LOSS OF PREVOICING IN MODERN JAPANESE /g, d, b/

Mieko Takada, Eun Jong Kong, Kiyoko Yoneyama, Mary E. Beckman

Aichi Gakuin University, Korea Aerospace University, Daito Bunka University, Ohio State University
takadam@dpc.agu.ac.jp, ekong@kau.ac.kr, yoneyama@ic.daito.ac.jp, beckman.2@osu.edu

ABSTRACT

In several dialects of Japanese, including the Tokyo-area dialects that are the basis for the standard variety, there is an apparent change in progress that affects the pronunciation of word-initial /g, d, b/. Although young speakers today often produce these stops with short-lag voice onset time (VOT), the distribution of VOT values measured in productions by several generations of speakers in a database that was recorded ten years ago suggests a phonetically gradual shift from very heavily voiced variants (with extra-long lead VOT) to less heavily voiced variants, rather than a phonologically gradual diffusion of a discrete change to "devoiced" values. The extremely long VOT values in the productions of the oldest speakers in the database are in keeping with philological evidence suggesting that the starting point for the change was a series of prenasalized stops. The suggestion is evaluated against the cross-generational corpus using two phonetic measures that might index the degree of nasal venting of oral air pressure during the closure of a pre-voiced stop.

Keywords: sound change, Japanese, voiced stops, prenasalization

1. INTRODUCTION

Modern Japanese is generally described as having voiced stops contrasting with voiceless stops, both word-medially and word-initially (Table 1).

	/g, d, b/		/k, t, p/
hige	'beard'	*hike*	'to pull (imp.)'
ga	'moth'	*ka*	'mosquito'
kadan	'flowerbed'	*katan*	'assistance'
doku	'poison'	*toku*	'to untie'
kanbu	'leader'	*kanpu*	'refund'
ban	'order'	*pan*	'bread

Table 1: Examples of minimal pairs contrasting voiced versus voiceless stops in word-medial and word-initial position in Modern Japanese.

However, phonetic details of the contrast vary across dialects and ages, as noted by [9] and [11], *inter alia*. For example, [9] notes that in the Kinki (Osaka, Kyoto, Nara) and in the Tokyo-area (Tokyo, Chiba, Saitama, Kanagawa) dialect regions, older speakers' productions of word-initial /g, d, b/ typically show "pre-voicing" – i.e., lead voice onset time (VOT) values – as in the tokens of the word *ga* 'moth' shown in the first and second panels of Fig. 1. By contrast, older speakers from the Tohoku area (Fukushima, Yamagata, Miyagi, Iwate, Akita, Aomori), as well as young adult speakers in many dialect areas, tend to "devoice" word-initial /g, d, b/. That is, they tend to produce the stops with short lag VOT values, as in the token of the word *ga* 'moth' in the third panel of Fig. 1.

The current study explores another type of variation that has not been described in as much detail – namely, variation in the duration of the voice bar when there is pre-voicing in word-initial position. As Figs. 1-2 show, in Tokyo-area speakers, these durations can range from very short (e.g., 30 ms in the production by the 57-year-old woman shown in the second panel of Fig. 1) to very long values (e.g., 130 ms in the production by the 78-year-old woman shown in the first panel of Fig. 1). The distribution of VOT values in the sample of word-initial stops in Fig. 2 shows a great deal of variation in female speakers' productions, with lead durations as long as 200 ms for some of the oldest females, shorter lead durations for the majority of middle-aged women, and predominately short lag values in all but two of the 20-35 year-old women.

2. A SOUND CHANGE IN PROGRESS

This progression of measured VOT values from extreme lead through short lead to short lag suggests a sound change in progress in the Tokyo area that is a phonetically gradual Neogrammarian change. That is, the progression from longer to shorter lead VOT values suggests a change in progress that in the "degree of voicing" of /g, d, b/, and by the typology of [5], this change is not a "change from above" which is advancing by the "diffusion" of discretely different ("devoiced") variants across regions or other social groups. Rather it seems to be a "change from below" that is happening internally to the Tokyo area and advancing by "incrementation" – i.e., by a phonetically gradual shift along some continuous phonological variable from more heavily voiced variants through less heavily voiced variants before reaching the endpoint "devoiced" variant.

