

# **Devoicing and its Environments in Perception: Kinki Japanese, or Tokyo?**

MIDORI Y. MORRIS  
*Gettysburg College*

## **0. Introduction**

Vowel devoicing in the Tokyo dialect is a common topic in Japanese phonology. The most general description of vowel devoicing is the one found in, e.g., Vance (1987) and Tsujimura (1996). That is, the high vowels /i/ and /u/ are devoiced between voiceless consonants and between a voiceless consonant and a pause. For example, /i/ and /u/ in /kikan/, /kukan/, and /hon desu/ are devoiced (italicized). Studies of different aspects of vowel devoicing in the Tokyo dialect have been extensively reported. Vowel devoicing in non-Tokyo dialects, like the Kinki dialect, which is spoken in the Kyoto-Osaka area, has also been studied, but not as fully as the Tokyo dialect, and studies on the perception of devoicing are even rarer. Some sociolinguistic studies have reported that people can detect language varieties based on his or her speech, and that their judgments can be affected by social information. In the case of vowel devoicing in Japanese, the results of perception experiments may not be predictable because of the allophonic status and the gap between the general belief about devoicing and actual devoicing in the Kinki dialect.

I conducted a perception experiment to examine how Tokyo and Kinki people judge a speaker as a local or a non-local person for them based on his or her devoicing variation and pitch accent patterns. The results show that both Tokyo and Kinki people tended to make judgments based on devoicing variation as well as on pitch accent, but the tendency of Kinki people to make use of devoicing variation is weaker. People also seem to use more covert knowledge of phonological factors in devoicing in production, and made judgements on that rather than audible pronunciation.

## **1. Previous studies**

### **1.1. Vowel devoicing in the Tokyo dialect**

In previous studies, different aspects of vowel devoicing in the Tokyo dialect have been reported. Those include physiological characteristics, phonology, and variability (Han 1962, Sugito 1969, Yoshioka 1981, Vance 1987, Jun and Beckman 1993, Kondo 1994, Imai 1997, and others). Devoicing is avoided when

a devoiced vowel is in an accented mora, in a mora that carries intonation, in a successive devoicing environment, and at a morpheme boundary. Devoicing shows variation. Imai (1997) collected data from natural conversation and found out environments and features that promote devoicing in production.

### **1.2. Vowel devoicing in the Kinki dialect**

Generally it is believed that vowel devoicing does not occur in the Kyoto and Osaka dialects (Horii 1972, Peng 1993). Data from previous studies show, however, that devoicing does occur there. In Tahara's (1998) database, 33 out of 40 tokens of devoiced vowels are devoiced. Nakai (1991) reported the sentence ending /u/ is devoiced by elementary school children. Sugito (1988) shows devoicing variations by both Tokyo and Kinki people.

Their data cover, however, very limited phonological environments or social variants. Another problem is that the devoiced vowel is not always compared with a vowel in the same phonological environment in another dialect, for example /kusa/ 'grass' in LH (Low-High of the pitch accent pattern) in Tokyo and /kusa/ in HL in Kinki. Kinki people may avoid devoicing here because of accentuation, just as Tokyo people do. Such comparison could lead to an inaccurate evaluation of overall devoicing rates.

Table 1 shows the devoicing rates in some studies. The rates by Tokyo speakers are quite similar, while the rate in Tahara's data is very high. Vowels in Tahara's data are all in the most general devoicing environment, that is, a high vowel between two voiceless consonants in an unaccented mora. Some re-calculated devoicing rates in such environments are shown in Table 2. Morris's data were obtained by asking Tokyo people to read a prepared passage. The devoicing rates of Tahara's data and two other data sets are quite comparable. It is necessary to collect a larger amount of more controlled Kinki production data to determine distribution and variation of devoicing before making comparison. Nevertheless, it is tempting to assume that the devoicing rate in Kinki is not all that different from that in Tokyo, at least in the most general environment, and I take this as a tentative assumption.

