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

Cover design by Nathan Poell and Stacey Stowers

Kansas Working Papers in Linguistics
Volume 26
2002

EFFECTS OF CONTEXT F0 RANGE IN PERCEPTION AND PRODUCTION OF A LEXICAL TONAL DISTINCTION[1]

Travis Wade
University of Kansas

Abstract: Fundamental frequency range is generally considered a variable, non-linguistic element of speech intonation. This study examined whether absolute F0 is predictable based on previous intonational context and is perceptually significant. Tokyo Japanese speakers produced sentences pairs differing lexically in the presence/absence of one pitch accent, as responses to speech cues. F0 placement of high tones was consistent across speakers and uniformly dependent on the cue intonation. Continuous manipulation of these sentences between typical accented and non-accent-containing versions were then presented to Japanese listeners for lexical identification. Perception was not significantly altered in compensation for artificial manipulation of preceding intonation. Results are generally consistent with the notion that pitch does not vary gradiently across speakers and situations but constitutes a predictable part of the phonetic specification of tones.

Introduction

It might be said that research involving phonetic description of intonational contours has required less quantitatively precise goals than similar study relating to segmental units. For speech segments, amplitude, spectral, and temporal properties have been observed with sufficient precision to characterize contrastively most or all the sounds of any particular language, across phonetic environments, utterances, and speakers. With pitch in speech intonation, however, descriptions in the fundamental frequency (F0) dimension tend to be exactly explicit enough to contrast the two generally accepted phonemically contrastive units: relatively high (H) and relatively low (L) pitch levels. The details of actual pitch realization (and perception) are explained in terms of the implementation of rules regarding heights of H and L accents relative to each other in various combinations and to their position in a phrase but assumed in general to be highly variable.

Perhaps, as summarized by Gussenhoven (1999), at least part of this variability is due to the participation of gradient, non-linguistic factors such as relative prominence in the "unused phonetic space" (e.g., exact F0 placement of a H tone peak) not required for contrasting between phonemic units in intonation. However, the relative success across languages with which studies of (for example) downstep have modeled typical relationships between tones within an utterance suggests that more precise measures of tonal F0 placement than relative high or low designations may contribute to their specification. This study deals with such a possibility, investigating (1) ways in which speakers may non-randomly vary their precise overall pitch range (i.e., the levels of their H tone peaks) to match or relate to those of other speakers in conversation, and (2) the extent to which these proposed relationships and also previously observed (within-utterance) downstep patterns may play a role in perception of the tones. For this latter question, the effects of previous context intonation on the perception of words in a lexical contrast in Japanese based substantially on the presence of a H tone are observed. Therefore, the intonational system of Japanese, as viewed in the influential Beckman and Pierrehumbert (1986, 1988) tradition (currently represented by the J_ToBI labeling system; see Venditti (forthcoming)) is briefly described below.

In J_ToBI analysis of Japanese intonation, prosodic contours are assumed to be largely underspecified for tone; contrary to earlier descriptions, each syllable or mora need not have its own (H or L) tone. Instead, there are a limited number of tones which are assumed to link directly to an accentual phrase node in the prosodic hierarchy: a phrasal H tone appears at the beginning of each phrase and is usually applied to the second mora of the phrase, and a boundary L% tone appears at the end of the phrase, attaching to the initial syllable of a subsequent phrase. (Additional boundary tones appear at higher levels of a prosodic hierarchy.) In addition to these phrasal tones, a separate, lexically determined pitch accent may attach to any mora[2] of a word. It has been variously labeled H+L, H*L, HL, and H* and is similar to the H* accent used for focus in English, except that its presence and location are entirely lexically determined. This pitch accent carries a clear distinctive functional load in current Japanese, evident in the relative number of minimal pairs where it is contrastive (see Kitahara and Amano 2000), its role in word recognition in auditory priming experiments (see Cutler and Otake 1999), and the fact that it is categorically perceived (Sugito 1982).

