Tsukuba English Studies (1997) vol. 16, 51-65

Syllable Subsidiary of the English [aɪ]

Takeshi SHIMADA*

1. Introduction

It is often said that the [ɪ] element in [aɪ] is not necessarily articulated accurately as the short vowel [ɪ] as in "hid."[1] Rather the movement of the tongue from [a] towards [ɪ] is sufficient for an RP native speaker to identify [aɪ] correctly, that is, [ae], or even [aɛ] is a possible realization form of /aɪ/ (cf. Jones (1960), Laver (1994), Takebayashi (1996), etc.). From this fact, the [ɪ] element is regarded as a goal in articulation of [aɪ]. However, one question arises. Is [aɪ] really pronounced as [ae] or [aɛ]? In order to answer the question, the acoustic characteristics of the syllable subsidiary of the English [aɪ] must be examined. To my knowledge, however, few acoustic studies which focus on the syllable subsidiary of the diphthongs in RP have been done.[2] The main purpose of this paper is to analyze [aɪ] in RP acoustically, and to show the acoustic nature of the [ɪ] element in [aɪ].

2. Experiment

2.1. Methods

To test the acoustic character of the [ɪ] element, data were collected and analyzed. Twenty-seven [aɪ] phonetic data were collected from the audio-cassette tapes included in Shimaoka and Wells (1992) and Fletcher (1990). All the data were pronounced by a male speaker of RP (Received Pronunciation). These data were measured by KAY CSL 4300™ for the first formant (F1) and the second formant (F2). The measuring procedure differed between [a] and [ɪ] in [aɪ], since the [a] element has the most stressed point in [aɪ] where formant values can be measured, whereas the [ɪ] element is the endpoint which is observed more easily by its spectrogram than by its intensity value. As for the [a] element, the maximal values for intensity were first measured. At the point where the maximum value for intensity was recorded, the values for F1 and F2 were measured. In the case of [ɪ], its spectrogram was obtained to specify the endpoint of [ɪ]. At that point, the values for its F1 and F2 were measured.

2.2. On Formants

In this experiment, the values for F1 and F2, were measured. F1 and F2 were used to estimate the tongue height and the tongue position (or advancement) of the syllable nucleus [a] and the syllable subsidiary [ɪ]. The tongue height is estimated with the value for F1 and the tongue position with the value for F2. The estimation

of the tongue height and the tongue position was made to plot the values for F1 and F2.

In estimating the tongue height and the tongue position, a simple rule is used to relate the vowel formant frequencies to vowel articulation. This rule states that F1 varies mostly with tongue height and F2 varies mostly with tongue position (cf. Kent and Read (1992)).[3] Specifically, low vowels have a high F1 frequency and high vowels have a low frequency. Back vowels have a low F2 frequency, and typically a small F2-F1 difference, whereas front vowels have a relatively higher F2 frequency and a large F2-F1 difference (Kent and Read (1992)). Let us take Figures 1a and 1b as an example and observe how this rule works.

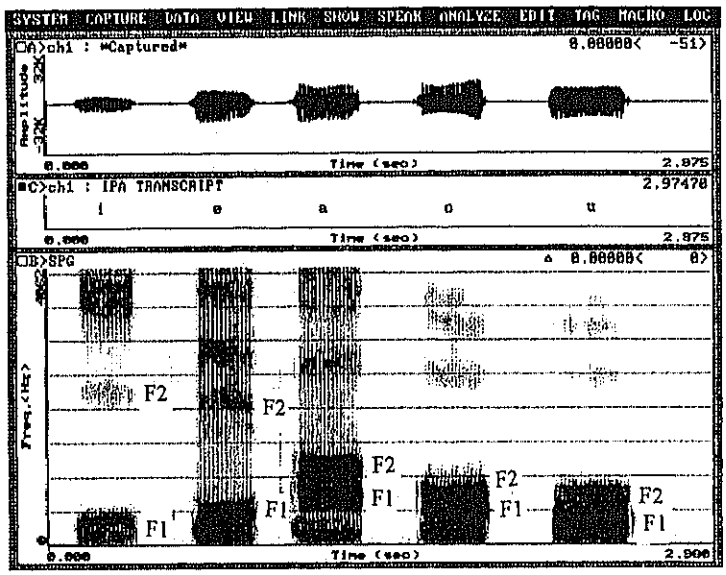

Figure 1a. A spectrogram of the Japanese five vowels [i, e, a, o, u]