Table 1: Variations of devoicing

	Tokyo subjects		Osaka subjects	
	Devoicing	Nondev.	Devoicing	Nondev.
Sugito (1969)	65.8%	34.2%	29.4%	70.6%
Sugito (1988)	55.6%	44.4%	32.2%	67.8%
Yoshioka (1981)	56.5%	43.5%	N/A	N/A
Tahara et al. (1988)	N/A	N/A	82.5%	17.5%

Table 2: Devoicing variations in different positions (Tokyo subjects)

	Unaccented		Accented		Consecutive Devoicing
	Dev.	Nondev.	Dev.	Nondev.	
Yoshioka (1981)	76.8%	23.2%	16.1%	83.9%	N/A
Morris	80.6%	19.4%	N/A	N/A	92.9%

### 1.3. Perception of dialects and attitude toward them

Labov (1972) discusses the benefits of sociolinguistic investigation derived from isolating a significant linguistic variant that may serve as an index to measure social behavior. There are many studies on dialect perception within this general framework. Kerswill (1985) shows that people judge those who speak a mixture of dialects in Norwegian correctly but cannot describe the differences they based their judgments on. Preston (1996) shows that people can identify the regions from which different speech samples came in accordance with their perception about distinctiveness of speeches. Purnell et al. (1999) show that people can discriminate the ethnicity of the speakers of different varieties of English without seeing their faces. Strand (1999) and Niedzielski (1999) show that people judge the speaker's pronunciation based on social information, rather than the actual pronunciation they hear. These studies support the following ideas: (i) respondents may identify someone's ethnicity or dialect region from their speech, (ii) they can do this based on forms they are not aware of and cannot describe accurately, (iii) such judgments are affected by social information, that is, stereotypes about the speaker who has it, as well as linguistic information.

Using Japanese, Warner (1997) shows that Tokyo and Kinki people can acquire pitch accent patterns in the other's dialect and that difficulties of acquiring such differences show the same pattern as in acquiring segmental differences. Pitch accent patterns in Japanese are suprasegmental but phonemic.

It seems quite reasonable then to collect quantitative data from Tokyo and Kinki people to examine how they perceive and judge variation of vowel devoicing because devoicing is not phonemic unlike accent patterns. In addition, since the devoicing rate could be similar in the most general devoicing environments in both dialects, people might, unconsciously, have some knowledge of that.

It is important to consider attitudes toward the standard language and dialects. Both the Tokyo and Kinki dialects are considered to be prestigious by the local people. People tend to recognize prestige in the Tokyo dialect because it is spoken in the capital metropolitan area and the standard language was based on it. Shibuya (1995) reports that Kyoto people give positive descriptions about their own dialect, and that they tend to speak their dialect, instead of trying to speak standard Japanese, in any situation asked.

My assumptions in this study are that a user of devoicing and Tokyo pitch accent is more likely judged as a Tokyo person by Tokyo people and as a non-Kinki person by Kinki people, and that a user of nondevoicing and Kinki pitch accent is more likely judged as a non-Tokyo person by Tokyo people and as a Kinki person by Kinki people. If the results confirm this, it suggests that both

Tokyo and Kinki people could make judgments based on stereotypes, roughly speaking, that Tokyo people devoice vowels and Kinki people do not.

## 2. Methods

In my experiment, I presented a test tape that consists of a word list to Tokyo and Kinki people and asked them to make judgments for each word whether the speaker is from the same region as their own. I chose words with only one devoiceable vowel in the most general devoicing environment, and prepared two tokens for each word with devoicing variation as much as possible as in (1a) and (1b). I also used words that contain no devoiceable vowels and are pronounced in different pitch accents in Tokyo and Kinki as in (1c) and words that contain no devoiceable vowels and are pronounced in the same pitch accents as in (1d).

(1)	Word	Gloss	Speaker	Devoicing/Accent
a.	atafuta	'hurriedly'	J	Devoiced
	atafuta	'hurriedly'	K	Nondevoiced
b.	nadeshiko	'a pink'	K	Devoiced
	michihide	'by high and low tides'	M	Nondevoiced
c.	kawari	'replacement'	H	LHH (Tokyo)
	kawari	'replacement'	G	HHH (Kinki)
d.	tabun	'probably'	E	

Table 3 below shows how the responses were tabulated (PA refers to 'pitch accent'). This manner of tabulation allows the overall results and comparisons between devoicing variation and accent patterns to be analyzed consistently with my assumptions for this study. The question to the respondents was "Is the speaker from the same region as yours?" The expected response is that Tokyo pitch accent and devoicing are judged as "from the same region" by Tokyo respondents and as "not from the same region" by Kinki respondents, and Kinki pitch accent and nondevoicing are judged as "not from the same region" by Tokyo respondents and as "from the same region" by Kinki respondents. The opposite response for each token is unexpected. Neutral tokens are expected to sound like their own.

