

# Kansas Working Papers in Linguistics

edited by

Stacey Stowers  
Nathan Poell

Volume 26  
2002

Partial funding for this journal is provided by the  
Graduate and Professional Association of the University of Kansas

ISSN 1043-3805

© Linguistics Graduate Student Association  
University of Kansas, 2002

Cover design by Nathan Poell and Stacey Stowers

**Kansas Working Papers in Linguistics**  
**Volume 26**  
**2002**

Mayan Morphosyntax Clifton Pye.....	1
Acquisition of Mayan Morphosyntax Clifton Pye.....	19
Spanish Present Subjunctive Usage by US Spanish Heritage Speakers Kristi Hislope.....	39
Light Verb just as a Little v Dukkyo Jung.....	59
Effects of Context F0 Range in Perception and Production of a Lexical Tonal Distinction Travis Wade.....	75
Acoustic and Perceptual Evidence of Complete Neutralization of Word-Final Tonal Specification in Japanese Kazumi Maniwa.....	93
Do Mass Nouns Constitute a Semantically Uniform Class? David Nicolas.....	113

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

Kazumi Maniwa  
The University of Kansas

**Abstract:** This study investigated the extent to which the 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.

### 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. 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. 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). Although in German, Polish, and Catalan, word-final obstruents do not contrast in voicing, some studies have shown small but significant acoustic differences in duration between underlyingly voiced and voiceless final obstruents, or their preceding vowels.

In early studies on this issue (Dinnsen & Charles-Luce, 1984 and Charles-Luce & Dinnsen, 1987 on Catalan, Slowiaczek & Dinnsen, 1985 on Polish), the authors found significant differences in production but effects of underlying voicing were restricted to certain speakers, environments, or specific final obstruents.

Using a large number of speakers and words read in isolation, Port & O'Dell (1985) found more reliable effects on vowel duration, closure voicing, and burst duration in German, all significant across speakers. Furthermore, the authors tested listeners' ability to identify the productions and found that listeners could tell which member of a 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 the phonetic processes implementing the German devoicing rule are somehow sensitive to the underlying voicing contrast. 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. 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 re-create 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 effects of pragmatics of speech style and underlying voicing in German by giving five speaking conditions to speakers. The results suggested that there were effects of underlying voicing that are not limited to careful speech though speakers can make more or less clear differences depending on style. 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: (1) the standard conception of neutralization, where no differences in either perception or production are observed between underlyingly contrasting forms, (2) a limited neutralization where (small) differences are maintained in production but are not perceptible, (3) as in German, incomplete neutralization where differences are observed in both production and perception, and (4) the impossible situation of perceptual differences occurring in the absence of differences in production. Dinnsen observed that type (3) is quite common, citing final devoicing in Catalan, Polish, and Russian as well as German. Type (2) is also entirely possible, though it is in many cases presently not distinguishable from type (3) as only production studies have been completed. He claims, however, that type (1) 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. 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 (1) 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) that incomplete neutralization in earlier studies of German resulted

from hypercorrect pronunciation of differences between minimal pair members in terms of orthography.

### Japanese pitch accent

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 predict what word is being uttered. 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.

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 accented phrase: each is pronounced entirely on a high pitch, except for the first mora, which is low-pitched" (1968, p.139).

Thus, in two-mora words, there are three pitch accent types, namely accented HL and LH, and unaccented LH, which are called accent-1, accent-2, and accent-0, respectively, and 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. For example, the pitch pattern on /hána+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 /hána/ and /hana/ being LH in isolation.

This 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 the 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) can 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 differences between /haná/ and /hana/ in isolation, and that more errors are 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 /hashi/ and /hashi/, and 36 subjects for a minimal pair of /kaki/ 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 final pitch accent and no accent are neutralized word-finally in Japanese or not. 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 /ki/ 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 Sugunami 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, those 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 Hosō 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 extracted from 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.

