1. Introduction

One of the goals of historical linguistics is to identify and describe sound changes systematically. A further goal is to provide an account of the causes of sound change. McCawley (1977) claims that the contemporary accentual system of the Tokyo dialect of Japanese is likely to have developed from a proto-system by a shift of the original accent one syllable to the right. The present study attempts to identify a trigger of this historical accent change in acoustic terms, following Ohala's (1981, 1983, in press) hypothesis of sound change. He claims that many sound changes result from errors of transmission of pronunciation from one speaker to another and that the "seeds" of such sound changes should occur in present-day speech, and thus be available for investigation.

In view of this hypothesis, it is intriguing to find that phonetically the accentual high tone in present-day pronunciation of Japanese frequently occurs on the post-accent syllable, apparently without listeners detecting any change in accent placement. We claim that this phenomenon, which is called delayed pitch fall, is a strong candidate for a seed of the rightward accent shift. In this paper, we sketch accentual changes in Japanese, then discuss acoustic and auditory characteristics of pitch accent, and finally report the results of our acoustic analysis of the pronunciation of the present-day Tokyo dialect.

2. Historical Accent Change

One of the major sources of information regarding the accentual system of Japanese is The Ruijumyogisho, a dictionary compiled in Kyoto about 1100 A.D. and substantially revised in the early 1200's. This document reflects the dialect of Late Old Japanese (LOJ), spoken by members of the Japanese aristocracy living in and around Kyoto. LOJ, as recorded in The Ruijumyogisho, had five accentual patterns for two syllable nouns. Another major document is The Bumoki of 1687, which records the Kyoto dialect of Middle Japanese (MJ). There are three accentual patterns for two syllable nouns (Hattori 1951; Kindaichi 1942, 1951, 1974; Komatsu 1977).
The most notable fact is that these accent changes are regular, i.e., words that belong to a certain class in LOJ shift accent in the same manner in MJ. Using these and other documents, many attempts have been made to reconstruct the accentual system of pre-historic Japanese (e.g. Hattori 1951; Hayata 1973; Okuda 1975).

The regularity in accent change generally holds in synchronic variations as well. Words which belong to class A in a given dialect together belong to class B in another dialect. McCawley attempted a reconstruction based solely on synchronic variations, assuming that the differences among the present-day dialects reflect the distinctions in the proto-language.

In the Tokyo dialect, the location of fundamental frequency (F0) fall from relatively high to relatively low is the only acoustic correlate of accent. The accent pattern of words, therefore, can be represented simply by marking the location of F0 fall, if there is one (McCawley 1968; Haraguchi 1977; Poser 1984; Beckman and Pierrehumbert 1986, Pierrehumbert and Beckman 1988). McCawley (1977) used an apostrophe to indicate this F0 fall. In order to represent the Kyoto dialect, which has more patterns than the Tokyo dialect, he used an apostrophe word-initially to distinguish initial low tone from initial high tone. If there is no apostrophe in initial position, the word begins with a high tone.
McCawley first compared two-syllable words in the Tokyo and Kyoto dialects. The Tokyo dialect has three patterns, whereas the Kyoto dialect has four. Since there are no segmental characteristics to account for this difference, he attributed it to the accentual system of the parent language. In the Akita dialect, the ‘bridge-flower’ class nouns divided into two sets of reflexes. Therefore, he added another pattern to the parent accentual system. As for the phonetic value of the ‘flower’ class nouns (class 3), he inferred an LL pattern from the Kagoshima dialect, in which Falling corresponds to initial high patterns, and Level corresponds to initial low patterns in the Kyoto dialect. He then checked his reconstruction against The Ruijumyogisho and confirmed that these two correspond quite well.

<table>
<thead>
<tr>
<th>Class</th>
<th>Example</th>
<th>Ruijumyogisho</th>
<th>Reconstruction</th>
<th>Tokyo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>usi ‘cow’</td>
<td>HH</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>2</td>
<td>hasi ‘bridge’</td>
<td>HL</td>
<td>0’0</td>
<td>00’</td>
</tr>
<tr>
<td>3</td>
<td>hana ‘flower’</td>
<td>LL</td>
<td>‘0’0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>hasi ‘chopstick’</td>
<td>LH</td>
<td>‘00’</td>
<td>0’0</td>
</tr>
<tr>
<td>5</td>
<td>mado ‘window’</td>
<td>LF</td>
<td>‘00’</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Ruijumyogisho and McCawley’s reconstruction

