Ann. Bull. RILP, No. 12, 141-148 (1978)

WORD ACCENT IN JAPANESE AND ENGLISH: A COMPARATIVE STUDY OF ACOUSTIC CHARACTERISTICS IN DISYLLABIC WORDS

Keikichi Hirose*, Hiroya Fujisaki, and Miyoko Sugito**

1. Introduction

The word accent plays an important role in speech communication, but its acoustic features usually differ from language to language. As the acoustic features we can list up fundamental frequency, segmental duration, acoustic power of vowels, formant frequencies, and others. While it is known that the voice fundamental frequency is the most important feature of word accent in Japanese, the segmental duration is known to be also important in English[1]. We have already analyzed the fundamental frequency contour (F_0-contour) of two-mora Japanese words in the Tokyo and Osaka dialects and revealed the relation between its pattern and accent types [2, 3]. In this paper the fundamental frequency and segmental duration of disyllabic English words are analyzed and compared with those of two-mora Japanese words, in order to determine both universal and language-specific characteristics of word accent.

2. Speech Materials

The English words chosen as the material for the present analysis are "permit", "record", "object", and others in which the change of part of speech from noun to verb is associated with that of accent position from the first to the second syllable. Four native speakers, one British, two American, and one Canadian, read randomized lists of these disyllabic words pausing about few seconds after each word, and the readings were recorded. For the sake of comparison, a native speaker of the Osaka dialect read a randomized list of four accent types of the two-mora words "ame" [2, 3]. Some personal data of the speakers are listed in Table 1. These speech materials are sampled at 10 kHz, quantized with 10 bit accuracy and stored in the magnetic tape memory.

Table 1. Personal data of speakers.

Speaker	Sex	Age	Country	Dialect
1	female	32	England	Hampshire
2	male	32	U.S.A.	New York
3	male	44	Canada	Ottawa
4	male	33	U.S.A.	Kansas
5	male	50	Japan	Osaka

*Faculty of Engineering, University of Tokyo
**Osaka Shoin Women's College

3. Analysis and Comparison of Fundamental Frequency Contour Characteristics

3-1 Extraction of F_0-contours

The fundamental frequencies are extracted by a method based on the short-term autocorrelation analysis and peak detection [4]. Once the fundamental period is extracted by the short-term autocorrelation analysis, the value is utilized to facilitate detection of the next period directly from the speech waveform. These fundamental periods, detected pitch-synchronously, are converted to fundamental frequencies which are further interpolated to produce an F_0-contour uniformly sampled at intervals of 12.8 msec.

3-2 Extraction of feature parameters from an F_0-contour

The F_0-contour can be decomposed into the voicing and the accent components which are assumed to be originally binary, but are smoothed by various neural, muscular, and pneumatic factors in the process of speech production. The basic concept of the model is illustrated in Fig. 1 [5, 6].

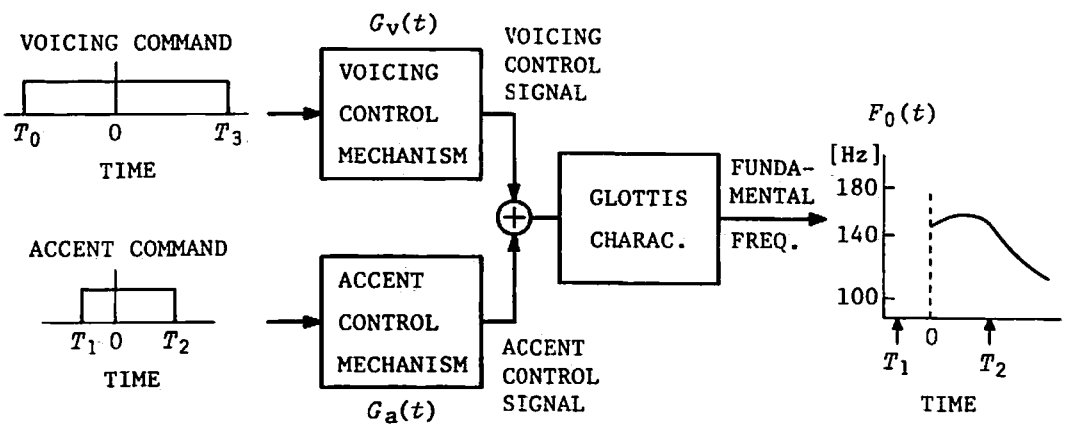

Fig. 1. A functional model for the process of generating an F_0-contour from voicing and accent commands.