A recent description of an accent-non-accent contrast characterizes the distinction rather qualitatively: "the accented phrase displays a precipitous fall in pitch starting near the end of the accented mora, while the unaccented phrase lacks such a fall" (Venditti, forthcoming); this summarizes much of the phonetic literature on pitch accent to date. Pierrehumbert and Beckman (1986, 1988) showed that a phrasal peak frequency is statistically higher, and its boundary L% lower, in phrases containing lexical accents; this indicates that the H*L is indeed distinct from a phrasal H tone. Sugito (1982) determined, among other things, that F0, rather than a combination of pitch and intensity, is the primary cue to accent perception. Similarly, Kawai (1999) reports that segmental differences between the two accent types do not play a role in their perception. Exploring Sugito's notion of *ososagari* (late fall) and variation in temporal location of the falling contour, Hasegawa and Hata (1992) suggest that, more specifically, the F0 peak and the slope of the following fall are the primary cues in perception. However, they only deal with the timing (not the F0 value) of the peak, establishing a possible compensatory relationship between peak location and F0 slope in accent location perception, and do not consider accent presence perception, where F0 levels are more relevant. Kitahara (2001) found *no* effect of slope between normalized H and L values and attributes Hasegawa and Hata's finding to methodological concerns. In a perceptual study involving variation in H and L F0 values and slope of falling contours across various degrees of devoicing, Kitahara suggests there are three perceptual cues to accent perception: pitch at the H and L tones and a falling contour between the two levels. When one or more of these is missing, the others must be used to determine presence and location of an accent.

It seems clear, then, that (F0-determined) presence of an H tone is substantially, but usually not conclusively, responsible for indicating the presence of a pitch accent. This seems to constitute an ideal situation for examining the extent of possible context effects on the perception of precise tone height; listeners expecting a certain H-to-H peak F0 relationship across phrases and/or utterances might alter a lexical distinction boundary in compensation for changes to context-providing pitch, but would not be forced to do so[3] as this peak does not constitute the only means of identifying the accent. For this reason, the production and perception studies discussed below make use of this pitch accent contrast in investigating (1) whether speakers place the H peak components of pitch accents and accentless phrases, and therefore the (lexical) distinction between the two, consistently with relation to the intonation of an interlocutor, and (2) the extent to which any such regularity might play a role in perception of the contrast.

Production Experiment

A production experiment was first conducted to observe possible influence of context-providing pitch range on the specific F0 placement of H tones. Its goal was to find regularity which could be predicted—in a manner generalizeable across speakers—by preceding intonation in a realistic, pseudo-discourse situation.

Methods: As with all studies presuming to address natural patterns in production of speech intonation, the method of elicitation was a major concern. As Beckman (1997) observes, current models of prosody were generally developed using "read speech", but such productions lack many aspects of spontaneous speech, suggesting methods of more realistic elicitation, from natural but unstructured narratives to more uniform but more artificial human-computer interactions. In looking for context-dependent, cross-speaker regularity in production of specific pitch contrasts, some spontaneity in this study had to be sacrificed in providing that all speakers produced the same words in the same context; with this limitation, elicitation was designed to be as realistic as possible. Recordings were made of native speakers producing a simple (but not unnaturally so) question or request, listed in appendix A. Subjects heard these prompting sentences, at a rate they determined interactively, and answered questions or complied with requests using information (pictures or text) appearing on a computer screen. The exact desired response in written form always accompanied (or comprised) this information, but subjects were instructed to answer or respond to the prompting sentences rather than simply read the text. Also, subjects were allowed to practice and familiarize themselves with the stimuli before the experiment, so that as little reading as possible was necessary during the test.

Materials consisted of 14 pairs of brief Japanese sentences or nominal phrases contrasting only in the presence or absence of a single (H*+L) pitch accent. These pairs (listed in appendix B) vary greatly in total length, location of accent contrast, accented syllable structure, and voicing and other features of phonemes surrounding accented vowels. Some of these contrasts were introduced for the purposes of the perception study described below; in general, the rest were present in order to account for various phonetic factors (segmental and otherwise) that might affect the F0 placement of particular H accents and ensure that a difference or similarity between groups was not due to such factors.