Evaluating this incrementation account requires a phonetically plausible hypothesis about how the most extreme lead VOT values are produced, so that a phonetically plausible measure of the progression can be devised. More specifically, such a hypothesis is necessary because the extremely long voice bars indicate productions that are not ordinary oral voiced stops. As noted by [8] and many others, there is a fundamental incompatibility between the oral air pressure build-up that is characteristic of the closure phase of an oral stop and aerodynamic requirements for voicing. However, sustained voicing during oral closure is compatible with the naso-pharyngeal venting of supraglottal air pressure during the nasal murmur phase of prenasalized stops, and there are several good reasons to suspect that there is at least some nasal venting during the longest voice bars observed in productions by older speakers of the Tokyo-area dialects (and by older speakers of many dialects other than the Tohoku-area dialects).

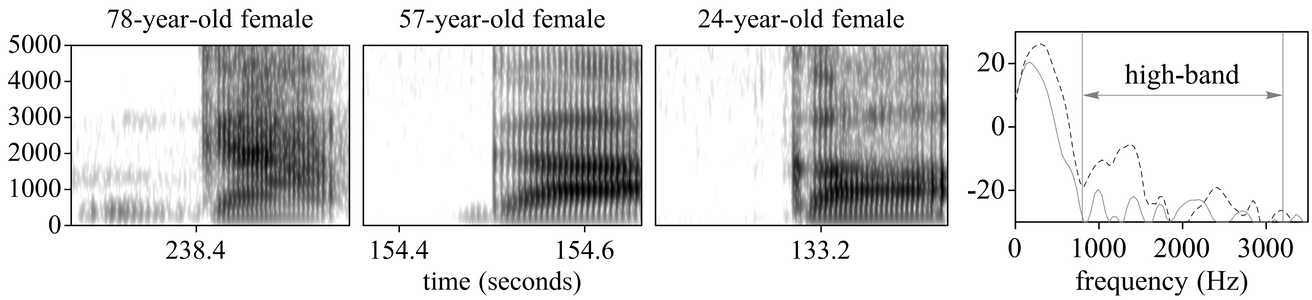


Figure 1: The first three panels show spectrograms of the stop closure and the (initial 160 ms of) the vowel in productions of the word *ga* ‘moth’ by 3 Tokyo dialect speakers from the corpus of recordings analysed in [9]. The fourth panel shows spectra estimated over a 20-ms window centred at the highest-amplitude pulse in the voice bar for the productions by the 78-year-old (dashed black line) and by the 57-year-old (solid grey line). The cursors delimit the frequency band over which energy was summed for the measure described in section 3.

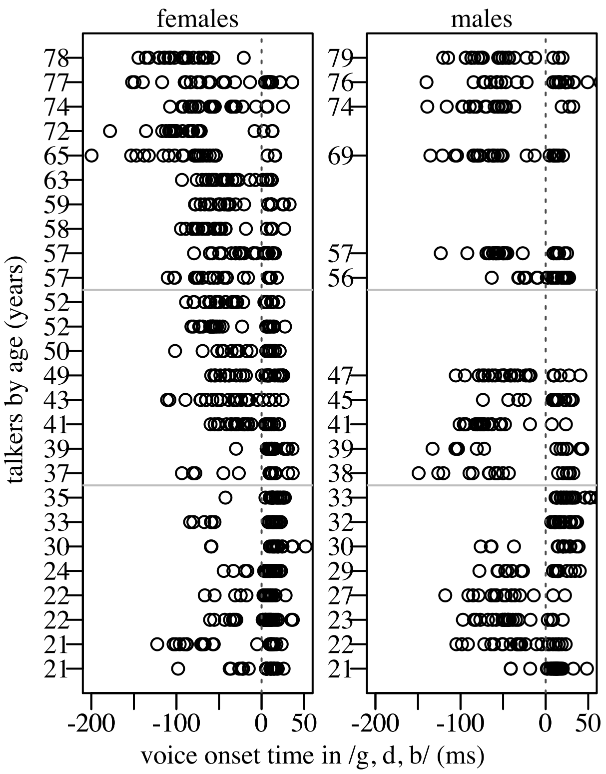


Figure 2: Voice onset time (VOT) values in tokens of word-initial /g, d, b/ produced by several generations of Tokyo speakers recorded by [9].

