An Acoustic Study of British [θriː] Misheard as [fɔː]

Kazuaki ICHIZAKI
Miyazaki Gakuen Jr. College, Japan
email: ichizaki@mwjc.ac.jp

［fɔː］と聞き間違われる英音［θriː］の音響的研究

市　﨑　一　章

要　旨

短期大学１年生の英語コミュニケーションのクラスで、筆者を含めて受講生全員が、イギリス英語の three [θriː]をfour [fɔː] に聞き間違うという事態が起こった。その原因を探るべく、聞き間違った three を同一話者が発音した four と対比させながら音声分析を行った結果、調音時の音の強さ不足、子音の調音位置のずれ、母音の調音点のずれ、継続時間の近似、より大きな卓立を有した後続音節が引き起こす（聴覚印象上の）先行音の弱化等が要因となったことが判明した。

1. Introduction

What caused the author to start the present research was an error in listening that occurred in an English communication class at college. The class was for the first year students taught by two teachers of English; one is British, and the other is Japanese, the author of the present paper; using a textbook for English communication with video published by Oxford University Press. There was a dialogue uttered by four British characters and eight students were asked to find some errors in the script summarising the dialogue. The following sentence in the dialogue became the target of the research: "*I went to Barcelona about three years ago*." The corresponding script was "*Helen went there three years ago with her family*." and all the students, and what was worse, the author too, pointed out that the *three* must have been pronounced *four* in the video. Only the British teacher listened to the statement correctly and he did not understand for a while why the other members responded with such a mysterious answer. He explained that British people, especially kids, sometimes pronounce th-sound [θ] as f-sound [f] and that might be why they recognized the *three* as *four*.

Although the explanation must be partly the answer, the author was not satisfied with it. The aim of this paper is to search for the reasons for the error in listening from an acoustic point of view and to find more persuasive, possibly more objective, answers.

2. Analysis

2.1. sentences for analysis

The textbook has six episodes in which conversation is carried out by four main British characters[1]. The sixth episode, having the target sentences, begins with the following statements:

Helen : *On Sunday mornings we usually just sit around and read the newspapers. It's so relaxing after a busy week.* (narration)
Jane : *This is good, 'a weekend in Barcelona, including flights and hotel, one hundred and twenty pounds.'*
David : *That's cheap.*
Helen : *I went to Barcelona about three years ago.*
Jane : *Did you have a good time?*
Helen : *It was brilliant. The food was wonderful and the nightlife...*
David : *How long did you stay?*
Helen : *For a week. We stayed in a four-star hotel.*
.
.
.

Besides the sentence first underlined, the last sentence underlined, including the word *four*, uttered by Helen was adopted as the other target to compare with the word *three* uttered by the identical character. Nobody in the class misheard the *four*.

2.2. procedure

The sentences were recorded onto digital audio tape (DAT) with a microphone from the original video tape and they were analyzed with software called *SUGI Speech Analyzer* (Sugito, 2000). The items concerning prosodic features measured and identified were : speech waveform, the spectrogram of wide-band analysis, the duration of segments, and the contour of sound pressure level. Concerning sound pressure level, as it is a unit used for convenience when referring to intensity, it will be described as sound pressure hereafter. Only a part of each sentence including the target word was focused and shown in Figure 1 and Figure 2 to make the comparison between them easier and clearer.

3. Results and Discussion

The target words can be dissected into a few constituent segments, i. e. independent speech sounds.

The word *three* can be described as [θriː] and the word *four* can be described as [fɔː] using IPA. From an articulatory viewpoint, both [θ] and [f] are fricatives and furthermore as a subdivision they belong to voiceless nonstridents. The contrast of strident-nonstrident is one of the distinctive features made by Jakobson *et al*. (1965). It is known that the overall noise energy of the nonstridents is obviously less than that for the stridents, which is recognized in the waveform of [θ] in Figure 1 and that of [f] in Figure 2. Both waveforms have very little amplitude and look like a straight line. The only difference between the two sounds is the place where they are articulated; [θ] is a dental sound and [f] is a labiodental.

Besides prosodic features shown in both Figure 1 and Figure 2, the contour of fundamental frequency (F_0) was also analyzed as a substitute intonation contour. Both the contour for *three* and that for *four* became level tone and no conspicuous difference was recognized between them.