Then I analyzed the responses, determining the significant factors for making judgments, among both phonological environments and social factors. I used a multivariate logistic regression program, which identifies insignificant factor groups and weight of each factor, and allows further analyses.

Table 3: Way of Tabulation

Respondent	Token type	Expected	Unexpected
Tokyo	Devoiced	from the same region	not from the same region
	Tokyo PA	from the same region	not from the same region
	Nondev.	not from the same region	from the same region
	Kinki PA	not from the same region	from the same region
	Neutral	from the same region	not from the same region
Kinki	Devoiced	not from the same region	from the same region
	Tokyo PA	not from the same region	from the same region
	Nondev.	from the same region	not from the same region
	Kinki PA	from the same region	not from the same region
	Neutral	from the same region	not from the same region

### 3. Results and discussion

#### 3.1. Overall results

Tables 4 and 5 below show the overall results by token types with Tokyo data and with Kinki data showing token numbers and percentages. The results show that devoicing tokens as well as Tokyo accent tokens are more likely judged as ‘Tokyo’ or ‘non-Kinki’ while nondevoicing tokens as well as Kinki accent tokens are more likely judged as ‘Kinki’ or ‘non-Tokyo’, as expected.

Table 4: Overall results by token types (Tokyo)

Token Type	Expected	Unexpected	Total
Devoiced	1712 (70.11)	730 (29.89)	2442 (100)
Tokyo PA	377 (67.20)	184 (32.80)	561 (100)
Nondevoiced	1174 (59.90)	786 (40.10)	1960 (100)
Kinki PA	507 (86.22)	81 (13.78)	588 (100)
Neutral	155 (73.81)	55 (26.19)	210 (100)
Total	3925 (68.13)	1836 (31.87)	5761 (100)

Table 5: Overall results by token types (Kinki)

Token Type	Expected	Unexpected	Total
Devoiced	1765 (52.17)	1618 (47.83)	3383 (100)
Tokyo PA	724 (86.84)	131 (15.32)	855 (100)
Nondevoiced	1505 (54.69)	1247 (45.31)	2752 (100)
Kinki PA	777 (87.60)	110 (12.40)	887 (100)
Neutral	275 (88.42)	36 (11.58)	311 (100)
Total	5046 (61.63)	3142 (38.37)	8188 (100)

Tables 6 and 7 show more clearly that devoicing variation affects the expected judgments in similar tendencies as accent patterns do. These results can be said to have moved the field from segmental and suprasegmental but phonemic features such as pitch accent patterns into allophonic features, and show that such lower level features can be recognized, have social meaning, and be used as clues in making judgments.

Table 6: Overall results by types of variants (Tokyo)

Token Type	Expected	Unexpected	Total
Pitch Accent	884 (76.94)	265 (23.06)	1149 (100)
Voicing Variants	2886 (65.56)	1516 (34.44)	4402 (100)
Total	3770 (67.92)	1781 (32.08)	5551 (100)

Table 7: Overall results by types of variants (Kinki)

Token Type	Expected	Unexpected	Total
Pitch Accent	1501 (86.17)	241 (13.83)	1742 (100)
Voicing Variants	3270 (53.30)	2865 (46.70)	6135 (100)
Total	4771 (60.57)	3106 (39.43)	7877 (100)

The results also show some differences between Tokyo and Kinki results. Tokyo and Kinki respondents tend to make expected judgments but Kinki respondents do so more weakly, as shown with lower percentages. The weaker tendencies in the Kinki results can support the speculation that devoicing occurs in Kinki as frequently as in Tokyo at least in the most general devoicing environments, and that is why Kinki respondents did not use devoicing variation efficiently as a criterion in making judgments. It is possible that Kinki people assume that devoicing is a non-Kinki feature based on the higher nondevoicing rate in other phonological environments (for example, in an accented mora), but confirmation of that interpretation would require a different study.