Figure 1a shows a spectrogram of the Japanese five vowels [i, e, a, o, u] pronounced by the author. Note that different vowels have different frequencies of the formants labeled "F1" and "F2." The differences among their frequencies of F1 and F2 correspond to the differences of tongue height and tongue position of a vowel, respectively. In Figure 1a, we find that the F1 for [a], a low vowel, is the highest of all the vowels and that the F1 for [i], a high vowel, is the lowest. The F1 for [u], which is also a high vowel, has almost the same F1 frequency as [i]. This confirms that the higher the frequency of F1, the lower the tongue height; and the lower the frequency of F1, the higher the tongue height. Next, consider the correspondence between F2 frequency and tongue position (or advancement). The front vowel [i] has the highest F2 frequency and [u], a back vowel, has the lowest F2 frequency. This

also confirms that the rule relative to tongue height and tongue position works well. The higher the F2 frequency is, the more the tongue is advanced, and the lower the F2 frequency is, the more the tongue is retracted.

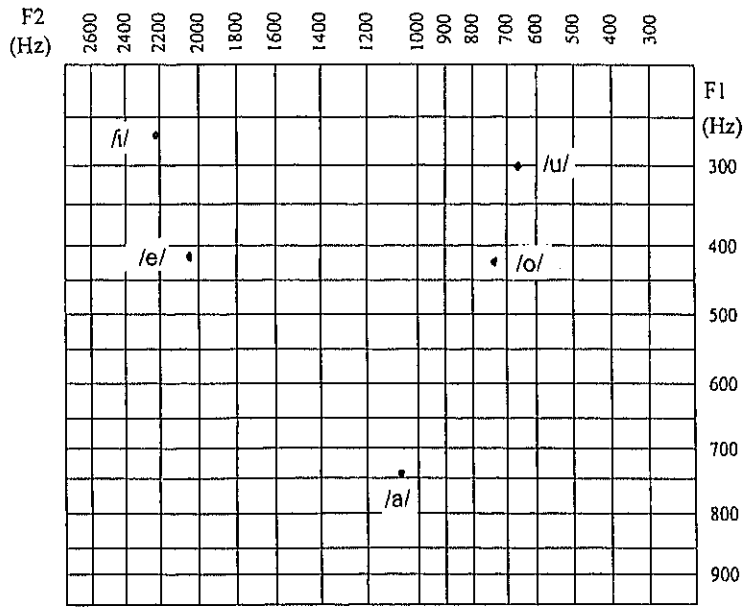

Figure 1b. A formant chart showing the relationship between the Japanese five vowels

In order to illustrate graphically the correspondence between F1 and tongue height, and between F2 and tongue position, the F1-F2 chart is often used like Figure 1b. Figure 1b shows the frequency of F1 on the vertical axis plotted against the frequency of F2 on the horizontal axis for the above five vowels produced by the author. Using this F1-F2 chart we can estimate the tongue height and the tongue position of a vowel (cf. Kent and Read (1992)), or the quality of a vowel (cf. Ladefoged (1993)), since the shape drawn by the frequencies of F1 and F2 of their vowels corresponds to a traditional vowel chart as illustrated in Figure 2 (from Okada (1991)).

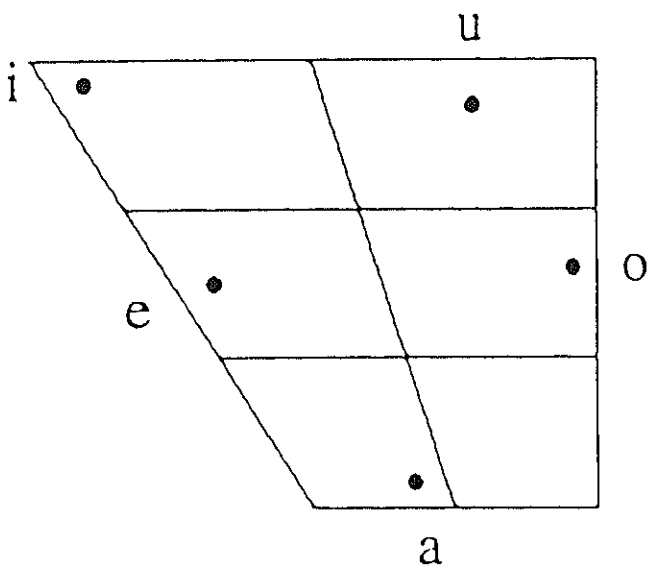

Figure 2. A vowel chart showing the tongue height and the tongue position of the Japanese vowels

Figure 2 shows the tongue height and the tongue position of the five Japanese vowels. Note that the relative position of the five vowels in Figure 1b and Figure 2 is almost the same, except the tongue position of [u]. The [u] in Figure 2 has a more advanced tongue position than that in Figure 1b, since the [u] in Figure 2 is an unrounded vowel, i.e. [ɯ] while the [u] in Figure 1b is a rounded one. This suggests that the tongue height and the tongue position of a vowel can be estimated fairly accurately by the frequencies of F1 and F2. In the next section, we will examine the characteristics of [aɪ] through the correspondence between acoustic features and articulatory features.