Class 1 has no accent, and no accent change has occurred. Classes 2 and 3, as well as classes 4 and 5, have merged in the Tokyo dialect. Notice, for example, that in class 2 nouns, the place of F0 fall is after the first syllable in the reconstruction, whereas it is after the second syllable in the Tokyo dialect. From this table, he concluded that the Tokyo dialect had developed from the proto-language by shifting the original accent one syllable to the right.5

3. Acoustic and Auditory Characteristics of Pitch Accent

The claim that Japanese is a pitch-accent language is based almost exclusively on native speakers’ introspection or impressionistic data. Onishi (1942) argued that since the function of accent is to differentiate the meaning of, or to make prominent a portion of, words or phrases, any features that can serve these purposes (e.g. loudness and duration) may be distinct. In the case of Japanese, he suggested that accent was an impressionistic sum of pitch and loudness.

Neustupný (1966) found positive evidence for this claim. He pointed out that accent, as conventionally known, and the real F0 fall often do not synchronize: F0 fall is delayed in relation to an accented syllable. He
called this phenomenon *oso-sagari* (delayed pitch fall). He therefore claimed that the F0 data by itself are not sufficient for determining the accent pattern, and that since, in his data, the amplitude peak falls on the accented syllable in the words in which F0 fall delays, both F0 and amplitude are distinctive features in the Japanese accentual system.

On the basis of acoustic and perceptual experiments using synthetic speech, Sugito (1972, 1982) refuted Neustupný's hypothesis. She found that native speakers perceive an accent on a vowel when the vowel is followed by a falling F0 contour, even though the F0 peak of the accented vowel is not higher than that of the following vowel. Her discovery is schematically represented in Figure 1. When the vowel /a/ in /ima/ 'now' has a falling F0 contour, native speakers perceive an accent on /i/ as if they heard the F0 contour indicated by the dashed line.

![Figure 1: F0 contour and perceived pitch accent](image)

4. Acoustic Study

The purpose of our acoustic study is to investigate whether this delayed pitch fall phenomenon in contemporary Tokyo Japanese may be a possible seed for the historical accent change. Applying instrumental analysis, we aimed to characterize delayed pitch fall in production and perception. We hypothesized that the delay of F0 fall can be compensated by a steeper fall. When the speaker utters a word with a delayed pitch fall, the listener normally is able to factor out this delay with the compensatory cue for the delay. When, on the other hand, some listeners fail to implement this perceptual compensation and rather take the location of the F0 fall at face value, the sound change could occur. We also investigated whether or not there is some limit beyond which the accent cannot be perceived on the preceding syllable even when the F0 fall is very steep.
4.1. Production Experiment

4.1.1 Method

Twenty-four words containing a /VmV/ sequence were prepared and embedded in semantically natural sentences. Those words have an accent on either the first or the second vowel of the /VmV/. There were two reasons for choosing a nasal as the intervening consonant. First, we wanted continuous F0 to examine where F0 starts falling. Second, there are indications that when a vowel which is expected to have an F0 peak is followed by a nasal, the F0 peak often appears in the nasal (Ladd and Silverman 1984). Because of the latter factor, we thought we might be more certain of getting tokens of delayed pitch fall with a nasal consonant. For example, the word /hanámìti/ ‘a passage way to the stage’ has an accent on the first vowel of the /VmV/ sequence, /a/, whereas /hanamí ni/ ‘for flower viewing’ has an accent on the second vowel, /i/.

Seven native speakers of the Tokyo dialect were asked to pronounce the words in sentences five times each for a total of 840 sentences. Segmentation was manually performed with the aid of both spectrograms and waveforms. The F0 and amplitude contours of the /VmV/ portions were then extracted.

Acoustical analysis was conducted by examining the location and the steepness of F0 fall. The tokens with the accent on the first vowel were categorized according to the actual location of an F0 fall. In this study, we were focusing on two situations: the F0 starts falling within the first (accented) vowel and within the second (post-accent) vowel. For convenience, we call the first type non-delayed tokens, and the second type delayed tokens.

4.1.2. Results

The results indicate that 24% of all tokens have an F0 fall delayed to the second vowel of /VmV/. The following figures show F0 contours of a non-delayed token and a delayed token for the same word /námída/ ‘tear (noun)’ spoken by two different female speakers. In Figure 2.1, the F0 starts falling within the accented /a/, whereas in Figure 2.2, the F0 fall occurs on the post-accent /i/.