Comparison of the Tokyo and Kinki results of the tokens by different pitch accent patterns also reveals differences. It is reasonable to assume that a non-local feature induces the response ‘non-local’ more easily than a local feature induces the response ‘local’. For this point, the results with Tokyo data are reasonable. The Kinki results do not show this pattern. The Kinki accent induces expected responses at as high a rate as the Tokyo accent does.

I believe the explanation lies in the different positions and values of the local language in Tokyo and Kinki. Kinki is an ancient capital region, so it is very natural for present-day people to share the Kinki dialect as a native one inherited from many previous generations. Tokyo, on the other hand, became the capital in 1868, and people started moving in after that. Consequently, people living in Tokyo may well have a smaller sense of solidarity, placing less value on local language. These results show a similar tendency to other studies and Preston's description of nonprestigious variety, which expresses local identity placing regional solidarity.

Areas with greater linguistic insecurity focus on regional solidarity ... to express local identity. Areas with considerable security do not use local speech to express such identity, for its 'uniqueness' is already taken up in the expression of status rather than solidarity matters. (Preston 1996: 317)

### 3.2. Results by the preceding consonants

The respondents seem to make judgments not simply based on devoicing variants but also based on some covert knowledge of how likely it is that vowels are to be devoiced in production.

Table 8 shows the Tokyo results by the consonant that precedes a devoiceable vowel after the first run of the statistics program. Roughly speaking, backness in terms of place of articulation and [-continuant] are promoters, that is, those factors help the respondents make the expected judgments.

Table 8: Results by the preceding consonants with features (Tokyo)

Consonant	Place of articulation	Manner of articulation	Weight
k	velar	-continuant	0.607
tʃ	prepalatal	±continuant	0.544
s	alveolar	+continuant	0.495
ɸ	labial	+continuant	0.440
ç	palatal	+continuant	0.416
ts	alveolar	±continuant	0.411
p	labial	-continuant	0.401
ʃ	prepalatal	+continuant	0.396

When separate results with devoiced tokens and nondevoiced tokens are compared as in Table 9 below, all the consonants except for [tʃ] and [ç] have the same tendencies in effects on the expected judgments. [k] is a promoter, and [s, ʃ, ɸ, p ts] are demoters in almost the same order. [tʃ] and [ç] are promoters in devoiced data, and their weights drop dramatically in nondevoiced data. In other words, when the preceding consonant is [tʃ] and [ç], the respondents chose 'Tokyo' regardless of the actual devoicing or nondevoicing of the devoiceable vowel. These two consonants are among those which Imai (1997) found as the best promoters of devoicing in production. According to her, /i/ with a preceding

fricative and a shared feature between the preceding consonant and the devoiceable vowel, that is, palatal for /i/ and labial for /u/, are the strongest promoters of devoicing. [ç] is exactly such a consonant, and [tʃ] is a pre-palatal consonant with [+continuant] feature. It seems that the respondents knew the best environments and assumed that the vowels were devoiced and chose ‘Tokyo’.

Table 9: Separate results by the preceding consonants (Tokyo)

Devoiced Data		Nondevoiced Data	
Consonant	Weight	Consonant	Weight
tʃ	0.749	k	0.654
k	0.591	tʃ	0.504
ç	0.540	ʃ	0.458
s	0.459	s	0.450
ʃ	0.399	ϕ	0.447
ϕ	0.399	p	0.419
p	0.367	ts	0.401
ts	0.311	ç	0.187

The Kinki data show simpler results, and again [tʃ] shows the similar effect, as in Table 10 below. When the preceding consonant is [tʃ], the respondents tend to judge the speaker as ‘non-Kinki’ regardless of actual voicing status. It seems that both Tokyo and Kinki people use this covert knowledge of best environments, and that they “hear” devoiced vowels in those best environments.

