (edge),	Ha-shi-ga (bridge),	Ha-shi-ga (chopsticks))
Ka-ki	(Ka-ki-gā (persimmon),	Ka-ki-ga (fence),	Ka-ki-ga (oyster))
A-ki	(A-ki-gā (space),	A-ki-ga (tired),	A-ki-ga (autumn)).

The accent positions of the homonyms presented above are in the case of the Tokyo dialect, though in Japanese, accent type of a spoken word often differs between

*e-mail: katsuse@fuk.kindai.ac.jp

dialects. For example, the standard accent type for "A-ka-ru-i (bright)," which is a highly familiar four-mora word, is Low-High-High-High. In practice, another accent type Low-Low-High-High is used in Hokkaido, Low-High-High-Low in Toyama, Low-High-Low-Low in Ishikawa, High-High-High-Low in Mie, and High-High-Low-Low in Ehime [4]. Further, there are areas where the "Ikkei accent" or "collapsed accent" is used. In these areas the meanings of homonyms are not distinguished by pitch accents. Therefore, it has been believed that pitch accent is generally of low importance for Japanese spoken-word recognition [5].

On the other hand, several researchers have shown that pitch-accent information indeed influences Japanese spoken-word recognition. Kitahara *et al.* [6] investigated the role of prosody in the cognitive process of spoken language. They demonstrated that intonation and pause information efficiently worked to facilitate language understanding while listeners were engaging simple operations. They reported that accent information especially contributed to it.

Minematsu and Hirose [7] manipulated the F_0 of spoken words in three ways using PARCOR analysis-resynthesis. In CASE 1, the F_0 was constant at 100 Hz and in CASE 2, the F_0 was translated into another accent type. In CASE 3, the original F_0 was used. A word intelligibility test was performed using speech stimuli generated by band-eliminating the synthesized speech. As a result, intelligibility in CASE 3 was higher than that in CASE 1 or CASE 2.

Speech communication is often practiced in the presence of background noise. When this occurs, phonetic information is often masked, although pitch information is more robust to the noise interference than phonetic information. It has been shown that listeners can catch pitch information well enough to recognize the accent type even in the presence of a high-level noise [8]. Therefore, it is expected that the contribution of F_0 information — especially accent-type information — to word recognition should be clearly observable, even in the presence of noise.

In the initial experiments discussed in this paper, the present author confirmed the contribution of pitch-accent information to Japanese spoken-word recognition in the presence of noise. She then examined in the secondary experiments, whether the contribution was observable even when there was no background noise and speech was not damaged by band-limitation and so on.

2. WORD LIST

Japanese four-mora words were used in the experiments. These have four standard types of accent. Each accent type is represented by high-and-low pattern of pitch as follows: low-high-high-high for accent type 0, high-low-low-low for accent type 1, low-high-low-low for accent type 2, and low-high-high-low for accent type 3.

The present author controlled the familiarity of words because a word's familiarity involves word intelligibility in the presence of noise [9]. The four-mora words were selected from the database included in Lexical Properties of Japanese [10] based on the following criteria: their familiarity is 6.0 or more in auditory presentation, their familiarity is 5.0 or more in visual presentation, and their familiarity is 5.0 or more in both auditory and visual presentation. The present author produced a list of 200 words selecting 50 words randomly from among the available words of each accent type. In addition, another word list was prepared for practice sessions. That word list was composed of words other than those in the former list and satisfied the same condition of word familiarity.

3. SPEECH STIMULI

3.1. Speech Materials

A male speaker read the word lists in a soundproof booth and the speech sounds were recorded on DAT. The speaker was instructed to read the lists monotonically and at a fixed speed and on beat at each mora. To make such artificial utterances easy, auditory and visual cues were presented to the speaker. As an auditory cue, a 140-Hz pure tone was presented for about 100 ms through a pair of headphones at intervals between moras, and as a visual cue, a circle was incrementally presented on a monitor at those same intervals. The mora intervals were determined so that the word duration became 850 ms.

