Final vs. no accent in interrogative melodies of Tokyo Japanese: 
Implications for tonal clash resolution

Mayuki Matsui¹, Hyun Kyung Hwang²

¹ National Institute for Japanese Language and Linguistics, Japan Society for the Promotion of Science
² RIKEN Brain Science Institute

Abstract (300-400 words)

Prosody occurs in the face of physical-time pressure. For example, multiple tonal events can clash within a single tone-bearing unit due to a drop of segment(s) and so on. Previous studies have found that in such a tonal clash context, speakers resolve tonal clash by modifying pitch contours in language-specific ways, such as compression and truncation. However, it remains unclear how lexical tones interact with post-lexical tones in tonal clash contexts, since previous studies have mainly focused on the languages without lexical tones.

To bridge this descriptive gap, the present study examines the tonal clash resolution in Tokyo Japanese, a pitch accent language having both lexical and a post-lexical tone in the prosodic system. We investigated how a lexical pitch accent interacts with a post-lexical tone. The study offers not only new empirical evidence of tonal clash resolution but also an interesting testing case of covert contrast at an accent-intonation interface.

In Tokyo Japanese, it has been widely acknowledged that the contrast between final-accented and non-accented words are either neutralised or reduced in utterance-final position, unless a particle follows. We examine (i) how speakers modify pitch contours when a falling lexical accent clashes with post-lexical rising tones in interrogative melodies and (ii) whether final-accented words are acoustically distinguishable from non-accented words in this context.

Four native speakers of Tokyo Japanese produced bimoraic bisyllabic words in interrogative melodies. The results demonstrate that final-accented words were realised as a rising contour without a pitch fall. As a result, final-accented words were largely
indistinguishable from non-accented words, suggesting a phonological deletion of the lexical pitch accent in the tonal clash context. At the same time, however, one speaker consistently distinguished final-accented words from non-accented words. Taken together, the results suggest that tonal clash can be resolved not only in language-specific ways but also in speaker-specific ways, interacting with (in)complete neutralisation of lexical contrast.

1. Introduction

1.1 Motivation of the study

Prosody conveys various lexical and post-lexical information including word meaning and sentence types. In actual implementation, however, there are cases where not enough physical time is given for both the lexical and post-lexical information to be realised. For instance, if prosodic units are too dense to manifest a lexical tone and a boundary tone, some sort of pitch modification is expected to resolve the clash.

Previous studies have reported that languages resolve tonal clash\(^1\) in language-specific ways. Two major ways to resolve tonal clash are compression (e.g. Grønnum 1991 for Danish, Casper and van Heuven 1993 for Dutch; Prieto 2005 for Catalan. See also Ladd 1996: 132–136, 2008: 180–184 for overview), by which the alignment between tones and tone-bearing units are reorganised to implement all the tones, and truncation, or curtailment, (Grice 1995 for Palermo Italian; Odé 2005 for Russian. See also Ladd 1996: 132–136, 2008: 180–184 for overview), by which the alignment between tone and tone-bearing units is kept intact, but instead, some edge tones are not realised due to lack of time. Such contour modification strategies may result in an apparent or complete loss of the contrast (Odé 2005 for Russian utterance-final intonation).

It is worth noting, however, that most previous studies have focused on the contour modification strategies in so-called ‘intonation’ languages, such as the languages mentioned

\(^1\) Some researchers may call it tonal crowding (e.g. Gussenhoven 2004).
above, where pitch functions exclusively to convey post-lexical information. Therefore, it remains unclear how lexical prosody interacts with a post-lexical prosody in a tonal clash context. To bridge this descriptive gap, the present study examines how a lexical pitch accent interacts with post-lexical boundary tones in a tonal clash context in Tokyo Japanese, where pitch functions to convey both lexical and post-lexical information (Pierrehumbert and Beckman 1988; Kubozono 1993 among others). We demonstrate that the most pervasive way to resolve a tonal clash in Tokyo Japanese is neither compression nor truncation, but by a phonological deletion of the lexical pitch accent. At the same time, however, we also demonstrate that one speaker resolves tonal clash without deleting lexical pitch accent. Taken together, we suggest that tonal clash resolution can be speaker-specific. The study offers not only new empirical evidence of tonal clash resolution but also an interesting testing ground of incomplete neutralisation, or covert contrast (Yu 2014 for review) at an accent-intonation interface, as reviewed in section 1.2.

