1. Introduction

The Tokyo dialect of Japanese is a typical pitch accent language in which accent is realized solely by a change in fundamental frequency (F0), not by a change in amplitude or duration such as found in English. Phonologically, the accented syllable in Japanese has a high tone, and the post-accent syllable a low tone. Phonetically, the accentual high tone is realized by a higher F0 value on the accented syllable than on the surrounding syllables (Pierrehumbert and Beckman, 1988). However, this F0 peak frequently occurs on the post-accent syllable, without listeners apparently detecting any change in accent placement—a phenomenon which is called delayed pitch fall.

Investigating this phenomenon, Sugito (1972) discovered that the real acoustic correlate of the Japanese accent is a falling F0 contour of the post-accent syllable, rather than the F0 peak location, i.e. native speakers of Japanese perceive an accent on a syllable when it is followed by a falling F0 contour.

Hasegawa and Hata (1988) also investigated the phenomenon using naturally-uttered sentences. Figure 1 shows F0 contours of non-delayed and delayed pitch fall tokens. Both of them are instances of the word /námida/ 'tear (noun)', where an accent as conventionally known falls, and native speakers do hear an accent, on the first syllable.

We found that there is a positive correlation between the steepness of F0 fall and the degree of delay in the delayed pitch fall tokens: a delayed pitch fall tends to be steeper the later it occurs. Table 1 provides the median of rated F0 fall in Hz/csec by subject. For example, Subject 4 shows 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 fall in comparison with non-delayed ones.
Figure 1: F0 contours of non-delayed and delayed tokens for /námida/ ‘tear’

<table>
<thead>
<tr>
<th>Subject</th>
<th>Non-delayed</th>
<th>Delayed</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 1: Median of F0 fall (Hz/csec) by subject

The present study examines the delayed pitch fall phenomenon from a perceptual point of view; viz. whether or not production and perception are correlated. Our aim is to determine the minimum F0 fall rate which is required for an accent to be perceived on the preceding syllable as the F0 peak location delays into a target vowel. We also investigate whether there is a limit in the degree of delay associated with this phenomenon.
2. Experiment

2.1. Materials and Procedure

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

![Sample F0 contour of the stimuli /mamama/]

Figure 2: Sample F0 contour of the stimuli /mamama/

The other variable was the rate of the F0 fall within the second vowel. The maximum F0 fall was 33 Hz, which is equivalent to 4 semitones of the peak value, 160 Hz. The fall rate was computed as follows:
Fall rate = $\frac{33}{t}$ Hz /csec (where $t = 1, 2, 3 \cdots 16$ csec)

<table>
<thead>
<tr>
<th>Stimulus No.</th>
<th>F0 fall rate</th>
<th>Stimulus No.</th>
<th>F0 fall rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33.0 Hz/csec</td>
<td>9</td>
<td>3.7 Hz/csec</td>
</tr>
<tr>
<td>2</td>
<td>16.5 Hz/csec</td>
<td>10</td>
<td>3.3 Hz/csec</td>
</tr>
<tr>
<td>3</td>
<td>11.0 Hz/csec</td>
<td>11</td>
<td>3.0 Hz/csec</td>
</tr>
<tr>
<td>4</td>
<td>8.3 Hz/csec</td>
<td>12</td>
<td>2.8 Hz/csec</td>
</tr>
<tr>
<td>5</td>
<td>6.6 Hz/csec</td>
<td>13</td>
<td>2.5 Hz/csec</td>
</tr>
<tr>
<td>6</td>
<td>5.5 Hz/csec</td>
<td>14</td>
<td>2.4 Hz/csec</td>
</tr>
<tr>
<td>7</td>
<td>4.7 Hz/csec</td>
<td>15</td>
<td>2.2 Hz/csec</td>
</tr>
<tr>
<td>8</td>
<td>4.1 Hz/csec</td>
<td>16</td>
<td>2.1 Hz/csec</td>
</tr>
</tbody>
</table>

Table 2: Type of stimuli

Approximately 50 stimuli were made for each vowel duration. They were separately randomized and presented to native speakers of Japanese, who were asked to mark the location of the accent—the first or the second syllable /ma/.