The same recordings of the two prompting sentences were used for all productions (across subjects, sentences, and repetitions) in the experiment. They each were spoken by the same adult female speaker of Japanese; multiple productions were recorded, and a natural-sounding token of each was chosen such that overall pitch range varied significantly between the two. In the experiment, each production was prompted by simultaneous visual and auditory cues presented by computer. Once subjects observed a cue, they responded by speaking an indicated word or sentence into a microphone and pressed a key in order to proceed to the next stimulus. Each item was elicited three times from each subject. Stimuli were randomized separately for each subject in three blocks representing groups of words indicated in appendix B, in a single recording session. Auditory prompts were administered through computer speakers (not headphones), so that subjects could hear and monitor their responses in a natural manner. Recording was done in the Kansas University Phonetics and Psycholinguistics Laboratory (KUPPL) in a sound-treated room using a microphone and high-quality cassette recorder, and recordings were digitized (and later manipulated) by computer using the speech analysis program Praat (Copyright 1992-2000 by Paul Boersma and David Weenink). Subjects were 7 native speakers (6 female, 1 male) of Tokyo Japanese, ranging in age from 19 to 34 years. None reported a history of speaking or hearing disability. One speaker's results were discarded due to inability to make several contrasts; results of the remaining six are discussed below.

Predictions: As a purpose of the experiment was to observe an effect of intonational context on the production particularly of H peak levels, the following null hypothesis was assumed:

> H_0-1a: H* peak frequencies will vary continuously across productions and subjects, with no significant differences depending on the prompting utterance.

To reject this hypothesis, then, peak frequencies would have to be similar or observably related across and within subjects and differ depending on the type of prompt used.

Addressing whether context affects the distinction location between accented and unaccented forms is slightly less straightforward; target syllables of unaccented forms, lacking pitch specification, should occur at a variety of possible pitch locations determined by their place in the phrasally defined contour and thus appear at variable distances from their more stable[4] H* counterparts. As, however, the F0 location of the phrasal H tone plays a large part in determining pitch level throughout the phrase, and many accent contrasts actually occur in the (second mora) H location, this value was considered a measure of F0 scaling in unaccented words. For this reason, the following additional null hypothesis was formed:

> H_0-1b: Phrasal H peak frequencies will vary continuously across productions and subjects, with no significant differences depending on the prompting utterance.

In order, then, for a shift in the location of the accent presence distinction to have taken place, both of these hypotheses had to be rejected, with context-determined differences occurring in the same direction.

Results: Table 1 shows individual speakers' average placement of H peak frequencies for each prompt, followed by standard deviations of means and a relative measure of overall variance.[5] Perhaps most immediately notable is the small within-subject variation in range. Within an accent or prompt condition, F0 location of contour peaks typically differs well under one semitone from a relevant mean, or by only a few percent of the height of the peak itself:

Subject		Peak for prompt 1			Peak for prompt 2		
		Mean	SD	Ratio (SD/Mean)	Mean	SD	Ratio (SD/Mean)
S1	H* accent	255	7.55	0.03	236	9.03	0.038
	phrasal H	231	3.13	0.014	205	6.48	0.032
S2	H* accent	150	9.44	0.063	122	10.04	0.082
	phrasal H	137	7.47	0.055	108	5.91	0.055
S3	H* accent	308	8.83	0.029	291	10.42	0.036
	phrasal H	266	6.35	0.024	252	6.51	0.026
S4	H* accent	276	6.8	0.025	257	9.54	0.037
	phrasal H	251	10.3	0.041	231	10.03	0.043
S5	H* accent	270	10.26	0.038	259	9.13	0.035
	phrasal H	248	9.31	0.038	225	9.29	0.041
S6	H* accent	256	9.75	0.038	235	10.83	0.046
	phrasal H	229	6.12	0.027	205	11.48	0.056

