ACOUSTIC AND PERCEPTUAL EVIDENCE OF COMPLETE NEUTRALIZATION OF WORD-FINAL TONAL SPECIFICATION IN JAPANESE

by

Kazumi Maniwa
B.A., Shimane University, Shimane, Japan

Submitted to the Department of Linguistics and Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirement for the degree of Master of Arts

Redacted Signature

Chief Advisor

Redacted Signature

Committee Member

Redacted Signature

Committee Member

Date submitted: 5/1/2002

ABSTRACT
Kazumi Maniwa, M.A.
Linguistics, April 2002
University of Kansas

This study investigates the extent to which Japanese lexical pitch-accent distinction is neutralized in word-final position. Native speakers of Tokyo Japanese produced minimal word pairs differing in final accent status. Words were produced both in isolation and in a sentential context, where neutralization would not be expected due to following tonal specification. Examination of pitch patterns on relevant moras revealed a clear distinction between accent-opposed pairs produced in context but no such difference between items produced in isolation. Both the words produced in isolation and the words excised from sentential contexts were then presented to Japanese listeners in a lexical identification task. Participants could clearly distinguish items extracted from sentences but identified words uttered in isolation at chance level. These results suggest that phonological neutralization of final pitch accent is complete, showing no effects of underlying specification in either production or perception.

Chapter 1
Introduction

A fundamental concept of phonological theory is neutralization, whereby phonemic distinctions are eliminated in certain phonological contexts. The phonological approach for merging distinctive phonemes into a single phoneme in particular phonological circumstances assumes that neutralization is phonetically manifested as complete. For example, the traditional analysis of German word-(syllable-) final obstruents discusses that only voiceless obstruents are allowed at the ends of words, or more generally, at the ends of syllables. Table I (Port & O'Dell, 1985; p.456) shows relevant data from German. A generative phonological rule of the type generally proposed to account for the neutralization would be;

[-sonorant] ⟶ [-voice] /____ #
where # means syllable boundary

	German word	English gloss
Initial position	der Back der Pack	mess, table pack, bundle
Medial and final Position	Plural Alben [albən] Alpen [alpən] Singular Alb [alp] Alp [alp]	 elves mountain pastures

Table I. Some Examples illustrating word-final devoicing in German

The rule stipulates that the obstruents must be voiceless at the end of a syllable, a position that normally includes morpheme boundaries. As can be seen,

voiced and voiceless obstruents contrast in initial position and intervocalically. However, in final position, for example, in the singular forms of *Alb* and *Alp,* both are pronounced as [p], despite the orthographic distinction. This justifies a phonological account postulating a rule that changes all obstruents (i.e. [-son] segments) to [-voice] at the end of syllables. Such an analysis claims that after the rule, the two segments have identical feature specifications. Thus it predicts that forms such as *Alb* and *Alp* should be pronounced identically in all respects, implying thereby also that there should be no perceptible differences between the two forms.

Acoustically, a voicing distinction in final stop consonants is generally seen in the duration of a stop closure (voiceless stops are longer), the amount of voicing into the closure (voiced stops have more), and the duration of a preceding vowel or other sonorant (vowels are longer before voiced stops) (Blumstein, 1991).

However, there have been a number of studies that question whether phonological neutralization is phonetically complete or incomplete. Many studies on neutralization have focused on word- (or syllable-) final consonant devoicing, with findings supporting either complete or incomplete neutralization.

In one early study on this issue by Dinnsen & Charles-Luce (1984) on Catalan, the authors found significant differences in duration of the stop closure, voicing during the stop closure, and of the preceding vowel, but found that different speakers marked those distinctions using different cues. Charles-Luce and Dinnsen (1987) reanalyzed a subset of data in response to criticism about the choice of items in the early study, and found that the only significant effect across

speakers was for closure voicing (significant across speakers). In those studies, significant effects of underlying voicing were highly restricted by circumstances.

Slowiaczek & Dinnsen (1985) and Port & O'Dell (1985) found more reliable effects of underlying voicing in Polish and German, respectively. The Polish study showed an effect of underlying voicing on vowel duration which was significant across speakers, and consistent across types of final obstruents. Effects of underlying voicing on closure duration and closure voicing duration were found, but these were limited to certain speakers, environments, or final obstruents. Using a large number of speakers and words read in isolation, Port and O'Dell (1985) found effects on vowel duration, closure voicing, and burst duration, all significant across speakers. Furthermore, the authors tested listeners' ability to identify the productions, and found that listeners could tell which member of the minimal pair was intended, with significantly greater than chance accuracy. They asserted that the apparent devoicing of final /d/ is due to an implementation rule somehow "warping or biasing [the] articulatory gesture" rather than actually changing its phonological specification.

Charles-Luce (1985) suggested similarly that a phonological devoicing rule of some type may take place in the German speakers but that the phonetic processes implementing the segments in production are somehow sensitive to the underlying voicing contrast. Additionally, the author showed that the phonetic and sentential environment of the devoiced segment affect the degree to which the voicing contrast is neutralized ---neutralization is complete in some contexts and clearly incomplete in others---so that this information must be available to an

implementation rule as well. The author concluded that final devoicing is not properly a neutralization rule that makes [+voice] obstruents [-voice] ones, but rather that it causes [+voice] obstruents to become unspecified for voice. Then, (context-sensitive) implementation rules similar to those proposed by Port and O'Dell (1985) must cause these unspecified segments to be realized as voiceless or nearly voiceless, depending on their environment.

On the other hand, some authors suggested that effects of underlying voicing in neutralization circumstances are due to orthographic differences or speaking style. Fourakis and Iverson (1984), in a study of German, discussed that the findings might be the result of 'hypercorrection' by subjects due to the orthographic differences between words, seen only because they were reading aloud, and thus not representative of more natural speech patterns. The authors found some significant effects of underlying voicing in the reading task, but not in the verb conjugation task, and concluded that incomplete neutralization occurs when speakers try to distinguish between words with differing orthography while reading. Jassem and Richter (1989), in a study on Polish using four speakers and seventeen minimal pairs, also avoided having speakers read the test words, and found no significant differences between underlyingly voiced and voiceless final segments.

However, Fourakis & Iverson's study has been variously criticized for using a small set of subjects and a small set of words in their conjugation task which did not involve actual minimal pairs of words but only phoneme sequences. It also does not seem likely that speakers of German should partially recreate a

neutralized distinction in the presence of orthography when speakers of Korean (Kim & Jongman, 1996) and probably Dutch (Jongman, Sereno, Raaijmakers, & Lahiri, 1992) do not make such differences even when reading from a list involving similar neutralized contrasts.

Port and Crawford (1989) reported an extensive investigation of speech style and effect of underlying voicing in German. The authors elicited the target words from speakers under five speaking conditions where it was determined that subjects can control the extent of final devoicing based on the pragmatics of a speaking situation. In conditions 1 and 2, the words were embedded in different semantically plausible sentences, with the prosody of the sentence closely matched for the minimal pairs. Filler sentences were included to disguise the minimal pair target words. Subjects read these sentences in condition 1, and repeated them after an experimenter in condition 2. Therefore, those tasks did not emphasize the possible distinction or promote careful pronunciation. In condition 3, subjects read sentences which contrasted the two members of the same minimal pairs, with the words disambiguated within the sentence (e.g. *Ich habe 'Rat', wie Ratschlag, gesagt; nicht 'Rad', wie Fahrrad.* = I said 'Rat', as in 'bit of advice'; not 'Rad', as in 'bicycle'.) In condition 4, subjects dictated sentences with no disambiguating information to a German experimenter who attempted to write the words (e.g. *Ich habe 'Rat'; nicht 'Rad' gesagt.* = I said 'Rat', not 'Rad'.). This speech style is expected to encourage speakers to produce a distinction between the members of the minimal pair. Finally, in condition 5, speakers read the target words from a list in isolation. The results suggested that there were effects of

underlying final voicing which are not limited to careful speech, although speakers can make more or less clear differences in the pairs depending on speech style. The authors also pointed out that listeners can make use of even the differences produced in less careful speech. They concluded, then, that German does not have an abstract phonological rule of neutralization despite almost a hundred years of assertions that it does, by accounting for the fact of practical neutralization in terms of phonetic implementation rules.

Thus, the acoustic and perceptual facts of final devoicing seem to suggest that the presumed neutralization is both observably incomplete and clearly variable in nature. Dinnsen (1985), examining numerous similar phenomena, offers a typology of four possible realizations of phonological neutralization: (A) the standard conception of neutralization, where no differences in either perception or production are observed between underlyingly contrasting forms, (B) a limited neutralization where (small) differences are maintained in production but are not perceptible, (C) as in German, incomplete neutralization where differences are observed in both production and perception, and (D) the impossible situation of perceptual differences occurring in the absence of differences in production.

Dinnsen observed that type C is quite common, citing final devoicing in Catalan, Polish, and Russian as well as German. Type C cases constitute non-neutralization, namely, rules that produce outputs with phonetic differences corresponding to underlying differences and those differences are discriminable. Type B would be an instance of neutralization limited to the perceptual domain where the listener treats two acoustically distinct tokens as perceptually

equivalent. The facts of production would not, however, be described by a neutralization rule. Type B cases are very similar to allophonic phenomena in that they involve production differences that are not generally discriminable by native speakers of the language. The difference is that the different sounds in the Type B cases occur in the same context. In any event, to the extent that sound changes in progress involve rules that are synchronically motivated. Type B is also entirely possible, though it is in many cases presently not distinguishable from type C as only production studies have been completed. One difficulty with Type B cases, according to Dinnsen, is that while it is claimed that they involve production differences that are not discriminable, it may well be that the perceptual tests were not sensitive enough to reveal perceptual salience. Thus it is not known whether the production differences were perceptually salient. More sensitive measures may result in the reanalysis of Type B cases as Type C.

Dinnsen (1985) claims, however, that Type A is not only unattested but also problematic in that there is always the possibility that a production study will fail to examine some aspect of an acoustic signal that would show relevant differences. Dinnsen discusses that the review of experimental studies examining putative neutralizations revealed in every case the existence of systematic production differences corresponding to underlying distinctions. In order for a rule to be denied Type A status, it is sufficient to find either production differences or perceptual differences. Type A cases also depend on the reasonable certainty that there are no other differences to be found in production and perception. Depending, then, on which phonetic parameters are selected for

examination, an instrumental study may show no differences. Given limited knowledge of all the factors involved in speech perception and production, it is virtually impossible to be sure that there are not some differences present somewhere in the signal that contribute to a production difference. While technically true, this last argument is not very useful in evaluating the extent to which very detailed perception/production studies may suggest that a neutralization is in fact complete, and more recent studies have shown instances where it is at least highly probable. For example, Lahiri, Schriefers & Kuijpers (1987) showed complete neutralization in their study of vowel length in Dutch; they found no differences in duration between long vowels served by an open-syllable lengthening rule and vowels that are underlyingly long. Kim and Jongman (1996) report Type A neutralization for manner of articulation in certain intervocalic Korean consonants, employing rigorous production and perception tests. The latter study is especially very important in that it investigated a different kind of neutralization from past research, namely that of manner of articulation, and moreover, it provided an instance of complete neutralization despite potential cues for underlying manner in the orthography. The latter finding challenges the claim by Fourakis and Iverson (1984) which argued that incomplete neutralization in earlier studies of German resulted from hypercorrect pronunciation of differences between minimal pair members in terms of orthography.

The present study will investigate a different type of neutralization, namely that of word-final pitch accent in Japanese. Both production and perception data will be presented.