A typical waveform and spectrogram of [θ] and [f] was listed in Kent and Read (1992) and they are given in Figure 3. The sound [θ] has larger amplitude than the sound [f]. Although there was no frequency scales two formant zones can be recognized in the spectrogram of [θ] in Figure 3: a darker zone in the upper area and a more or less dark zone in lower area. Comparing the waveform and the spectrogram of [θ] in Figure 1 and those of [f] in Figure 2 with their counterparts in Figure 3, Helen's [θ] and [f] have poorer amplitude and there is a difference in the spectrogram between the two [θ]s while no apparent difference in the spectrogram was noticed between the two [f]s. The equivalents for the two separate formant zones mentioned above can be seen at the beginning of [θ] articulation in Figure 1 but they disappear in the middle of [θ] segment. A sound with no formant zone suggests that it sounds just like a noise without any particular timbre or quality.

Duration of each speech sound was added onto the display of sound pressure in Figure 1 and Figure 2. There was very little durational difference between [θ] and [f]; the former lasted for 71ms while the latter lasted for 68ms. Nor was any difference in the change of sound pressure noticed throughout the articulation of [θ] and [f]. The sound pressure analyzed with SUGI Speech Analyzer is represented in the range of -60dB to 0dB on the display, which means 0dB is recognized as the maximum value of sound pressure. The range of [θ] was from -41dB to -26dB, a change of 15dB; while that of [f] was from -40dB to -27dB, a change of 13dB.

Let us look at the following speech sounds of each target word. The waveform still shows small amplitude at [r], a voiced sound, of *three* in Figure 1. The sound [r] lasted for only 24ms and transferred to [iː], which lasted for 70ms. Very little change (only 2dB) in sound pressure is recognized throughout [riː]. The other following speech sound [ɔː] of *four*, on the other hand, also shows small amplitude in its waveform and lasted for 96ms. As the speaker Helen was British she doesn't seem to have pronounced [r] at the end of the word *four* followed by a consonant [s] of *star*. Both [r] and [iː] are voiced sounds and the duration of the compounded voiced sound lasted for 94ms, which becomes nearly as long as the duration of [ɔː]. As for sound pressure, there was a 6dB decrease throughout the sound [ɔː].

For another way to compare two vowels, [iː] and [ɔː], the frequencies of the first formant (F_1) and

the second formant (F_2) were measured in the middle of each sound and were shown in Figure 4. The vertical line in Figure 1 and Figure 2 shows the point of the measurement. The averaged values of the two sounds pronounced by American women given by Peterson and Barney (1952) were added for comparison. It has been pointed out in earlier studies that F_1 and F_2 are the main element to form any vowel and that F_1 varies mostly with tongue height and F_2 varies mostly with tongue advancement (that is with variation on the antero-posterior position of the tongue) (Kent and Read, 1992). The F_1-F_2 frequency values of the two vowels pronounced by Helen were : 300Hz-1170Hz for [iː] and 580Hz-1610Hz for [iː]. Although the tongue position of her [ɔː] was close to that of other women's [ɔ], it was surprising to see that Helen's [iː] had a very small F_2 value, which was about the same as the mean value of [u] or [ɔ]. The result suggests that Helen articulated [iː] with her tongue being high and at a posterior position, which overlaps the tongue position of [u] articulation. And in the middle of the mean position of [u] and the mean position of [ɔ] comes the mean position of cardinal vowel No.7 [o] that Americans use for the vowel of *four*, which was not shown in Figure 4 though. In other words, Helen's [iː] could have sounded like [o] to Japanese ears who had not been accustomed to some British-like accents. That seems to be one of the reasons why Japanese misheard Helen's [iː] as [oː/ɔː].

There is another point to pay attention to. It is how much difference in prominence there was between the target word and the following segments. As shown in the waveform and the pressure contour in Figure 1, more prominence was given to the word *years* than to *three* while very little difference was recognized in the waveform and the pressure contour of *four-star* in Figure 2. Although it was not an objective measure, the author played the recorded material *about three years*, with *years* and without *years*, repeatedly and listened to them again and again. The *three* with *years* had a very vague resonance in his auditory impression every time he listened. Relatively greater prominence of *years* seems to have caused an impression of sound reduction on the preceding syllable *three* regressively, which resulted in appearing pseudo-obscure vowel and pseudo-weakened friction of [r] and [θ].

4. Concluding words

It was not intonation but sound quality and the similar duration of the target words and possibly a larger prominence of the following item relative to the target that caused the mishearing this time. It was suggested from the waveforms having little amplitude that the friction for the [θ] of Helen's *three* was not strong enough to have the intrinsic quality of its own. With incomplete formant zones the [θ] did not keep enough dental resonance but had a vague quality in the last half of the segment. The [iː] of Helen's *three* had a resonance similar to [uː] or [oː] as a result of an articulation with her front tongue being down and back tongue being up, which may be the manner sometimes adopted by British people when they pronounce *three*.

