Regional and generational variation of VOT in Japanese word-initial stops

Mieko TAKADA

Aichi Gakuin University, Japan
takadam@dpc.agu.ac.jp

1. Introduction

In this study, Voice Onset Time (VOT) in word-initial1 plosives in Japanese is analyzed to reveal regional variations and changes across generations.2

Many languages have a contrast between voiced and voiceless obstruent consonants, and this contrast has been studied in many languages, both phonetically and phonologically (see, e.g., Kent & Read 1992, Borden et al. 2002). However, few studies have considered language internal variation in VOT.

VOT is an acoustic feature usually defined as the time interval between the onset of vocalization and the release of occlusion (Lisker & Abramson 1964). It has been widely accepted as one of the basic determinant features of the distinction between voiced and voiceless consonants (e.g., Lisker & Abramson 1964, Shimizu 1996). Lisker and Abramson (1964) report that in many languages which phonologically have a voicing contrast in stops basically it has been found that the VOT of voiceless plosives takes a positive value (the occlusion is released before vocalization starts), while that of voiced plosives takes a negative value.

In Japanese, as in many other languages, this tendency in VOT values has been supported by previous studies (e.g., Kobayashi 1981, Shimizu 1993, 1996). However, other studies have reported that the VOT values of Japanese phonologically voiced plosives vary widely and may even be positive (e.g., Homma 1980, Kobayashi 1981).

Voiced plosives without prevoicing are also observed in English and German, and are referred to as “devoiced” (Crystal 2008), or “half-voiced” (Hattori 1984, Kamei et al. 1996). In this paper, we call this sound in word-initial plosives “devoiced (DV)” and the sound with prevoicing “fully voiced (FV).”

2. Methodology

2.1. Materials

The speech data analyzed in this study consist of two sub-corpora: Corpus 1 and Corpus 2. The former, called the “high school corpus,” was constructed by Inoue and colleagues from 1986 to 1988 (see Inoue 1989). Data were tape-recorded, and were gathered from each prefecture of Japan. Data for approximately 700 speakers were collected, but only 443 speakers are analyzed in this study, due to sound quality issues. The subjects were chosen from two generations: high school students aged 15–18, who were born around 1970, and their grandparents, born between 1895 and 1940 (most of these participants were born between 1900–1929).

Corpus 2, the “indexical region corpus,” was constructed by Takada (the present author) from 2006 to 2007. The data were collected in five regions from northern to southern Japan: Tohoku, Kita-Kanto, Kanto, Kinki, and Kyushu. The speakers were chosen from a range of generations. The data were recorded into a personal computer directly by the author. In the present study, data from only three regions—Tohoku, Kinki, and Kita-Kanto—are used. These regions are shown with larger circles on the map in Figure 1.
Figure 1 also shows the locations of the speakers in Corpus 1, with dots; these data were collected from approximately 100 places, distributed widely throughout Japan. Finally, this map also shows the Japanese dialect regions.

Figure 1. Speakers’ locations and dialect regions of Japan (as suggested by Tojo 1954). Small dots indicate locations of speakers in Corpus 1, and black circles indicate those in Corpus 2.

<table>
<thead>
<tr>
<th>Word-initial Consonants</th>
<th>Words</th>
</tr>
</thead>
</table>

The speakers in the corpora were generally born and bred in their communities of residence. Corpus 2 was designed on the basis of the analysis of Corpus 1, and these two corpora are conceived as a “real-time” (Labov 1963) corpus, that is, in such a way as to allow comparison across generations alive at the same time. The speakers were asked to read aloud a list of over 100 words, but only the 30 words that began with plosives /b, d, g, p, t, k/ (shown in Table 1) were analyzed in this study.
2.1. Methods of acoustic analysis

The analogue data were converted into digital data (22050 Hz sampling rate, 16 bit quantization rate) and analyzed with acoustic software (WaveSurfer version 1.8.5). VOT was judged by measuring the length between the burst (release) and the voice onset, from the waveforms and the spectrograms; when the voice onset occurred before the burst, the VOT was considered to take a negative value, and when it occurred after the burst, the VOT took a positive value. Figure 2 illustrates examples of the waveforms and spectrograms of word-initial voiced plosives from the current data. Figure 2a is an example of a negative VOT value, while Figure 2b is an example of a positive value.