Chapter 2
Japanese pitch accent

2.1 Comparison of competing phonological theories on pitch contour

Japanese is considered to be a pitch accent language: pitch functions to make lexical distinctions so that the presence or absence of an accent on a particular syllable can determine what word is being uttered. This lexical accent has been described as a rapid fall from a relatively high pitch to a relatively low pitch, while other lexical items lacking this fall are said to be unaccented. The accent patterns on short phrases in Tokyo Japanese (standard Japanese) are traditionally described as follows: (1) one characteristic pitch pattern, namely a high-low tonal sequence, marks the word accent (2) a word has at most one accent on any mora or can be unaccented (3) thus, *n*-mora words have *n*+1 possible accentuations (4) phrase-initial morae have a low tone and second morae have a high tone unless the word in that position has an initial accent (Kitahara, 2001). Conventionally, accent location is counted from the beginning of a word in the literature of Japanese accentology. Thus, the initial-accented form is called accent-1, the final-accented form of a 2-mora word and the penultimate-accented form of a 3-mora word are called accent-2, and so on. In addition, the unaccented form is called accent-0.

The possible accent assignments for two-mora words therefore have been exemplified by the traditional theory as below[1]:

accent-1 (accented on the 1st mora)	accent-2 (accented on the 2nd mora)	accent-0 (unaccented)
há shi **H** L *'chop sticks'*	ha shí L **H** *'bridge'*	ha shi L H *'edge'*

Table 2. Possible accent locations for two-mora items in the Tokyo Standard dialect

As the table shows, this traditional understanding assumes a fully specified surface phonological representation in which every tone-bearing unit is specified as being produced at a high or low pitch level. McCawley (1968) derived this output within an early generative phonology framework of linear models. Haraguchi (1977) offered an autosegmental analysis with the same output, except for a small difference in non-initial phrases, and produced the schematized pitch patterns shown in table 2.

According to considerable literature (McCawley, 1968; 1977; Weitzman, 1970; Sugito, 1982; Higurashi, 1983; Poser, 1984; Vance, 1987; 1995), for nouns with a short final syllable, the difference between final accent and no accent is typically manifested when nouns are followed by a grammatical particle such as /ga/ (Nominal), /wa/ (Topical), and /o/ (Accusative). Otherwise, words with the accent on their final mora and words with no accent at all have the same F0 pattern within the word. McCawley made this explicit when he stated that "a final-accented phrase...is indistinguishable from an unaccented phrase: each is pronounced entirely on a high pitch, except for the first mora, which is low-pitched" (1968, p.139).

---

[1] The sequences of LL and HH do not exist.

If another syllable, such as a grammatical particle, follows the final-accented or unaccented word within the same prosodic phrase, then the underlying difference between the two types of LH becomes evident with tonal sequences LHL for an accented form and LHH for an unaccented form.[2] For example, the pitch pattern on /haná+ga/ 'flower+Nominative' is described as LHL, whereas that on /hana+ga/ 'nose+Nominative' is described as LHH. The distinction is said to be neutralized utterance-finally, both /haná/ and /hana/ being LH in isolation.

accent-2	accent-0
ha ná ga	ha na ga
L **H** L	L H H
'flower+Nominative'	*'nose*+Nominative'

Table 3. Possible accent locations for two-mora words with a Nominative particle in the Tokyo Standard dialect

That is, a final-accented word will be followed by a low tone mora, since any mora after the accent is low, while an unaccented word will be followed by a high tone mora, since there is no accent to trigger the fall to low pitch. It should be noted here that within the framework of the traditional theory, the tones assigned to the two types are still the same for the words themselves, with the same H label for both accented second mora /ná/ and unaccented second mora /na/, and diverge only on the following mora.

However, the traditional theory makes no explicit predictions about what fundamental frequency will do during any of the tones assigned by the theory. F0 does not progress in sudden jumps between low and high, stair-step fashion, and any mapping from high and low tones to F0 is not straightforward. Without

[2] Bold letter H was introduced in the present study to differentiate pitch accent high (H) and

instrumental methods, linguists working in the traditional theory could do little toward a more explicit description. Several recent approaches to Japanese pitch accent use instrumental methods and deal with F0 contour itself. Poser (1984) investigated F0 extensively, and some of the conclusions were reproduced in one of the most comprehensive characterizations of the phonological and phonetic instantiation of pitch accent in Tokyo standard Japanese to date by Pierrehumbert & Beckman (1986, 1988). Japanese has a rule of catathesis, Japanese has a L at every accentual phrase boundary rather than at a subset of such boundaries, and the high to which the F0 rises in the accented case is higher than in the unaccented case. However, the study was still conducted within some form of the traditional theory.

Pierrehumbert & Beckman (1988) introduced an entirely new method of describing Japanese pitch accent by using only a few tones per phrase, with interpolation between them. In their study, grouping of words into prosodic phrases occurs at three levels in Japanese: the accentual phrase (AP), the intermediate phrase (IP), and the utterance (utt). The accentual phrase (AP) is typically characterized by a rise to a high around the second mora, and subsequent gradual fall to a low at the right edge of the phrase. The degree of perceived disjuncture between words within an accentual phrase is less than between sequential words with an accentual phrase boundary intervening.[3] The second type of prosodic grouping in Japanese is the higher-level intermeditate phrase (IP), which consists of a string of one or more accentual phrases. Like accentual

---

phrasal high (H).

phrases, this level of phrasing is also defined both tonally and by the degree of perceived disjuncture within/between the groups. However, the tonal markings and the degree of disjuncture are different from those of the accentual phrase. The intermediate phrase is the prosodic domain within which pitch range is specified, and thus at the start of each new phrase, the speaker chooses a new range which is independent of the former specification. The utterance consists of one or more intermediate phrases (IP).

The tones assigned to a phrase are limited to a boundary low tone (%L) at the beginning of an utterance, a 'phrase peak' high tone (H) which is normally attached to the second mora, an 'accent peak' high-low tone (HL) on the accented mora, and a boundary low tone (L%) at the end of each phrase. The HL composite label placed within the accented mora is used to mark the lexical accent in accented accentual phrases (AP). The H portion indicates that the high part of the falling tone is associated with the accented mora itself, and the following L indicates that a low occurs at some fixed point afterward, usually within the following mora. This HL accent label is absent in unaccented words.

The J_ToBI model of Japanese intonation distributed by Venditti (1995, 2000) is a transcription model of intonational patterns developed for Tokyo Japanese. This system relies heavily on the model of Japanese tone structure put forth by Beckman and Pierrehumbert (1986, 1988), which uses a tone-sequence approach to intonation modeling as mentioned above. The most noticeable difference between the model of Japanese tone structure ('JTS') by Pierrehumbert

---

[3] In Tokyo standard Japanese, it is most common for unaccented words to combine with adjacent words to form accentual phrases (Venditti, 2000).

& Beckman (1988) and J_ToBI is the reduction in the number of prosodic phrase levels. JTS proposed three levels above the word in the hierarchy of Japanese, namely the accentual phrase (AP), the intermediate phrase (IP), and the utterance. The accentual phrase was defined exactly as it is in J_ToBI, but the JTS intermediate phrase and utterance have been merged into one level of phrasing in J_ToBI, namely, the intonation phrase (IP). The J_ToBI also introduced H*+L accent label instead of HL. Thus, the complete tonal transcription of the APs is:

Unaccented AP	%L H- L%
Accented AP	%L (H-) H* +L L%

The most significant feature of the JTS model by Pierrehumbert & Beckman (1988) and J_ToBI model by Venditti (2000) is the sparse specification of tones, compared to the models by McCawley (1977) and Haraguchi (1977) where tones spread to all tone-bearing units. In addition, the JTS and J_ToBI models separate the pitch accent (H*+L) from the phrasal H while in the traditional theory, just a single type of high tone (H) is assigned to every mora between the second mora of a phrase up to the accent or the accentual high tone spreads. Several experimental studies (Poser, 1984; Kubozono, 1983) have revealed that an accent peak (at the H*+L tone) is higher than a phrasal peak (H tone only). This phonetic fact cannot be captured by theories with just a single type of high tone.

The JTS model by Pierrehumbert & Beckman (1988) and J_ToBI model by Venditti (2000) also include several purely phonetic factors which affect F0. One such factor is final lowering; the last several morae of a declarative utterance have lower F0 than in a corresponding question, and therefore posit a final lowering

effect for statements. A second phonetic factor is declination, that is, a completely phonetic unconditioned lowering by which F0 falls by some small number of Hertz per second in all utterances. One additional phonological factor which makes those models different from other traditional models is the effect of catathesis. Pierrehumbert & Beckman's (1988) data showed that the value of a final boundary low tone depends on whether there is an accent in the prosodic phrase it terminates or not: a final L% is at considerably higher F0 at the end of an unaccented phrase than at the end of a phrase containing an accent. Accent H*+L tones trigger catathesis, which decreases the pitch range after the H*+L tone by lowering the high line. This has the effect of making everything after an accent lower than otherwise expected until pitch range is reset at the next intonational phrase boundary.

Figure 1 and Figure 2 show examples of the tonal representation predicted within the framework of Pierrehumbert & Beckman's (1988) theory and J_ToBI model proposed by Venditti (2000). These representative productions include two pitch accent types of two-mora words, /haná/ *'flower'* and /hana/ *'nose'*, respectively, in sentential context. These two productions, namely stimuli with final accented /haná/ and unaccented /hana/, are also reintroduced in Chapter 5: General Discussion, including the pitch contour from the actual data for more a detailed discussion of the issue of neutralization.

Figure 1 is the example of a second-mora accented item (accent-2), including the token /haná/ *'flower'*. The following is the prosodic structure of the carrier

sentence and the tones as predicted by the JTS model of Pierrehumbert & Beckman's (1988) model and J_ToBI model by Venditti (2000).

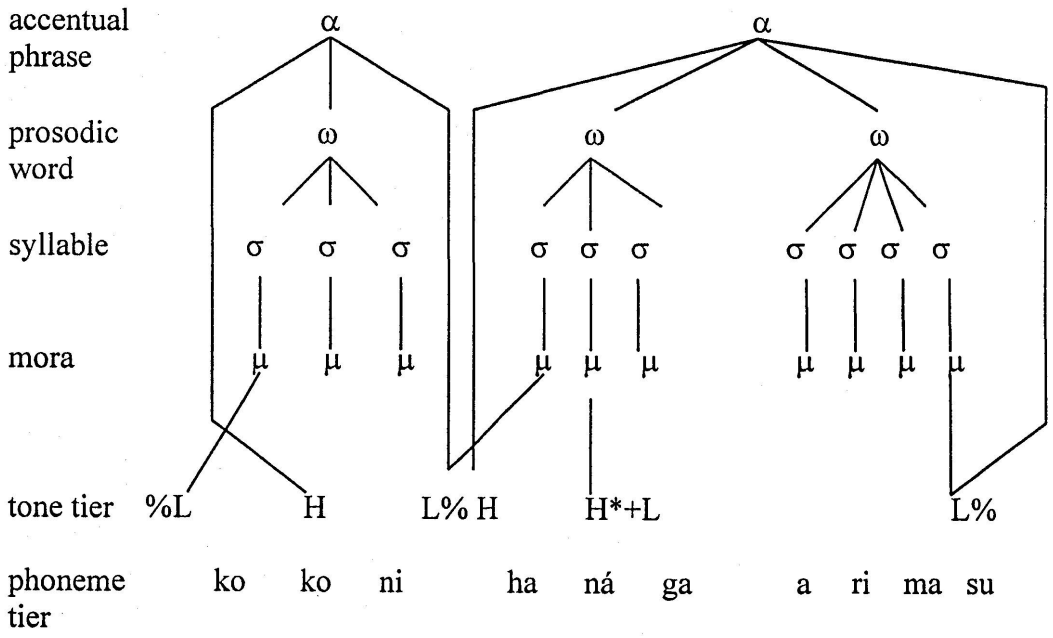

Figure 1. Prosodic and tonal structure of an utterance [kokoni hanága arimasu] 'here is a flower' based on Pierrehumbert & Beckman's framework and J_ToBI.