3.2. Representative F_0 Contours

Even if we handle accent type as an experimental variable, we must manipulate the F_0 of stimuli in practice to manipulate accent type of the stimuli. However, because Japanese is not an intonation-based language as mentioned in the Introduction, strictly speaking, the accent type does not provide an F_0 contour. Therefore, the present author prepared a representative F_0 contour for every accent type. First, the same speaker uttered a nonsense word /oeoe/ with four types of accents several times and the utterances were recorded. The same auditory and visual cues used in recording the speech materials were presented to the speaker. F_0 contours were then extracted from these utterances and were smoothed. The present author could classify those F_0 contours into two groups by their temporal characteristics and choose two representative contours for every accent. Next, in order to select a representative F_0 contour from them, the present author preliminarily conducted a subjective evaluation examination. To generate the speech stimuli for the preliminary experiment, 100 spoken words were re-synthesized by transforming the original F_0 contour into one of the smoothed F_0 contours. The alternated F_0 contour represented a standard accent type for the word. Three listeners, who were from different regions of Japan, were presented a pair of speech stimuli,

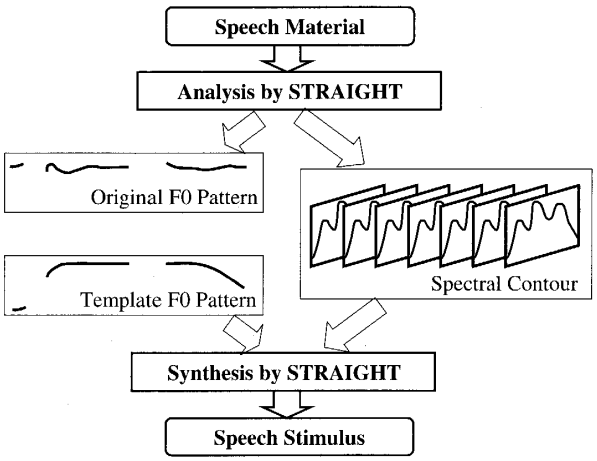

Fig. 1 Generation of stimuli by analysis-synthesis system.

and were required to select the more natural one. As a result, the F_0 contour of the elected stimulus was determined as the representative of the accent type. These F_0 contours are called "representative F_0 contours."

3.3. Re-synthesis of Speech Stimuli

Speech stimuli for the experiments were re-synthesized as shown in Fig. 1. The original speech sounds were analyzed by STRAIGHT [11] to extract spectral information and F_0 information separately. Speech stimuli were then re-synthesized using STRAIGHT by transforming the original F_0 into an alternative F_0 while normalizing the speech duration to 850 ms. The alternative F_0 was generated by adding a small F_0 fluctuation, which was obtained by eliminating a trend component from the original F_0, to one of the representative F_0 contour, because each F_0 fluctuation influences the sound quality of consonants [12–14]. Thus, all speech stimuli used in the experiments were artificially generated with high sound quality. All stimuli were tapered for 10 ms at their beginning and end. The sampling frequency was 32 kHz.

4. EXPERIMENTAL DESIGN

4.1. Introducing a Rating Scale for Adequacy of Accent Type

In Japanese, a spoken word has a standard accent (that is, Kanto accent) type, though various accent types are used in practice, depending on the dialect. For this reason the previous studies conventionally adopted Kanto dialect speakers as participants in their experiments. Nowadays, with the advent of media such as television and people actively moving from place to place, it is not unusual for the participants to regard plural accent types as adequate for one spoken word. At least, most participants may have aurally heard other kinds of dialects than the Kanto dialect. Consequently, strict control of perceptual experiences in spoken language seems to be difficult.

Furthermore, when we treat the stimuli accent types as experimental variables, there remains the problem of how we measure the subjective difference between standard and other accent types.

To avoid such problems, the present author introduces a new subjective scale for evaluating the adequacy of an accent type in a word. This scale ranges from inadequate to adequate. When a word is presented as an auditory stimulus, participants rate the adequacy of the word's accent using the scale. They don't rate the accent based on whether it is a Kanto accent or not, but rather on another "standard" that is formed based on the participants' linguistic experience. By having a scale with several steps, a subjective difference between the "standard" accent type and the presented accent types is measurable.

