f0 enhancement in Japanese voicing contrast by Japanese native speakers and its implication to L2 perception/production

Keiji Iwamoto

Indiana University Bloomington
kiwamoto@iu.edu

ABSTRACT

This study investigated how f0 was enhanced when Japanese native speakers try to disambiguate minimal pairs of voicing contrast (i.e., the directions of f0 enhancement: up or down). A phonological focus task was utilized (e.g., it's not ***t****enki*, 'weather', but ***d****enki*, 'electricity'). Speakers should maximize phonological contrast when disambiguating the minimal pair. The results showed that f0 was raised for voiceless items and depressed for voiced items when the items were in focus (i.e., when they tried to maximize the phonological contrast) in comparison to 'unfocus' items. The results indicate that speakers have control over the manipulation of f0 (support for the concept of controlled phonetics [8]). These findings also contribute to our understanding of L2 perception and production, especially of learners whose L1 are 'aspirating languages' (e.g., Mandarin Chinese).

Keywords: voicing contrast, phonological focus, Japanese

1. INTRODUCTION

The motivation for the current study is to investigate the acquisition of the L2 phonological system. However, this study aims to examine what the target language speakers do in the first place. The linguistic target is Japanese voicing contrast and the target learner's L1 is Mandarin Chinese. First, the relationship between Japanese voicing contrast and f0 will be reviewed and the results of the L1 production study (the current experiment) will be discussed. Second, based on the results and the past literature, some implications for L2 perception of Japanese voicing contrast will be discussed.

Japanese voiced stops have pre-voicing (negative VOT), whereas voiceless stops have short-lag aspiration (positive VOT). It might be this difference in VOTs that Japanese speakers rely on to distinguish the voicing contrast in Japanese. However, studies of modern Japanese have indicated that the phonetic cue of pre-voicing is playing a less important role in voicing contrast due to an overlap of the VOTs between voiced and voiceless sounds. That is, there has been no prominent difference between voiced and voiceless stops in terms of VOTs [11]. If it is the case that pre-voicing does not help much in terms of Japanese people's perception of voicing contrast, there must be other acoustic cues that Japanese people recruit to distinguish voiced and voiceless sounds. Previous studies have suggested that in addition to VOT, other acoustic cues, such as consonantally induced f0 (CF0) might be what has been employed in voicing contrast by Japanese native speakers [2, 4], especially in the word-initial contexts.

This study investigated the importance of the f0 cue for Japanese voicing contrast by having speakers disambiguate minimal pairs of /t/ and /d/, where they must maximize the voicing contrast. The phonological focus task was employed [3]. The phonological focus task has subjects disambiguate minimal pairs differing only in one phoneme (e.g., *ma*_**d**_ vs. *ma*_**t**_). The task has been used to diagnose which acoustic cues or phonetic features are employed for the phonemic contrast in a language [7]. This method allows us to examine which acoustic cues would be perceived as more important by speakers.

There are two conditions: focus and unfocus. In the focus condition, they need to disambiguate minimal pairs, whereas, in the unfocus condition, they do not need to differentiate minimal pairs. The followings are examples.

(1) Focus
It's not **t**enki but **d**enki.
It's not **d**enki but **t**enki.

(2) Unfocus
She didn't say **t**enki but ***he*** did.
She didn't say **d**enki but ***he*** did.

Since devoicing of voiced stops has been observed in modern Japanese, the pre-voicing cue is less robust. In addition, positive VOT cue also seems to be overlapping between voiced and voiceless stops. Therefore, f0 might play a key role in the maximization of the voicing contrast.

The previous studies on CF0 revealed that phonologically voiceless stops have higher CF0 than voiced counterparts [6]. This study assumes the controlled phonetics that speakers have the capability to control f0 manipulation [8]. That is, f0 is employed intentionally by speakers to maximize the voicing contrast. The question is whether there would be a difference in f0 between the focus and unfocus conditions.

If native speakers would like to manipulate f0 to maximize the contrast, the direction of the f0 enhancement would be *raising* (up) after voiceless stops, whereas it would be *depressing* (down) after voiced stops. The degree of f0 enhancement would be larger in the focus condition than in the unfocus condition.

2. METHODS

2.1. Speakers

Three monolingual native Japanese speakers (3 males) with a Tokyo dialect were recruited remotely. Because the VOT shift is happening among younger generations [11], participants were between the ages of 22 and 32. They had not been exposed to extensive second language training and linguistics. This was to avoid L2 influence on the L1.