2.2. Results and Discussion

Figure 3 shows the relationships between the percent of identification of the accent on the first syllable of /mamama/ and the fall rate (Hz/csec) in a log scale. In this perceptual experiment, as in the previous 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, rather than the second 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 F0 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 never judged the first 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. About two-fifths of the subjects perceived an accent on the first syllable at the 70% location. In other words, the majority of the subjects heard the second syllable as accented. These results demonstrate that the delay cannot be limitless. A delayed pitch fall may be compensated for by a steep fall up to a certain limit, but it cannot be compensated beyond that limit. We suggest tentatively that the limit of delay in male speech is somewhere between 60% and 70% of the duration of a target vowel at normal speed of utterance.
Another characteristic was observed across vowel durations. For the 130-msec vowel stimuli, even when the fall was as steep as 33 Hz/csec beyond the 60% location, an accent was never perceived on the preceding syllable. In other words, a ceiling effect exists somewhere between the 50% and 60% locations. 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. 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.

![Figure 4: Relationships between F0 fall rate and fall location](image)

Figure 4 shows the relationships between the fall rate and the location in the /mámama/ identification (i.e. an accent on the first syllable). The broken lines indicate the identification of stimuli with 100-msec and 130-msec vowels. The solid lines are regression lines which are fitted to each set of stimuli. As can be seen in the above semi-log figure, the F0 fall rate increases linearly the longer the delay is in both types of stimuli. This finding is mathematically represented as follows:
• stimuli with 100-msec vowels

\[
\text{fall rate of the 50\% cutoff point } = 1.483 \, e^{p/29} \\
(20 \leq p < 70)
\]

• stimuli with 130-msec vowels

\[
\text{fall rate of the 50\% cutoff point } = 1.581 \, e^{p/36} \\
(20 \leq p < 60)
\]

where \( p \) = the F0 fall location in percent of the vowel duration.

Note that although the constant before \( e \) is larger for the 130-msec vowel stimuli than for the 100-msec vowel stimuli, the constant dividing the F0 peak location \( (p) \) has a greater effect on the minimum fall rate of both types. At the same location (in ratio) of an F0 peak, the minimum fall rate required for 50\% identification of an accent on the first syllable is always greater for the 100-msec vowel stimuli than for the 130-msec vowel stimuli.

3. Conclusions

As Sugito claims, native speakers of Japanese perceive an accent on a vowel when the vowel is followed by a falling F0, and therefore, the F0 peak value of the accented vowel is not necessarily higher than that of the following vowel. This fact indicates that there are at least two orthogonal parameters involved in perception of pitch accent: the F0 peak location and the F0 fall rate.

This study has determined: (1) these two parameters are correlated in such a way that the later the F0 peak occurs, the steeper the F0 fall must be, without the listener's detecting a shift in accent placement; (2) there is a limit to the degree of delay. When a vowel duration is between 100 and 130 msec, the limit of delayed F0 peak is approximately between 60\% and 70\% of the vowel duration.

Based on this work and these results, we claim that in order to determine the place of accent in Tokyo Japanese, it is necessary to examine both the location of F0 peak measured from the onset of the vowel where it occurs and the F0 fall rate within the vowel. We provide tentative formulae which can be used for this purpose. More investigation is necessary for the generalization of these formulae. When the measured fall rate is close to the value calculated according to the formulae, special attention should be paid before determining accent placement.
Acknowledgements

We would like to thank Ted Applebaum, Hector Javkin, Jean-Claude Junqua, John Ohala, Hisashi Wakita, and Helen Wheeler for their comments and suggestions. Any errors of facts or reasoning are of course our responsibility.

References