The present paper may have achieved its aims to some extent with some objective data by using sound analyzing software. The paper would be useful if learners of English would understand the

reasons for the mishearing and would be able to avoid it. It must be confirmed, however, that the sound Helen made is not unnatural but natural in British English and that the British teacher identified it correctly. In addition to acquiring phonetic knowledge and information, learners of English, including the author, are expected to understand the importance of their ear training.

Resource Material of the Research

New Headway Video Elementary, Oxford: Oxford UP, 2003.

Note

1) No detailed information of the characters is given in the textbook. All of them seem to be in their twenties and they are playing roles of British people; and their pronunciation was recognized as British English with slight variations of accent by the British teacher using the textbook.

REFERENCES

[1] R. Jakobson, G. M. Fant, M. Halle. *Preliminaries to Speech Analysis*. Cambridge, Mass.: MIT, 1965.

[2] Ray D. Kent and C. Read. *The Acoustic Analysis of Speech*. San Diego, Cal.: Singular, 1992.

[3] G. E. Peterson and H. E. Barney. "Control methods used in a study of vowels," *Journal of the Acoustical Society of America* 24, 175-184, 1952.

[4] M. Sugito. *SUGI Speech Analyzer* CD-ROM, Yokohama: Fujitsu Animo, 2000.

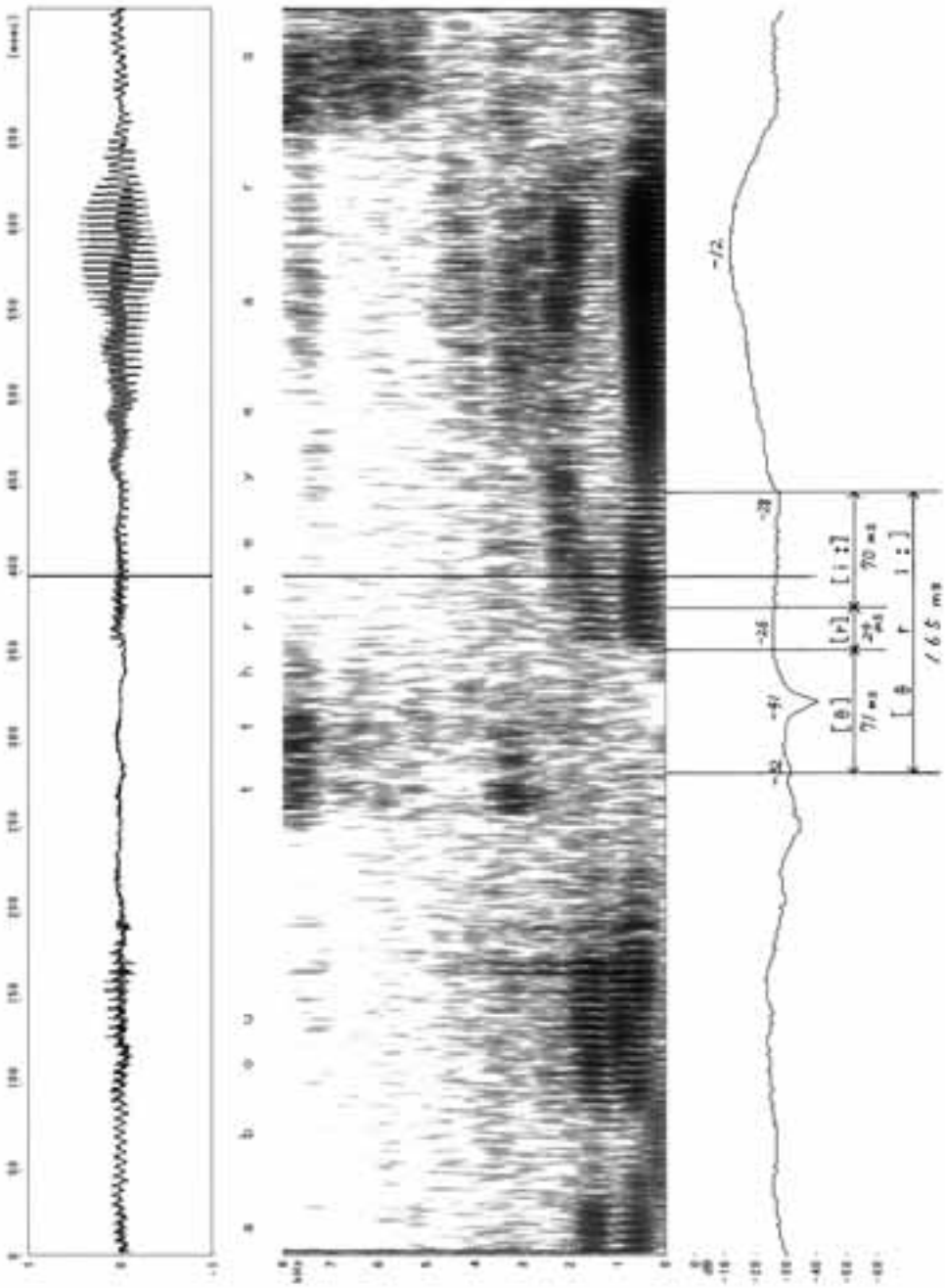

Figure 1 Waveform, Spectrogram, and Pressure Contour of "about three years"