There are two accentual phrases (APs), namely [koko+ni (Locative)] and [haná+ga arimasu]. The first AP is unaccented but there are two tones (%L and H) associated with it. The utterance- and accentual phrase-initial %L tones attach to the 1st mora of the first word of the accentual phrase unless that attachment is blocked by a lexical pitch accent tone on the 1st mora of the accentual phrase. The accentual phrase-initial H generally attaches to the 2nd mora of the accentual phrase, second [ko] in [koko+ni] in this case. The second AP consists of two prosodic words, namely [haná+ga (Nominative)] and predicate [arimasu]. At the word level, the phrase-boundary L% tone is associated with the 1st mora of the second accentual phrase. A pitch accent (H*+L) links to a lexically specified

mora: [ná] in [haná + ga (Nominative)] in this case. The H* portion indicates that the high part of the falling tone is associated with the accented mora itself, and the following +L indicates that a low occurs at some fixed point afterward, within the following particle, in this case. After this sharp fall, the following tones are lowered until the next AP. The final low boundary tone, L%, is placed at the phrasal edge, namely, on the third mora [ma] in this case, because the last mora of the second AP, [su], is devoiced.

As mentioned above, what makes these models different from other generative studies is the sparse specification of tones and separation of the pitch accent H*+L from the phrasal H. In traditional studies, just a single type of high tone is assigned to every mora between the second mora of a phrase up to the accent (McCawley, 1977) or the accentual high tone spreads (Haraguchi, 1977). Therefore, it was almost impossible to predict prosodic and tonal patterns above the word level in traditional generative theory. These models also illustrate the difference in F0 of the phrasal H and the pitch accent H*+L revealed by several experimental studies.

For an unaccented token, the association of tones will produce the following structure in Figure 2.

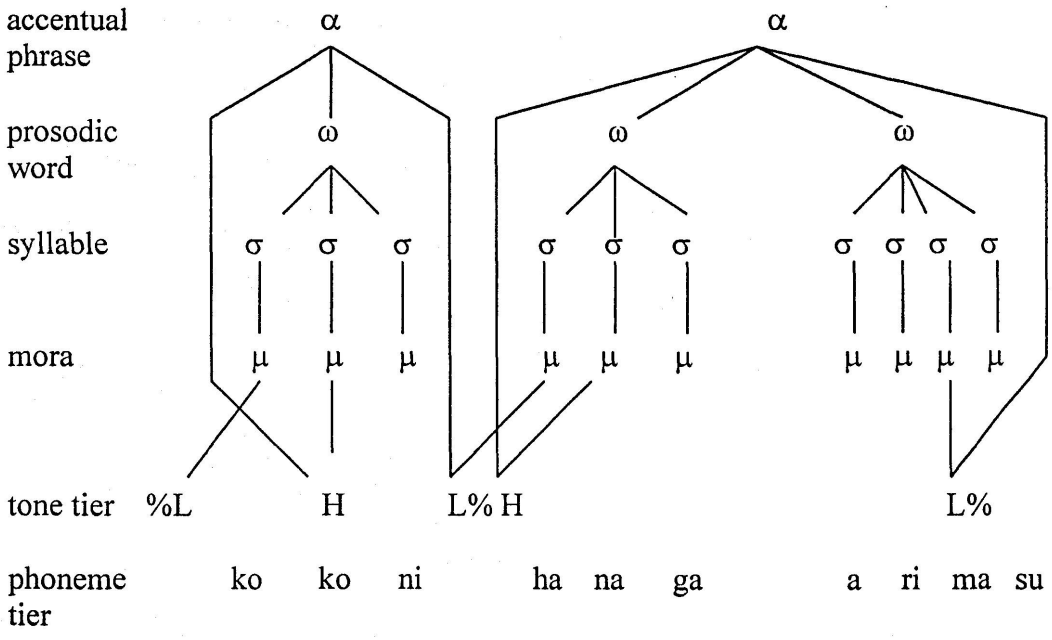

Figure 2. Prosodic and tonal structure of an utterance [kokoni hanaga arimasu] 'here is nose' based on Pierrehumbert & Beckman's framework and J_ToBI.

Again, there are two accentual phrases (APs), namely [koko+ni (Locative)] and [hana+ga (Nominative) arimasu]. The first AP is unaccented but is associated with two tones, the utterance-and accentual initial low tone %L and phrasal H. The second AP consists of two prosodic words, [hana+ga] and predicate [arimasu]. The boundary L% is attached to the first mora of the second AP, /ha/, and phrasal H is associated with the second mora, /na/. There is no sharp pitch fall from high tone to low as in H*+L, and therefore the tone shows a gradual fall to the final boundary L%.

As shown in Figure 1 and 2, the JTS model by Pierrehumbert and Beckman (1988) and the J_ToBI model by Venditti (2000) predict that final accented and unaccented words will be different in tone on the second mora when followed by the grammatical particle. The pitch accent marker H*+L which has a low tone portion for the sharp fall on the following mora and phrasal H with no such fall

following, are introduced for the second mora of the accented AP and unaccented AP, respectively. In contrast, the tonal specification predicted by traditional theory did not capture this difference, simply using H tone for both the pitch accent and the phrasal H. The instrumental methods, and explicit acoustic measurements and perceptual experiments are introduced in the present study to examine purely phonetic factors and to discuss those issues.

2.2 Neutralization in word-final pitch accent

The question of whether or not final-accented and unaccented words have the same pitch contour in isolation has been investigated before (e.g., McCawley, 1968; 1977; Weitzman, 1970; Uwano, 1977; Neustupny, 1978; Sugito, 1982; Higurashi, 1983; Poser, 1984; Vance, 1987; 1995). The claim of neutralization, however, has been challenged by some studies. For instance, Uwano (1977) claimed that the pitch pattern on pairs like /haná/ and /hana/ are not identical for all speakers on all occasions. He suggested that an accented final mora might differ from an unaccented one by having a higher pitch or a falling contour. Neustupny (1978) claimed that the distinction is neither clearly maintained nor entirely neutralized. Although he proposed that it is realized acoustically by some inconsistent set of interacting features, the author explicitly mentioned only pitch and intensity as possibilities. These studies suggested that the neutralization of Japanese word-final pitch accents is incomplete, and that it is restricted by speakers and circumstances. However, Uwano's study was originally aimed at dialectal comparisons, and some of his subjects were not Tokyo native speakers. Neustupny's own experiment was limited to a single speaker since his study was

more focused on perception of pitch accent than on production. In addition, the results only showed that listeners tend to identify all isolated tokens as accented. As almost all other traditional studies, neither of these studies made explicit predictions about what F0 will do during any of the tones assigned by traditional theory, or provided acoustic measurements with instrumental methods.

Several more recent approaches to Japanese pitch accent have employed instrumental methods and dealt with the F0 contour itself. A study by Sugito (1982), which is one of the pioneering and significant works on the issue, investigated F0 extensively, and observed in acoustic measurements that some subjects (three out of 14) could make a clear distinction between accented /haná/ and unaccented /hana/ in isolation. In those cases, maximum F0 on the vowel of the second mora of accented /haná/ is slightly higher than that of unaccented /hana/. Sugito also conducted perception tests and found that even the subjects who made a clear distinction in production could not tell the difference between /haná/ and /hana/ in isolation, and that more errors were made for the perception of unaccented /hana/ than for the perception of /haná/. Unaccented /hana/ is recognized as accented /haná/ when the magnitude of rise from minimum F0 on a vowel of the first mora to maximum F0 on a vowel of the second mora is greater; if the magnitude of rise is relatively small, unaccented /hana/ is perceived as unaccented. Sugito therefore concluded that although maximum F0 on a vowel of the second mora is distinctive between accented and unaccented syllables acoustically in some speakers, it is the magnitude of rise in F0 that is more relevant to the acoustic and perceptual distinction. Sugito's study is significant in

that it provided explicit F0 values with a large number of speakers and synthesized sounds in one of the perception tests to examine the cues for perception. However, the author's conclusion on the issue of neutralization in the production of final-accented and -unaccented words is questionable. Sugito claimed that some speakers could make a distinction even when words were produced in isolation and that neutralization is speaker-dependent. It should be noted here that all of the speakers who clearly maintained a distinction were professional newscasters, and it is not implausible to suppose that such speakers are likely to produce careful, precise speech.

Vance (1995) corroborated Sugito's (1982) study by employing not only disyllabic but also monosyllabic words as stimuli. In preliminary tests, Vance compared four speakers in production and found that one speaker made a clear distinction. In a perception test with 40 listeners, some of the subjects could not perceive the difference between final-accented words and final-unaccented ones even in a carrier sentence (20 subjects for a minimal pair of /é/ and /e/, 8 subjects for a minimal pair of /ná/ and /na/, 14 subjects for a minimal pair of /hashí/ and /hashi/, and 36 subjects for a minimal pair of /kakí/ and /kaki/), and most of them could not distinguish final accent from no accent in isolated words. In follow-up experiments, two subjects, who also participated in the production test, were compared both in production and perception. Acoustically, there were significant differences in maximum F0 between accented and unaccented tokens in both monosyllables and disyllables for Speaker 2 but not for Speaker 1 when the words were produced in isolation. Neither speaker maintained a distinction between

accented and unaccented forms in terms of the minimum F0 on the first mora. On the other hand, Speaker 2 maintained a clear distinction between the maximum F0 on the second mora in disyllabic stimuli and on the first (and only) mora in monosyllabic stimuli with higher pitch for accented words. In a perception test, both speakers listened to their own and each other's productions. Speaker 2 showed high accuracy for both monosyllabic and disyllabic tokens for her own speech and performed above chance level for Speaker 1's tokens. Speaker 1 could distinguish only the disyllabic words produced by Speaker 2. With these data, Vance suggested that Speaker 1 and Speaker 2 might not rely on the same perception cue; some listeners may respond to magnitude of rise as Sugito (1982) claimed while others may respond to pitch-contour, amplitude, or vowel quality.

Vance's study, however, is inconclusive on the issue of whether or not final pitch accent and no accent are neutralized word-finally in Japanese. The author suggested individual variation of the sort that Sugito (1982) indicated as one of the possible explanations for the difference in production; some Tokyo native speakers may partially maintain a word-final distinction while others do not. However, the number of subjects Vance used was limited, and only one speaker out of four made a distinction. In addition the author provided F0 values for only two minimal pairs produced in isolation, namely, /haná/ and /hana/ and /kí/ and /ki/.

Vance questioned whether this speaker maintained the distinction in a sentential context. Although he recorded minimal pairs in carrier sentences, Vance did not provide the F0 values for those tokens so that it is not clear if and

how F0 of those words in isolation changes when pronounced in a sentence, or if it affects listeners' perception.

The author also made reference to the speakers' dialects as another possible explanation for the data; Speaker 1 was raised in Suginami Ward, a part of western Tokyo proper, and in Mitaka City, a suburb just to the west of Suginami Ward while Speaker 2 was raised in Katsushika Ward, a part of eastern Tokyo proper. According to the author, speakers from the peripheral Tokyo areas may be influenced by the neighboring prefecture, Chiba, where Kato (1970) claimed that an accent distinction between the isolation forms of /haná/ and /hana/ is maintained in several locations. However, considerable research on dialects in Japan (e.g. Kindaichi, 1981; Uwano, 1989, Nihon Hoso Kyokai 1998) recognizes the Chiba prefecture and those peripheral areas as Tokyo standard Japanese-speaking areas.

In the present study, two experiments are reported. The original motivation for this work was to explore Sugito's (1982) and Vance's (1995) findings. The first experiment consists of acoustic measurements to examine if Tokyo native speakers distinguish accented and unaccented tokens in isolation forms and in words produced in carrier sentences with a particle. The second experiment is a perception test designed to analyze how accurately subjects can identify accented and unaccented words in both isolation forms and tokens extracted from sentences. Tokens extracted from a carrier sentence have not been extensively studied in previous research on either production or perception.