The organisation of this paper is as follows. In the remainder of this section, we review the essential background of Tokyo Japanese prosody (section 1.2), and then propose the research questions and hypotheses (section 1.3). Section 2 describes the method. Section 3 is devoted to an acoustic study. Section 4 discusses the results of the acoustic study. Throughout this paper, we adopt X-JToBI model (Maekawa et al., 2002; Venditti 2005, Venditti et al., 2008; See also, Venditti 2005, for the former JToBI model), which is developed from the autosegmental-metrical (AM) theory (Goldsmith 1976, Pierrehumbert and Beckman 1988, Gussenhoven 2004, Ladd 1996, 2008, among others). In this framework, every tonal event, including both lexical and post-lexical ones, are represented as a sequence of level tones H (high) and L (low). Throughout the paper, we follow the typographical conventions of X-JToBI (as well as the AM theory); a tonal constituent containing an asterisk represents a pitch accent, while a tonal constituent containing ‘%’ refers to a boundary tone.
1.2 Background

1.2.1 Lexical pitch accent

Tokyo Japanese is classified as a pitch-accent language (McCawley 1977, Hyman 2001) with a lexical pitch accent $H^*+L$ (Pierrehumbert and Beckman 1988): Lexical items are distinguished by the presence or absence of an abrupt pitch fall (e.g. a’mé ‘rain’ vs. amé ‘candy’, where the location of the pitch accent is indicated by an apostrophe). Also, words having a pitch accent are further distinguished by the location of the pitch accent within a word (e.g. ha’sí ‘chopsticks’ vs. hasí ‘bridge’). This means that, for instance, a bimoraic bisyllabic word, which is the minimum required prosodic length to exhibit both accentedness and the accent location contrasts, can logically exhibit three accent patterns: accent in the first mora (initial-accented), accent in the second mora (final-accented), or non-accented. These three accents are contrastive if a particle follows\(^2\). Typically, an accentual phrase (AP) in this language thus consists of one lexical word plus one (or more) particle(s) following the lexical word (Igarashi 2015: 529). Henceforth, an AP containing an initial-accented lexical word, final-accented lexical word and non-accented lexical word is simply referred to as initial-accented phrase, final-accented phrase and non-accented phrase, respectively.

1.2.2 Post-lexical boundary tones

In addition to the lexical pitch accent reviewed in section 1.2.1, post-lexical boundary tones are also a major prosodic feature in Tokyo Japanese. The boundary tones include a simple low tone (L\%) and so-called ‘boundary pitch movements’ in X-JToBI, which may or may not occur together with L\%, as reviewed more in details below. As in other languages, the boundary tones contribute to the pragmatic interpretation of the utterance (Igarashi 2015: 527), including questioning (Maeda and Venditti 1998). Although the exhaustive inventory of

\(^{2}\) More precisely, those accents are contrastive if at least one functional word, including a particle, follows. For example, those accents are also contrastive if a copula follows.
the boundary tones in Tokyo Japanese is yet understudied, the current inventory at least includes L% (low tone only), L%+H% (simple rise), L%+LH% (scooped rise), L%+HL% (rise-fall), L%+HLH% (rise-fall-rise) (Igarashi 2015 for review). The pragmatic functions of those tones include, but are not limited to: conveying an insisting attitude of the speaker (L%+H%), interrogative meaning (L%+LH%), explanatory and emphatic meaning (L%+HL%), and wheeling or cajoling in an infant-directed speech (L%+HLH%), in addition to a neutral, declarative meaning (L%) (Maeda and Venditti 1998 for a perceptual study; Venditti et al. 2008, Igarashi 2015 for review).

To understand the basic pattern of those tones, consider a single AP declarative (or neutral) utterance as a default. In this pragmatic context, only an L% is associated with the final mora (i.e. the right edge of the AP and the utterance). On the other hand, in the interrogative utterance, for instance, LH% is associated with the right edge of a prosodic unit (typically, utterance) immediately after the L%. Thus, the resulting phrase-final tones in the interrogative context are L%+LH%³. Throughout this paper, we use the term ‘boundary tone(s)’, to refer to a tonal sequence of an AP-final boundary tone L% and a boundary pitch movement, if any.

To summarize, in Tokyo Japanese, (i) an accented phrase contains a lexical pitch accent and one or more post-lexical boundary tones; (ii) a non-accented phrase contains no lexical pitch accent but does contain one or more boundary tones⁴.

---

³ In this paper, we assume L%+LH% as a typical boundary tones in interrogative context in Tokyo Japanese, following the criteria described in Venditti (2005:184), Venditti et al. (2008) and Igarashi (2015). It should be noted, however, that the interrogative may not retain the L tone (Gussenhoven 2004: 203, 2018). This L-tone debate is beyond the scope of this paper, which does not affect the main conclusion of this paper in that H tone is undoubtedly involved in either analysis at least. See also Hatano and Ishi (2017) for the recent discussion of the acoustic boundary of the L%+LH% and L%+H%.