#### Acoustic study

**Speakers:** Four college-educated Japanese women ranging in age from 24-47 years who were born and raised in Tokyo served as speakers. None of them had any known speech or hearing disorders.

**Stimuli:** Four minimal pairs of monosyllabic words and four minimal pairs of disyllabic words listed in Table 1 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 Hosō 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			
/ki/	"tree"	/ki/	"spirit"
/é/	"picture"	/e/	"handle"
/hi/	"fire"	/hi/	"day"
/ná/	"green"	/na/	"name"
Disyllabic			
/haná/	"flower"	/hana/	"nose"
/hashí/	"bridge"	/hashi/	"edge"
/kaki/	"fence"	/kaki/	"persimmon"
/murá/	"village"	/mura/	"unevenness"

Table 1. 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.

Procedure: Five repetitions of each word were randomized and presented in Japanese kanji characters 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 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.

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.

#### Results:

Words in isolation: Table 2 shows 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 mora		unaccented mora	
	1	2	1	2
Speaker 1	205	221	207	222
2	190	216	191	214
3	199	220	199	220
4	213	232	214	233
Mean	202	222	203	222

Table 2. 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 disyllabic words for each speaker averaged across 5 repetitions.

Mean minimum F0 of all of four accented tokens is 202 Hz and that of all of four unaccented tokens is 203 Hz. There is no significant difference between those two values [ $t(158)=-.682, p>.496$ ]. Mean maximum F0 of all of 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 [ $t(158)=.237, p>.813$ ].

Table 3 illustrates the mean values of maximum F0 on the vowel of the first (and only) mora in monosyllabic minimal pairs for each speaker.

	accented		unaccented
Speaker 1	221	Speaker 1	226
2	214	2	214
3	225	3	224
4	220	4	223
Mean	222	Mean	221

Table 3. 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 F of all of four accented tokens is 222 Hz and that of all of four unaccented tokens is 221 Hz. There is no significant difference between those two values [ $t(158)=-.059, p>.953$ ].

Words in context: Measurements of disyllabic stimuli are shown in Table 4. The data for F0 values on the vowel of the third mora, namely the vowel of the grammatical particle /ga/, have not been explicitly reported in previous studies.

	accented				unaccented		
	mora				mora		
	1	2	3		1	2	3
Speaker 1	203	257	206	Speaker 1	205	233	225
2	193	238	179	2	192	212	204
3	218	272	215	3	217	245	234
4	211	264	218	4	208	232	229
Mean	207	257	204	Mean	205	231	223

Table 4. 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 207 Hz and that for unaccented tokens is 205 Hz. There is no significant difference between those two values [ $t(158)=.580, p>.562$ ]. For the second mora, mean maximum F0 of accented tokens is 257 Hz and that of unaccented tokens is 231 Hz. The difference between those two values is significant [ $t(158)=10.407, p<.001$ ].

Mean maximum F0 on the third mora (ga) of accented tokens is 204 Hz and that of unaccented tokens is 223 Hz. The difference between those two values is significant [ $t(158)=-6.627, p<.001$ ].

Table 5 illustrate the measurements of monosyllabic stimuli produced in context.

	accented			unaccented	
	mora			mora	
	1	2		1	2
Speaker 1	267	214	Speaker 1	205	225
	2	243		2	180
	3	275		3	215
	4	271		4	199
Mean	264	211	Mean	200	224

Table 5. Minimum 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.

Mean maximum F0 of accented tokens is 264 Hz and that of unaccented tokens is 200 Hz. There is a statistically significant difference between those two values [ $t(158)=23.600, p<.001$ ].

Mean maximum F0 of accented tokens is 211 Hz and that of unaccented tokens is 224 Hz. As is the case for disyllabic tokens, there is a statistically significant difference between those two values [ $t(158)=-4.180, p<.001$ ].

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 disyllabic and monosyllabic 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 four subjects showed consistent neutralization for all tokens. In addition, contrary to previous studies, the subjects were not influenced by the research method. Complete neutralization was observed even when the subjects read a list of words in Japanese kanji orthography. This challenges Vance's (1995) explanation of some of his results.