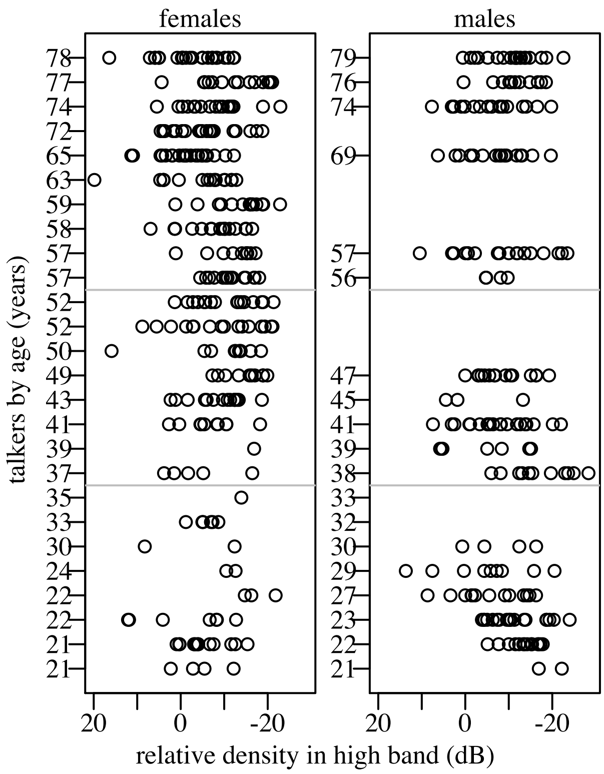


Figure 3: Energy in a high-frequency band relative to energy in the burst in productions of /g, d, b/ in Fig. 2 that have voicing lead longer than 20 ms.

First, there is philological evidence (summarized in [7, 10]) suggesting that /g, d, b/ in Old Japanese developed first in clusters of nasal with /k, t, p/, that these sounds originally were prenasalized stops, and that the nasal element was maintained at least into the 16th century (in at least those dialects that were targeted in Portuguese missionary translations and wordlists). More recently, surveys of pronunciation variation across dialects in the middle part of the 20th century (summarized in [11]) transcribe the sounds in many dialects as being nasals or prenasalized stops in word-medial position. And [2] describe an "on-glide nasal" for /g/ and /d/ even in word-initial position in productions by older speakers of some Shikoku dialects.

In support of their description, [2] show spectrograms of word-initial productions in which faint bands of energy are evident near the frequencies of the velar or dental locus of energy in the following stop burst spectrum. Spectrograms of Tokyo productions with long lead (as in the first panel of Fig. 1) similarly show faint formant bands during the closure.

Taken together, these different pieces of evidence support the hypothesis that the starting point for the current Tokyo-area sound change that is devoicing word-initial /g, d, b/ was either a series of prenasalized stops or a series of "heavily voiced" stops in which the velo-pharyngeal constriction is loose enough to vent supra-glottal air pressure so as to sustain audible vocal fold vibration over intervals as long as 200 ms.

3. ENERGY ABOVE THE FUNDAMENTAL

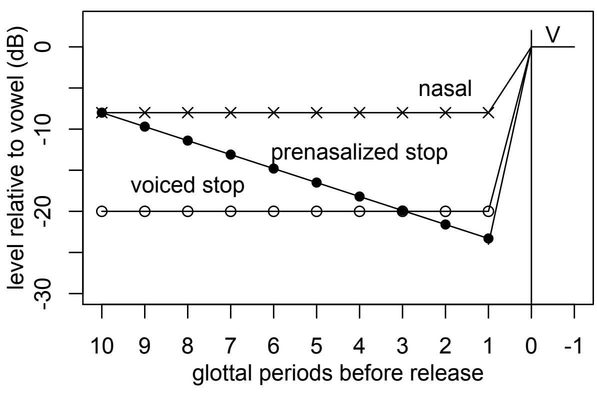
This section describes an acoustic measure based directly on the kind of spectrographic evidence shown in [2]. Specifically, to quantify the degree to which formants are present during closure, we measured the energy in a band between 800 Hz and 3200 Hz in a DFT spectrum estimated over a 20-ms Hamming window centred at the highest-amplitude pulse in the voice bar of any tokens of /g, d, b/ with a lead VOT starting more than 20 ms before the burst. We normalized the energy relative to the energy in a 6-ms window centred on the stop burst.

Fig. 3 shows this measure applied to all relevant tokens in Fig. 2 – i.e., all tokens where there was a voice bar during closure that was at least 20 ms in duration. The figure shows relatively little energy in the voice bars for the tokens produced by four of the women in the 20-35 year-old band, who also had the shortest lead VOT values (and many devoiced tokens). However, the tokens by the four older women who produced the longest voice bars do not show appreciably higher energy values than others.

While we are not ready to abandon the idea of developing a measure that can be related to the spectrographic evidence in Fig. 1, this first attempt at developing a quantitative measure of the spectral evidence for nasal venting in [2] does not capture the seemingly gradual progression of the sound change.