⁴ Strictly speaking, the following tonal events are also observed within a single AP: a phrase-initial boundary tone (L%) and a phrasal high tone (H+) immediately after L%, which are not directly relevant to the discussion here. Those tones occur within an AP, regardless of the difference in the lexical and post-lexical conditions we examine. For the sake of simplicity, those context-free tones are not considered in the current paper.
1.3 Research questions and hypotheses

An open question is how a lexical pitch accent (see section 1.2.1) interacts with post-lexical boundary tones (see section 1.2.2) in a tonal clash context. As reviewed in section 1.2.1, a typical AP consists of one lexical word plus one (or more) particle(s) following the lexical word. In other words, a typical AP has a particle at the right edge of AP. In everyday communication, however, the particle can be dropped, and leads to the emergence of an atypical AP with a lexical word at the right edge. As a consequence of particle drop, a final-accented phrase encounters a potential risk of clashing with a boundary tone in the final mora.

Interestingly and importantly, there is a debate regarding phonetic differences between final-accented and non-accented phrases in this context. Some studies claim that final-accented phrases (e.g. hana’ ‘flower’) are not distinguishable from non-accented phrases (e.g. hana ‘nose’) unless a particle follows (Sugito 1968, Weitzman 1970, McCawley 1977, Abe 1980, Higurashi 1983, Poser 1984, Vance 1987, Haraguchi 1991). In contrast, other studies argue that final- and non-accented phrases have different surface representations (Pierrehumbert and Beckman 1988), thus predicting no neutralisation of the contrast. Instrumental studies also show mixed results: the contrast may be distinguishable for some speakers (Sakuma 1929, Sugito 1982, Vance 1995) but not for others (Sugito 1968, Higurashi 1983, Poser 1984, Sugiyama 2012). Other than those studies, Warner (1997), who directly tested Pierrehumbert and Beckman’s (1988) prediction, showed that (i) the phrasal peak of fundamental frequency (F0) of final-accented phrases was higher than that of non-accented phrases, and (ii) the difference between the phrasal peak and the F0 value of the penultimate mora of final-accented phrases was smaller than that of non-accented phrases. These patterns were also attested in the reiterant speech, where all the syllables were substituted with [ma] to minimise micro-prosody. It should be noted, however, that Warner (1997) compared final-accented phrases with non-accented phrases with a frame sentence $X$ da (“It is X”), where $X$
is the target lexical word. Therefore, the pitch accent was located in a penultimate mora. In this sense, the phonetic condition tested by Warner (1997) is different from other instrumental studies, such as Sugiyama (2012, Chapter 2), who tested the pitch accent in the last mora, among other phonetic conditions, showing no difference between final- and non-accented phrases.

To summarise, there is a debate as to whether or not final- and non-accented phrases are distinguishable when no particle follows. The situation where two or more categories are apparently indistinguishable, yet there is a small but consistent phonetic difference, is often referred to as covert contrast or incomplete neutralisation, which has been one of the central issues at the phonetics-phonology interface literatures for decades (Port 1996, Ernestus 2011, Yu 2014 for review). While previous instrumental studies have made great contributions to uncovering a case of covert lexical pitch contrast in Tokyo Japanese, all the previous instrumental studies hitherto have paid little attention to a potential tonal clash behind this covert lexical contrast.\(^5\)

It should be noted that, to the best of our knowledge, all the previous instrumental studies examined the covert contrast in isolation or in declaratives, in which case, as mentioned above, only a low boundary tone (L%) is associated with the final mora. This prosodic context is not an ideal context for testing tonal clash since a low tone in H*+L may clash with another low tone, i.e. L%. Therefore, it is difficult to see how a lexical tone interacts with a post-lexical tone in this L-tone sequence. To resolve this problem, it is ideal to test the contrast in a high tone context.

The present study thus examines final- and non-accented words, as well as initial-accented words, in interrogative melodies. In investigating the interaction between the lexical

---

\(^5\) One notable exception is Gussenhoven (2004: 202–203), where the possibility of tonal clash in interrogative context was suggested, showing the utterances of monomoraic and bimoraic monosyllabic words produced by one speaker. However, it should be noted that in this extremely short context, the accented mora can in fact be interpreted as word-initial as well as word-final. This point should be kept in mind since our topic of discussion here is on the contrast between final vs. no accent.
and post-lexical prosody in Japanese, we believe that interrogatives provide another testing
ground of this covert contrast, as a Japanese question melody typically involves a rising pitch
contour as reviewed in Section 1.2.2. Specifically, we examine how the covert lexical
contrast is prosodically implemented in the interrogative melody.