2.2. Items

Two minimal pairs of /t/ and /d/ were used: *tenki*, ‘weather’, & *denki*, ‘electricity’, and *temae*, ‘front’, & *demae*, ‘catering’. Both /t/ and /d/ occur in the word-initial context where f0 is most likely to be used [4]. The former /t/ and /d/ occur in the syllable with a high tone, whereas the latter pair occurs in the syllable with a low tone. As well as the two minimal pairs of /t/ and /d/, minimal pairs of nasals (/m/ and /n/ i.e., sonorants) for each tonal context were employed to establish a baseline to which we can compare f0 values of post-plosive vowels. Sonorants, unlike obstruents, do not create f0 perturbation effects in the following vowel [4]. The minimal pairs of /m/ and /n/ were these two pairs: *maasu*, ‘Mars’ *& naasu*, ‘nurse’ (word-initial & high tone), and *makeru*, ‘to lose’ *& nakeru*, ‘to make/let you cry’ (word-initial & low tone). The following vowels were not matched between the voicing items and nasal pairs because of the author’s mistake. However, since the purpose of having nasal pairs is to set the baseline of the *directions* of f0 enhancement, the following vowel should not play a significant role.

2.3. Task

As discussed earlier, the phonological focus paradigm was employed in the current study [3]. The lexical items were placed in carrier sentences. In addition to making the onsets and closures of the target plosives more distinguishable, reciting words in a carrier sentence makes the overall acoustics of the sentence (and thus the items) more stable. Two types of sentences are used to elicit different focusing effects. First is the lexical focus condition in which the speaker corrects a word from the list of minimal pairs (i.e., the focus is on the target word; hereafter referred to as ‘focus’). Second is the postnuclear focus condition in which the speaker also employs correction, but in a frame preceding the target word (hereafter referred to as ‘unfocus’). This condition will act as a baseline for the lexical focus condition. The sentence structures are simply Japanese translations of de Jong’s [3]; in both English and Japanese, the position of the target words and emphases of these sentences are approximately the same.

(3) The sentence frame for 'focus’ conditions
これは X じゃなくて、Y です。
‘This is not X, it is Y.’
これは Y じゃなくて、Xです。
‘This is not Y, it is X.’

(4) The sentence frame for ‘unfocus’ conditions
彼女じゃなくて彼が X と言った。
‘She didn’t say X, but he did.’

Data collection was done remotely, without supervision. Participants were given instructions prior to the start of the experiment, such as recording in a quiet environment with minimal electronic noise. The recording was done by the participants themselves on their phones or laptop, and the audio files were sent directly to the researcher upon completion.

2.4. Measurement

Praat was used to analyze the speech data. After annotating, f0 on post-stop vowels was extracted through Praat scripts [12]. Semitone was used as a unit for f0. f0 differences in hertz are not a good measurement because f0 needs to be perceived on a relative scale. For example, absolute value means nothing when male and female speech are compared (male voice is inherently lower than female voice). It is said that a logarithmic scale of f0 can approximate human perception and the semitone is used as a unit for the logarithmic scale of f0 [1]. One semitone difference can be consciously perceived by people.