Chapter 3
Acoustic study

3.1. Speakers

Eight college-educated Japanese women ranging in age from 22-47 years who were born and raised in Tokyo served as speakers. None of them had any known speech or hearing disorders.

3.2. Stimuli

3.2.1. Words in isolation and words in context

In the present paper, the expressions "words in isolation" and "words in context" are often used. Considerable literature have argued that the underlying distinction between final accent and no accent for nouns with a short final syllable emerges when words are followed by some grammatical units such as a particle and a copular within the same prosodic phrase. The words in these circumstances are often called "words in context". Otherwise, words with underlying final accent and words with no final accent have the same F0 pattern within the words themselves. In the studies on neutralization of word-final pitch accent and non-accent, words pronounced alone without any words or phrases before and after were employed to examine the difference in F0 pattern between words with underlying accent on the final mora and with no accent on the final mora. They are called "words in isolation" contrasted to "words in context".

The definition of the "context" has been ambiguous in most previous studies. It is generally implicated as simply "a carrier sentence" or "a grammatical particle". However, in these previous studies, it was not clear if the underlying

distinction is neutralized when words are produced within the carrier sentence but without any grammatical units such as a particle or a copular following right after the words. In Japanese, grammatical particles such as *ga* (Nominative) and o (Accusative) are often omitted in conversation. Little arguments have been made about if words pronounced by dropping a particle in a sentence are considered to be produced "in isolation" or "in context".

However, these words without a grammatical particle or a copular appear more in the middle of the sentence where a particle or a copular is dropped (e.g. akai hana kaita. 'I draw a red flower/nose.') than at the end of the sentence where a copular is dropped (e.g. watashi-ga kaita-noha akai hana. 'What I draw (was) a red flower/nose.').

Figures 3, 4, and 5 show the prosodic and tonal structure of the utterances with targeted words in the middle of the sentence without a particle after the word based on models of the Pierrehumbert & Beckman and the J_ToBI.

First, Figure 3 is the example of the sentence in which the word after the target word, namely the verb *kaita,* the past tense of the verb *kaku* 'to write', has the pitch accent on the first mora [káita].

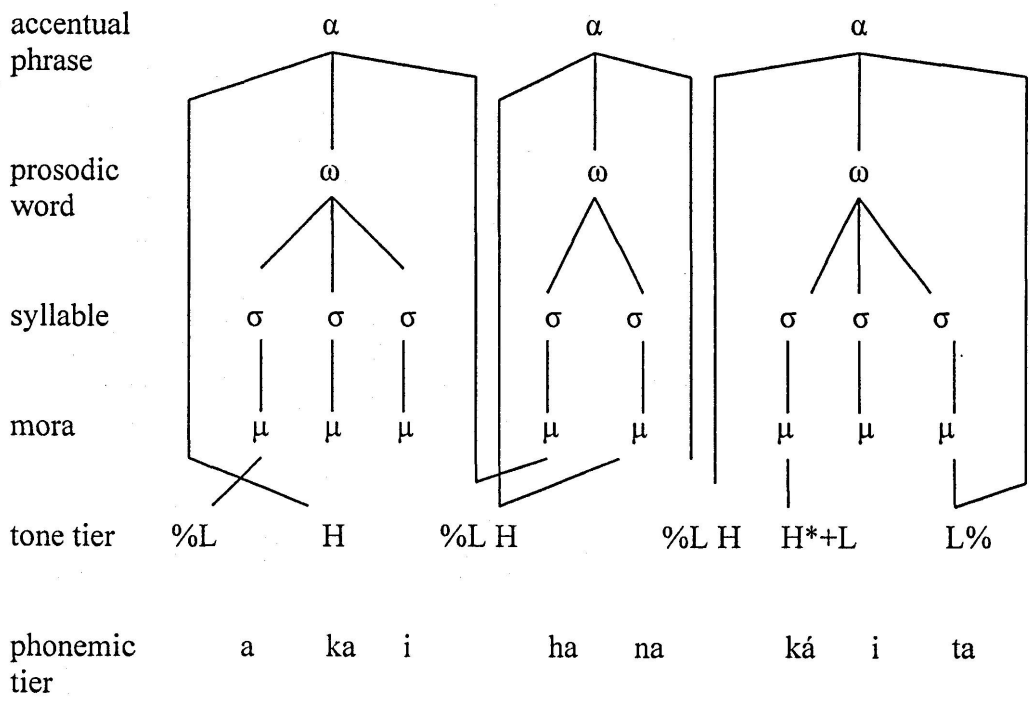

Figure 3. Prosodic and tonal structure of an utterance [akai haná kaita] 'I draw a red flower' based on models of the Pierrehumbert & Beckman's framework and the J_ToBI.

It should be noted here that the underlying accented /ná/ in /haná/ is pronounced without a pitch accent, and that the utterance with underlying final-unaccented word /hana/ 'nose' has the same structure and F0 patterns as in Figure 3. In summary, these two /haná/ and /hana/ are neutralized in these carrier sentences.[3]

Figure 4 is again the example of the sentence having the target word with no particle in the middle of the sentence. However, the word after the target word, namely the verb *funda,* the past tense of the verb *fumu* 'to step on', has no pitch accent on the first mora [funda].

[3] The author asked four native Tokyo speakers how to pronounce those two. They made no distinction in accent patterns for these two sentences. They made distinction in pitch patterns on

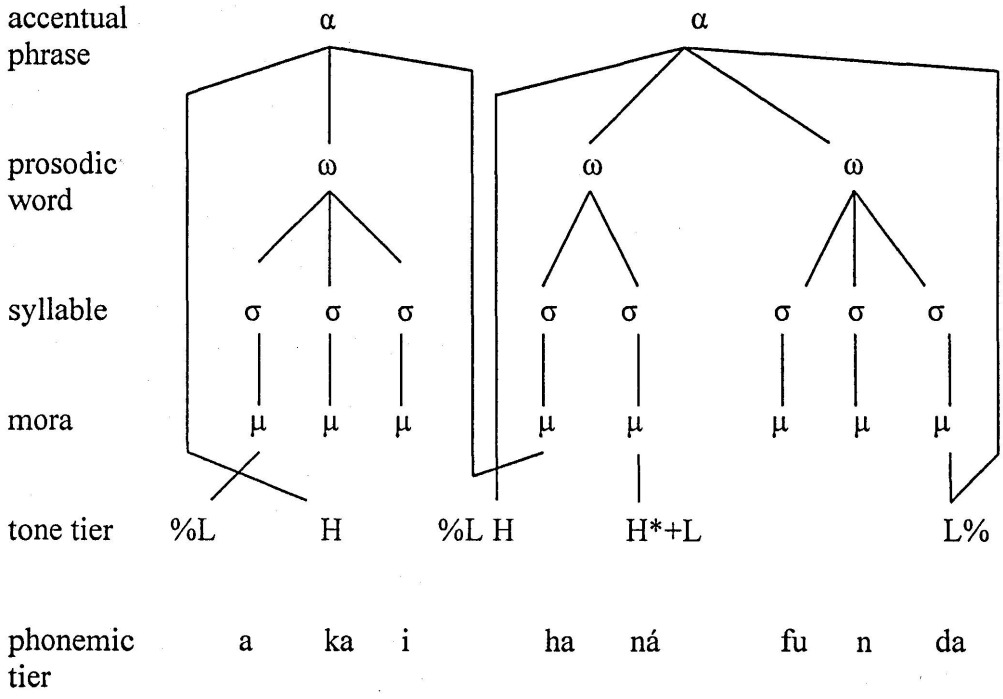

Figure 4. Prosodic and tonal structure of an utterance [akai haná funda] 'I stepped on a red flower' based on models of the Pierrehumbert & Beckman's framework and the J_ToBI.

Figure 4 shows that the underlying accent on the second mora of the word /haná/ appears on the surface in this carrier sentence.

On the other hand, Figure 5 illustrates that the second mora of the word /hana/ has no pitch accent and the underlying difference is clearly maintained in this carrier sentence.

---

the second mora of the accented /haná/ and unaccented /hana/ for the utterance [akai haná funda] and [akai hana funda], respectively.

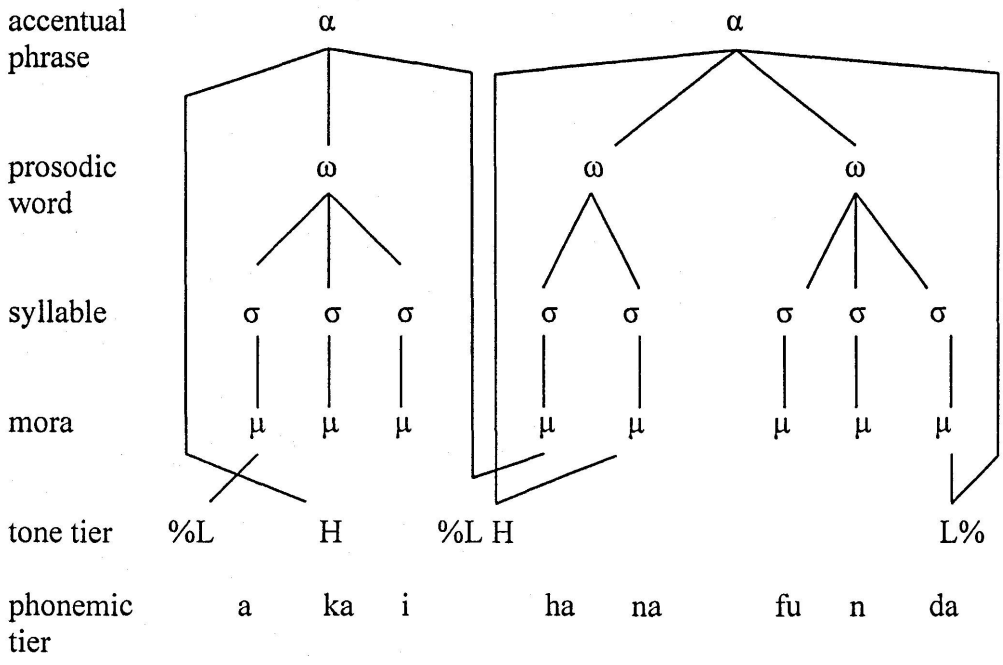

Figure 5. Prosodic and tonal structure of an utterance [akai hana funda] 'I stepped on a nose' based on models of the Pierrehumbert & Beckman's framework and the J_ToBI.

In summary, it is plausible that the effect of the context is actually the effect of the accentual phrase (AP) boundary. If the targeted word is placed at AP boundary, it is considered to be pronounced in 'isolation'. On the other hand, when the target word is not placed at the boundary even if it is immediately followed by a particle or a copular, it is considered to be placed in 'context'.

Words pronounced at the end of the utterance are also placed at AP-finally, and therefore, they are regarded the same as being produced in isolation.

In the present study, these utterances in which a grammatical particle or a copular is omitted are not introduced. Words in isolation are the stimuli pronounced by themselves and not placed in a sentence or a phrase. Words in context are the stimuli produced with a grammatical particle after the target word

in a carrier sentence so that the final mora of the target word is not placed at the AP boundary.

3.1.2. Data sets

Four minimal pairs of monosyllabic words and four minimal pairs of disyllabic words listed in Table 4 were chosen for recording with an additional four monosyllabic and four disyllabic words as fillers. All 24 words are nouns. The two words in each pair differ only in that, according to a standard accent dictionary (Nihon Hoso Kyokai, 1998), one has final accent while the other is unaccented. These 24 words were recorded in isolation.