To begin with, we consider the utterance from a single AP as the most fundamental
prosodic structure. Each AP consists of a bimoraic bisyllabic lexical word, which is the
minimum required prosodic length to exhibit an accent location contrast (initial vs. final) as
well as accentedness (accented vs. non-accented). Assumed autosegmental representations of
three accent types in a question melody are given in (1). The final-accented phrase with a
rising tone in (1b) displays the tonal clash between a lexical accent (H*+L) and boundary
tones (L%+LH%).

(1)  

\[
\begin{align*}
\text{a. initial-accented} & \quad \{[\sigma \sigma]_{\text{AP}}\}_{\text{utterance}} & \quad \text{b. final-accented} & \quad \{[\sigma \sigma]_{\text{AP}}\}_{\text{utterance}} & \quad \text{c. non-accented} & \quad \{[\sigma \sigma]_{\text{AP}}\}_{\text{utterance}} \\
& \quad H^*+L \quad L%^+LH% & \quad H^*+L \quad L%^+LH% & \quad L%^+LH%
\end{align*}
\]

Recall that cross-linguistically, two major ways to resolve tonal clash are compression
and truncation (section 1.1). Based on the cross-linguistic findings, we constructed three
hypotheses regarding tonal clash resolution in Tokyo Japanese: compression, truncation and
deletion hypotheses.

The compression hypothesis predicts that the alignment between tones and tone-
bearing units are reorganised to implement all the tones. Therefore, if the compression
strategy was used to resolve tonal clash, both a falling (H*+L) and a rising (L%+LH%)
would be realised within the final-accented phrases. The truncation hypothesis predicts that
the alignment between tone and tone-bearing units remains intact, but instead, that some edge
tones are not realised due to lack of time. Therefore, if the truncation strategy was employed, the tones would be cut off from the end of the utterance (i.e. from the H in L%+LH%). The deletion hypothesis predicts that the deletion of a pitch accent from a phonological output leaves only the L%+LH% tone to be realised. If this strategy is employed in Japanese to resolve the tonal clash, it is predicted that final-accented and non-accented words are indistinguishable in this context. The case of phonological deletion (or phonological replacement) is mentioned for German and Dutch in Ladd (2008: 183–184), though it is presumably less attested crosslinguistically than the other two strategies.

The current production study explores reiterant speech (Larkey 1983) together with actual Japanese words, of which the validity and reliability were convincingly reported in previous research (Warner 1997, Matsui and Hwang 2018). Specifically, every mora of the tested words was replaced by [ma] in order to circumvent the segmental effect on prosody. In comparing the pitch patterns of original speech and reiterant speech, the results corroborate the validity of reiterant speech in prosodic studies.

2. Method

2.1 Participants

Four native Japanese speakers, who grew up in Tokyo (Mean age: 39 years old, ranging from 29 to 49) participated in the recordings. Of these, two were females and the others were males. All of them were former or current staffs at the National Institute for Japanese Language and Linguistics, Tokyo, Japan.

2.2 Speech materials

The targets were eight initial-accented words (e.g. /o’no/), eight final-accented words (e.g. /imo’/) and eight non-accented words (e.g. /ume/) at the right edge of an AP. All the target
words consisted of two light syllables (i.e. two moras). The list of words appears in Appendix A. In addition to the target words, we also recorded (i) the same set of words followed with a nominative particle -ga (e.g. /ono/+/ga/), where initial-, final- and non-accented words are contrastive (e.g. Sugiyama 2012) and (ii) reiterant speech (Warner 1997), the details of which are described in section 2.3.

2.3 Recording procedures
The participants were tested individually in quiet rooms at the National Institute for Japanese Language and Linguistics and Tokyo University of Foreign Studies, Tokyo, Japan. In the recording session, participants were asked to produce the target words in a frame dialogue. In each dialogue, the experimenter said a phrase containing a target word such as (2a), where X is the target word. In response to the (2a), the participant was instructed to ask the experimenter back (2b). After that, the participant was also required to produce the reiterant version of the preceding speech (2c), in which the prosody of the preceding utterance is reiterated but segments are substituted with /ma/ to minimise micro prosody (See Warner 1997 as a previous study applying reiterant speech in Tokyo Japanese). In addition to the targets, we also recorded non-clash versions with a particle /ga/ as a control (See (3)).

(2) Target
a. Experimenter: X-ga suki. ‘I like X.’

b. Participant: X? ‘(Do you like) X?’

c. Participant: ma.m.a. (reiterated version of 2b.)
(3) Control

a. Experimenter: X-ga suki. ‘I like X.’

b. Participant: X-ga? ‘(Do you like) X_{NOM}?’

c. Participant: ma.ma.ma. (reiterated version of 3b.)