3. RESULTS

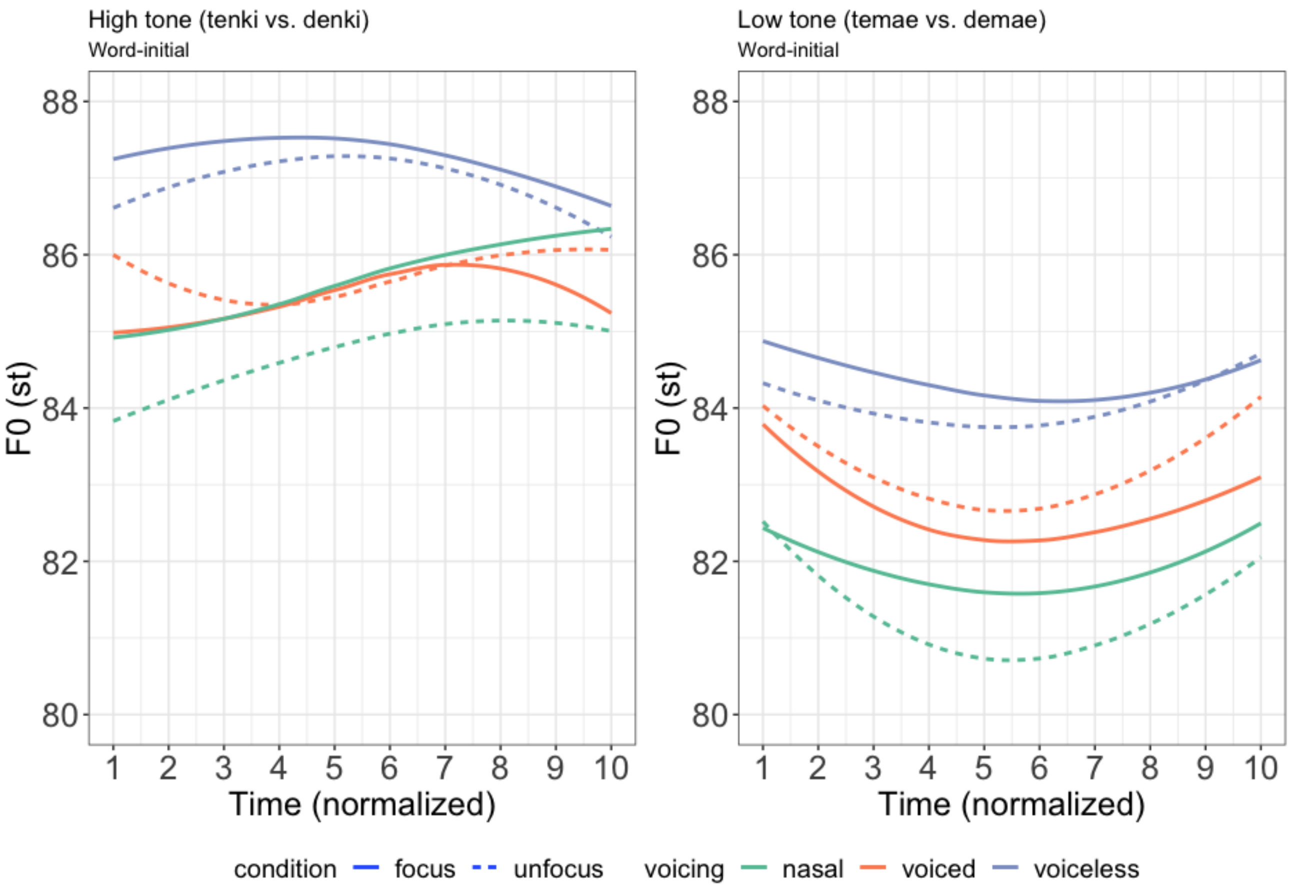

Figure 1: The f0 contours on post-stop vowels in semitone (st)

Figure 1 shows the f0 contours on post-stop vowels. Y-axis shows f0 in semitone and X-axis shows normalized time. There are two conditions: focus (Solid line) and unfocus (Dashed line). What we are interested in is the directions of f0 enhancement (up or down).

The *directions* of f0 enhancement would be analyzed; f0 depression for voiced stops and f0 raising for voiceless stops. Speakers are expected to enhance f0 in this manner so that they can maximize the contrast. The direction of f0 enhancement for nasals between focus and unfocus conditions will also be analyzed to check the default f0 enhancement effect due to the focus effect. The f0 direction of the nasals will be the baseline of the analysis.

First, let’s look at the nasals (baseline). If we compare the focus and unfocus conditions, both plots show that when items are in focus, f0 tends to be raised. In a high-tone context, the difference between the conditions is clear, whereas, in a low-tone context, there is no difference at the f0 onset but there is a clear difference at the middle time points. In sum, when items are in focus, f0 is raised (even when there is no CF0 effect).

The left plot shows a high-tone context (*tenki* vs. *denki*). As expected, there are visually significant differences between voiced and voiceless stops (i.e., CF0 effect), especially at the f0 onset. There is also a focus effect. The direction of the f0 enhancement is ‘up’ for voiceless items, as predicted. Since the direction of the enhancement for nasals is also ‘up’, we are still not sure if the enhancement is the enhancement for voicing contrast or for the overall saliency of the syllable. However, the direction is ‘down’ for voiced items, which is different from the baseline. Therefore, this f0 depression due to the voiced items could be interpreted as an intentional manipulation by speakers. They might have depressed f0 to maximize the voicing contrast on purpose.

The right plot shows a low-tone context (*temae* vs. *demae*). The same tendency can be observed. The direction of the f0 enhancement is ‘up’ for voiceless items and ‘down’ for voiced items.

4. DISCUSSION

4.1. Support for Controlled Phonetics

This study investigated the directions of f0 enhancement. The target context was the word-initial context. In both high-tone and low-tone contexts, the directions were the same: f0 rising for voiceless stops and f0 depression for voiced stops.

The results provide support for controlled phonetics [8]. It seemed that speakers deliberately depressed f0 in the post-voiced-stop vowels (and possibly raise f0 in the post-voiceless-stop vowels) to maximize the voicing contrast.

4.2. Implications to L2 Perception/Production

Based on the past literature and the results of the current study, some implications can be made for second language acquisition.

4.2.1 Implications to L2 Perception

The results suggest that f0 seems to play a role in Japanese voicing contrast when speakers try to disambiguate the contrast. In addition, the results of the past literature indicate that positive VOT cues and pre-voicing cues are less robust for the voicing contrast.