Final-accented word	Gloss	Unaccented word	Gloss
Monosyllabic			
/kí/	"tree"	/ki/	"spirit"
/é/	"picture"	/e/	"handle"
/hí/	"fire"	/hi/	"day"
/ná/	"green"	/na/	"name"
Disyllabic			
/haná/	"flower"	/hana/	"nose"
/hashí/	"bridge"	/hashi/	"edge"
/kakí/	"fence"	/kaki/	"persimmon"
/murá/	"village"	/mura/	"unevenness"

Table 4. Minimal pairs used as stimuli for recording

The same words were also recorded in a simple carrier sentence, "koko-ni___ ga arimasu" (here is ———), containing the grammatical particle "ga" followed by the predicate "arimasu". As noted above, the difference between final accent and no accent is supposed to be typically realized when followed by a grammatical particle. A list of 24 sentences was then prepared for recording.

3.3. Procedure

Five repetitions of each word were randomized and presented in Japanese kanji characters (e.g. 花 /haná/ *'flower'* and 鼻/hana/ *'nose'*) on a computer screen after a voice prompt recorded by a Tokyo speaker saying "kore-wa nandesuka?" (what is this?). Each speaker was instructed to read each word aloud. Next, after a short break, speakers were instructed to read 5 repetitions of 24 randomized sentences, following prompts consisting of a recorded instruction "koko-ni nani-ga arimasuka?" ("what is here?") and each sentence written in ordinary Japanese orthography with kanji and hiragana (e.g. ここに花があります。/koko-ni haná-ga arimasu/ *'here is a flower'* and ここに鼻があります。/koko-ni hana-ga arimasu/ *'here is a nose'*) on the computer screen. Recordings were made in the KU Phonetics and Psycholinguistics Laboratory (KUPPL) using a cardioid microphone (Optimus) and high-quality cassette recorder (Marantz PMD221). Before recording, speakers practiced reading a few randomly chosen test words and sentences to familiarize themselves with the materials. Materials were read at a comfortable speed with 1000ms ISI throughout the recording sessions.

3.4. Analysis

All recordings were digitized onto a PC using the speech analysis program Praat at a sampling rate of 22050 Hz with 16-bit resolution. The words in sentence context were extracted with the particle /ga/ from the carrier sentence by examining waveforms and spectrograms. For monosyllabic words, the maximum F0 of the vowel was measured. For disyllabic words in isolation, the minimal F0

of the vowel of the first mora and the maximum F0 of the vowel of the second mora were measured. For both monosyllabic words and disyllabic words in context, the maximum F0 of the vowel in /ga/ was also measured.

3.5. Results

3.5.1. Words in isolation

Table 5 shows the mean values of maximum F0 on the vowel of the first mora of monosyllabic words for each speaker.

	accented		unaccented
	1st mora		1st mora
speaker 1	221	speaker 1	226
2	214	2	214
3	225	3	224
4	220	4	223
5	234	5	235
6	217	6	217
7	231	7	231
8	235	8	236
Mean	225	Mean	226

Table 5. Maximum F0 (Hz) on the vowel of the first mora of 4 minimal pairs of monosyllabic tokens for each speaker averaged across 5 repetitions.

Mean F0 of all four accented tokens is 225 Hz and that of all four unaccented tokens is 226 Hz. There is no significant difference between those two values [$F(1,318)=.001, p>.978$].

Table 6 illustrates the mean values of minimum F0 on the vowel of the first mora and maximum F0 on the vowel of the second mora of disyllabic words for each speaker.

	accented			unaccented	
	1st mora	2nd mora		1st mora	2nd mora
speaker 1	205	221	speaker 1	207	222
2	190	216	2	191	214
3	199	220	3	199	220
4	213	232	4	214	233
5	200	225	5	199	225
6	187	216	6	186	216
7	204	226	7	204	226
8	200	220	8	200	221
Mean	200	222	Mean	200	222

Table 6. Mean values of minimum F0 (Hz) on the vowel of the first mora and maximum F0 (Hz) on the vowel of the second mora of the minimal pairs of disyllabic words for each speaker averaged across 5 repetitions.

Mean minimum F0 of all four accented tokens is 200 Hz and that of all four unaccented tokens is 200 Hz. There is no significant difference between those two values [$F(1,318)=.022, p>.882$]. Mean maximum F0 of all four accented tokens is 222 Hz and that of all of four unaccented tokens is 222 Hz. Again, there is no significant difference between those two values [$F(1,318)=.027, p>.869$].

Table 7 shows the results of the analysis of variance (ANOVA) to examine if there is a significant difference between two pitch accents (accented and unaccented) on the vowel of the first mora of monosyllabic tokens for each speaker.

	F(1,38)	sig.
Speaker 1	1.078	$p>0.306$
2	0.448	$p>0.507$
3	0.368	$p>0.500$
4	0.115	$p>0.737$
5	0.000	$p>0.992$
6	0.010	$p>0.921$
7	0.010	$p>0.923$
8	0.003	$p>0.957$

Table 7. Results of ANOVA for differences between maximum F0 on the first mora of monosyllabic tokens in isolation with each speaker as a factor.

The results are consistent for each speaker; no significant differences are found for the values of maximum F0 between accented and unaccented tokens.

Table 8 illustrates the results of analysis of variance to analyze the speaker variances for minimum F0 on the first mora and maximum F0 on the second mora of disyllabic tokens in isolation.

	F(1,38)	sig.	F(1,38)	sig.
	1st mora		2nd mora	
Speaker 1	0.006	$p>0.941$	0.005	$p>0.943$
2	0.208	$p>0.651$	0.438	$p>0.512$
3	0.008	$p>0.928$	0.008	$p>0.931$
4	0.058	$p>0.811$	0.030	$p>0.864$
5	0.578	$p>0.452$	0.000	$p>0.999$
6	0.067	$p>0.798$	0.003	$p>0.957$
7	0.019	$p>0.890$	0.018	$p>0.895$
8	0.001	$p>0.977$	0.004	$p>0.952$

Table 8. Results of ANOVA for differences between minimum F0 on the first and maximum F0 on the second mora of disyllabic tokens in isolation with each speaker as a factor.

Table 8 again indicates that there are no significant differences of F0 on both first and second mora between accented and unaccented disyllabic words for all speakers.

3.5.2 Words in context

Measurements of monosyllabic stimuli are shown in Table 9.

Mean maximum F0 of accented tokens is 264 Hz and that of unaccented tokens is 197 Hz. There is a statistically significant difference between those two values [$F(1,318)=1512.099, p<.001$].

The data for F0 values on the vowel of the second mora, namely the vowel of the grammatical particle /ga/, have not been explicitly reported in previous

studies. Mean maximum F0 for the *ga* particle of accented tokens is 208 Hz and that of unaccented tokens is 223 Hz. The difference between those two values is significant [$F(1,318)=90.107, p<.001$].

	accented		unaccented	
	1st mora	ga	1st mora	ga
speaker 1	267	214	205	225
2	243	189	180	206
3	275	224	215	236
4	271	220	199	231
5	269	208	204	226
6	250	190	186	208
7	267	205	195	228
8	273	210	196	222
Mean	264	208	197	223

Table 9. Mean values of maximum F0 (Hz) on the vowel of the first mora and maximum F0 (Hz) on the vowel of the second mora in 4 minimal pairs of monosyllabic tokens for each speaker averaged across 5 repetitions.

Measurements of disyllabic stimuli are shown in Table 10.

	accented			unaccented		
	1st mora	2nd mora	ga	1st mora	2nd mora	ga
speaker 1	203	257	206	205	233	225
2	193	238	179	192	212	204
3	218	272	215	217	245	234
4	211	264	218	208	232	229
5	205	276	214	205	229	229
6	192	243	180	192	217	206
7	216	270	220	216	234	234
8	200	266	213	199	227	231
Mean	205	260	205	204	229	224

Table 10. Minimum F0 (Hz) on the vowel of the first mora, maximum F0 (Hz) on the vowel of the second mora, and maximum F0 (Hz) on the vowel of the third mora of 4 minimal pairs of disyllabic tokens for each speaker averaged across 5 repetitions.

Mean minimum F0 for accented tokens is 205 Hz and that for unaccented tokens is 204 Hz. There is no significant difference between those two values [$F(1,318)=.139, p>.709$].

For the second mora, mean maximum F0 of accented tokens is 260 Hz and that of unaccented tokens is 229 Hz. The difference between those two values is significant [$F(1,318)=406.311, p<.001$].

Mean maximum F0 on the third mora (ga) of accented tokens is 205 Hz and that of unaccented tokens is 224 Hz. The difference between those two values is significant [$F(1,318)=103.766, p<.001$].

Table 11 shows the results of the analysis of variance (ANOVA) to examine if there is significant difference between two pitch accents (accented and unaccented) on the vowel of the first mora and also on the vowel of the second mora, namely a grammatical particle, of monosyllabic tokens in context for each speaker.

	F(1,38)	sig.	F(1,38)	sig.
	1st mora		ga	
Speaker 1	373.013	$p=.000$	16.616	$p=.000$
2	436.487	$p=.000$	16.233	$p=.000$
3	162.65	$p=.000$	8.674	$p<.005$
4	446.59	$p=.000$	14.356	$p<.001$
5	453.249	$p=.000$	68.3	$p=.000$
6	345.012	$p=.000$	47.814	$p=.000$
7	515.291	$p=.000$	67.295	$p=.000$
8	314.401	$p=.000$	86.36	$p=.000$

Table 11. Results of ANOVA for differences between maximum F0 on the first and maximum F0 on the second mora of monosyllabic tokens in context with each speaker as a factor.

The results are consistent for each speaker; significant differences are found for the values of maximum F0 on both the vowels of first and second morae between accented and unaccented tokens for each speaker.

Table 12 illustrates the results of analysis of variance to analyze the speaker variances for minimum F0 on the first mora, maximum F0 on the second mora

and maximum F0 on the third mora, namely a grammatical particle, of disyllabic tokens in context.

Table 12 indicates that there is no significant difference of minimum F0 on the first mora between accented and unaccented disyllabic tokens in context for each speaker. However, there are significant differences of maximum F0 on the second and third morae for each speaker.

	F(1,38)	sig.	F(1,38)	sig.	F(1,38)	sig.
	1st mora		2nd mora		ga	
Speaker 1	0.064	*p*>.802	60.117	p=.000	15.863	p=.000
2	0.168	*p*>.684	73.684	p=.000	41.755	p=.000
3	0.207	*p*>.652	75.499	p=.000	68.011	p=.000
4	0.636	*p*>.430	68.456	p=.000	14.84	p=.000
5	0.051	*p*>.822	302.488	p=.000	79.088	p=.000
6	0.055	*p*>.815	114.729	p=.000	98.399	p=.000
7	0.093	*p*>.762	380.455	p=.000	59.17	p=.000
8	0.275	*p*>.603	466.597	p=.000	80.748	p=.000

Table 12. Results of ANOVA for differences between minimum F0 on the first, and maximum F0 on the second and third mora of disyllabic tokens in context with each speaker as a factor.

3.6. Discussion

The data reported above demonstrate some new important findings which are different from previous studies. First, for the words in isolation, there are no statistically significant differences in terms of F0 between accented and unaccented words. This is true for both monosyllabic and disyllabic tokens, and the results are consistent across all speakers. It should be concluded that the distinction between final accented and unaccented words is neutralized when they are uttered in isolation. Unlike previous studies (e.g.Uwano, 1977; Neustupny,

1978; Sugito, 1982; and Vance, 1995), speaker variance was not found in the present study. All eight subjects showed consistent neutralization for all tokens.