The experiments consisted of a word intelligibility test and a rating test for the adequacy of accent type in a word. The word intelligibility test was coupled with the rating test. In the word intelligibility test, accent types in words were manipulated; that is to say the experimental variable is accent type. In the rating test, the adequacy of accent type for the same words as used in the intelligibility test was evaluated by the same subjects who participated in the intelligibility test. The intelligibility was evaluated based on the rating values for adequacy of accent type in the word.

4.2. Excluding the Linguistic Learning Effect

Four kinds of stimuli were generated for every word with the four accent types: type 0, type 1, type 2, and type 3. In addition, there were five signal-to-noise ratio (SNR) conditions in the intelligibility test. That is to say, there were 20 experimental conditions for every word in the intelligibility test. To exclude the linguistic learning effect, the experiments were performed not in a repeated-measures design but in a randomized block design. Each word with one of four accent types was presented once to a participant; four participants comprised one participants' group. It was randomly decided which of the four participants would be allocated the word and with which of the four accent types. One group consequently covered 800 stimuli (200 words $\times$ 4 accent types). Five SNR conditions were allocated to different groups, and forty participants participated in the experiments. That is to say, there were ten participants' groups with two groups allocated to each SNR condition.

5. EXPERIMENTS: RELATION BETWEEN WORD INTELLIGIBILITY AND ADEQUACY OF THE ACCENT TYPE IN A WORD

The experiments comprised a word intelligibility test and a rating test to determine the adequacy of accent type in a word. All subjects participated in the intelligibility test before the rating test. Stimuli were diotically presented to

the participants at 70 dB SPL though a pair of headphones (STAX SR-Λ PRO) in a sound-proof booth. The participants were college students or graduate students and had normal hearing. Most students have lived in more than one region of Japan. Prefectures where they have lived longest are Hokkaido, Aomori, Akita, Yamanashi, Chiba, Tokyo, Hyougo (four people), Osaka, Nara (three), Hiroshima (six), Yamaguchi (two), Kagawa, Ehime (two), Fukuoka (twelve), Saga, Nagasaki, and Miyazaki. They were paid a small amount of money for their participation.

5.1. Experiment I: Word Intelligibility Test

There were five different SNR conditions involving four noise conditions (SNRs: −2, 2, 6, and 10 dB) and a clean condition. Under noisy conditions, a pink noise was added 200 ms before the start of speech and continued until 200 ms after the end of speech. The participants were presented the stimuli in a random order and required to write a word as they listened. A 500-Hz beep tone was presented 500 ms before each stimulus. Intervals between the stimuli were about 10 s. A practice session was held prior to the experiment. In that session, the present author used 20 stimuli generated based on the word list for the practice sessions.

5.2. Experiment II: Rating Test for the Adequacy of Accent Type in a Word

The participants were presented with the speech stimuli used in Experiment I but in a different random order and without any background noise. They were required to rate the adequacy of the word's accent type. The rating scale has five steps ranging from "inadequate" to "adequate." The intervals between stimuli were about 6 s.

The participants were instructed as follows:

(A) "Accent type" means the appearance of a change in voice pitch. "Adequacy of accent type" means the extent to which the accent type of the presented stimuli fits the word's accent type that the participants experientially recognize when the word is presented as an auditory stimulus.

(B) The participants should rate the adequacy of each word's accent type when the word is presented alone; they should not assume any context surrounding the word.

(C) The participants should rate adequacy based not on how they speak the word but on their perceptual image when they hear the word as an auditory stimulus.

Prior to the experiment the present author held a practice session. In that session, she used 20 stimuli generated based on the word list for practice sessions.

5.3. Results

The frequency of correct or incorrect responses in the word intelligibility test was counted for every rating value with respect to the adequacy of accent type in the word.

Table 1 Cross table of frequency of correct or incorrect responses in the word intelligibility test according to rating categories.