The target and control items were randomised, and each dialogue was recorded four times in a different order. The potential effect of the interlocutor (i.e. experimenter) on the participants was counterbalanced. If the experimenter and/or participant felt that they failed, the dialogue was recorded again. The resulting number of the intended tokens was 384 per participant (24 target words + 24 controls, two speech modes, four repetitions). The tokens were recorded onto a portable recorder (MARANTZ PMD661) with a microphone (SHURE SM10A) with 44.1 kHz sampling rate and a 16-bit quantification level.

2.4 Analysis procedures

The target words were acoustically analysed using the Praat speech analysis software (Boersma and Weenink 2010). F0 values were extracted using the autocorrelation algorithm.

In the acoustic analysis, we measured the following F0 values focusing on the second syllable (mora) in both normal and reiterant speech: Maximum F0 value (Peak), Minimum F0 value (Valley) and the F0 range (Max–Min; Range). In the second syllable, we looked at the F0 properties at the vowel interval (e.g. the /i/ part in /usi/), where F0 showed minimum deviation. This is because the words in normal speech contained a voiceless segment in the given syllable interval (e.g. /usi/), where F0 was invisible. An example of these measurement points is given in Figure 3 in Section 3.2.

During the acoustic analysis, one token was excluded due to a pronunciation error. The tokens in which the F0 was not correctly extracted due to irregular vocal fold vibration
were excluded from the analysis. More than 90% of the reiterant version of initial-accented tokens produced by speaker FMN was realised with creaky voice, which made tracking accurate F0 contours problematic. Therefore, we systematically excluded this condition (i.e. reiterated initial-accented phrases produced by speaker FMN) from the analysis.

3. Results
3.1 Holistic shape of the contours

In the interrogative melody with a particle (i.e. control), the words with initial and final accent were realised with a pitch fall in the first or second syllable and a rise in the second syllable, while the words with no accent were realised with a simple rise in the second syllable. The representative contours are shown in Figure 1. In the interrogative melody without a particle (i.e. target), the words with initial accent were realised with a pitch fall in the first syllable and a rise in the second syllable, while the words with final and no accent were realised with a simple rise in the second syllable. The representative contours are shown in Figure 2. To summarise, unless a particle follows, the final accent and no accent, which are under debate, are both realised as a rising contour. Therefore, the next section takes a closer look at the phonetic details of this rising contour, focusing on the final and no accent.

![Figure 1](image1.png)
![Figure 2](image2.png)

**Figure 1.** A representative F0 contours of initial-accented words (a), final-accented words (b) and non-accented words (c) in an interrogative context with a particle (control condition).
Figure 2. A representative F0 contours of initial-accented words (a), final-accented words (b) and non-accented words (c) in an interrogative context without a particle (target condition).

3.2 Phonetic details of the rising contour

To examine whether there are any significant differences among initial, final and no accents in the target condition, linear mixed-effect regression models were constructed using lme4 package (version 1.1-7, Bates et al. 2014) in R 3.1.0 (R Core Team, 2014). In our models, the fixed effect was lexical accent type (three levels: ‘initial’, ‘final’, ‘no’). Items (24 items) and repetitions (4 times) are random effects with a random intercept. To address inter-speaker variation, models were constructed per speaker. Also, we created the models with different intercept specification, thereby we tested all three contrasts. The dependent variables were F0 maximum (i.e. the Peak of the rising contour), F0 minimum (i.e. the Valley of the rising contour) and the Range (the difference between Peak and Valley) as described in Section 2.4 above and illustrated in Figure 3 below. Since three acoustic dimensions in the same dataset were assessed at multiple times, the α-level was adjusted for multiple comparisons by using Bonferroni correction: $\alpha_{\text{adjusted}} = 0.017$ (0.05/3, rounded to 3 decimal places).

Figure 3. Acoustic measurements for the rising contours.
Figures 4, 5 and 6 respectively show mean values of Peak, Valley and Range in normal speech and in reiterant speech, broken down by accent type and speakers. The speakers distinguished initial accent from final accent, and initial accent from no accent by Peak, Valley, and/or Range: The F0 peak and valley of the initial-accented words were significantly lower than that of the final-accented words for all the speakers. The F0 range of the initial-accented words was significantly lower for the majority of our speakers (3 out of 4), though one speaker (MYM) showed a marginal effect (see Figure 6). Likewise, the F0 peak and valley of the initial-accented words were significantly lower than that of the non-accented words for all speakers. The F0 range of the initial-accented words was significantly lower for most of our speakers, except for one speaker (MYM).