Linguistic factors of voicing and accent are both assumed to have the form of stepwise binary commands to the control mechanism of the fundamental frequency. Commands for voicing and accent are smoothed separately by the low-pass characteristics of their respective control mechanisms, each being approximated by a critically damped second-order linear system. The output fundamental frequency F_0 as a function of time t is given by

$$\ln [F_0(t)/F_{min}] = G_v(t - T_0) + G_a(t - T_1) - G_a(t - T_2) - G_v(t - T_3), \quad (1)$$

where

$$G_v(t) = [A_v \alpha t \exp(-\alpha t)]\, u(t),$$

$$G_a(t) = [A_a \{ 1-(1+\beta t) \exp(-\beta t)\}]\, u(t),$$

u(t): step function,

T_0 : instant of onset of voicing command,

T_1 : instant of onset of accent command,

T_2 : instant of offset of accent command,

T_3 : instant of offset of voicing command.

The accentual features are well represented by T_1 and T_2. Using the above mentioned model, characteristic parameters of an observed F_0-contour can be extracted by the method of Analysis-by-Synthesis.

3-3 Results

Examples of the extracted fundamental frequencies (marked with "+") and the best approximations by the model (solid curves) are shown in Fig. 2 (a)-(d) for the voiced segments of "pérmit/permít". Panels (a) and (b) are for Speaker 1 (Hampshire, England), and panels (c) and (d) are for Speaker 2 (New York, U.S.A.). Examples of F_0-contours of "āme/amẽ" are also shown in Fig. 2 (e) and (f) for comparison. While a general similarity can be observed between the F_0-contour characteristics of "pérmit/permít" and "āme/amẽ", respectively, individual differences are found to be much greater in the timing of the accent command for "pérmit" than in other cases. In (a) the offset of the accent command is located close to the voicing onset while in (c) it lags behind the voice onset by about 250 msec, producing a marked similarity to the F_0-contour of "āme" shown in (e). The offset of accent command of "pérmit" is located at several tens of seconds after the voice onset in utterance samples by Speakers 3 and 4.

In Table 2 the characteristic parameters extracted from utterance samples are summarized for Speakers 1 and 2. The parameters are averaged over three samples in Speaker 1 and over four samples in Speaker 2.

Table 2. Parameters of fundamental frequency contours of "permit" extracted by Analysis-by-Synthesis.

Parameters / Accent Types	F_{min} (Hz)	T_0 (sec)	T_3 (sec)	A_v	α (sec^{-1})	T_1 (sec)	T_2 (sec)	A_a	β (sec^{-1})
"pérmit" (Speaker 1)	129	-0.19	0.30	1.18	5.27	-0.27	0.02	0.24	22.0
"permít"	119	-0.30	0.34	1.58	5.40	0.09	0.26	0.45	19.7
"pérmit" (Speaker 2)	78	-0.24	0.45	1.16	4.93	-0.10	0.22	0.43	28.3
"permít"	77	-0.32	0.36	1.51	4.20	0.13	0.27	0.49	26.5

4. Analysis and Comparison of Segmental and Syllabic Durations

4-1 Measurement of segmental and syllabic durations

The segmental and syllabic durations of words are measured directly from the waveforms, referring to the sound spectrogram.