3. Results and Discussion

3.1. Mean values for F1 and F2 of [aɪ].

Table 1 and Table 2 show the frequencies of F1 and F2 of the [a] element and the [ɪ] element in [aɪ] respectively.

	MAX.	MIN.	MEDIAN	SD	Token
F1(Hz)	766	404	611	87	27
F2(Hz)	1897	766	1234	212	27

Table 1. The numerical data of the [a] element in [aɪ]

	MAX.	MIN.	MEDIAN	SD	Token
F1 (Hz)	401	218	306	107	27
F2 (Hz)	2226	1167	1808	277	27

Table 2. The numerical data of the [ɪ] element in [aɪ]

Table 1 indicates that the values for F1 of [a] vary between 404 and 766 Hz and that the values for F2 vary between 766 and 1897 Hz. The value of the median for F1 of [a] in [aɪ] is 611 Hz and for F2 1234 Hz. The values for standard deviation (SD) of F1 and F2 are 87 and 212. Since the greater value for SD means a greater degree of dispersion, these SD values indicate that the range of the tongue position of the [a] element is wider than that of its tongue height. In fact the [a] element in [aɪ] can display its quality between [a] and [ɑ] as has been pointed out by Jones (1960), and Gimson (1994), among others.

Table 2 indicates that the values for F1 of the [ɪ] element in [aɪ] vary between 218 and 401 Hz, and that the values for F2 vary between 1167 and 2226 Hz. The values for SD of F1 and F2 are 107 and 277 Hz. These values suggest that the range of the tongue position of the [ɪ] element is wider than that of its tongue height.

It is noteworthy that the F1 values of the [ɪ] element are 401 Hz maximum and 218 Hz minimum. The maximum value enables us to estimate that the tongue height is the same as that of a short vowel [ɪ] as in "hid"; on the other hand, the minimum value enables us to estimate that the tongue height is the same as that of [i] (cf. Gimson (1994)).[4] This shows that the tongue height of the [ɪ] element varies from [ɪ] to [i]. The fact is opposed to the native speaker's intuition that the [ɪ] element can be pronounced as [e]. Rather the [ɪ] element should be considered as pronounced as [j] as in American English (cf. Olive, Greenwood and Coleman (1993)) as long as the results in the present study are concerned. The same is true for F2, i.e., the tongue position of the [ɪ] element. Its F2 maximum value is 2226 Hz. This compares to the F2 value of [i], or [j]-element of [eɪ] in American English. However, the F2 minimum value of 1167 Hz enables us to estimate that its tongue position is the same as a central vowel like [ɨ]. In this case too, the [ɪ] element is not considered to be [e] numerically. Figure 3, in which the mean [aɪ] is plotted, clarifies this point.[5]
The mean [aɪ] begins with [a] running to [ɪ]. The second element has the quality of [ɪ] not [e]. Why do previous studies regard [ɪ] in [aɪ] as [e]? One reason may be that the loudness of [ɪ] is very small and its duration is very short (cf. Roach 1991). Another reason may be that the gliding nature, rather than the actual frequencies of [ɪ], is a clue to the perception of the diphthongs (cf. Gay (1970)).

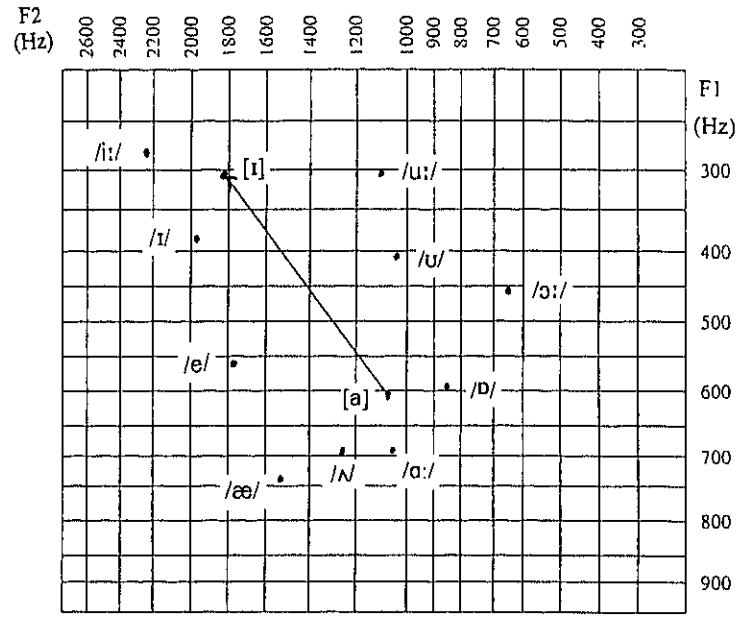

Figure 3. A formant chart for the mean [aɪ]