Adequacy rating in rating test	Frequency in word intelligibility test		
	Incorrect	Correct	Total
	Clean		
1	7	334	341
2	6	436	442
3	0	207	207
4	2	236	238
5	1	371	372
Total	16	1,584	1,600
	SNR = 10 dB		
1	63	339	402
2	54	355	409
3	17	139	156
4	14	179	193
5	32	408	440
Total	160	1,420	1,600
	SNR = 6 dB		
1	76	162	238
2	132	320	452
3	64	202	266
4	43	239	282
5	56	306	362
Total	371	1,229	1,600
	SNR = 2 dB		
1	235	162	397
2	158	200	358
3	93	146	239
4	73	134	207
5	142	257	399
Total	701	899	1,600
	SNR = −2 dB		
1	412	109	521
2	206	93	299
3	112	62	174
4	122	74	196
5	225	185	410
Total	1,077	523	1,600

The results are presented in Table 1. In the table, adequacy rating 1 means "inadequate" and adequacy rating 5 denotes "adequate."

Figure 2 illustrates the relationship between word intelligibility and adequacy of the words' accent type. The figure depicts the conditions where SNRs are 10, 6, 2, and −2 dB, where the higher the adequacy of the accent type is rated, the higher the word intelligibility is. In this way, under all noise conditions, there are significant positive correlations between word intelligibility and the adequacy of the accent type.

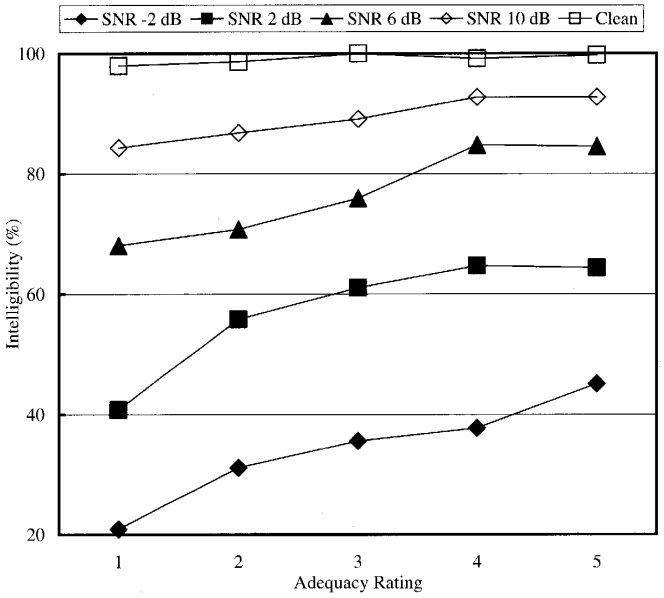

Fig. 2 Relationship between percentage of word intelligibility and adequacy ratings for each accent type.

The present author calculated Goodman-Kruskal's rank measure of association and performed a statistical test for it. The results reveal that $\gamma = 0.248$ ($p < 0.01$) under the condition that the SNR was 10 dB, $\gamma = 0.261$ ($p < 0.01$) when the SNR was 6 dB, $\gamma = 0.251$ ($p < 0.01$) where SNR was 2 dB, and $\gamma = 0.306$ ($p < 0.01$) under the condition that the SNR was -2 dB. In this way, under all noise conditions, there are significant positive correlations between word intelligibility and the adequacy of the accent type in the words.

When there was no background noise, the intelligibility reached almost 100% regardless of the adequacy rating. This indicates that the quality of stimuli used in the experiments was very high.

5.4. Discussion

Where there was no background noise, there was no apparent correlation observed between word intelligibility and adequacy ratings; however, the present author supposes that the same mechanism works in speech perception as in noisy conditions, even in that case. In the present experimental design, the correct rate in the clean condition saturates; therefore, the effect of adequacy of accent type might not appear in the difference of the correct rate.