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. The minimum F0 on the vowel of the first mora of the disyllabic words is not significantly different for accented and unaccented tokens. However, for the maximum F0 on the vowel of the second mora in disyllabic words and on the vowel of the first mora in monosyllabic words, there are significant differences between underlying distinctive pitch patterns. 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 disyllabic and monosyllabic words, F0 on the vowel of the grammatical particle is much lower in accented than in unaccented tokens. This finding is in agreement with the phonological theory of Japanese pitch accent as proposed by Beckman and Pierrehumbert (1986, 1988) and the J\_ToBI model of Japanese intonation (Venditti, 2000). These issues will be discussed in General discussion.

### 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.

**Subject:** 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.

**Materials and procedure:** For words in isolation, all five productions of the four minimal disyllabic pairs, and the four minimal monosyllabic pairs, produced by each speaker in the acoustic study (see Table 1), were used in the perception experiment.

In order not to make the perception experiment too long, the stimuli were divided into two tests. The same words had originally also been produced in a sentential context. For the perception experiment, these words were excised from the context using Praat software.

One consisted of the disyllabic tokens, /haná/ & /hana/ and /hashí/ & /hashi/, and the monosyllabic tokens /ki/ & /ki/ and /é/ & /e/. Test 1 consisted of 8 test blocks, and each block consisted of (1) accented /haná/ and unaccented /hana/ in isolation, (2) accented /hashí/ and unaccented /hashi/ in isolation, (3) accented /ki/ and unaccented /ki/ in isolation, (4) accented /é/ and unaccented /e/ in isolation,

(5) accented /haná/ and unaccented /hana/ from a context, (6) accented /hashi/ and unaccented /hashi/ from a context, (7) accented /ki/ and unaccented /ki/ from a context, and (8) accented /é/ and unaccented /e/ 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 disyllabic tokens, /kaki/ & /kaki/ and /murá/ & /mura/, and the monosyllabic tokens /hi/ & /hi/ and /ná/ & /na/.

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. Listeners were to indicate which word they perceived, accented or unaccented, by pressing one of two buttons. Stimuli were randomized separately for each subject with a 1000ms ISI.

Results: Results of the perception experiment are shown in Table 6.

	ISOLATION			CONTEXT		
	Accented	Unaccented	Mean	Accented	Unaccented	Mean
Monosyllabic	53	49	51	82	80	81
Disyllabic	49	48	48.5	63	63	63
Mean	51	49	50	73	72	72

Table 6. Correct identification (%) of disyllabic and monosyllabic words as a function of accent and context.

First, perception of both accented and unaccented disyllabic words in isolation is not significantly different from chance [ $t(11)=-.584$   $p>.571$ ] and [ $t(11)=-2.075$   $p>.062$ ], respectively. On the other hand, perception of these disyllabic tokens in context is significantly better than chance level. The accuracy for accented words and unaccented words is above chance level [ $t(11)=7.826$   $p<.001$ ] and [ $t(11)=6.060$   $p<.001$ ], respectively.

The data from the monosyllabic stimuli are very similar: perception of both accented and unaccented tokens in isolation is not significantly different from chance [ $t(11)=1.429$   $p>.181$ ] and [ $t(11)=-.679$   $p>.511$ ], respectively. In addition, perception of both accented and unaccented words in context is significantly better than chance [ $t(11)=14.099$   $p<.001$ ] and [ $t(11)=31.739$   $p<.001$ ], respectively.

Discussion: The perception results indicate that the distinction between word-final accented and unaccented morae is completely neutralized in Japanese. The listeners performed at chance level in their identification of the distinctive underlying word-final pitch accents when words were not followed by a grammatical particle. That is, listeners were unable to reliably distinguish the differences between accented words and unaccented words in isolation. On the other hand, the subjects remarkably improved their accuracy of perception for both disyllabic and monosyllabic tokens produced in context. Independent t-tests indicates that words produced in context are perceived significantly more accurately than words in isolation in both disyllabic and monosyllabic tokens ( $[t(1150)=-11.708 p<.001]$  for disyllabic tokens;  $[t(1150)=-23.220 p<.001]$  for monosyllabic 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".