Now we are in a position to examine the exact frequencies of the second elements, since the frequencies reflect the articulatory characteristics of the diphthongs, i.e., the motor control for the production for the diphthong. In the next subsection, we consider the actual frequencies of the [ɪ] element in [aɪ] in several contexts and identify the [ɪ] element as [ɪ] not [e] or [ɛ].

3.2. The [ɪ] element in [aɪ] in several contexts

In this subsection, we will discuss the influence of the consonants [t], [s], and [n]. Since these consonants occur in high frequency (Gimson (1994)), the [ɪ] element followed by these consonants can be regarded as representative of the syllable subsidiary of [aɪ]. The influence of degree of stress on the [ɪ] element will be examined. Note that the [ɪ] element is always pronounced weaker than the nucleus [a]. In English, if a vowel is weakened, then it is reduced. In other words, its quality becomes closer to [ə]. We will observe whether this reduction undergoes the [ɪ] element.

First of all let us consider the examples followed by no consonants. It seems that in English the example which occurs most frequently is the singular first person pronoun "I." The F1 value of the [ɪ] element is 218 Hz, and the F2 value is 2043 Hz as shown in Table 3.

		F1(Hz)	F2(Hz)
a.	I	218	2043
b.	high	218	2080

Table 3. The F1 and F2 frequencies of the [ɪ] element in [aɪ] followed by no consonants.

The F1 value of the [ɪ] element in "I" has the minimum value of all examples in the present experiment. Its value indicates that the tongue height of the [ɪ] element in "I" is the same as [i] or [j], not [e]. The F2 value shows that it also has the same tongue position as that of [i] or [j], not [e].

Let us consider another example, the word "high." The [aɪ] in "high" is preceded by the onset [h] but not followed by any coda consonant. Thus we predict that the [ɪ] element in "high" shows the same behavior as that in "I"; that is, it is articulated as [ɪ] or [i], not [e]. As illustrated in Table 3, the F1 value of the [ɪ] element in "high" is 218 Hz, which is the same as that of the [ɪ] element in "I," and its F2 value is 2080 Hz. Both of the values for F1 and F2 of the [ɪ] element in "high" indicate that the [ɪ] element is the same tongue height and tongue position as those of [ɪ] in "I" as we predict. The examples which we have observed here indicate that their second element of [aɪ] is pronounced as [ɪ] or [i] and not pronounced with the quality of [e] nor [ɛ] as shown in Figure 4.

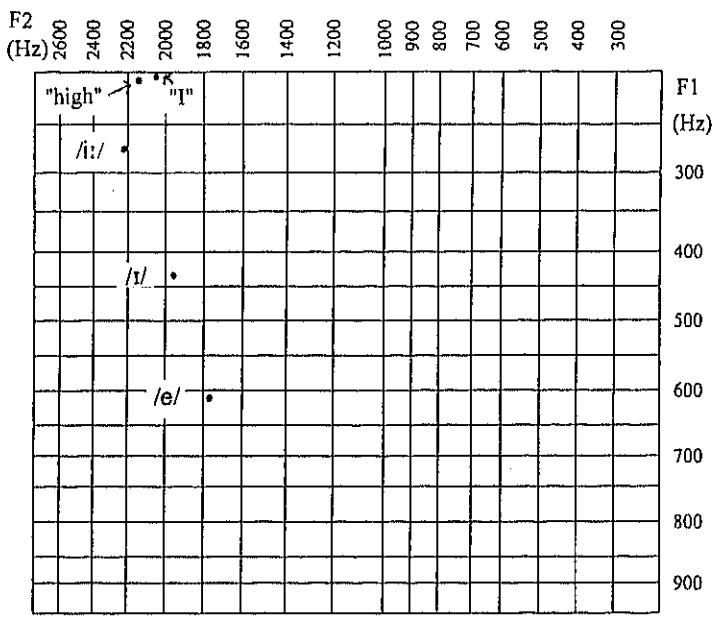

Figure 4. A formant chart for the [ɪ] elements in "I" and "high"

Figure 4 indicates that the positions of the [ɪ] elements in "I" and "high" are close to [ɪ] rather than [e] or [ɛ].