As Sugito pointed out, the amplitude is not a cue for determining the location of accent. Figures 3.1 and 3.2 indicate the amplitude contour for the same tokens as the previous figures. Notice that in Figure 3.2, the amplitude for the accented /a/ is lower than for the post-accent /i/. We contend that neither F0 peak nor amplitude peak signals accent location in this token.
The following table shows the median of the rate of F0 fall computed in Hz/csec for each subject (csec = 10 msec). For example, Subject 4 has 10.0 Hz/csec for delayed tokens as opposed to 6.5 Hz/csec for non-delayed tokens. The overall tendency is for delayed tokens to show a steeper F0 fall in comparison with non-delayed tokens.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Non-delayed tokens</th>
<th>Delayed tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(Male)</td>
<td>3.5</td>
<td>5.3</td>
</tr>
<tr>
<td>2(Male)</td>
<td>3.5</td>
<td>4.8</td>
</tr>
<tr>
<td>3(Female)</td>
<td>4.6</td>
<td>6.2</td>
</tr>
<tr>
<td>4(Female)</td>
<td>6.5</td>
<td>10.0</td>
</tr>
<tr>
<td>5(Female)</td>
<td>5.7</td>
<td>7.5</td>
</tr>
<tr>
<td>6(Female)</td>
<td>7.2</td>
<td>10.2</td>
</tr>
<tr>
<td>7(Male)</td>
<td>2.1</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Table 4: Median of F0 fall (Hz/csec) by subject

The above finding is clearly demonstrated by the following figure, where the solid line indicates /ámi/ with delayed pitch fall and the broken line indicates /ami/ without delayed pitch fall. Note that the F0 peaks occur in virtually the same place and that the rising F0 contours have virtually the same shape. But fall rates are different in these two words. The word /ami/ shows a much gentler slope in comparison with /ámi/.

![Figure 4: F0 contours of a delayed token /ámi/ and a non-delayed token /ami/]()  

4.2. Perceptual Experiment

4.2.1. Method

The second experiment examines the delayed pitch fall phenomenon from a perceptual point of view: viz. whether or not the longer delay of F0 fall
is compensated by the steeper fall in perception as well. And when this is so, we attempt to determine the minimum F0 fall rate required for an accent to be perceived on the preceding syllable as the F0 peak location delays into a target vowel.

We synthesized nonsense 3-syllable stimuli /mamama/, using a male speaker's pitch range. The duration of the vowel /a/ was either 100 or 130 msec, whereas the duration of /m/ was fixed to 70 msec. The F0 contour of the stimuli is a rise-fall shape with the starting F0 at 125 Hz, linearly ascending to 160 Hz, and ending at 80 Hz. These stimuli were prepared with two variables: F0 peak locations, and F0 fall rates. The peak occurred at several different locations: at approximately 20, 30, 50, 60 or 70% of the second vowel of /mamama/.

The 100-msec and 130-msec tokens were separately randomized and presented to 15 native speakers of Japanese, who were asked to determine

![Figure 5: Sample F0 contour of the stimuli /mamama/](image)

The other variable was fall rates (from 2.1 Hz/csec to 33 Hz/csec), which was computed as follows:

$$\text{Fall rate} = \frac{33}{t} \text{ Hz/csec} (\text{where } t=1,2,3\ldots 16 \text{ csec})$$

The 100-msec and 130-msec tokens were separately randomized and presented to 15 native speakers of Japanese, who were asked to determine
whether the accent pattern was like /námida/ 'tear' (accent on the first vowel), or like /okási/ 'sweets' (accent on the second vowel) for each /mamama/ token.

4.2.2. Results

The following figures indicate the subjects' judgment for the first vowel to be perceived as accented. The horizontal axis shows the pitch fall rate in Hz/csec, while different curves show the different locations of the beginning of F0 fall within the second vowel. In this perceptual experiment, as in the production experiment, we found that the later the F0 fall occurs in the second vowel of the stimuli, the steeper the F0 fall required in order for the listener to identify accent on the first syllable. For the 100-msec vowel stimuli, more than half of the subjects perceived the first syllable as accented, even when the F0 fall was as mild as 3-4 Hz/csec at the 20% or 30% peak location (i.e. the peak at 20 msec or 30 msec from the onset of the second vowel). However, when the F0 peak occurred at the 50% location, approximately 8 Hz/csec were necessary for the majority of the subjects to perceive an accent on the first syllable. At the 60% location, a much steeper fall of 16 Hz/csec was needed. Furthermore, at the 70% location, a fall as steep as 33 Hz/csec failed to compensate for the delay. In this case, the majority of subjects perceived the second syllable as accented.