Minematsu and Hirose [7] and Cutler and Otake [5] showed using the gating paradigm [15] that the initial F_0 contour of spoken words enabled listeners to identify the accent types of those words, thus facilitating word recognition. In other words, there seems to be a "uniqueness point [16]" on an F_0 contour. Minematsu and Hirose [7] asserted that prosodic information is used to help access the mental dictionary in long-term-memory (LTM) by limiting the search space. If the search space is limited, search time should be shortened. Moreover, if recognition of a spoken word with an inadequate accent type takes longer time than that with an adequate accent, reaction time in shadowing a spoken word with an inadequate accent may be longer than that with an adequate one.

In the following experiments, the present author measured reaction time in a shadowing task to investigate the contribution of pitch accent information to the recognition of spoken words without any background noise or intentional damage.

6. EXPERIMENTS: RELATION BETWEEN REACTION TIME IN SHADOWING AND ADEQUACY OF A WORD'S ACCENT TYPE

This set of experiments comprised of a shadowing test and a rating test for the adequacy of accent type in a word. The shadowing test was coupled with the rating test. All subjects participated in the shadowing test before the rating test. Stimuli were diotically presented to the participants at 70 dB SPL in a sound-proof booth. Eight normal-hearing participants who did not participate in Experiments I and II participated in these experiments. The eight participants formed themselves in groups of four because the experiments were designed according to a randomized block design. Prefectures where they have lived longest are Shizuoka, Tottori, Hiroshima (two), and Fukuoka (four). They were paid a small amount of money for their participation.

Again, a practice session was held prior to experiments. In the session, the present author used about 20 stimuli generated based on the word list for practice sessions.

6.1. Experiment III: Shadowing Task

In the shadowing test, the participants were presented with the speech stimuli through a pair of headphones without any background noise and were required to shadow quickly and clearly the word as soon as they identified it. At the same time, they were instructed that they didn't need to shadow the word with the same accent type they heard. They were also instructed not to stutter or make a slip of the tongue, although they could start their utterances before the presentation of the stimuli had finished.

A 500-Hz beep tone was presented before each stimulus. The interval between the beep tone and the stimulus was randomly set between 500 and 900 ms.

The utterances spoken by the participants were recorded and digitalized at an 8-kHz sampling rate. The present author segmented the speech period and recorded the interval between the beginning of a stimulus and the beginning of a participant's utterance as the reaction time. When the utterance was incorrect or stuttered, it was excluded from the data.

Table 2 Frequency of stimuli classified into rating categories and the mean reaction time.

Evaluation	Frequency	Mean of RT (s)	SD of RT (s)
1	226	0.9877	0.2134
2	325	0.9466	0.1904
3	304	0.9205	0.1900
4	312	0.9104	0.1628
5	414	0.8789	0.1858

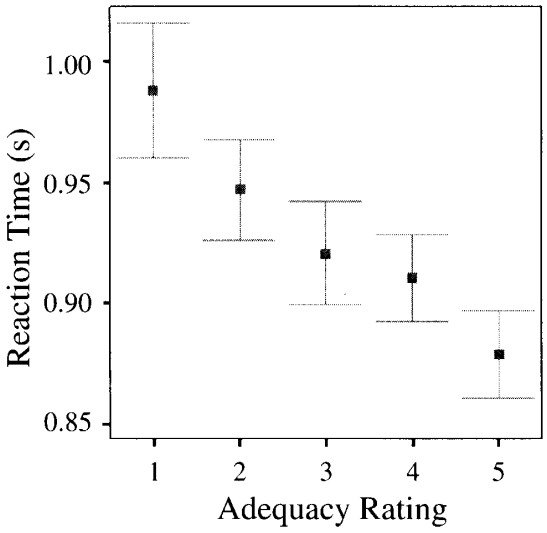

Fig. 3 Relationship between adequacy ratings and mean reaction time with a 95% confidence interval for the mean.

6.2. Results

Table 2 shows the frequency of stimuli classified into rating values and the mean reaction time. A rating of 1 means "inadequate" and a rating of 5 denotes "adequate." The number of excluded utterances was only 19, which amounts to less than 1.2% of all utterances.