Next consider the examples which are followed by [t]. The F1 value of the [ɪ] element in "fight" is 364 Hz and its F2 value is 2153 Hz as shown in Table 4.

		F1(Hz)	F2(Hz)
a.	fight	364	2153
b	light	328	2226
c.	frightening	255	1167

Table 4. Examples of [aɪ] followed by [t]

These values indicate that the tongue height of [ɪ] in "fight" is the same as that of the short vowel [ɪ] and the tongue position is more advanced than that of the short vowel [ɪ]. The F1 value of the [ɪ] element in "light" is 328 Hz and its F2 value is 2226 Hz as in Table 4. As the F2 value indicates, the tongue position of the [ɪ] element in "fight" is more advanced than the short vowel [ɪ] and the [ɪ] element in "fight." Their relative positions are shown in Figure 5.

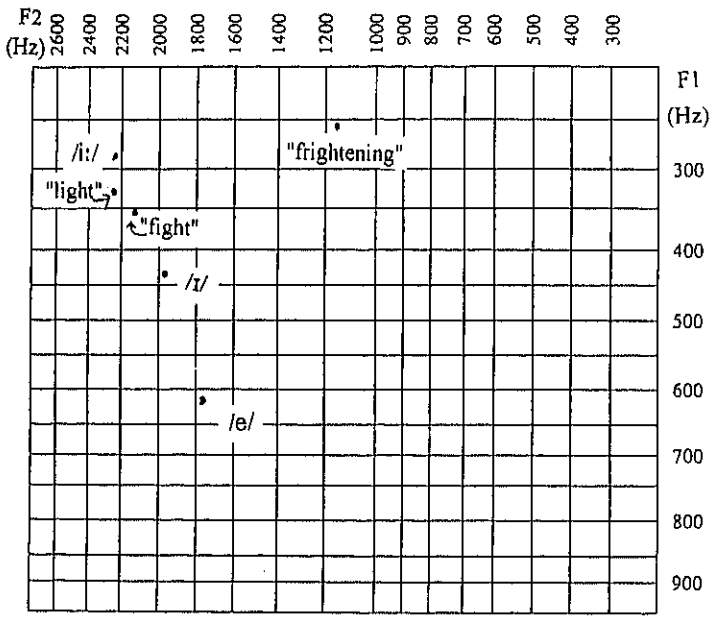

Figure 5. A formant chart for the [ɪ] elements in "fight," "light," and "frightening"

Figure 5 shows that both [ɪ] elements have a higher tongue height and a more advanced tongue position than [ɪ]. This means that they are precisely articulated as [ɪ] not [e].

Figure 5, however, shows that the [ɪ] element with the cental tongue position even if it is in the vicinity of [t]. Consider [aɪ] in "frightening." The [ɪ] element in [aɪ] in "frightening" has the F1 value of 225 Hz and the F2 value of 1167 Hz. The F1 value is satisfactory for it to be regarded as [ɪ]. However, its F2 value is too small to be regarded as [ɪ]; rather it should be regarded as [ɨ] or [ɯ]. One reason why the [ɪ] element of [aɪ] in "frightening" is centralized is that it is influenced by [ə] in the second syllable. Since [ə] has a central tongue position, the preceding [ɪ] is elementy assimilated in anticipation of the cental tongue position. However, the assimilation is incomplete, since the preceding [ɪ] element has a high tongue height, whereas [ə] is a mid tongue position.

In sum, as long as the examples in the neighborhood of [t] are concerned, their [ɪ] elements of [aɪ] have the characteristics of high vowel, not mid vowel such as [e]: the [ɪ] elements of [aɪ] in "fight" and "light" have a tongue height and a tongue position between [ɪ] and [i]; [aɪ] in "frightening" has a centralized [ɪ] element which is close to [ɨ] or [ɯ], but it is a high vowel, not a mid vowel.

Next consider the examples of [aɪ] followed by [s] in Table 5.

		F1(Hz)	F2(Hz)
a.	price	255	2153
b.	decisive	328	1788

Table 5. Examples of [aɪ] followed by [s]

The F1 value in "price" is 255 Hz and its F2 value is 2153 Hz. These values indicate that the [ɪ] element in "price" has the quality of [i], that is, its tongue height is high and its tongue position is front as shown in Figure 6. No feature is found which characterizes [e], that is, [aɪ] is pronounced as [aɪ] rather than [ae].