![Waveform and spectrogram examples](image)

Figure 2. Examples of the acoustic analysis

3. Results and discussion

3.1. Patterns in the Regional and Generational Distribution of VOT in Voiced Stops

Figure 3 shows the VOT distribution of voiced stops by prefecture, using column graph. Only Corpus 1 is analyzed here. Figure 3a gives results for the older generation (the grandparents), while 3b shows results for the younger generation (the grandchildren).

Figure 3a clearly illustrates a difference in the older generation in VOT between Northeast Japan (hereafter, “the Tohoku region”) and Southwest Japan (hereafter, “the non-Tohoku region”). In the Tohoku region, voiced stops predominantly carry positive VOT values, whereas in the non-Tohoku region, they usually have negative values. In other words, it can be said that there are two regional varieties of VOT in Japanese word-initial voiced stops: the Tohoku variety, which is dominated by DV and the non-Tohoku variety, which is dominated by FV. Furthermore, this clear geographical pattern can be regarded as a kind of the “East–West distribution,” which is a well-known distribution pattern in Japanese dialectology (see, e.g., Onishi 2002).

However, this regional pattern is not so clear among the younger generation, as shown in Figure 3b. While in the Tohoku region the VOT of voiced stops is consistently positive from the older to the younger generation, in the non-Tohoku region VOT begins to take larger—and positive—values among the younger generation. Thus, the regional difference in VOT values between Tohoku and non-Tohoku becomes less clear.

In the non-Tohoku region, another geographical pattern can be found: a kind of “peripheral distribution,” of which the Kinki area can be considered the most typical: there, voiced stops take smaller
values (i.e., exhibit longer prevoicing) than in surrounding areas such as Kyushu and Kanto. This geographical pattern persists in the younger generation, though VOT values become bigger overall.

a. Older generation (born before 1940)
b. Younger generation (born around 1970)

Figure 3. VOT distribution map of word-initial voiced stops
Each color in the bar charts indicates the proportional breakdown of VOT by prefecture. As shown in the key, black to grey indicate negative VOT values and white indicates positive values.
3.2. Change of Devoicing Rate in Individuals

Figure 4 shows each speaker's devoicing rate (DV rate) in three areas: Tohoku, Kita-Kanto, and Kinki. Here, Corpus 1 and Corpus 2 are coordinated with reference to the speakers' birth years, and different markers express the difference in the Corpus.

a. Tohoku

b. Kita-Kanto

c. Kinki

In these figures, three areas with characteristic patterns emerge. Voiced stops in Tohoku (Figure 4a) realize predominantly as DV, so that most speakers across all generations in Tohoku show almost 100% for DV rate (with many markers overlapping), and the lower end is not less than 25%.

In contrast, in Kinki (Figure 4c), DV rate varies according to the speaker’s year of birth. It is 0% in 1920, but reaches 100% in 1980. Interestingly, the DV rate in a given year does not take one definite value. In Figure 4c, the markers are distributed in the lower right, and the upper left quadrant is blank; the lower end does not change from 0%, but the upper limit goes up according to the speaker’s year of birth. It can be said that this indicates a diversification of DV rate within individuals: in the younger generation, the flexibility of devoicing allowed to the speaker is higher. In fact, this result for Kinki is shared across the non-Tohoku region (see, Takada 2011), although the starting points and gradients of DV rate change are slightly different among the areas.
Kita-Kanto (Figure 4b) has a profile between those of Tohoku and Kinki. DV rate markers distribute from 0% to 100% over the entire period, although more markers exist around 100% in the younger generation than in the older.

Notably, the upper limit of DV rate in Kinki (Figure 4c) appears to grow exponentially, at an accelerated pace, and the pace continues at a high pace though the higher limit reaches around 100%. This pattern differs from common patterns of change in sociological phenomena, where the pace of change is one of “initial stasis, rapid rise, and tailing off,” forming an s-shaped curve (Inoue 1997, Chambers 2002). This difference may result from the nature of the data considered in this research, which does not contain speakers born in what would be expected to be the ending phase of the change. In general, variation increases at the start of the change, but converge on a target at the end, so that the pace of the rise of the lower limit should be rapid initially and then slow down when it reaches around 100%. In this way, the DV rate average for each generation should trace an s-shaped curve. Figure 5 illustrates this idea.