#### 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 all tokens, the result is observed consistently in each speaker despite speaker variability in pitch (e.g., speaker B has a lower pitch than any other speaker). 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. However, for disyllabic words, this difference is restricted to the second and third moras. While F0 of the first mora does not differ as a function of accent, F0 of the second mora in accented tokens is increased strikingly compared to unaccented tokens. For monosyllabic words in context, F0 is substantially higher for accented words relative to unaccented words. However, this distinction does not show up in isolation. As mentioned above, 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 disyllabic and monosyllabic words.

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. Any theory which assigns these same tones, without discussing details of the phonetic F0 contour, such as what F0 will do during any of the tones assigned by the theory, can only predict that final-accented and unaccented words will be the same: there is simply no mechanism in such theories for giving the same LHH string different F0 values. Of course, no one would claim that F0 changes in a sudden jump between low and high. However, without instrumental methods, linguists working in traditional theory could do little toward a more explicit description.

Pierrehumbert and Beckman (1988) introduced an entirely new method of describing Japanese pitch accent by using only a few tones per phrase, with interpolation between them. The tones they assign 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. An accent peak (at the HL tone) is higher than a phrasal peak (H tone only). The HL accent tone does not occur in an unaccented phrase.

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. According to Venditti (2000), there is only one type of pitch accent in Japanese: a sharp fall from a high occurring near the end of the accented mora to a low in the following mora. Accentual phrase (AP) is a lower-level of prosodic phrasing and is typically characterized by a rise to a high around the second mora, and subsequent gradual fall to a low at the right phrase edge. This tonal pattern on the unaccented phrase is demonstrated in the way that the initial phrasal high tone is marked by placing a H-label on the second mora of the phrase, while the final low boundary tone is indicated by %L placed at the phrase edge. When the accentual phrase follows a pause, an additional delimitative %L tone is marked at the phrase onset, to provide an anchor from which the F0 rises. Thus, the complete tonal transcription of the APs is:

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

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

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

Similarly, the pitch patterns in monosyllabic accented /ki/ and unaccented /ki/, accented /ki+ga/ and unaccented /ki+ga/ can be described as follows:

Accented AP	k i
	%L H
Unaccented AP	k i
	%L H
Accented AP	k i g a
	%L H*+ L%
Unaccented AP	k i g a
	%L H

This model as illustrated above appears to mark pitch pattern and onset more accurately than a traditional description that places L or H on each mora. The description H\*+L clearly illustrates the sharp fall from a high pitch within a vowel of the second (or the first in the case of monosyllabic tokens) mora.

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.

**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 any speaker variances that other studies had suggested. Another new finding is that neutralization occurs even when the speakers read a list of words in Japanese traditional orthography, which some researchers consider an unnatural test circumstance.

A perception experiment was 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 in production. The present findings provide support for the theory of Japanese pitch accent by Beckman & Pierrehumbert (1986, 1988) and Venditti (1995, 2000).

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.

## REFERENCES

- 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.
- 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 obstruents, *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, 2<sup>nd</sup> edition.
- Kitahara, M. (2001). *Category structure and function of pitch accent in Tokyo Japanese*, Doctoral dissertation, University of Indiana.

- 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.
- McCawley, J.D. (1977). Accent in Japanese. In *Studies in stress and accent* (L.M. Hyman, editor), pp. 261-302. Los Angeles: 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 (1989). *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 neutralizing 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 Japanese phonology*. Albany: State University of New York Press

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

Venditti, J.J. (1995). Japanese ToBI Labelling Guidelines. Unpublished manuscript, Ohio State University. (Downloadable from: [ling.ohio-state.edu/phonetics/J\\_ToBI/jtobi\\_homepage.html](http://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](http://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.