Another example is "decisive." There is one difference between "price" and "decisive": the position to which [s] belongs. The consonant [s] in "price" is in coda position while that in "decisive" is in onset position; i.e., [praɪs] vs. [dɪ.saɪ.sɪv] (cf. Jones (1997)). The F1 value of the [ɪ] element in [aɪ] in "decisive" is 328 Hz and its F2 value is 1788 Hz. These values indicate that the [ɪ] element in [aɪ] in "decisive" has the tongue height and the tongue position of the short vowel [ɪ] as

illustrated in Figure 6. It is plotted in the more back and more raised position than the typical [ɪ]. As seen in the examples followed by [s], the [ɪ] elements have the quality of [ɪ] whether [s] belongs to coda or onset. Again [aɪ] is pronounced as [aɪ] rather than [ae].

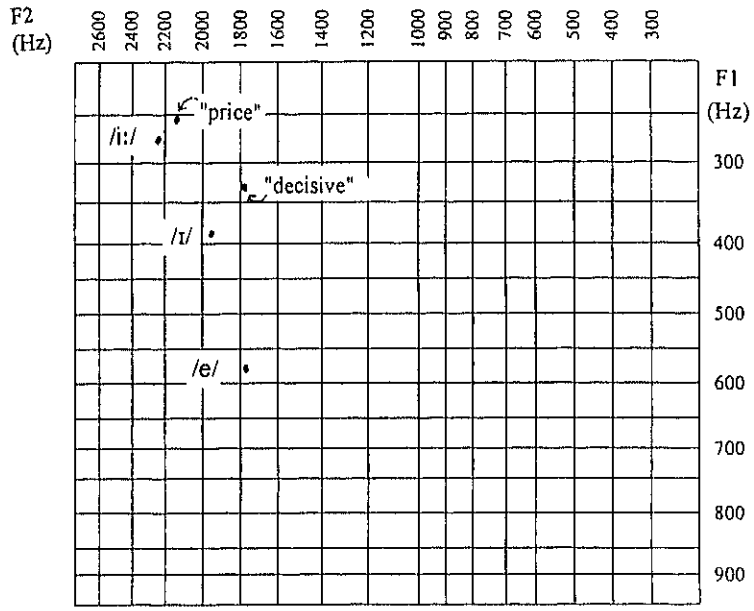

Figure 6. A formant chart for "price" and "decisive"

We next turn to the examples of [aɪ] followed by [n]. The consonant [n] is a coronal consonant as [t] and [s]. Thus we predict that the [ɪ] elements in the vicinity of [n] behave the same as those neighboring [t] and [s], since they are also coronal consonants. In Table 6, we give two examples.

	F1(Hz)	F2(Hz)
kindness	364	2164
nineteen	291	1715

Table 6. Examples of [aɪ] followed by [n]

The F1 value of the [ɪ] element of [aɪ] in "kindness" is 364 Hz and its F2 value is 2164 Hz. These values indicate that the [ɪ] element of [aɪ] in "kindness" has a quality of [ɪ]. As shown in Figure 7, its position is between [ɪ] and [i], not close to [e] and not centralized.

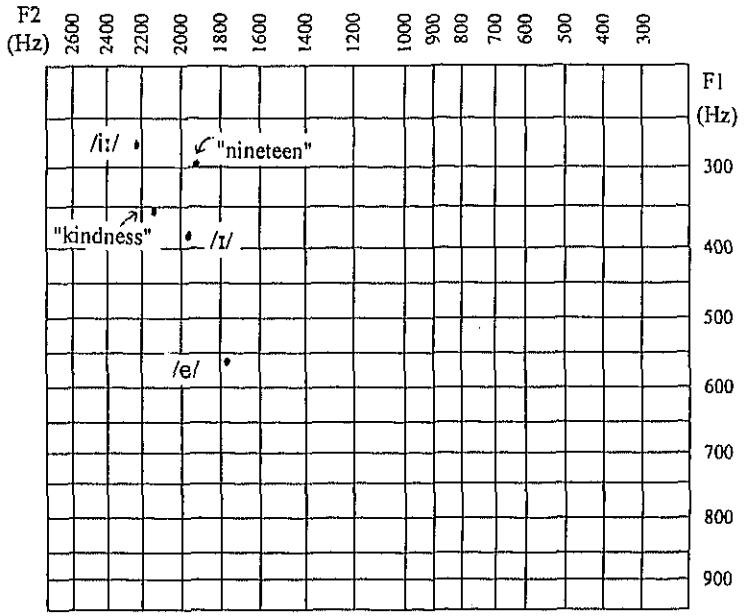

Figure 7. A formant chart for "nineteen" and "kindness"

The [ɪ] element of [aɪ] in "nineteen" has the F1 value of 291 Hz and the F2 value of 1715 Hz. As illustrated in Figure 7, the [ɪ] element in "nineteen" is more raised and more slightly retracted than the short vowel [ɪ]. It can be regarded as [ɪ] not [e]. As is evident from the examples in Table 6 and their tongue heights and tongue positions in Figure 7, the [ɪ] element of [aɪ] in the vicinity of [n] has a quality between [ɪ] and [i] not [e]. That is, the [aɪ]s are pronounced as [aɪ] not [ae].