3.3. Change in the VOT range of [+/-voice] stops

The results shown in section 3.1 above illustrate that there are both regional and generational variations in VOT in Japanese voiced stops. The interesting question now is how their counterparts, voiceless stops, vary across regions and generations. To answer this question, two regions were chosen from Corpora 1 and 2: the Tohoku and Kinki regions. The Kinki region, which is located in the middle of the non-Tohoku region, can be considered representative of the non-Tohoku variety.

Figure 6 shows relative frequency distributions by region and generation. When we look at the sequence of these small graphs in order of birth year for each region, it can be seen that VOT range is changing in a different manner in each region.

First, we analyze the Kinki region, where change in voiced stops has already been observed. In Figure 6a1, for speakers born before 1920 voiced and voiceless stops are clearly distinguished by VOT range. However, among the younger generation, the VOT ranges of voiced and voiceless stops overlap at around 0–20 ms on the VOT axis. The shifting of voiced stops is what causes this overlap: the VOT range for voiced stops has shifted to higher, and positive, values in the Kinki region (and the non-Tohoku region as a whole). However, this is not the case for voiceless stops, which remain almost the same.
Figure 6. VOT distribution of voiced and voiceless stops in Tohoku and Kinki varieties from the generation before 1920 to the generation after 1980. Each small graph shows a relative frequency distribution. The horizontal axis indicates VOT values, and the vertical axis indicates the number of items for each VOT value. The results for voiced stops are indicated by a solid line, and those for voiceless stops by a dashed line.
On the other hand, in the Tohoku region, where a change in voiced stops was not observed, a change in the relation between voiced and voiceless stops can be observed. In Figure 6b1, the ranges of voiced and voiceless stops already overlap to some extent for the generation born before 1920, and for the later generation, the overlap increases. However, the overlap in the Tohoku region behaves differently from that in the Kinki region; in the Tohoku region, while the VOT range for voiced stops remains the same, the VOT range for voiceless stops shifts to lower values.

These results suggest that the distinctive function of VOT—to allow the identification of voiced versus voiceless stops phonologically—may be weakening or disappearing in Japanese initial stops, if it has ever actually functioned in Japanese. It may be that some other phonetic features are replacing VOT and assuming this function—for example, pitch, which is known as another important feature in making voicing distinctions (e.g., Haggard et al. 1970, Abramson & Lisker 1985, Shimizu 1996).

4. Conclusions
This study analyzed VOT in Japanese word-initial stops and revealed regional and generational differences and distribution patterns. The geographical pattern of voiced stops was clear in the older generation, and can be interpreted as an East–West distribution with two regional varieties, the Tohoku and the non-Tohoku. With regard to generation, it was found that younger speakers in the non-Tohoku region gave voiced stops higher (more positive) VOT values. This tendency can be interpreted as a diachronic sound change in the non-Tohoku variety moving voiced stops from fully voiced to devoiced. This sound change also can be viewed from the perspective of DV rate, which became higher over time in Kinki (non-Tohoku), but not in Tohoku. Furthermore, there is a diversification of DV rate within individuals. In the younger generation, the flexibility of DV rate allowed to the speaker is higher. It was also found that in the younger generation the VOT ranges for voiced and voiceless stops overlap in both the Tohoku and non-Tohoku varieties. This phenomenon resulted from the shifting of VOT in voiced stops in the non-Tohoku variety but in voiceless stops in the Tohoku variety. These results lead us to conclude that the distinctive function of VOT for voicing may be weakening or disappearing in Japanese initial stops.

Areas for future research will include the investigation of VOT and other acoustic features in order to determine what actually differentiates voiced from voiceless phonologically in Japanese initial stops. It is also important to pursue our research on VOT in both regional varieties to draw an overall picture of this sound change in Japanese.

---

1 “Word-initial” here refers to the phonetic context at the start of an utterance (the same as the start of a word in the present data), with a pause in front of it.

2 The distribution of DV and FV in Japanese is mainly conditioned by region and generation of the speaker, as opposed to linguistic factors like place of articulation and following vowel (see Takada 2011).

3 For further details on this regional distribution, see Takada (2011).

References