We are not sure if the f0 cue is a primary cue or a secondary cue (i.e., a redundant cue) for the perception of Japanese native speakers, but it seems that f0 is used when VOT and pre-voicing cues are ambiguous [5]. However, the redundant cue could be a primary cue for L2 learners, especially for L2 learners whose L1 does not employ pre-voicing as a phonologically distinctive cue (e.g., Mandarin Chinese). This is because Mandarin Chinese, for example, employs aspiration for the voicing contrast. L1 Mandarin Chinese learners of L2 Japanese have difficulty in the Japanese voicing contrast [9]. The previous research on the L2 perception of Japanese voicing contrast did not manipulate f0 but given the results of the current study and the past literature [5], further L2 research with the variable of f0 is warranted. Having Mandarin Chinese learners of L2 Japanese as participants would be interesting because it employs f0 for lexical contrast (i.e., tone) and thus they are sensitive to f0 even in non-native languages [10].

The hypothesis is that although VOT and pre-voicing cues may not be readily available to L1 Mandarin Chinese learners of L2 Japanese, they may be able to distinguish the Japanese voicing contrast when f0 is perceptually different enough, because of their sensitivity to f0 [10]. If this is the case, Mandarin learners should be able to distinguish Japanese voicing contrast in the word-initial contexts where f0 seems to be employed for the contrast [4], whereas they should have difficulty in the intervocalic (word-medial) contexts where f0 is not employed for the contrast [4].

4.2.2 Implications to L2 Production

In a similar vein, the implications for L2 production can be discussed too. If Mandarin learners of L2 Japanese perform the same task (the phonological focus task), and if they are sensitive to f0 in L2 perception (in other words, the input to learners would lead to learning of L2 perception), they should be able to produce the Japanese voicing contrast distinctively in the word-initial contexts (f0 is available), whereas they may not be able to in the intervocalic (word-medial) contexts (f0 is not available). Other learners whose L1 employs aspiration for L1 voicing contrast but is not a tonal language (e.g., English) can be recruited for the comparison.

5. CONCLUSION

In conclusion, this study aimed to examine what Japanese native speakers do in their production of the voicing contrast to draw implications for L2 production/perception, especially for Mandarin Chinese learners of L2 Japanese. More specifically, this study investigated whether f0 is enhanced when Japanese native speakers attempt to disambiguate minimal pairs of /t/ and /d/. The results showed the deliberate manipulation of f0 to maximize the contrast, suggesting that f0 plays a role in the voicing contrast, at least in the word-initial contexts. Based on the results, hypotheses, and predictions were generated regarding L2 perception and production

These hypotheses on L2 perception and production should be applicable to other typologically similar language pairs. Specifically, when tonal and ‘aspirating’ language speakers learn ‘true voicing’ languages (e.g., Japanese, Russian, Spanish, Swedish, etc.), their sensitivity to f0 differences learned in their L1 may play a role in their L2 perception and production. A seemingly redundant cue (i.e., f0) may become a precious cue when viewed from a different perspective (i.e., L1).

6. ACKNOWLEDGEMENTS

I thank Lily Kawaoto for help as a team member of the Japanese L1 production study, and Professor Kenneth de Jong for his guidance. The shortcomings in this paper are certainly my own.

6. REFERENCES

[1] Baart, J. 2010. *A field manual of acoustic phonetics*. SIL International.

[2] Byun, H. 2021. Acoustic characteristics for Japanese stops in word-initial position: VOT and post-stop f0. *Journal of Phonetic Society of Japan*, 25, 41-63.

[3] de Jong, K., J. 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32. 493-516.

[4] Gao, J., & Arai, T. 2019. Plosive (de-)voicing *f*0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics*, 77, 1-33.

[5] Gao, J., Yun, J., Arai, T. 2019. VOT-F0 coarticulation in Japanese: Production-biased or misparsing?. *Proc. 19th ICPhS* Melbourne, 210–214.

[6] Guo, Y., Kwon, H. 2022. Production and Perception of Mandarin Laryngeal Contrast: The Role of Post-plosive F0. *Frontiers in Communication*, 7, 1-15.

[7] Kim, S., Kim, J., Cho, T. 2018. Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics*, 71, 65-80.

[8] Kingston, J., Diehl, R. L. 1994. Phonetic knowledge. *Linguistic Society of America* 70(3), 419-454.

[9] Liu, J., Zeng, T., Lu, X. 2019. Challenges in multi-language pronunciation teaching: A cross-linguistic study of Chinese students' perception of voiced and voiceless stops. *Círculo de Lingüística Aplicada a la Comunicación*, 79. 99-118.

[10] Schaefer, V., Darcy, I. 2014. Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology*, 5(4), 489–522.

[11] Takada, M., Kong, E.J., Yoneyama, K., Beckman, M.E. 2015. Loss of prevoicing in modern Japanese /g, d, b/. *Proc. 18th ICPhS* Glasgow.

[12] Xu, Y. 2013. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France. 7-10.