On the other hand, the results for words embedded in a carrier sentence suggest that the underlying distinctive pitch patterns are preserved when words are spoken in a context followed by a grammatical particle. For the maximum F0 on the vowel of the first mora in monosyllabic words and on the vowel of the second mora in disyllabic words, there are significant differences between underlying distinctive pitch patterns (67 Hz for disyllabic tokens and 31 Hz for monosyllabic tokens). The minimum F0 on the vowel of the first mora of the disyllabic words is not significantly different for accented and unaccented tokens. F0 on the vowel of the second mora in accented words rises abruptly from the first mora while F0 on the vowel of the second mora in unaccented words shows a much smaller rise. This result suggests that the phonemic distinctions are maintained in words in context followed by a particle.

As mentioned above, previous research has not reported acoustic measurements of the grammatical particle itself. The present results indicate a significant difference in F0 on the vowel of the grammatical particle following accented versus unaccented tokens. For both monosyllabic and disyllabic words, F0 on the vowel of the grammatical particle is much lower in accented than in unaccented tokens (15 Hz higher for the vowel on the grammatical particle of unaccented monosyllabic tokens and 19 Hz higher for the vowel on the particle of unaccented disyllabic tokens).

Chapter 4
Perception study

Acoustic analysis established that there was no phonetic difference in terms of F0 between words in isolation which are underlyingly accented word-finally and words which are underlyingly unaccented word-finally. While no such differences were found in the present study, it is possible that underlying distinctions might be preserved through other phonetic parameters (e.g. amplitude, vowel quality, pitch contour). In order to investigate this possibility, a perception experiment was conducted. Both tokens originally produced in isolation and in context were included in the experiment.

4.1. Subjects

Twelve native Japanese listeners (9 female, 3 male) who were born and raised in Tokyo or Kanto area (suburb area of Tokyo) were selected from the KU student population. None of the listeners had any known hearing disorders.

4.2. Materials and procedure

For words in isolation, all five productions of the four minimal monosyllabic pairs, and the four minimal disyllabic pairs, produced by eight speakers in the acoustic study (see Table 1), were used in the perception experiment. The same words had originally also been produced in a sentential context. For the perception experiment, these words were extracted without a grammatical particle from the context using Praat software, and were provided to listeners.

In order not to make the perception experiment too long, the stimuli were divided into two tests. One consisted of the monosyllabic tokens /kí/ & /ki/ and /é/ & /e/, and the disyllabic tokens, /haná/ & /hana/ and /hashí/ & /hashi/. Test 1 consisted of 8 test blocks, and each block consisted of (1) accented /kí/ and unaccented /ki/ in isolation, (2) accented /é/ and unaccented /e/ in isolation, (3) accented /haná/ and unaccented /hana/ in isolation, (4) accented /hashí/ and unaccented /hashi/ in isolation, (5) accented /kí/ and unaccented /ki/ from a context, (6) accented /é/ and unaccented /e/ from a context, (7) accented /haná/ and unaccented /hana/ from a context, and (8) accented /hashí/ and unaccented /hashi/ from a context. Subjects took a short break after block 4.

Test 2 was organized in the same way as Test 1 and included the monosyllabic tokens /hí/ & /hi/ and /ná/ & /na/, and the disyllabic tokens, /kakí/ & /kaki/ and /murá/ & /mura/.

Each test consisted of eight blocks of 40 stimuli (5 productions of 2 tokens of one minimal pair, accented and unaccented, by 4 speakers). Six subjects took Test 1 and another group of six subjects took Test 2.

Using the subject-testing software package SuperLab, listeners were presented with five productions of each word in randomized order. Two Japanese kanji characters representing two words in each minimal pair were shown at the same time on the computer screen for underlyingly accented tokens (1) and for unaccented tokens (2). For example, the first block of Test 1 has prompt kanji characters on screen as follows; (1) 木 'tree' and (2) 気 'spirit'. Listeners were to indicate which word they perceived, accented or unaccented, by

pressing one of two numeric buttons, 1 or 2, respectively. Stimuli were randomized separately for each subject with a 1000ms ISI.

4.3. Results

Results of the perception experiment are shown in Table 13.

	ISOLATION			CONTEXT		
	Accented	Unaccented	**Mean**	Accented	Unaccented	**Mean**
Monosyllabic	53	49	51	82	80	81
Disyllabic	49	48	49	63	63	63
Mean	51	49	50	73	72	72

Table 13. Correct identification (%) of monosyllabic and disyllabic words as a function of accent and context.

First, perception of both accented and unaccented monosyllabic words in isolation is not significantly different from chance [$t(11)=1.429$ $p>.181$] and [$t(11)=-.679$ $p>.511$], respectively (mean 51% correct identification). On the other hand, perception of these monosyllabic tokens in context is significantly better than chance level. The accuracy for accented words and unaccented words is above chance level [$t(11)=14.099$ $p<.001$] and [$t(11)=31.739$ $p<.001$], respectively (mean 81% correct identification).

The data from the disyllabic stimuli are very similar: perception of both accented and unaccented tokens in isolation is not significantly different from chance [$t(11)=-.584$ $p>.571$] and [$t(11)=-2.075$ $p>.062$], respectively (mean 49% correct identification). In addition, perception of both accented and unaccented words in context is significantly better than chance [$t(11)=7.826$ $p<.001$] and [$t(11)=6.060$ $p<.001$], respectively (mean 63% correct identification).

4.4. Discussion

The perception results indicate that the distinction between word-final accented and unaccented morae is completely neutralized in Japanese. That is, listeners were unable to reliably distinguish the differences between accented words and unaccented words in isolation. The listeners performed at chance level in their identification of the distinctive underlying word-final pitch accents when words were produced in isolation. On the other hand, the subjects remarkably improved their accuracy of perception for both monosyllabic and disyllabic tokens produced in context. The analysis of variance indicates that words produced in context are perceived significantly more accurately than words in isolation in both disyllabic and monosyllabic tokens ([$F(1,46)=-234.410$ $p<.001$] for monosyllabic tokens; [$F(1,46)=79.426$ $p<.001$] for disyllabic tokens.) No previous studies have investigated the perception of words in context. The present results clearly indicate that listeners can perform above chance level in their identification of word-final pitch accent differences in context. However, they cannot perceive those underlying differences in words produced in isolation. These results, thus, support the idea of complete neutralization in both production and perception (referred to as Type A neutralization by Dinnsen, 1985), and challenge Dinnsen's (1985) claim that this classic type of neutralization is "unfortunately without empirical support".

Chapter 5
General discussion and conclusions

5.1. General Discussion

Acoustic data show that Tokyo standard Japanese speakers do not distinguish final accent and non-accent in either monosyllabic or disyllabic words in isolation (for monosyllabic tokens, mean 225 Hz on the vowel of the first mora in accented words and mean 226 Hz in unaccented words. for disyllabic tokens, mean 200 Hz on the vowel of the first mora in accented words and mean 200 Hz in unaccented tokens, and mean 222 Hz on the vowel of the second mora in accented words and mean 222 Hz in unaccented words.) For all tokens, the result is observed consistently in each speaker despite speaker variability in pitch (e.g., speakers 2 and 6 have a lower pitch than any other speakers).

On the other hand, F0 changes when the target word is uttered in a sentence, and a significant difference in pitch pattern between final accented and non-accented words emerges. For monosyllabic words in context, F0 is substantially higher (67 Hz higher) for accented words relative to unaccented words. For disyllabic words, while F0 of the first mora does not differ as a function of accent (205 Hz for accented tokens and 204 Hz for unaccented tokens), F0 of the second mora in accented tokens is increased strikingly compared to unaccented tokens (31 Hz higher). However, this distinction does not show up in isolation. As mentioned earlier, another result that should be noted here is that F0 on the vowel of the particle /ga/ is significantly lower in accented words compared to

unaccented words. This was true for both monosyllabic and disyllabic words (15 Hz lower for monosyllabic tokens and 19 Hz lower for disyllabic tokens).

These results may challenge the traditional theory (McCawley, 1977; Haraguchi, 1977; 1991) which assigns High and Low tone for each mora in a binary way. For example, the traditional theory describes final-accented and unaccented words as LHL and LHH, respectively. According to the traditional theory, the pitch accents on vowels of the second mora are the same, both H, and the difference appears only when a grammatical particle follows the word. However, the present study indicates that the F0 on the vowel of the second mora of an accented word in context is significantly higher than that of an unaccented word.

The analysis of Japanese tone structure introduced by Pierrehumbert & Beckman (1986, 1988) and the J_ToBI model put forth by Venditti (2000) may explain pitch contours in a way consistent with the current phonetic data. Figure 6 shows the waveform in the top panel, and the pitch contour in the middle panel for the final-accented /haná/ in a carrier sentence [koko-ni haná-ga arimasu]. The placement of the tones that are specified in Figure 1 (Chapter 2) has been superimposed on the pitch contour following the models by P&B and J-ToBI.

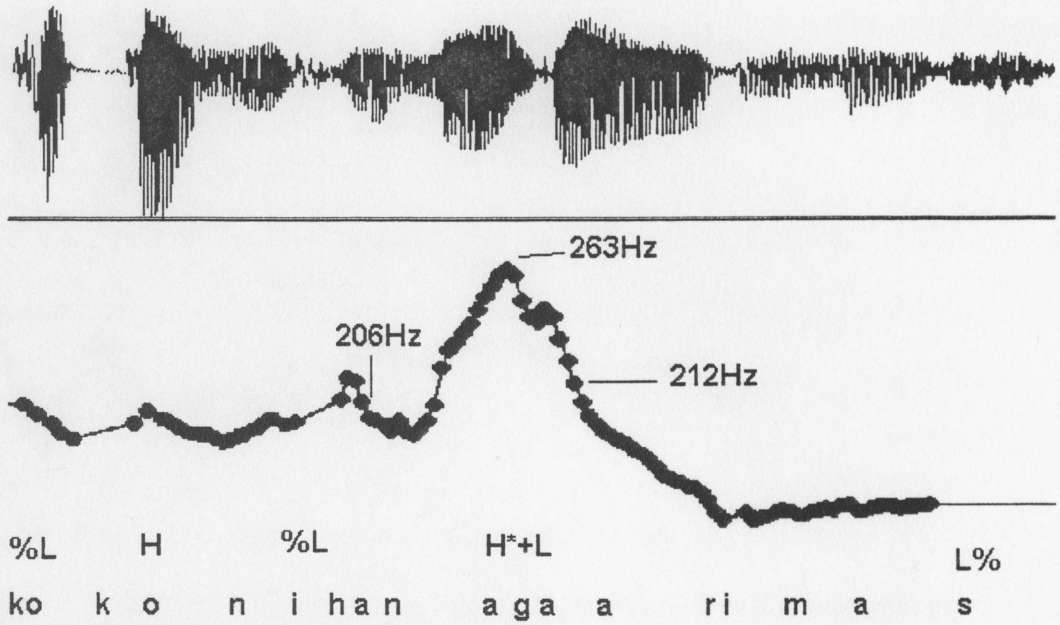

Figure 6. Wave form, F0 contour, and J_ToBI transcription of the accented /haná+ga/ 'flower+Nominal' in the context [koko-ni haná-ga arimasu] 'here is a flower' Speaker 1.

There are two accentual phrases (AP), [koko-ni] and [haná-ga arimasu]. In the first AP, the utterance-initial low tone (%L) is associated with the first mora, and the accentual phrasal high (H) is attached to the second mora. This AP does not have pitch accent. The AP final boundary low tone (%L) is placed at the phrasal edge, and also continuously links to the first mora of the second AP /ha/. This low pitch rises up abruptly to a higher pitch to the second mora and sharply falls to the third mora. The H*+L composite label is used to mark this sharp fall from a high tone, namely, lexical accent, in the accented AP, and is associated with the second mora. The utterance-final boundary low tone (L%) links to the mora before the last mora /s/ because the last mora is devoiced.

Figure 7 shows the waveform in the top panel, and the pitch contour in the middle panel for the final unaccented /hana/ in a carrier sentence [koko-ni hana-ga arimasu].