For the final vs. no accent contrast in question, the maintenance of the contrast was highly limited. Only one speaker (FMN) distinguished final accent from no accent: the estimated F0 peak of the final-accented words was 59.19 Hz higher, compared with non-accented words [Std error = 17.01, df = 26.98, $t = -3.48$, $p < 0.01$]. This pattern was replicated in the reiterant version of speech as well, though the effect was slightly weaker than in normal speech. The estimated F0 peak of the final-accented words was 33.27 Hz higher, compared with non-accented words [Std error = 15.02, df = 53.53, $t = -2.216$, $0.017 < p < 0.05$]. Also, the estimated F0 range of the final-accented words was 58.86 Hz higher, compared with non-accented words [Std error = 19.03, df = 29.45, $t = -3.092$, $p < 0.01$]. However, this pattern was not replicated in reiterant speech [Std error = 15.1, df = 53.52, $t = -1.99$, $p = 0.0517$]. Thus, it appears that F0 peak pattern is more consistent and robust than the F0 range pattern for this speaker.

For three other speakers, no significant difference was found between final- and non-accented words for any acoustic dimensions measured. It should be noted that, as illustrated in Figure 4a, speaker FMH marginally distinguished final-accented words from non-accented
words in normal speech [Std error = 22.09, df = 28.67, \( t = 2.198, 0.017 < p < 0.05 \)], but this pattern was not attested in reiterant speech [Std error = 17.13, df = 22.7, \( t = -1.448, p = 0.161 \)]. The summary of the statistical analysis is provided in the Appendix B.

To summarise, acoustic analyses showed that three speakers (out of four) failed to distinguish final-accented words from non-accented words, while one speaker showed a distinct F0 peak pattern for final and no accents both in normal and reiterant speech.

![Figure 4](image_url)

**Figure 4.** Mean F0 peak values in normal speech (a) and in reiterant speech (b), broken down by accent type and speakers. Error bars represent one standard deviation. \( \alpha_{\text{adjusted}} = 0.017 \). *, **, *** respectively indicate significance at 0.017, 0.01, 0.001 levels. † indicates a marginal p-value \( (0.017 < p < 0.05) \).

![Figure 5](image_url)

**Figure 5.** Mean F0 valley values in normal speech (a) and in reiterant speech (b), broken down by accent type and speakers. Error bars represent one standard deviation. \( \alpha_{\text{adjusted}} = 0.017 \). *, **, *** respectively indicate significance at 0.017, 0.01, 0.001 levels.
4. Discussion

The present study had two aims. One was to examine how speakers modify pitch contours in the context where lexical accent (H*+L) potentially clashes with post-lexical boundary tones (L%+LH%). Another was whether or not final-accented phrases are distinguishable from non-accented phrases in this context. The results are discussed in the following sections.

4.1 Final vs. no accent in interrogative melodies of Tokyo Japanese

As reviewed in section 1, the AM analysis of Tokyo Japanese prosody predicts that a final lexical accent (H*+L) and post-lexical boundary tones (L%+LH%) can clash in the final mora in interrogative melodies, unless a particle follows. Based on cross-linguistic studies, we constructed three hypotheses (section 1.3). The first hypothesis was that the lexical accent and post-lexical boundary tones were both realised in the tonal clash context (Compression hypothesis), as in Danish (e.g. Grønnum 1991). If this was correct, both a pitch fall associated with lexical accent (H*+L) and a rise associated with post-lexical boundary tones (L%+LH%) should be realised within a phrase. The second hypothesis was that tonal events were not realised from utterance-final due in lack of time (Truncation hypothesis), as in Russian (Odé 2005). If this was the case, the tonal events should be cut off from the right...
edge of the utterance, namely from the high tone of LH%. Finally, the third hypothesis was that the lexical accent was deleted (or delinked) from the phonological output (Deletion hypothesis). In this case, we would expect that the pitch contour of the final-accented phrases should be indistinguishable from that of non-accented phrases.

The results of our acoustic study show that speakers realised the final-accented phrases as a rising contour without a pitch fall (Figure 2b), apparently being identical to the non-accented phrases (Figure 2c). More detailed analysis indicated that the majority of our speakers (3 out of 4) failed to reliably distinguish final-accented phrases from non-accented phrases. This supports the Deletion hypothesis: as illustrated in (4), it can be interpreted that a final lexical pitch accent is deleted from phonological output in the tonal clash context, thereby accounting for the fact that speakers cannot distinguish final-accented phrases from non-accented phrases.