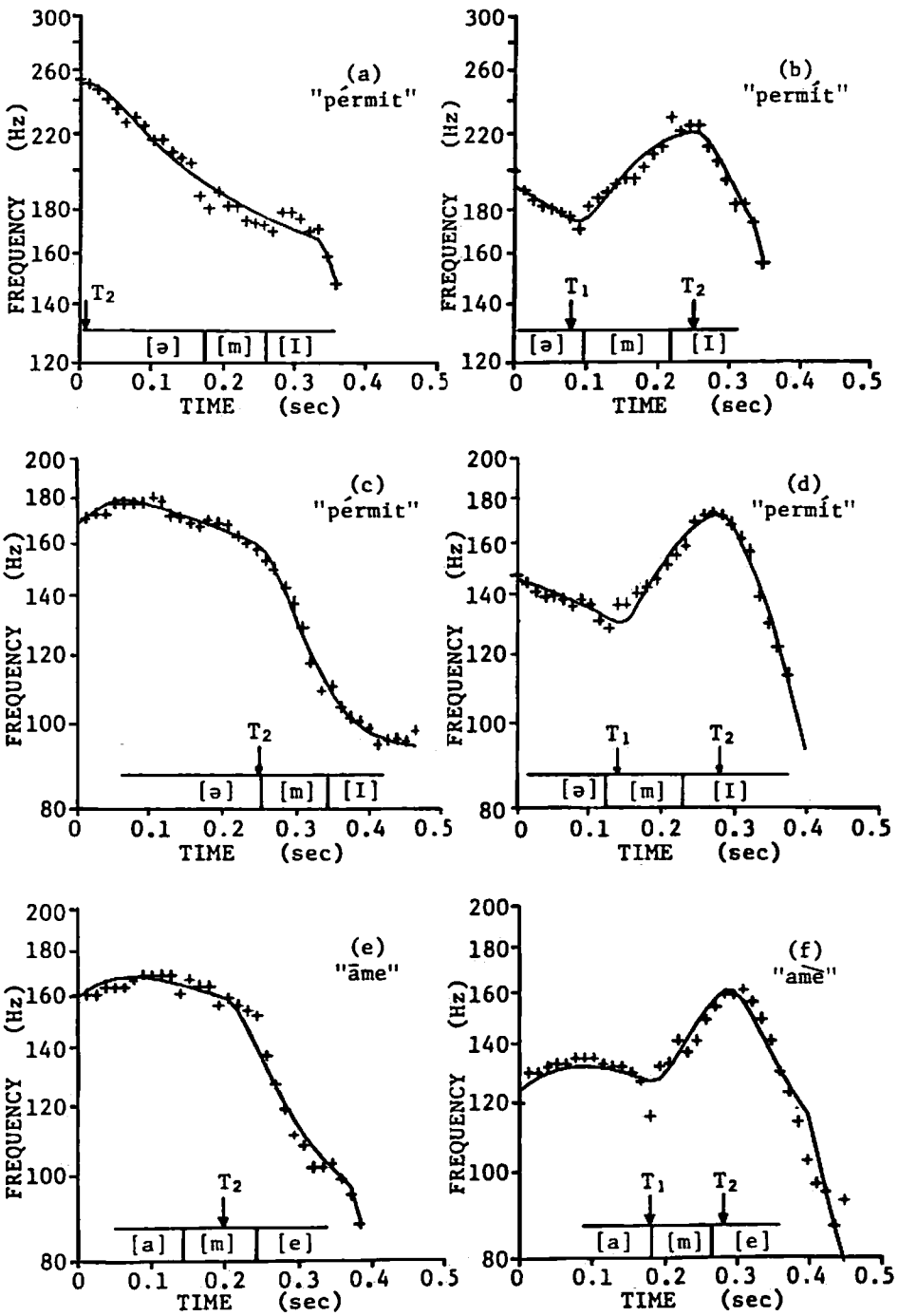

Fig. 2. Analysis-by-Synthesis of F_0-contours in (a) "pérmit" by Speaker 1, (b) "permít" by Speaker 1, (c) "pérmit" by Speaker 2, (d) "permít" by Speaker 2, (e) "āme" by Speaker 5, and (f) "amẽ" by Speaker 5.