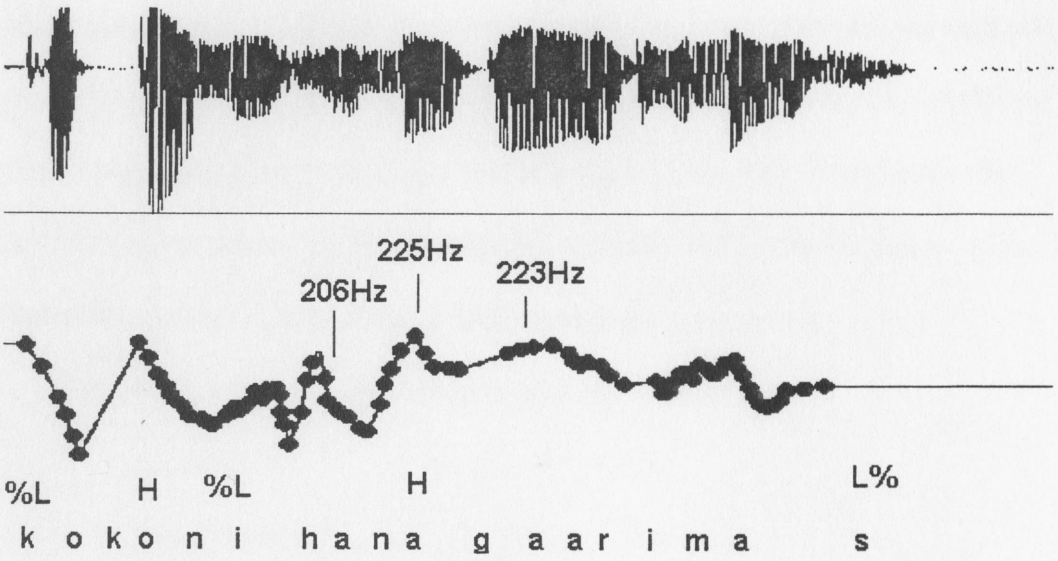

Figure 7. Waveform, F0 contour, and J_ToBI transcription of the unaccented /hana+ga/ 'nose+Nominal' in the context [koko-ni hana-ga arimasu] 'here is a nose' by Speaker 1.

The difference between the J_ToBI transcription in Figure 6 and Figure 7 is that in Figure 7 the second mora in the second AP is marked with the phrasal high (H) instead of the lexical pitch accent marker (H*+L). The unaccented mora does not have a sharp fall and therefore there is no low tone part of the high-low tone sequence between the second mora /na/ and the third mora /ga/. It is in accordance with the actual pitch contour and F0 data illustrated above[4].

At the word level, a pitch accent marked with high-low sequence indicating a sharp fall from a high tone, namely H*+L, links to a lexically specified mora. The unaccented word is associated with two tones, an accentual phrase-initial %L,

[4] Another phonetic factor included in the theories by Pierrehumbert & Beckman (1988) and Venditti (2000) appears in these current data sets. A final L% is at considerably higher F0 at the end of unaccented phrase than at the end of a phrase containing an accent. Accent H*+L tones trigger catathesis, which decreases the pitch range after the H*+L tone by lowering the high line. The present data in Figure 3 and Figure 4 show two effects of catathesis; first, as mentioned above, F0 has a higher value at the end of an unaccented phrase than at the end of an accented one. Second, the phrasal high in the second AP is slightly lower than the phrasal high in the first AP. On the other hand, the lexical accent has a higher F0 than the phrasal high in the first AP. It also indicates that the F0 of the accent H*+L is higher than that of the phrasal H.

and a phrasal high H. These tones come from the accentual phrasal level, and one or more words constitute an accentual phrase where the phrasal high tone (H) links to the second sonorant mora, and the boundary low tone (L%) links to the last mora of the phrase. The L% boundary tone also links to the first mora of an upcoming phrase when the first syllable is short and unaccented.

Thus, the complete tonal transcription of the APs is:

Accented AP	%L (H-) H* +L L%
Unaccented AP	%L H- L%

For instance, the pitch patterns in disyllabic accented /haná/ and unaccented /hana/, and accented /haná+ga/ and unaccented /hana+ga/ can be marked as follows[5]:

Accented AP	h	a	n	á			
	%L			H			
Unaccented AP	h	a	n	a			
	%L			H			
Accented AP	h	a	n	á	g	a	
	%L			H*L			
Unaccented AP	h	a	n	a	g	a	
	%L			H			

As illustrated above, accented /haná/ and unaccented /hana/ in isolation have the same contour, namely, %L and H. On the other hand, accented /haná+ga/ and unaccented /hana+ga/ in context may be described differently; accented /ná/ is associated with a high-low sequence H*+L which indicates a sharp fall from a

high tone, and unaccented /na/ has the phrasal high H without any falling contour. Venditti (2000) suggests that the position of the H*+L label will coincide with the location of the actual F0 maximum in many cases[6].

Figures 8 and 9 provide the waveform and pitch contour for the accented word /haná+ga/ and unaccented /hana+ga/ extracted from the context with J_ToBI transcription.

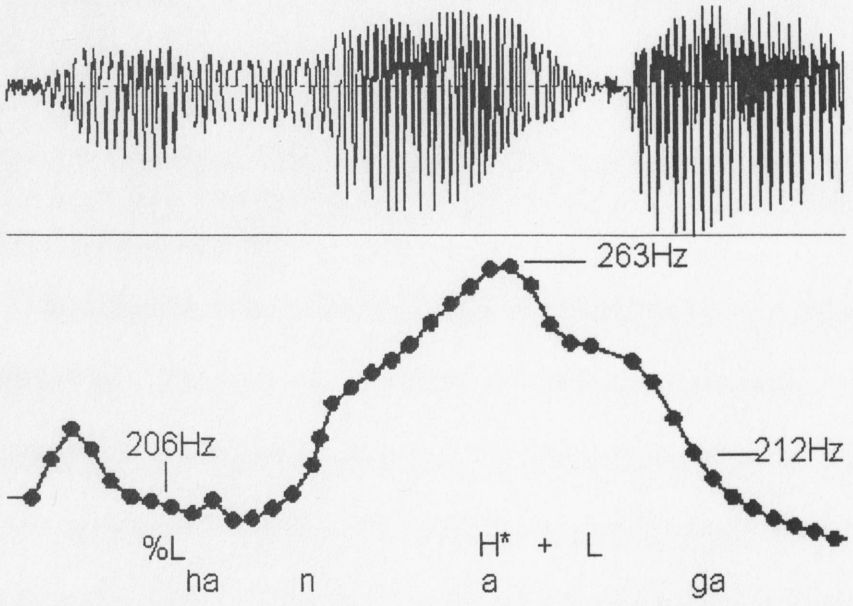

Figure 8. Waveform, F0 contour, and J_ToBI transcription of the accented /haná+ga/ 'flower+Nominal' extracted from the context [koko-ni hana-ga arimasu] 'here is a flower' by Speaker 1.

[5] Either isolated word tokens/haná/ & /hana/ or words with a particle /haná+ga/ & /hana+ga/ constitute the first half of the AP [hana+ga arimasu] as many accentual phrases (AP) consist of more than one words. Therefore, the AP-final boundary low tone, L%, is not described here.

[6] It is not uncommon for the peak to occur after the accented mora but still be perceived as occurring on the accented mora if the accented mora is a devoiced vowel (Sugito, 1981; Hata & Hasegawa, 1988; Kitahara, 2001).

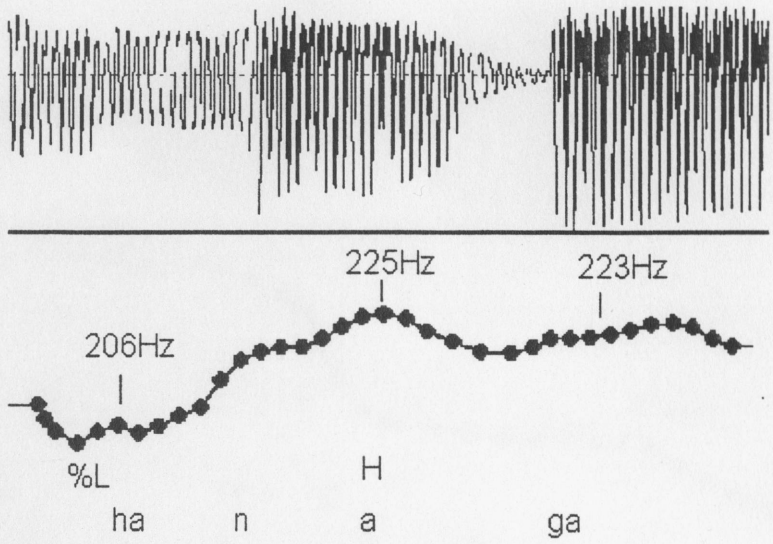

Figure 9. Waveform, F0 contour, and J_ToBI transcription of the unaccented /hana+ga/ 'nose+Nominal' extracted from the context [koko-ni hana-ga arimasu] 'here is a nose' by Speaker 1.

These figures indicate that the models by Pierrehumbert & Beckman (1988) and Venditti (2000) are more accurate to describe the pitch contour in standard Japanese. The difference in maximum F0 between second morae with a pitch accent and an unaccented phrasal high tone is given by two different markers, (high portion of) H*+L and H. The low portion of a high-low sequence H*+L, also differentiates maximum F0 on the vowel of the third mora of an accented token from that of an unaccented token; the accented token has a significant lower F0 on the third mora than the unaccented token.

The models accurately illustrate the pitch contour of monosyllabic tokens, too. Figures 10 and 11 are the waveform and the pitch tier with the J_ToBI transcription superimposed for the placement of tones for accented monosyllabic word /kí+ga/ and unaccented /ki+ga/ extracted from a carrier sentence, respectively.

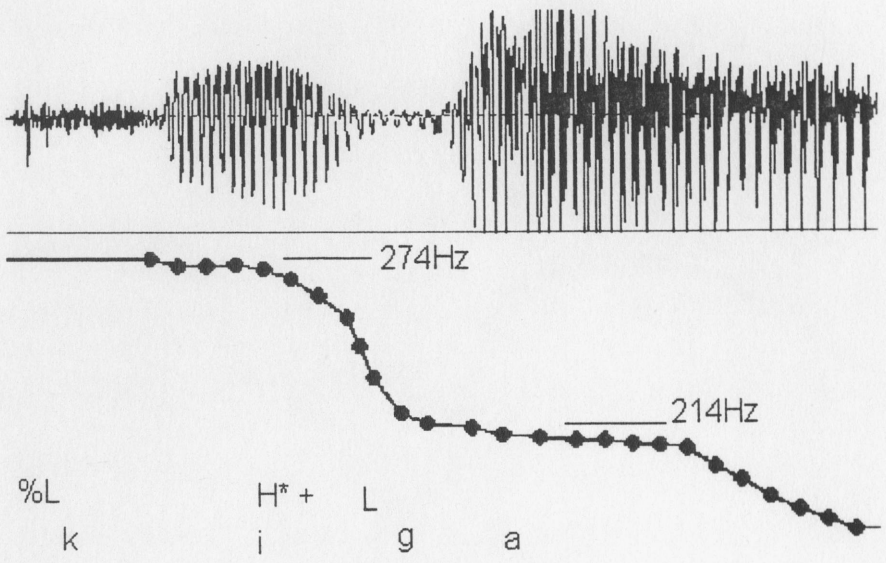

Figure 10. Waveform, F0 contour, and J_ToBI transcription of the accented /kí+ga/ 'tree+Nominal' extracted from the context [koko-ni kí-ga arimasu] 'here is a tree' by Speaker 1.