4. AMPLITUDE TRAJECTORY SHAPE

This section describes a second measure that we applied to the same tokens of /g, d, b/. This measure is based on a model of voicing energy during stop closures in prenasalized stops. The model is schematized in the left panel of Fig. 4. It was originally proposed by [1] for languages that have phonological contrasts between prenasalized and voiced oral stops and adapted as the basis for a measure of phonetic prenasalization in [3] and [4].



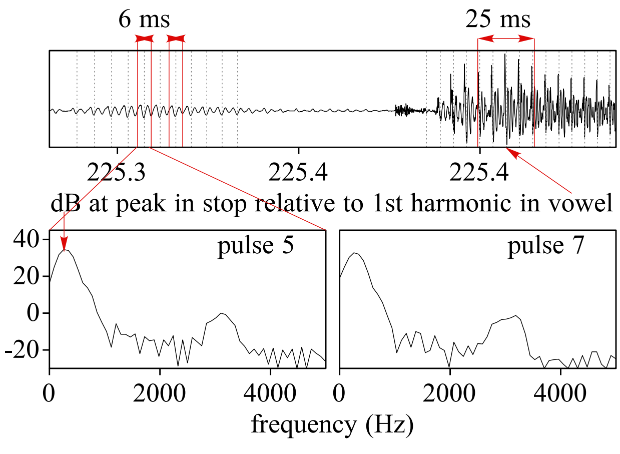


Figure 4: The left panel shows schematic drawings of amplitude trajectories predicted by the model in [1] for nasals, voiced stops, and prenasalized stops. The top panel on the right shows the windows centred at pulses 5 and 7 and in the vowel, to illustrate the calculation of pulse-by-pulse energy in the adaptation of the model by [3] and [4].

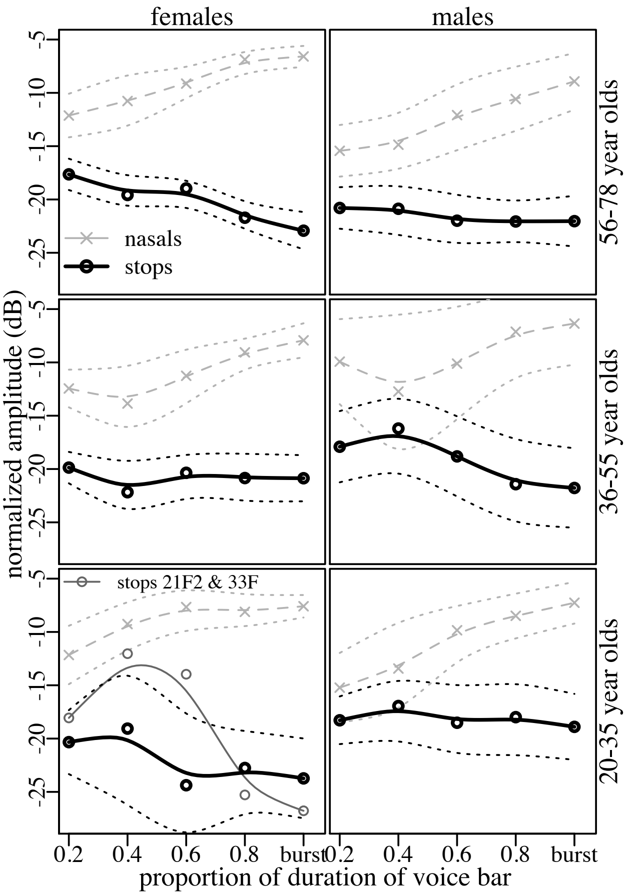


Figure 5: Average amplitude trajectories in the voice bar of the prevoiced tokens of /g, d, b/ shown in Fig. 3.

As the schematic diagram in Fig. 4 shows, energy in the nasal murmur of a nasal (points marked with x) is predicted to be higher than energy in the voice bar of an ordinary oral stop (open circles). Since a prenasalized stop begins with an interval of nasal murmur, energy should be high at the beginning of the voice bar of prenasalized stop before falling to a much lower level prior to the plosive release (closed circles). Given this model, degree of nasal venting in more or less heavily voiced stops can be evaluated by examining the amplitude trajectory pulse-by-pulse over the course of the voice bar.

The calculation of the pulse-by-pulse amplitudes is shown in the right panels of Fig. 4. First, times for the succession of glottal pulses are estimated using the autocorrelation-based periodicity detector in Praat. A spectrum is then estimated over a 6-ms Hamming window centered at the time of each pulse, and the amplitude is measured at the peak at the fundamental. Amplitudes are normalized to the amplitude of the first harmonic in the spectrum of a 25 ms analysis window starting at the third pulse after the burst. Average trajectories can then be estimated by regressing the normalized amplitudes against time in mixed-effects polynomial fits, as described in [6].