Before closing the discussion on the syllable subsidiary of [aɪ], let us observe whether the degree of stress influences the tongue height and the tongue position of the [ɪ] element. The [ɪ] element in "ice cream" has the F1 value of 219 Hz, the F2 value of 2043 Hz, and the value for intensity of 71 dB as shown in Table 7.

	F1(Hz)	F2(Hz)	intensity(dB)
ice cream	219	2043	71
I scréam	255	2043	69

Table 7. Examples of [aɪ] with primary stress and secondary stress
: "ice cream" and "I scréam"

The values for F1 and F2 indicate that the [ɪ] element in "ice cream" has a quality of [i] or [j]. The [ɪ] element in "I scréam" has the F1 value of 255 Hz, which is larger than that of "íce cream." However, the F2 value of "I scréam," 2043 Hz, is the same as that of "íce cream." The quality of the [ɪ] element in "I scréam" is nearly the same as that of [i] as shown in Figure 8.

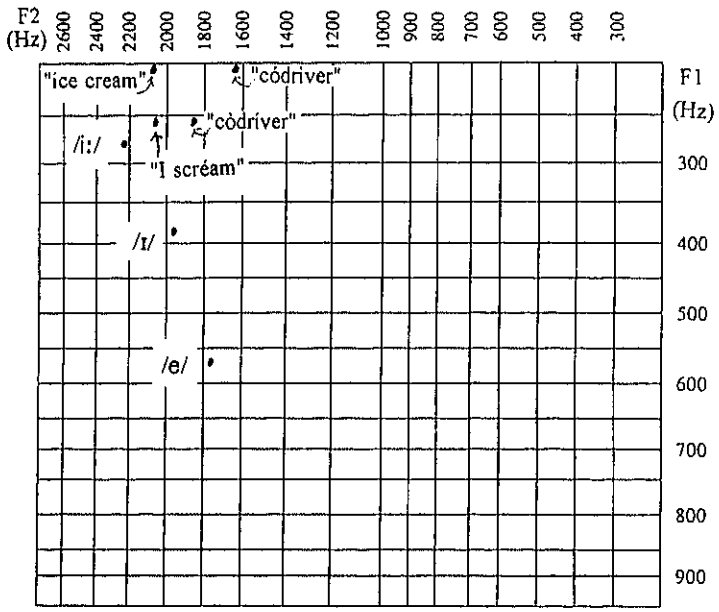

Figure 8. A formant chart for "íce cream," "I scréam," "còdríver," and "códríver"

Now let us compare their tongue heights and their tongue positions in Figure 8. Figure 8 shows that the [ɪ] element in "íce cream" and that in "I scréam" have the same tongue position as is also indicated by the fact that they have the same F2 value, 2043 Hz. On the other hand, the tongue height of the [ɪ] element in "íce cream" is higher than that in "I scréam" as shown in the difference between the F1 value of 219 Hz for "íce cream" and that of 255 Hz for "I scréam." Does the difference in tongue height between "íce cream" and "I scréam" depend on their difference in stress? To answer this question, let us consider another pair which differs only in stress.

	F1(Hz)	F2(Hz)	intensity(dB)
còdríver	256	1751	65
códrìver	219	1642	55

Table 8. Examples of [aɪ] with primary stress and secondary stress : "còdríver" and "códrìver"

In Table 8, the [ɪ] element in "còdríver" has the F1 value of 256 Hz and the F2 value of 255 Hz. These values indicate that the [ɪ] element in "còdríver" has the quality of [ɪ̈]. On the other hand, the [ɪ] element in "códriver" has the F1 value of 219 Hz and the F2 value of 1642 Hz. These values indicate that the [ɪ] element in "códrìver" has the quality of [i]. They are more centralized than those in "ice cream" and "I scréam." This is because the following schwa influences the [ɪ] element in "códrìver" and "còdríver" just as that in "frightening" is affected by the following schwa.

Let us now compare the pair "ice cream" and "I scréam" and the pair "còdríver" and "códriver." The pair "ice cream" and "I scréam" shows that the [ɪ] element with primary stress has a smaller F1 value (219 Hz) than that with secondary stress (255 Hz). As for the pair "còdríver" and "códriver," the [ɪ] element with primary stress has a larger F1 value (256 Hz) than that with secondary stress (219Hz). Thus we can conclude that there is no relation between stress and F1 values.