The accent H*+L tone on the first mora of this AP (the second AP in the carrier sentence) prevents the association of the AP initial boundary low tone (%L) to the first mora. Instead, this low tone actually links to the edge of the last mora of the first AP [koko-ni]. Then, the pitch abruptly rises up to the first mora of the second AP, namely /kí/ which is described by the high portion of H*+L tone sequence. The sharp fall from the high tone marked with L follows within this mora to the third mora and F0 stays low.

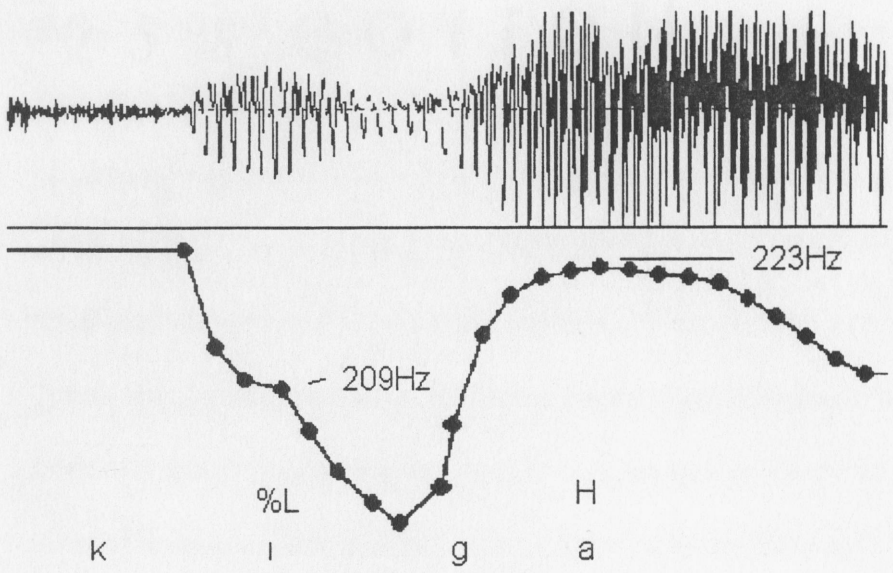

Figure 11. Waveform, F0 contour, and J_ToBI transcription of the unaccented /ki+ga/ 'spirit+Nominal' extracted from the context [koko-ni ki-ga arimasu] 'here is a spirit' by Speaker 1.

The boundary initial low tone (%L) is attached to the first mora of the second AP, /ki/, and since this AP does not have a pitch accent, the phrasal accentual high tone (H) is associated with the second mora of the word. The difference between pitch contours in Figures 10 and Figure 11 is that F0 on the first mora of accented /kí+ga/ has a very high tone and then sharply falls down on the way to the second mora. On the other hand, F0 on the first mora of unaccented /ki+ga/ starts with a low tone and rises gradually to the second mora. Therefore, the first mora of the accented /kí+ga/ has a falling contour within one syllable /kí/ wherease the first mora of unaccented /ki+ga/ shows a gradual rising contour within one syllable /ki/.

This fact may explain why subjects were more accurate in the perception of monosyllabic tokens than in the perception of disyllabic tokens (mean 81% correct identification for monosyllabic tokens and mean 63% correct

identification for disyllabic tokens). As the figures above illustrate, accented /kí/ has a high tone (high portion H* of H*+L) because of the pitch accent but unaccented /ki/ is a low tone (%L), not a phrasal high. Therefore, the difference between H*+L and L% is bigger than the difference between H*+L in accented /haná/ and the phrasal H in unaccented /hana/. F0 measurements reported in Tables 9 and 10 support this explanation; mean values of maximum F0 on the first mora of accented monosyllabic tokens are similar to those of F0 on the second mora of accented disyllabic tokens (mean 264 Hz on the first mora of monosyllabic tokens and mean 260 Hz on the second mora of disyllabic tokens), and mean values of maximum F0 on the first mora of unaccented monosyllabic tokens are much smaller than those on the second mora of unaccented disyllabic tokens (mean 197 Hz on the first mora of monosyllabic tokens and mean 229 Hz on the second mora of disyllabic tokens). Therefore, the difference on F0 on the vowel of the first mora between accented and unaccented monosyllabic tokens is much bigger than that on the vowel of the second mora between accented and unaccented disyllabic tokens (67Hz difference for monosyllabics and 31 Hz difference for disyllabics). Native speakers can tell the differences, and it may affect the results of the perception experiments.

F0 values on the grammatical particle *ga* of monosyllabic tokens and those values on the grammatical particle of disyllabic tokens are similar in both accented and unaccented cases (208 Hz for accented and 223 Hz for unaccented monosyllabic words, and 205 Hz for accented and 224 Hz for unaccented disyllabic words).

The present study, therefore, challenges the traditional phonological analysis, and instead, supports those theories (e.g., Beckman and Pierrehumbert,1986;1988, and Venditti, 2000) that are based on instrumental methods and explicit acoustic measurements and perceptual experiments, and thus include several purely phonetic factors.

5.2. Conclusions

This study investigated the acoustic and perceptual correlates of neutralization of pitch accent (F0) of word-final accented and unaccented words in Japanese. In an acoustic experiment, F0 values were measured to examine whether Tokyo standard Japanese speakers distinguished accented and unaccented words in isolation and in context. Findings suggest that they do not make any distinction in isolated words, but do produce clearly different pitch patterns in a sentence. In addition, those results were consistent across all speakers and tokens, and did not show the speaker variability that other studies had suggested.

A perception experiment was also conducted to determine whether listeners could distinguish two members of a minimal pair even when there was no clear surface distinction in pitch in isolated word tokens, and whether they could distinguish minimal pairs extracted from context. Subjects demonstrated above chance accuracy for tokens extracted from context but did not distinguish the tokens produced in isolation. Thus, the results showed complete neutralization for isolated words in perception as well as in production.

In summary, while most of the phonetic debate regarding neutralization has focused on the voicing distinction, the present results show that neutralization of

the word-final pitch accent distinction in Japanese is phonetically complete for isolated tokens.

In the present study, the utterances where a particle or a copular is omitted following the targeted word in a sentence are not focused. The context means a prosodic word/phrase, namely, AP. Therefore, even if the target word is embedded in a sentence, when a particle or a copular is omitted and the word is followed by a word with a pitch accent on the first mora, it is placed at the AP boundary and thus may be neutralized in word-final pitch accent. The underlying difference in a pitch accent may emerge on the surface when the target word without a particle or a copular is followed by the word with no accent on the first mora. The target word is placed in the middle of the AP, and thus is supposed to be in 'context'. In the future study, it will be meaningful contribution to investigate F0 on each mora of the underlying final-accented word and final-unaccented word in these environments and to examine the possible effects of word placement on the AP boundary.

Bibliography

Beckman, M.E. and J.B. Pierrehumbert (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* 3: 255-309.

Blumstein, S.E. (1991). The relation between phonetics and phonology, *Phonetica*, 48:108-19.

Charles-Luce, J. (1985). Word-final devoicing in German: Effects of phonetic and sentential contexts, *Journal of Phonetics*, 13: 309-324.

Charles-Luce, J. & Dinnsen, D. (1987). A reanalysis of Catalan devoicing, *Journal of Phonetics*, 15:187-190.

Dinnsen, D. (1985). A re-examination of phonological neutralization, *Journal of Linguistics*, 21:265-279.

Dinnsen, D.A. & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation, and individual differences, *Journal of Phonetics*, 12:49-60.

Fourakis, M. & Iverson, G. (1984). On the 'incomplete neutralization' of German final obstruents, *Phonetica*, 41: 140-149.

Haraguchi, S. (1977). *The tone pattern of Japanese: an autosegmental theory of tonology*. Tokyo: Kaitakusha.

Haraguchi, S. (1991). *A theory of stress and accent*. Foris: Dordrecht.

Higurashi, Y. (1983). *The accent of extended word structures in Tokyo standard Japanese*. Tokyo:Educa.

Jassem, W. & Richter, L. (1989). Neutralization of voicing in Polish obstrunets, *Journal of Phonetics*, 17:317-325.

Jongman, A., Sereno, J., Raaijmakers, M. & Lahiri, A. (1992). The phonological representation of [voice] in speech perception, *Language and Speech*, 35:137-152.

Kato, M. (1970). Henka suru kogai no kotoba: Tokyo no higashigawa [Suburban language change:east of Tokyo], *Gengo seikatsu*, 225:64-72.

Kim, H. & Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean, *Journal of Phonetics*, 24: 295-312.

Kindaichi, H. (1981). *Meikai Nihongo Akusento Jiten*. Sanseido, Tokyo, 2nd edition.

Kitahara, M. (2001). *Category structure and function of pitch accent in Tokyo Japanese*, Doctoral dissertation, University of Indiana.

Kubozono, H. (1993). *The organization of Japanese prosody*. Tokyo: Kuroshio.

Lahiri, A., Schriefers, H., & Kuijpers, C. (1987). Contextual neutralization of vowel length: Evidence from Dutch, *Phonetica*, 44, 91-102.

McCawley, J.D. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McCawely, J.D. (1977). Accent in Japanese. In *Studies in stress and accent* (L.M. Hyman, editor), pp. 261-302. Los Angels: University of Southern California Department of Linguistics.

Neustupny, J.V. (1978). *Post-structural approaches to language: language theory in a Japanese context*. Tokyo:Univeristy of Tokyo Press.

Nihon Hoso Kyokai (1998). *Nihongo hatsuon akusento jiten, kaitei shinpan* [Japanese pronunciation and accent dictionary, new revised edition]. Tokyo: Nihon Hoso.

Pierrehumbert, J. B. and Beckman, M.E. (1988). *Japanese Tone Structure*. Cambridge: MIT Press.

Port, R. & Crawford, P. (1989). Incomplete neutralization and pragmatics in German, *Journal of Phonetics*, 17:257-282.

Port, R. & O'Dell, M. (1985). Neutralization of syllable-final voicing in German, *Journal of Phonetics*, 13: 455-471.

Poser, W.J. (1984). *The phonetics and phonology of tone and intonation in Japanese*. MIT dissertation

Slowiaczek, L. & Dinnsen, D. (1985). On the neutralizating status of Polish word-final devoicing, *Journal of Phonetics*, 13:325-341

Sugito, M. (1982). Tokyo akusento ni okeru "hana" to "hana" no seisei to chikaku [The production and perception of "hana" and "hana" with Tokyo accent] In *Nihongo akusento no kenkyu*, pp. 182-201. Tokyo:Sanseido.

Uwano, Z. (1977). Nihongo no akusento [Japanese Accent]. In *Iwanami koza Nihongo* 5: *on in* [Iwanami course on Japanese 5: phonology] (S. Ono and T. Shibata, editors), pp. 281-321. Tokyo: Iwanami.

Uwano, Z. (1989). Nihongo no akusento [Japanese accent]. In Sugito, M., editor, *Kouza Nihongo to Nihongo Kyouiku [Lectures on Japanese and Japanese Language Education]* 2, 178-205. Meiji shoin, Tokyo.

Vance, T. J. (1987). *An introduction to Japnese phonology*. Albany:State University of New York Press

Vance, T. J. (1995). Final accent vs. no accent : utterance-fianl neutralization in Tokyo Japanese. *Journal of Phonetics* 23, 487-499.

Venditti, J.J. (1995). Japanese ToBI Labelling Guidelaines. Unpublished manuscript, Ohio State Univerisity. (Downloadable from: ling.ohio-state.edu/phonetics/J_ToBI/jtobi_homepage.html).

Venditti, J.J. (2000). The J_ToBI model of Japanese intonation. Unpublished manuscripts, Ohio State University. (Downloadable from: cs.rutgers.edu/~venditti/pubs/jtobichapt_wordNEWFIGS.doc)

Weitzman, R.S. (1970). Word accent in Japanese. *Studies in the phonology of Asian languages* 9. Acoustic phonetic research laboratory, University of Southern California.