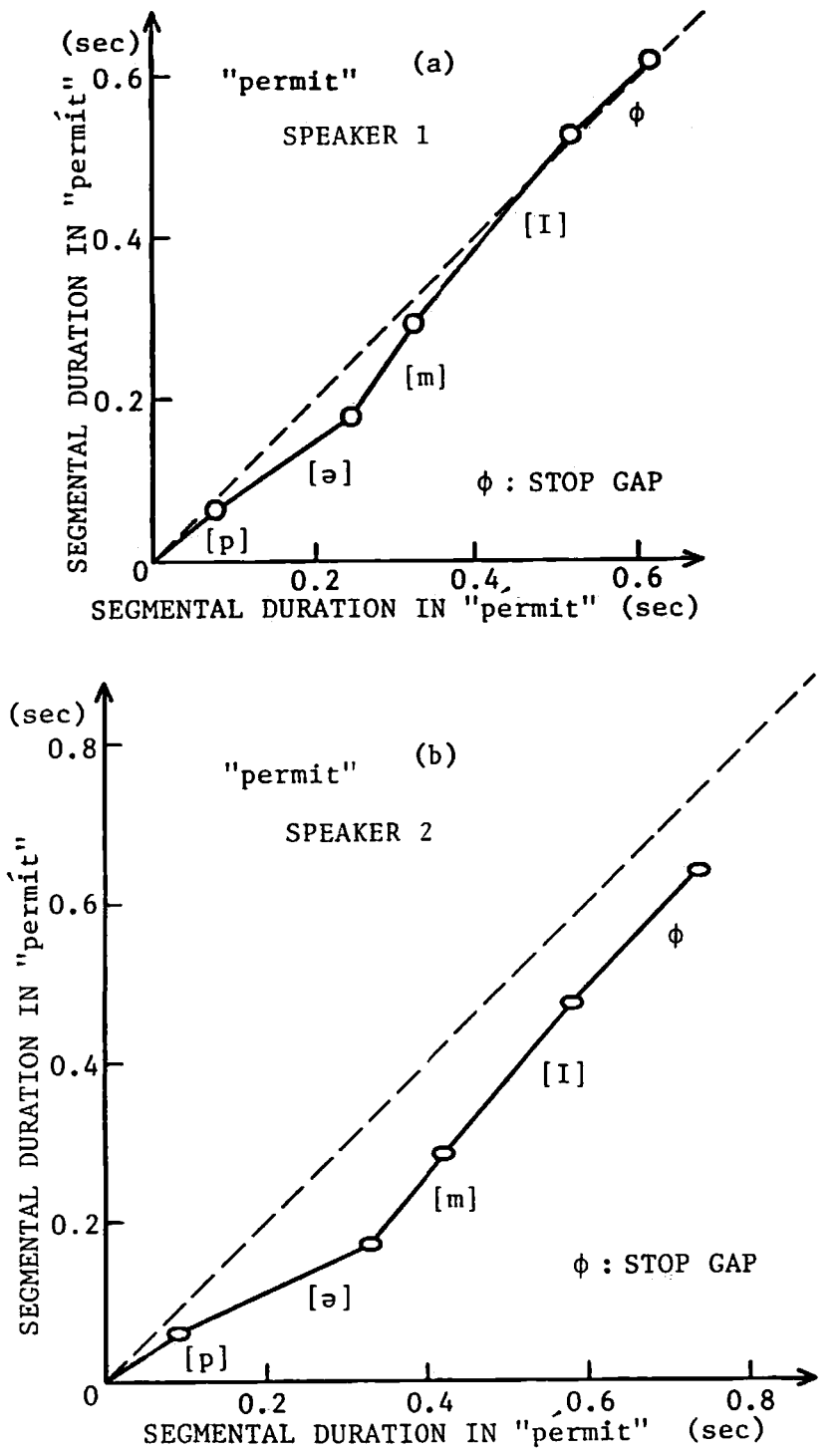

Fig. 3. Comparison of segmental durations in the two accentual types of "permit" by (a) Speaker 1 and (b) Speaker 2.

4-2 Results

Segmental durations of "permit" are shown in Fig. 3 (a) and (b) for Speakers 1 and 2, respectively. The durations are averaged over several typical samples. Definitions of segmental durations are as follows:

[p]; from noise burst to voice onset,

[ə]; from voice onset to onset of oral occlusion,

[m]; from onset to offset of oral occlusion,

[I]; from offset of oral occlusion to voice offset,

ϕ ; from voice offset to noise burst.

The solid line indicates the relationship between cumulative durations of "pérmit" (ordinate) and "permít" (abscissa), and the deviation of the solid line from the dotted line indicates the change of segmental durations due to the accentual change.

The total duration of "pérmit" is nearly the same as that of "permít" in Speaker 1 and slightly longer in Speaker 2.

The segmental durations in "ame" are shown in Fig. 4. Contrary to the case of "permit" the accentual position has hardly any effect upon the durations of [a] and [m].

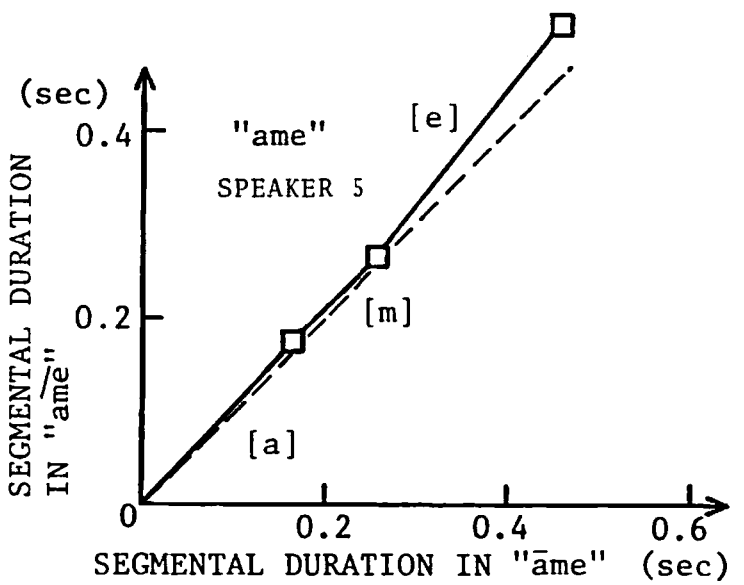

Fig. 4. Comparison of segmental durations in the two accentual types of "ame" by Speaker 5.