Table 1: Individual speaker H peak F0 for 3 repetitions of sentences 1-5 (prompt 1) and 6-9 (prompt 2)

More critically, it can be shown that the type of prompting utterance played a predictable role in determining pitch range in both accent conditions and across speakers. Univariate analysis of variance was performed on peak frequencies for all speakers, and main effects were found for both Accent and Prompt type. No significant Prompt x Accent type combined effects were observed, indicating that the Prompt effect was stable across prompting type and vice-versa:

	F	Sig.
Prompt	(1, 324) 12.91	<.001
Accent	(1,324) 25.67	<.001
Prompt x Accent	(3,324) 0.077	0.781

Table 2: Effects (ANOVA) of prompting and accent type across speakers

This means that both null hypotheses concerning production data may be reasonably rejected, and it may be assumed that context played a predictable role in the assignment of pitch contours to the productions. Also important, however, is the degree to which these patterns extend across speakers. Based on conversation with subjects, it was observed that subjects S1, S4, S5, and S6 had medium female pitch ranges similar to that of the prompting speaker, while subject S3 had a relatively high female range and subject S2 a medium male range. These last two speakers, who deviated most drastically from the others in produced intonation, might have used different mechanisms for adapting to the pitch context provided in the test. (Subject S2's H values, e.g., were consistently approximately one octave below those of S1, S4, S5, and S6.) Excluding subjects S2 and S3, then, notable homogeneity is to be found in the F0 contours of the remaining four, as shown below:

	Peak for prompt 1			Peak for prompt 2		
	Mean	SD	Ratio	Mean	SD	Ratio
H* accent	247	14.89	0.06	264	12.65	0.048
phrasal H	216	14.82	0.069	239	12.74	0.053

Table 3: Average peak values for subjects S1, S4, S5, and S6 combined

Again, typical differences are quite small, especially when compared to absolute means. While additional experiments lacking the speech prompts described above were not carried out, it was assumed that the regularity observed would not be expected from a set of productions assigned pitch based on continuously varying, speaker-defined or otherwise arbitrarily assigned values.

Perception Experiment

Production results suggest that at least both types of H tone levels in Japanese, and therefore the level associated with the categorical distinction between them, may be related to the intonation of preceding context. As discussed above, H peak values are of course not the only cues to pitch accent perception. A following L value and an intervening fall are also key features, and it would be expected of such a locally oriented contrast that there exist sufficient acoustic invariance among these cues for accent identity to be unambiguous with or without information from previous context. Still, context is known to have effects on the perception of categorical distinctions in speech even when there is sufficient local information for a categorization to be made (Moore and Jongman 1997, e.g.). It seemed appropriate, then, in evaluating how functional a role context effects may play in intonation perception, to conduct a perception study based on these same contrasts. It was hypothesized that subjects attending to intonation contours might expect correspondence between H tone levels as uniform as that observed in the production study. Altering context-providing pitch contours (or portions of contours) could, then, cause the subjects to compensate for this change when making lexical judgments based substantially on following H tone information. Two typical productions taken from the production experiment were manipulated for F0 in various ways to construct six perception experiments to address this question.

Methods: A first set of experiments involved the single utterance *Róndon-no ame/ áme desu*, as produced by speaker S-4. This sentence contains two pitch accents, the second of which is involved in an accent contrast. From this second word was created a continuum of F0 values between typical accented and typical unaccented productions. This continuum then appeared in three different conditions, one (A-1) with only the word and no preceding context , one (A-2) with the word in a natural sentence context, and one (A-3) where the preceding accent peak had been lowered significantly. (In all cases, the final *desu* was removed.)