Table 10: Separate results by the preceding consonants (Kinki)

Devoiced Data		Nondevoiced Data	
Consonant	Weight	Consonant	Weight
p	0.656	p	0.656
k	0.555	ts	0.588
ts	0.514	k	0.546
tʃ	0.478	s	0.509
ʃ	0.449	ʃ	0.499
ç	0.430	ç	0.476
ϕ	0.393	tʃ	0.331
s	0.276	ϕ	0.297

### 3.3. Results by the following consonants

In the results by following consonants as shown in Table 11 below, there is a tendency that [+continuant] promotes the expected judgments with devoiced data and demotes them with nondevoiced data. In other words, just as with palatal [+continuant] preceding consonants, [+continuant] in the following consonants helps the respondents judge a speaker ‘Tokyo’ regardless of actual devoicing or nondevoicing.

Table 11: Separate results by the following consonants (Tokyo)

Devoiced Data		Nondevoiced Data	
Consonant	Weight	Consonant	Weight
ç	0.813	k	0.910
s	0.693	p	0.649
t	0.650	ts	0.621
ʃ	0.579	t	0.550
h	0.541	ʃ	0.468
φ	0.483	s	0.423
p	0.441	ç	0.326
tʃ	0.347	tʃ	0.302
k	0.283	h	0.220
ts	0.244	φ	0.077

The Kinki results show similar tendencies to the Tokyo results, but less clearly, as in Table 12 below. The results with devoiced data show a random ordering of weights. The results with nondevoiced data show that [+continuant] tends to be a demoter and [-continuant] a promoter, just as in the Tokyo results. That means that fricatives following the vowel generally help Tokyo and Kinki respondents judge a speaker as ‘Tokyo’ or ‘non-Kinki’ regardless of actual voicing status.

Table 12: Separate results by the following consonants (Kinki)

Devoiced Data		Nondevoiced Data	
Consonant	Weight	Consonant	Weight
ç	0.677	h	0.681
k	0.626	k	0.664
ʃ	0.553	p	0.634
s	0.543	t	0.621
t	0.516	s	0.477
tʃ	0.489	ʃ	0.476
φ	0.408	tʃ	0.397
h	0.339	ç	0.314
ts	0.331	φ	0.301
p	0.282	ts	0.264

The different effects of preceding consonants and following consonants seem to be consistent with a temporal order of perception. After noticing a preceding consonant, the respondent still has a chance to assess the voicing status of the vowel, and judges the speaker accordingly. Different consonants have different effects, but the same consonants help the respondents make expected judgments in the devoiced data and nondevoiced data. Only the best preceding consonants for vowel devoicing make them ignore the voicing status. On the other hand, on hearing following consonants, the respondents cannot retrieve their perception of

the vowel. So once they miss it they cannot help but assume the vowel is devoiced by the series of segments that form devoicing environments, and go on to reason that a devoicer is from Tokyo or non-Kinki. Kinki respondents may well have the same knowledge of the devoicing environments, if their devoicing rate is almost the same as the one by Tokyo people.

When the respondents judge the speaker as ‘Tokyo’ or ‘non-Kinki’ using knowledge of good environments for vowel devoicing in production, it is suggested that they “hear” devoiced vowels in those environments. This means the phonological system of its speakers rather than the acoustic information affects their perception and this matches previous studies, for example, Beddor, et al (2002). They conducted perceptual experiments and report that the perception of vowels is affected by language-specific patterns of coarticulation, and they “hear” coarticulation when there is no coarticulation in the coarticulatory context in the respondent’s language. My study is not directly asking the devoicing status but the speaker’s region assuming they use the devoicing status, but this interpretation does not seem unreasonable.

#### **4. Conclusion**

The results of this perception experiment using vowel devoicing and pitch accent patterns in Japanese suggest the following:

First, although devoicing is allophonic and people are not aware of it in natural conversations, its variation contributes to the respondents’ judgments that a devoicer is judged as Tokyo or non-Kinki and a nondevoicer is judged as non-Tokyo or Kinki. Both Tokyo and Kinki results show similar tendencies, but the tendencies among Kinki people are weaker. This supports the idea that they devoice vowels as frequently as Tokyo people do at least in the most general devoicing environments while Kinki people are alleged to be nondevoicers.