(4) Tonal clash resolution by Speakers FMH, MYM, MSH

a. initial-accented
\[ \{\sigma \sigma\}_\text{AP} \text{utterance} \]
\[ \begin{array}{c}
H^*+L \\
L\%+LH%
\end{array} \]

b. final-accented
\[ \{\sigma \sigma\}_\text{AP} \text{utterance} \]
\[ \begin{array}{c}
H^*+L \\
L\%+LH%
\end{array} \]

\[ \begin{array}{c}
H^*+L \\
L\%+LH%
\end{array} \]

At the same time, however, one of our speakers produced a significantly higher F0 peak in final-accented phrases than in non-accented phrases. This pattern was observed both in normal speech and in reiterant speech, suggesting that this speaker shows the case of covert lexical contrast. Also, this suggests that if we assume that the underlying representations are the same among these speakers, then this speaker does not employ deletion to resolve a tonal clash, since the lexical contrast is not completely neutralised. A
natural question then arises: How does the speaker resolve tonal clash, by compression, truncation, or another strategy?

One might speculate that the pattern produced by the Speaker FMN can be seen as a case of compression, simply because both the lexical and post-lexical contrasts are in fact preserved for this speaker. However, this interpretation is not plausible, since some tones do not surface at all.

Another possibility is that the pattern is a consequence of truncation, by which only a high tone in the lexical pitch accent (i.e. $H^*$) surfaces. This is what is in fact predicted in an L-tone context in a previous study (Pierrehumbert and Beckman 1988). In this context, final-accented words can be realized with slightly higher F0 peak than non-accented words in this context (e.g., Warner 1997). Therefore, it is reasonable to interpret the parallel situation occurs for speaker FMN in the current context ($L%+LH\%$). However, if this happened in interrogative melodies as well, we would predict that the post-lexical contrast, i.e. the one between interrogative and declarative, is neutralised due to lack of boundary tones. Contrary to this prediction, such a post-lexical neutralisation is unlikely as the native speakers can distinguish the difference between interrogatives from declaratives. Therefore, the pattern of the speaker FMN cannot fall into the case of truncation.

To summarize, it is difficult to classify the F0 pattern produced by Speaker FMN either as the consequence of deletion, compression or truncation. This type of tonal clash resolution can possibly be situated as an extreme case of the F0 undershoot (e.g. Prieto 1998 for Mexican Spanish), where some F0 inflection points are reduced due to compression (Ladd 2008). To the end, as Ladd (2008: 183) points out, the boundary of truncation, compression, (and other strategies) is not clear-cut in nature, whereas all the strategies are conspicuous in a physical-time pressure. The interpretation of this type of tonal clash resolution has much

---

6 This point is also confirmed by our pilot perceptual study, where a native listener distinguished interrogatives from declaratives with almost a perfect accuracy.
room to be revisited both from empirical and formal points of view in future studies. Further production studies are necessary to investigate to what extent this type of tonal clash resolution is pervasive in Tokyo Japanese. At the same time, a follow-up perceptual study will inform us as to whether the learners (i.e. listeners of language) can use the contrastive F0 peak height as a consequence of tonal clash resolution produced by FMN, as a cue to the identification of lexical accents.

4.2 Can tonal clash resolution be speaker-specific?
Our results entertain the possibility that tonal clash can be resolved in speaker-specific ways. A recent comparative study showed that tonal clash resolution strategies are variable not only across languages but also within a language (Rathcke 2016). For example, research on German and Russian, both of which have been previously reported to use truncation (Ladd 1996), has shown a variety of strategies within a language by examining how various (non-lexical) pitch accents, say, H*+L, L+H*, and H*, clashed with L%. This suggests that speakers can resolve tonal clash in context-specific ways within a language. The current study extended the previous study’s findings by showing that speakers can also resolve tonal clash in speaker-specific ways within a single language; Some speakers resolve tonal clash by deleting pitch accent, while others resolve it without deleting pitch accent, resulting in either complete or incomplete neutralization of lexical contrast.

4.3 Conclusion
Through a case study of Tokyo Japanese, the present study offers new evidence of tonal clash resolution in a pitch-accent language with both lexical and post-lexical tones. We demonstrated that final-accented phrases are largely indistinguishable from non-accented phrases in the interrogative context, where final lexical accent clashes with post-lexical
boundary tones. The results support the idea that lexical accent is deleted from phonological output to resolve tonal clash, resulting in the complete neutralisation of lexical contrast. However, one speaker distinguished final-accented phrases from non-accented phrases by slightly higher F0 peak. Taken together, we entertained the possibility of speaker-specific tonal clash resolution; Tonal clash can be resolved not only in language-specific ways but also in speaker-specific ways, interacting with the (in)complete neutralisation of lexical contrast.