The second set of conditions was similar, except that no such within-utterance downstep relationship existed; the target sentence *shitai/shitài desu* (as produced by speaker S-6) contained only one possible accent, and context came in the form of the interlocutor request immediately preceding the sentence. As in the first set of conditions, the sentence occurred after a natural (S-2) or an altered (S-3) version of the same interlocution or else without preceding context (S-1). In this case, the context in condition S-3 consisted of the same utterance transposed (in its entirety) down significantly from its original pitch range.

In all conditions, artificially manipulated contours consisted of straight-line segments including sufficient detail that endpoint stimuli were virtually indistinguishable from the natural tokens used to generate them. In each case, all endpoints of the contour were varied simultaneously so that a single (ten-member) continuum similar to the schematic representation in figure 1 resulted, as opposed to a set of cross-related continua with each target varied independently of the others.

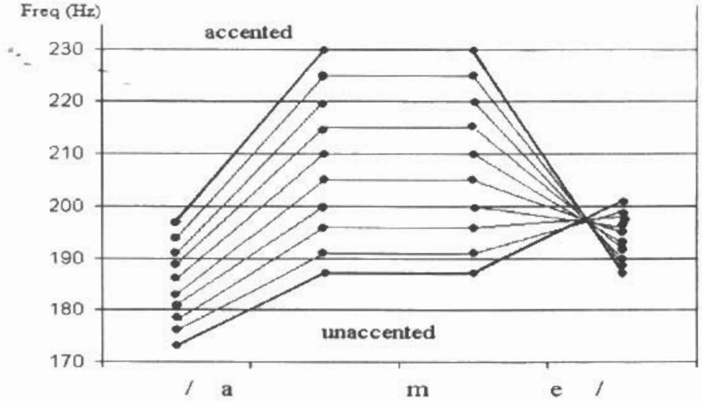

Figure 1: Schematic diagram of *ame* parameters varied simultaneously

This was both in the interest of keeping stimulus numbers at a minimum and in order to vary pitch in a more realistic, natural manner. Pilot studies confirmed previous reports that both H and L values contribute to accent perception and suggested additionally that when only one of these cues is varied and the other remains in an optimal accented or unaccented position, observably unnatural contours may result. It was also independently determined that the accent type of the token used as the base for pitch manipulation did not affect perception of the continuum, so originally accented types were used in all conditions as they were generally clearer.[6]

Each of the six conditions listed above was used to create a separate test; tests involved an identification task across a ten-member continuum; ten repetitions of each item resulted in 100 total stimuli for each test. Subjects were first allowed to listen to the two continuum end members until they were confident of their ability to make the distinction, and then listened to the 100 test stimuli in random order, after each responding whether the stimulus sounded like the first or the second endpoint lexical item. Subjects heard each stimulus only once but could progress at any rate, clicking a mouse button in order to hear each stimulus and even stopping to review example stimuli if necessary, though they were encouraged not to do so in the instructions. Tests were created in web-page format; approximately half were administered in a controlled laboratory setting (KUPPL), while the rest were completed via internet by subjects elsewhere in the United States and in Japan. Subjects were adult native Japanese speakers aged 19 to 50 years. None reported a history of speaking or hearing disability. Subjects were allowed to participate in as many separate experiments as they wished; some took only one or a few tests while others took several.

Predictions: As the goal was to observe possible differences in perception based on H* accent context, the following null hypothesis was adopted:

> H_0-2: Perception of the pitch accent contrast will not significantly differ between the altered preceding accent case, the natural sentence context, or sentences lacking intonational context.

If listeners' expectation of a consistent peak-to-peak downstep relationship was influential in their lexical decision, then, location of the accent contrast would differ, presumably being somewhat lower in the case of the lowered previous accent condition.