Also, beyond voicing variants, more covert phonological knowledge contributes to respondents’ judgements. Features that form best environments for vowel devoicing in production, namely, [ç] and [tʃ] in preceding consonants and [+continuant] in following consonants, tend to aid in distinguishing between ‘Tokyo’ or ‘non-Kinki’ regardless of actual voicing status. These results illustrate the reasonable coordination of the perception and indicate utilization of the knowledge of the phonological system to “hear” what is not in the acoustic information.

The Tokyo pitch accent is not a good clue for Tokyo people, while the Kinki accent is as good as the Tokyo one for Kinki people. This indicates another example of a nonprestigious variety that expresses local identity and regional solidarity.

## References

- Beddor, Patrice Speeter, James D. Hansberger, and Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30: 591-627.
- Han, Mieko Shimizu. 1962. Unvoicing of vowels in Japanese. *Onsei no Kenkyu [Study of Sounds]* 10: 81-100.
- Imai, Terumi. 1997. Vowel devoicing in Japanese. MA thesis, Michigan State University.
- Jun, S.-A., and M. Beckman. 1993. A gestural-overlap analysis of vowel devoicing in Japanese and Korean. Paper presented at the 1993 Annual Meeting of the Linguistic Society of America, Los Angeles, January 7-10.
- Kerswill, Paul E. 1985. Native dialect and dialect mixing in Bergen: A perception experiment. *Cambridge Papers in Phonetics and Experimental Linguistics* 4.
- Kondo, Mariko. 1994. Mechanisms of vowel devoicing in Japanese. Proceedings of the 1994 International Conference of Spoken Language Processing 1: 61-66. Yokohama, Japan.
- Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Nakai, Yukihiro. 1991. Terebi shutsuen-ji no akusento: Kinki chihô no shôgakusei ga Kinki rookaru no bangumini shutsuen suru baai [Accents in appearing on a TV program: The case of elementary school children in Kinki appearing in a program broadcast locally in Kinki]. Script for the presentation at the 52nd Conference of Nihon Hôgen Kenkyûkai [Japan Dialect Studies Association].
- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, vol. 18, no. 1, 62-85.
- Preston, Dennis R. 1996. Where the worst English is spoken. In E. W. Schneider (ed.) *Focus on the USA*, 297-360. Amsterdam: John Benjamins.
- Purnell, Thomas, William Idsardi, and John Baugh. 1999. Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18(1): 10-30.
- Shibuya, Katsumi. 1995. Shinjô to wakimae ishiki no shôtotsu suru tokoro [A place where emotion and discretion clash]. *Gengo* 24(12): 110-121 [Special November Issue]. Tokyo: Taishukan.
- Strand, Elizabeth A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1): 86-100.
- Sugito, Miyoko. 1969. Tokyo, Osaka ni okeru musei-boin ni tsuite [A study of voiceless vowels in Tokyo and Osaka]. *Onsei no Kenkyu [Study of Sounds]* 14: 249-264.
- Sugito, Miyoko. 1988. Nihon no 8 toshi ni okeru boin no museika [Vowel devoicing in 8 cities in Japan]. *Osaka Shoin Women's College Collected Essays* 25: 1-10.
- Tahara, Hiroshi, et al. 1998. *Jinmonkon Database vol. 2 (CD-ROM)*. Kanagawa, Japan: Sôgô Kenkyû Daigakuin Daigaku.

*Devoicing and its Environments in Perception*

- Tsujimura, Natsuko. 1996. *An Introduction to Japanese Linguistics*. Cambridge, MA: Blackwell.
- Vance, Timothy. 1987. *An Introduction to Japanese Phonology*. Albany: SUNY Press.
- Warner, Natasha. 1997. Recognition of accent patterns across dialects in Japanese. *Proceedings of the 23rd Annual Meeting of the Berkeley Linguistics Society*, 364-375.
- Yoshioka, Hirohide. 1981. Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38: 236-251.

Midori Y. Morris  
Department of Asian Studies  
Campus Box 441  
Gettysburg College  
Gettysburg, PA 17325

myonezaw@gettysburg.edu