The tendency for the longer delay of F0 fall to require the steeper fall was also observed for the 130-msec vowel stimuli. There is a difference in subjects' judgment across vowel durations, however. For the 100-msec vowel stimuli, the fall rate of 33 Hz/csec yielded 93% identification of the accent on the first syllable at the 60% location. In this case, the ceiling effect occurs somewhere between the 60% and 70% locations. For the 130-msec vowel stimuli, beyond the 60% location, an accent was never perceived on the preceding syllable, even when the fall was as steep as 33 Hz/csec. In other words, a ceiling effect existed somewhere between the 50% and 60% locations. We speculate that as the vowel becomes shorter, a somewhat longer delay (in terms of ratio to the vowel duration) is permitted to be compensated by a steeper fall.

We conclude that there exists a positive correlation between steepness of F0 fall and the degree of delay in delayed tokens. Moreover, the F0 fall delay was found to have some limit beyond which the accent was never perceived on the preceding syllable even when the fall was very steep.
Figure 6.1: Responses of /mámama/ in percent (Vowel = 100 msec)

Figure 6.2: Responses of /mámama/ in percent (Vowel = 130 msec)
5. Conclusions

Hypothesizing that the contemporary accentual system of the Tokyo dialect has developed from a proto-system by a shift of the original accent one syllable to the right, we looked at acoustic data to determine a possible seed of this accent change. We confirmed Sugito's claim that the F0 peak of the accented vowel is not necessarily higher than the F0 peak of the following vowel because native speakers perceive an accent on a vowel when the vowel is followed by a falling F0 contour. We also found that there is a positive correlation between degrees of delay and steepness of F0 fall in delayed pitch fall tokens. Furthermore, our data show that this delay cannot be limitless.

How do these results, then, lead to the historical accent change we mentioned at the beginning? The scenario would work in the following way: the speaker pronounced a word with delayed pitch fall. If the delay was compensated by a steep fall, the accent would be perceived on the conventionally accented syllable, even though the potential incipience of the accent shift existed. If, on the other hand, the listener interpreted the delay as a rightward shift of accent, a sound change would occur and might spread into the larger community.

Notes

We would like to thank Haruo Aoki, Ted Applebaum, Michelle Caisse, Hiroya Fujisaki, Glen Grosjean, Brian Hanson, Hector Javkin, Jean-Claude Junqua, Hideo Komatsu, Kenji Matsui, John Ohala, Hisashi Wakita, Raymond Weitzman, and Helen Wheeler for their comments and suggestions. Any errors of fact or reasoning are our responsibility. This study was supported in part by a Sloan grant to the Berkeley Cognitive Science Program.

1 Hyman (1978) also observes that when tones and syllables desynchronize, it is almost always the case that the tones last too long. In other words, tones almost always spread rightwards rather than the reverse.

2 Strictly speaking, it is not pitch but fundamental frequency which delays (cf. note 5).

3 Although small in number, there are some two-syllable words with the rise-low accentual pattern recorded in The Ruijumyogisho (Hattori 1951).

4 Even though The Ruijumyogisho and The Bumoki record five and three patterns, respectively, for two-syllable nouns, more patterns existed, which were manifest only when a noun was followed by an enclitic particle. For example, Hattori (1951) assumes that there were at least six patterns for two-syllable nouns in LOJ; viz. HH(H), HH(L), HF, HL, LH(H).
LH(L).

The term *tone* refers to a particular way in which pitch is utilized in language; the term *pitch* refers to how a hearer places a sound on a scale ranging from low to high without considering the physical properties of the sound; the term *fundamental frequency* refers to the frequency of repetition of a sound wave of which, when analyzed into its component frequencies, the fundamental is the highest common factor of the component frequencies (Ladefoged 1962).

McCawley notes that there are no neat correspondences for three syllable nouns, except those found between the Kagoshima dialect and the *Ruijumyoogisho*.

Fujisaki and Sugito (1977) found that amplitude has little influence on perception of pitch.

References