The accentual change of syllabic durations are shown in Fig. 5. It is to be noted that the final segment [t] is not included in the duration of the second syllable. In "ame", accentual changes are only found in the duration of the second syllable, while in "permit" they are found to be complementary in the first and the second syllables.

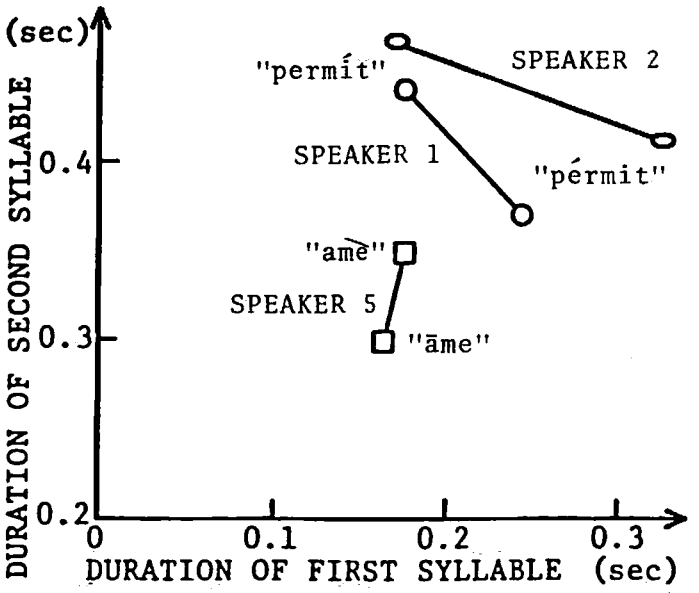

Fig. 5. Durations of the first and the second syllables in "permit" and "ame".

5. Temporal Relationship between Segmental and Suprasegmental Features

Temporal relationships between segmental and suprasegmental features can be investigated by Fig. 2 where the timing of [m] is indicated at the bottom of each panel. The offset of accent command of "pérmit" differs individually to a large extent and its location is distributed from the beginning to the end of [ə] while that of "āme" is located within [m]. In the case of "permít" and "amẽ" the onset and offset of the accent command are also located around [m].

6. Conclusions

Suprasegmental features of the disyllabic words in Japanese and English are investigated mainly in regards to the F_0-contour and the segmental duration. In "ame" only the duration of the second syllable changes distinctly, but in "permit" the durations of the two syllables change complementarily. The F_0-contours of "permit" are similar to those of "ame" in both cases of first-syllable accented and second-syllable accented. However in "pérmit" large individual differences are found in the offset of the accent command. It is concluded that the functional model for the F_0-contour generation is also valid for the English words, though their accentual features are somewhat different from those of the Japanese words.

In the present study every word samples are spoken separately, but the contextual change of acoustic features needs to be examined when the samples are in the sentence context. Changes of the acoustic features also need to be examined in words where the change in accent position is accompanied by phonemic changes like in "récord/recórd".

Acknowledgment

The research reported here was supported by a Grant-in-Aid for Scientific Research (No. 310707) from the Ministry of Education.

References

1. Fry, D. B. (1955); Duration and Intensity as Physical Correlates of Linguistic Stress, J. Acoust. Soc. Am., 27, 765-768.

2. Fujisaki, H., Y. Mitsui and M. Sugito (1974); Analysis, Synthesis and Perception of Accent Types, Transactions of the Committee on Speech Research, Acoust. Soc. Japan, S73-51.

3. Fujisaki, H. and M. Sugito (1976); Acoustic and Perceptual Analysis of Two-Mora Word Accent Types in the Osaka Dialect, Ann. Bull. RILP, No. 10, 157-171.

4. Fujisaki, H. (1960); Automatic Extraction of Fundamental Period of Speech by Autocorrelation Analysis and Peak Detection, J. Acoust. Soc. Am., 32, 1518.

5. Fujisaki, H. and S. Nagashima (1969); A Model for the Synthesis of Pitch Contours of Connected Speech, Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 28, 53-60.

6. Fujisaki, H. and H. Sudo (1971); A Model for the Generation of Fundamental Frequency Contours of Japanese Word Accent, J. Acoust. Soc. Japan, 27, 445-453.

7. Fujisaki, H. and H. Sudo (1971), Synthesis by Rule of Prosodic Features of Connected Japanese, Proc. 7th ICA, 3, 133-136.