4. Conclusion

In this paper we have discussed the quality of the syllable subsidiary of [aɪ]. Although previous studies say that the [ɪ] element in [aɪ] is not always realized as [ɪ], and it can surface as [e] or even [ɛ], it is evident from the present study that the [ɪ] element in [aɪ] is pronounced as a high vowel between [ɪ] and [i] and that the consonants in the neighborhood of [aɪ] do not affect the tongue height of [ɪ] so strongly as the tongue is lowered to the position of [e]. With respect to its tongue position, it can vary between front and back depending on the following vowel as in "frightening."

The present study has focused on the F1 and F2 values only. However, other characteristics such as F0 (i.e. fundamental frequency or pitch) and duration must be examined. On this issue, we should await further research.

Notes

* I am grateful to Naotsugu Nakashima, Hideki Zamma, Hideki Tanaka, Koichi Nishida, Hiroyuki Tahara and Yoshitake Nagao for valuable comments and discussion. I also thank Joe Morita for acting as a go-between to ask Naotsugu Nakashima to proofread a draft of this paper. I also appreciate Yvonne Stapp, who proofread a draft and rectified stylistic errors. All remaining errors and inadequacies are of course my own.

[1] In this paper, the term "the [ɪ] element" stands for the sound in the syllable subsidiary of [aɪ] and the term "the short vowel [ɪ]" stands for the vowel as in "hid," respectively.

[2] Bauer (1994) points out "there are surprisingly few independent study of the acoustic nature of RP vowels; much more is available on the vowels of General American English."

[3] There are, however, exceptions. For example, F2 is lowered by lip-rounding. But several studies confirm the general accuracy of this rule (cf. Fox (1983)); Rakerd and Verbrugge (1985)).

[4] The frequencies for F1 and F2 in the short vowel [ɪ] is 382 Hz and 1958 Hz (from Gimson (1994)).

[5] All the values for the vowels /iː/, /ɪ/, /e/, /æ/, /ʌ/, /ɑː/, /ɒ/, /ɔː/, /ʊ/, and /uː/ are cited from Gimson (1994).

References

Bauer, L. (1994) *Watching English Change: An Introduction to the Study of Linguistic Change in Standard Englishes in the Twentieth Century*, Longman, London.

Fletcher, C. (1990) *Longman Pronunciation Dictionary Study Guide*, Longman, Essex.

Fox, R. A. (1983) "Perceptual Structure of Monophthongs and Diphthongs in English," *Language and Speech* 26, pp. 21-60.

Gay, T. (1970) "A Perceptual Study of American English Diphthongs," *Language and Speech* 13, pp. 65-88.

Gimson, A. C. (1994) *An Introduction to the Pronunciation of English*, Fifth Edition, Revised by Alan Cruttenden, Edward Arnold, London.

Jones, D. (1960) *An Outline of English Phonetics*, Ninth Edition, Cambridge University Press, Cambridge.

Jones, D. (1997) *English Pronouncing Dictionary*, Fifteenth Edition, Peter Roach and James Hartmann (eds.), Cambridge University Press, Cambridge.

Kent, R. D. and C. Read (1992) *The Acoustic Analysis of Speech*, Singular, San Diego.

Knowles, G. (1987) *Patterns of Spoken English*, Longman, London.

Ladefoged, P. (1993) *A Course in Phonetics*, Third Edition, Harcourt Brace Jovanovich, Fort Worth.

Laver, J. (1994) *Principles of Phonetics*, Cambridge University Press, Cambridge.

Okada, H. (1991) "Japanese (Illustrations of IPA)," *Journal of the International Phonetic Association* 21:2, pp. 94-96.

Olive, J., A. Greenwood and J. Coleman (1993) *Acoustics of American English Speech*, Springer-Verlag, New York.

Rakerd, B. and R. R. Verbrugge (1985) "Linguistic and Acoustic Correlates of the Perceptual Structure Found in an Individual Differences Scaling Study of Vowels," *Journal of the Acoustical Society of America* 77, pp. 296-301.

Roach, P. (1991) *English Phonetics and Phonology*, Second Edition, Cambridge University Press, Cambridge.

Shimaoka, T. and J. C. Wells (1992) *Jishin ga Tsuku Eigo Hatsuon*, Fukutakeshoten, Tokyo.

Takebayashi, S. (1996) *Eigo Onseigaku*, Kenkyusha, Tokyo.

Wells, J. C. (1990) *Longman Pronunciation Dictionary*, Longman, Essex.

Doctoral Program in Literature and Linguistics
University of Tsukuba
KHF04562@niftyserve.or.jp