Figure 3 illustrates the mean reaction time every rating value with a 95% confidence interval of the mean. Here we can read the tendency that reaction time shortens as the rating value of adequacy increases.

Spearman's rank correlation coefficient was calculated and statistically tested to clarify the relationship between the rating of the adequacy of accent type and the reaction time. The result indicated a statistically significant negative relation [$\rho = -0.177, p < 0.01$] between them.

Marslen-Wilson [17] classified shadowers into two types. Close shadowers can begin to shadow before they understand the word, and the reaction time ranges from about 250 to about 300 ms. Distance shadowers begin to shadow after they have perfectly analyzed the perceived object, and the reaction time is longer than 500 ms. Only eight among 65 participants were close shadowers. In the present experiment, the participants were instructed to shadow when they identified the word; therefore, the task seems to require distance shadowing. As a result, the number of utterances for which reaction time was shorter than 500 ms was only 19 among 1,600, and the shortest reaction time was 409 ms. Thus we could conclude that most of the shadowing in the present experiment was distance shadowing.

6.3. Discussion

The results showed that reaction time in shadowing words with inadequate accent was longer than that for shadowing words with adequate accent. One hypothesis for explaining the result is that recognition of spoken words with an inadequate accent type takes longer even if phonetic information enables listeners to recognize the word. A simple explanation is that pitch-accent information facilitates spoken-word recognition [7]. Another, more complicated, explanation may be possible. It has been demonstrated that nonsense words are processed in a different channel in the brain than normal words [19]. Marslen-Wilson [18] measured the difference of reaction time between the shadowing of normal words and shadowing of nonsense words. He showed that reaction time in the case of normal words was significantly shorter than that for nonsense words. Although "normal" and "nonsense" have been distinguished from a lexical viewpoint, pitch-accent information might play some part in judging whether a word is nonsense or normal at early stage of speech processing. In this case, the difference of reaction time in the present experiment might mean the difference of channel of spoken-word processing in the brain. What role does pitch-accent information take in spoken-word recognition? When does it operate? [20] These interesting problems still remain.

Now, another hypothesis may explain the difference of reaction time: A longer time may be taken not in the perception process but in the production process. It is known that speech perception and production link together. For example, in the experiments with altered auditory feedback the correlation between a perceived fluctuated signal and produced speech signal was observed at about the 150-ms point of reaction time. Thus, can we interpret this as meaning the difference of reaction time observed in the experiment resulted from the perception-production link?

In the experiment, the participants mostly shadowed a word with an adequate accent even if they heard the word with an inadequate accent. Therefore, in the case that the perceived stimuli had more inadequate accents, there were consequently greater differences in pitch pattern between perceived speech and produced speech. A speech motor plan may take longer in that case because of the perception-production link. It is worth noting, however, that the participants were permitted to utter words with any accent

in the experiment. If the perception-production link had any strong influence, it seems rather natural that produced speech would tend to have the accent of perceived stimuli because the participants were required to shadow as soon as possible. Indeed, in a preliminary examination, the present author required one participant to shadow the words with the same accent type as that of the perceived stimuli. Consequently, it was difficult for him to shadow the words and many incorrect utterances were observed. This suggests the possibility that the participant had to freshly re-construct the motor plan to utter the word with the perceived accent. This is no longer called a perception-production link. Therefore, the difference of reaction time in the experiment seems to originate in the perceptual process, although we could not conclude this due to insufficient proof.

7. CONCLUSIONS

In this research, the present author examined the contribution of pitch-accent information to Japanese spoken-word recognition. Using speech stimuli to manipulate those pitch accents, the present author conducted a word intelligibility test under different SNR conditions and a rating test for the adequacy of accent types in words. The word intelligibility scores were examined based on the adequacy rating of the accent types. Results reveal that under noisy conditions, the higher the adequacy of the accent type was rated, the higher the word intelligibility was.