**Acknowledgements**

This study is a part of the NINJAL collaborative research project (project name: Word-prosody and sentence prosody), and partially supported by KAKENHI grants (#17J08493 to the first author; #26244022 to the second author). The authors are benefited from the audience at NINJAL collaborative research meeting (September 2016), APAP-3 (June 2017) and the Phonetics/Phonology Group meeting at the University of Toronto (Dec 2017) for valuable feedback. We would also like to thank Haruo Kubozono, Tim Vance and Carlos Gussenhoven for their comments on earlier versions of this paper. All the remaining errors are of course on our own. Finally, we thank our speakers for their participation in the experiment.

**References**


### Appendix A: Speech materials (Targets)

<table>
<thead>
<tr>
<th>Initial accent</th>
<th>Final accent</th>
<th>No accent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthography</td>
<td>Phonemic transcription</td>
<td>Orthography</td>
</tr>
<tr>
<td>奈良？</td>
<td>nara</td>
<td>花？</td>
</tr>
<tr>
<td>屋根？</td>
<td>jane</td>
<td>村？</td>
</tr>
<tr>
<td>文字？</td>
<td>mozi</td>
<td>波？</td>
</tr>
<tr>
<td>斧？</td>
<td>ono</td>
<td>海苔？</td>
</tr>
<tr>
<td>春？</td>
<td>haru</td>
<td>指？</td>
</tr>
<tr>
<td>麦？</td>
<td>mugi</td>
<td>網？</td>
</tr>
<tr>
<td>罍？</td>
<td>nabe</td>
<td>馬？</td>
</tr>
<tr>
<td>夜？</td>
<td>joru</td>
<td>苾？</td>
</tr>
</tbody>
</table>
### Appendix B: Statistics summary

#### Dependent variable: Peak

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Contrast</th>
<th>Intercept Estimate (β) Std error df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FMN</strong></td>
<td>Initial vs. final</td>
<td>268.29 27.18 43.85 9.876 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>209.2 27.18 43.85 7.698 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>-59.19 17.01 26.98 -3.48 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FMH</strong></td>
<td>Initial vs. final</td>
<td>200.88 23.69 23.74 8.48 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>249.44 23.69 23.74 10.53 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>48.56 22.09 28.67 2.198 †</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MYM</strong></td>
<td>Initial vs. final</td>
<td>79.395 11.791 23.34 6.733 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>69.281 11.831 23.69 5.856 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>-10.114 10.561 38.6 -0.958 0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSH</strong></td>
<td>Initial vs. final</td>
<td>67.11 11.64 28.44 5.765 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>81.56 11.58 27.84 7.043 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>14.46 10.73 34.58 1.548 0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Dependent variable: Valley

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Contrast</th>
<th>Intercept Estimate (β) Std error df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FMN</strong></td>
<td>Initial vs. final</td>
<td>82.45 16.18 37.26 5.097 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>79.76 16.18 37.26 4.931 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>-2.692 10.457 33.59 -0.257 0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FMH</strong></td>
<td>Initial vs. final</td>
<td>62.18 17.71 27.26 3.511 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>84.64 17.71 27.26 4.779 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>22.45 15.65 43.66 1.435 0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MYM</strong></td>
<td>Initial vs. final</td>
<td>52.87 6.664 26.301 7.935 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>40.65 6.696 26.832 6.581 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>-8.813 6.113 38.15 -1.442 0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSH</strong></td>
<td>Initial vs. final</td>
<td>22.127 5.003 29.043 4.423 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>23.776 4.969 28.381 4.785 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>1.649 4.647 34.17 0.355 0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Dependent variable: Range

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Contrast</th>
<th>Intercept Estimate (β) Std error df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FMN</strong></td>
<td>Initial vs. final</td>
<td>185.28 30.13 43.29 6.149 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>126.42 30.13 43.29 4.196 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>-58.85 30.13 43.29 -3.992 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FMH</strong></td>
<td>Initial vs. final</td>
<td>139.94 23.96 23.77 5.841 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>165.47 23.96 23.77 6.906 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>25.252 22.454 27.896 1.137 0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MYM</strong></td>
<td>Initial vs. final</td>
<td>26.009 11.956 23.495 2.175 0.999 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>24.458 12.006 23.925 2.037 0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>-1.551 10.782 37.68 0.144 0.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSH</strong></td>
<td>Initial vs. final</td>
<td>45.691 10.053 78.03 4.545 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial vs. no</td>
<td>57.044 9.886 78.08 5.713 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final vs. no</td>
<td>11.352 9.398 78.14 1.209 0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>