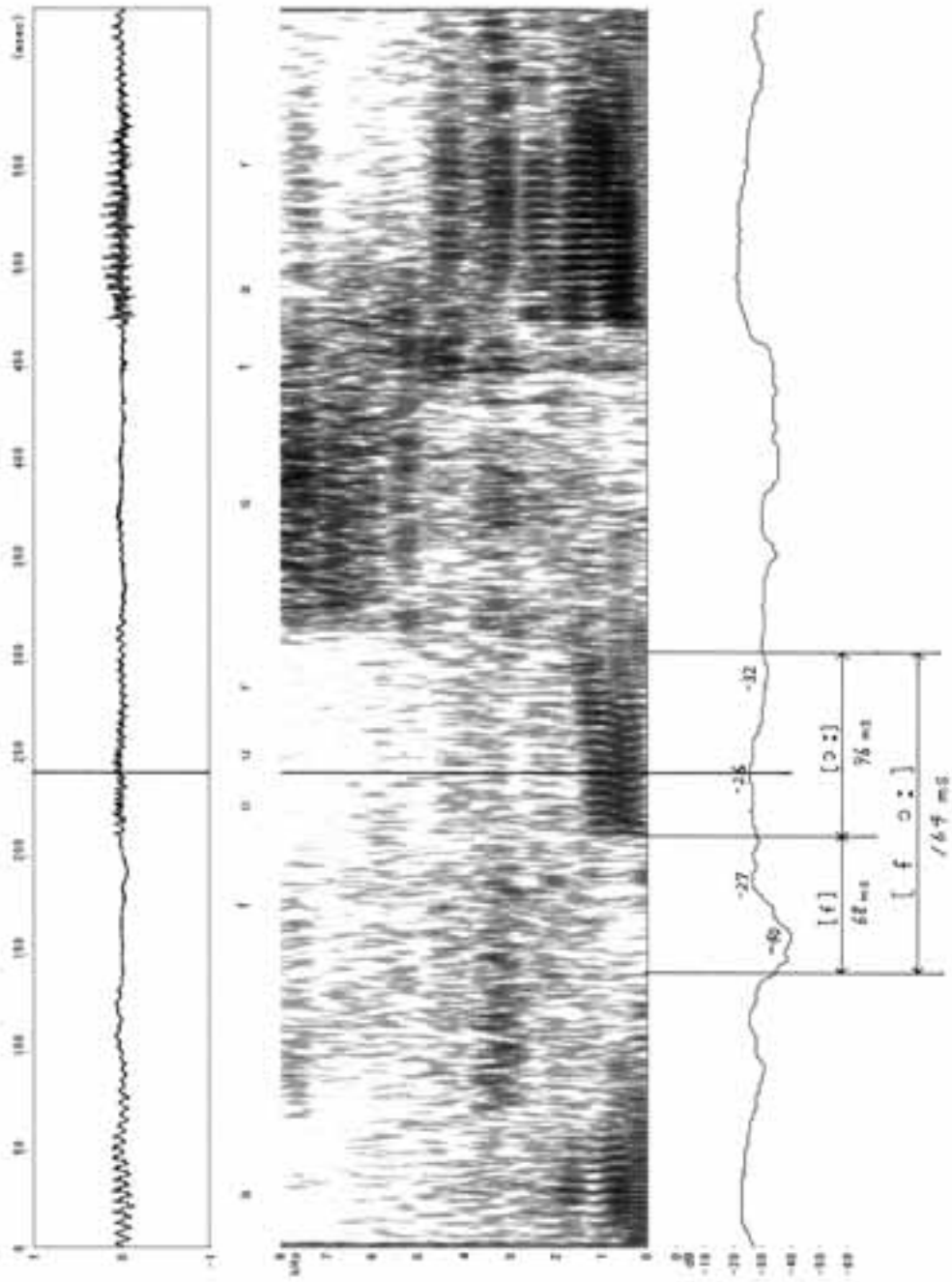

Figure 2 Waveform, Spectrogram, and Pressure Contour of “a four-star”

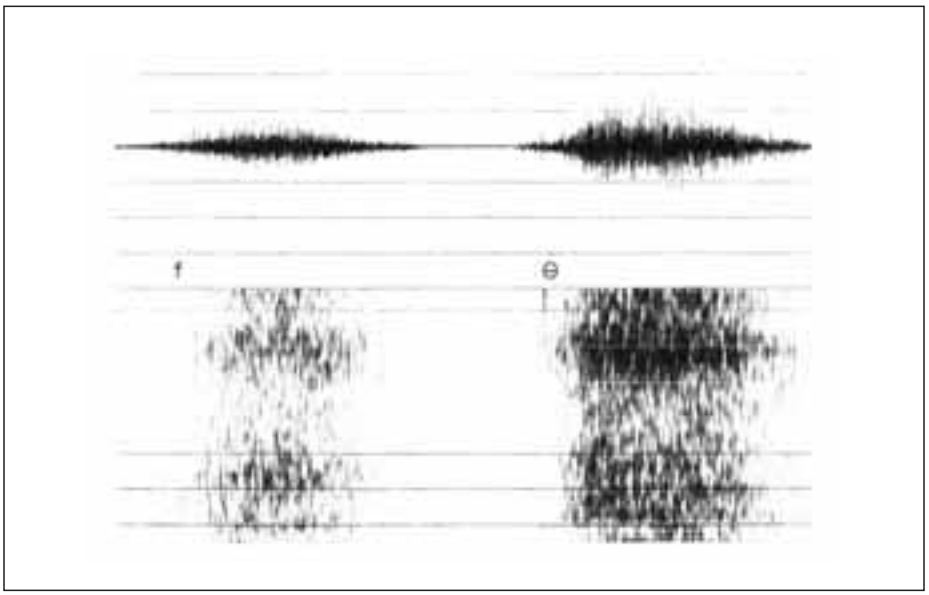