This method has been used to differentiate the prenasalized /g, d, b/ of the Thessaloniki dialect of Greek from the voiced stops of speakers from Crete as well as to differentiate heavily voiced Greek stops from the much more weakly voiced /g, d, b/ of young Tokyo-area Japanese speakers in a database of productions by 20-year-olds (which was recorded ten years after the database of recordings in [9] which provided the productions in Fig. 2).

We applied the metric to the productions shown in Fig. 2, separating the speakers generally into age- and gender-defined groups (except that we built a different model for productions by 21F2 and 33F, the two females in the youngest group who showed much longer voicing lead values than any of their age peers). For reference, we also applied the metric to productions of word-initial /m/ produced by the same groups of adult speakers who were older, middle-aged, or young at the time of the recordings of the database in [9]. The resulting model curves and confidence intervals are shown in Fig. 5.

The model curve for the /g, d, b/ tokens produced by females in the oldest group begins at a fairly high amplitude value and has a steeply falling shape that is like the trajectory for the prenasalized stops of Thessaloniki speakers in [3]. By contrast, the model curve for females in the 36- to 55-year-old group begins at a somewhat lower amplitude and is quite flat, in keeping with the suggestion that the typically shorter durations of their voice bars reflect less nasal venting to sustain voicing. The model curves for the males generally look more like the curve for the females in this middle group. The exception is the model curve for the males in the 36- to 55- year old group. However, this model also shows much more variability (larger confidence intervals), a pattern that also could be in keeping with a sound change in progress. There is a steep fall from a higher initial amplitude also in the model curve for speakers 21F2 and 33F, the two young females who produced the longest voicing lead values, whereas the model curve for stop tokens produced by other females in this youngest group begins at a lower amplitude value than the model curve for 21F2 and 33F and is more level, as if there were no nasal venting at all.

While considerably more work is in order before we can be confident of the interpretation of these patterns, the results of this second set of analyses suggest that we will be able to find acoustic measures to provide an effective quantitative evaluation of the hypothesis that the starting point for the sound change from voiced to devoiced /g, d, b/ was a series of prenasalized stops that might have been maintained into the 20th century in the Tokyo area (as well as in some other regions such as the Shikoku and the Kinki-area dialects).

5. REFERENCES

[1] Burton, M., Blumstein, S., Stevens, K. N. 1992. A phonetic analysis of prenasalized stops in Moru. *Journal of Phonetics* 20, 127–142.

[2] Kishie, S., Yoshihiro, A. 2006. On-glide nasal /d/ and /g/ in the Shikoku dialects: A focus on the Tokushima dialect. *Journal of the Phonetic Society of Japan* 10, 49–59.

[3] Kong, E.-J., Syrika, M., Edwards, J. 2012. Voiced stop prenasalization in two dialects of Greek. *Journal of the Acoustical Society of America* 132, 3439–3452.

[4] Kong, E.-J., Beckman, M. E., Edwards, J. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics* 40, 725–744.

[5] Labov, W. 2007. Transmission and diffusion. *Language* 83, 344–387.

[6] Mirman, D., Dixon, J. A., Magnuson, J. S. 2008. Statistical and computational models of the visual world paradigm: Growth curves and individual differences," Journal of Memory and Language 59, 475–494.

[7] Ramsey, S. R., Unger, J. M. 1972. Evidence of a consonant shift in 7th century Japanese. Papers in Japanese Linguistics 1, 278–295.

[8] Solé, M.-J. 2009. Acoustic and aerodynamic factors in the interaction of features: The case of nasality and voicing. In Vigário, M., Frota, S. Freitas, M. J. (eds), *Phonetics and Phonology: Interactions and Interrelations*. Amsterdam: John Benjamins, 205–234.

[9] Takada, M. 2011. *Nihongo no Gotô-heisaon no Kenkyû: VOT no Kyôjiteki Bunpu to Tsûjiteki Henka* [Research on the word-initial stops of Japanese: Synchronic distribution and diachronic change in VOT]. Tokyo: Kuroshio.

[10] Vance, T. J. 1983. On the origin of voicing alteration in Japanese consonants. *Journal of the American Oriental Society* 102, 333–341.

[11] Yamane-Tanaka, N. 2005. The implicational distribution of prenasalized stops. In van de Weijer, J., Nanjo, K., Nishihara, T. (eds), *Voicing in Japanese*. Berlin: Mouton de Gruyter, 123–156.