In the clean condition, the participants were presented speech stimuli with manipulated accent type and shadowed the words as quickly as possible. The present author measured the reaction time in shadowing, with the results indicating a significant negative correlation between reaction time and adequacy rating of accent types.

These results showed the contribution of pitch-accent information to word recognition.

ACKNOWLEDGEMENTS

Most of the experiments were performed as a collaboration of Kyushu Institute of Design (Kazuo Ueda) and Institute of Systems and Information Technologies (Ikuyo Masuda-Katsuse) from 2003 to 2004. Dr. Kazuo Ueda, Dr. Hideki Kawahara, Dr. Shigeaki Amano, and Dr. Kazuhiko Kakehi kindly gave the present author plenty of constructive advice. The present author could use the STRAIGHT system due to the courtesy of Dr. Hideki Kawahara. Mr. Yuichiro Katsuse patiently continued uttering artificial utterances for a long time in recording the original speech sounds. The present author thanks Mr. Ryuji Furuyama and Dr. Kunikazu Hirosawa for their assistance in all of the experiments. Finally, she thanks many students at Kyushu Institute of Design for participating in the experiments.

REFERENCES

[1] Y. Saito, *Introduction to Japanese Phonetics* (Sanseido, Tokyo, 1997), pp. 112–123.
[2] K. Clark and C. Yallop, *An Introduction to Phonetics and Phonology*, 2nd ed. (Blackwell Publishers, Oxford, 1995), pp. 347–348.
[3] A. Cruttenden, *Intonation* (Cambridge University Press, Cambridge, 1986), pp. 12–14.
[4] Y. Ebata, M. Kato and H. Hondo, *A Dictionary of Japanese Dialects* (Gakken, Tokyo, 1998), p. 13.
[5] A. Cutler and T. Otake, "Pitch accent in spoken-word recognition in Japanese," *J. Acoust. Soc. Am.*, **105**, 1877–1888 (1999).
[6] Y. Kitahara, S. Takeda, K. Ichikawa and Y. Tohkura, "Role of prosody in cognitive process of spoken language," *Trans. IEICE*, **J70-D**, 2095–2101 (1987).
[7] N. Minematsu and K. Hirose, "Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese," *J. Acoust. Soc. Jpn. (E)*, **16**, 311–320 (1995).
[8] I. Masuda-Katsuse, "Relation between word intelligibility in the noise and validity of accent pattern of the word," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 439–440 (2002).
[9] K. Kato, S. Amano and K. Kondo, "Familiarity effects on Japanese spoken words with/without noise," *Tech. Rep. Psychol. Physiol. Acoust. Acoust. Soc. Jpn.*, H-99-8 (1999).
[10] S. Amano and T. Kondo, *Lexical Properties of Japanese* (Sanseido, Tokyo, 1999).
[11] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigne, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Commun.*, **27**, 187–207 (1999).
[12] H. Sato, "Relations between pitch-contour and phoneme," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 435–436 (1974).
[13] K. Ochiai, "A Perceptual study on the effect of the pitch perturbation by the voiceless stop consonant," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 157–158 (1968).
[14] Y. Kato and K. Ochiai, "Rapid transition of glottal pulse amplitude and pitch frequency as a cue of voiced plosives," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 391–392 (1967).
[15] F. Grosjean, "Spoken word recognition processes and the gating paradigm," *Percept. Psychophys.*, **28**, 267–283 (1980).
[16] S. Amano, "A word recognition point estimated by gating task," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 355–356 (1992).
[17] W. D. Marslen-Wilson, "Linguistic structure and speech shadowing at very short latencies," *Nature*, **244**, 522–523 (1973).
[18] W. D. Marslen-Wilson, "Speech shadowing and speech comprehension," *Speech Commun.*, **4**, 55–73 (1985).
[19] S. D. Newman and D. Twieg, "Difference in auditory processing of words and pseudowords: An fMRI study," *Hum. Brain Mapp.*, **14**, 39–47 (2001).
[20] S. Amano, "Contemporary models of word perception," *J. Acoust. Soc. Jpn. (J)*, **48**, 20–25 (1992).