Figure 3 Waveforms and Spectrograms for Isolated Production of the Fricatives [f] and [θ] in Kent and Read (p.127) 1992

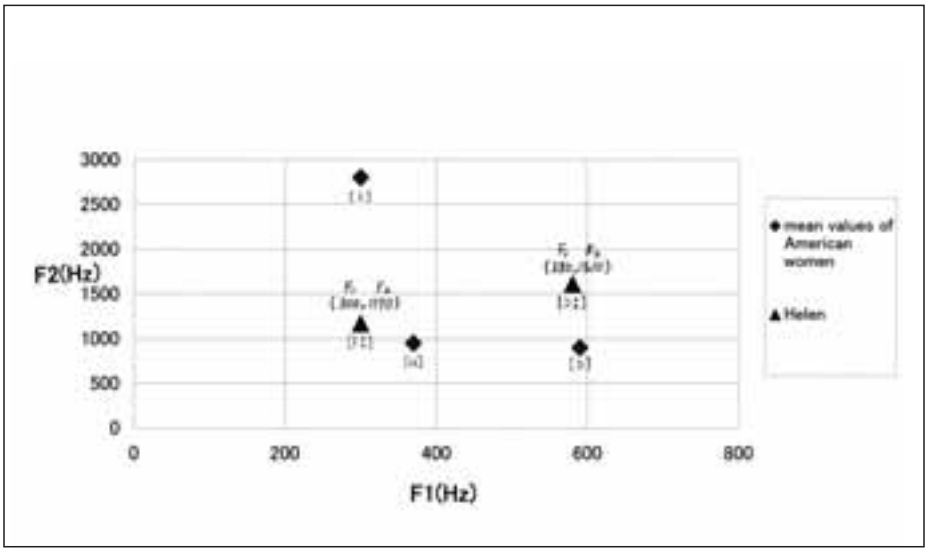

Figure 4 F1-F2 Chart