Results: Perception results for tests A-1 - 3 and S-1 – 3 are shown in figures 2 and 3, respectively (plotted response proportions are averaged across participants):

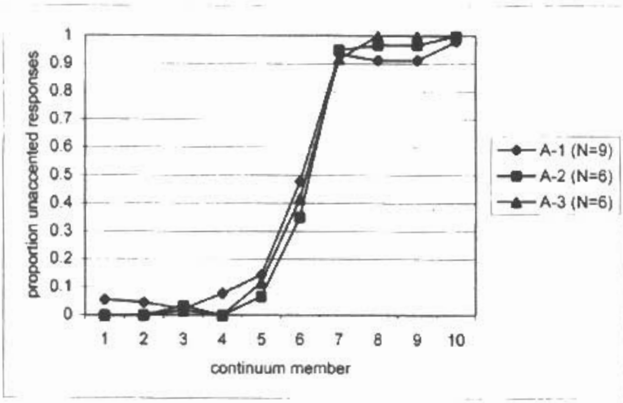

Figure 2: Averaged responses to *ame* / *áme* stimuli[7]

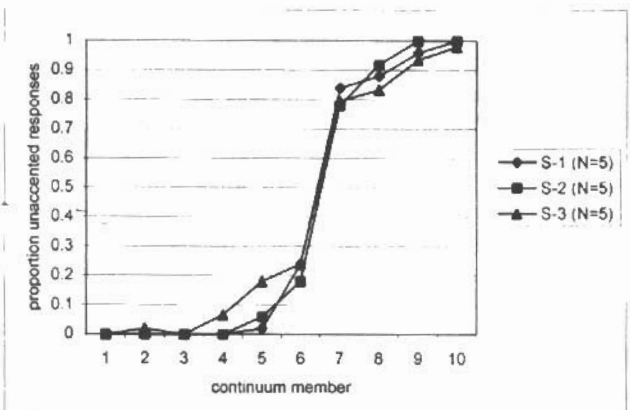

Figure 3: Averaged responses to *shitai* / *shitái* stimuli

These rather clear distinctions seem to suggest that subjects did *not* perform differently across testing conditions. Between-test correlation values of response proportion curves as shown above were nearly perfect in each case, and, critically, test 3 in each set, in which the preceding intonation was altered and thus any context effects would be observed, did not vary appreciably from either of the other two tests.

In order to look more precisely at possible differences between tests, it was necessary to use a single measure of distinction location that could be treated as a perception score for each subject. Probit regression analysis was first used to transform individual subject response data into curves which used all observed proportions in predicting continuum locations for response frequency values. 0.50 crossover points on these curves, then, were used to represent the category boundary for each subject. A summary of Probit-predicted distinctions is given below:

test	average Probit 0.50 value	SD	test	average Probit 0.5 value	SD
A-1	5.94	0.47	S-1	6.58	0.71
A-2	6.18	0.48	S-2	6.55	0.6
A-3	6.02	0.57	S-3	6.48	0.75
combined average	6.03	0.49	combined average	6.54	0.64

Table 4: Probit-defined distinction locations

As shown above, standard deviations of within-test means were all much smaller than one continuum member, indicating that there was agreement between subjects as to the location of the distinction. A one-way ANOVA failed to reveal a significant effect of test (*Ame* tests: $F(2,16)=.340$; $p=.717$; *Shitai* tests: $F(2,12)=.031$; $p=.970$).

It does not appear, then, that subjects were allowing context to influence their perception of the contrast in any way, and the results failed to reject null hypothesis H_0-1. The high cross-test agreement shown by the small standard deviations in distinction location suggest strongly that this null finding was not due to the limited number of participants but reflected a genuine perceptual tendency.

Discussion

The present production study results suggest strongly that the pitch context provided by a preceding (prompting) utterance does indeed play a role in the scaling of at least H tone peak frequencies in produced intonation. This is evidenced in that both types of peaks observed appeared in limited ranges which were predictable by prompting type. Such uniformity would not be expected from random or arbitrary assignment, nor would the highly consistent differences between peaks in responses to differing prompting types. This finding is consistent with a view that at least the H level component of overall pitch range is, in the presence of previous F0 context, *not* an entirely gradient, continuous, non-linguistic variable.

Perception results helped to reveal the limits of the importance of such relationships by failing entirely to show differences in the boundaries of a categorical perception based on compensation for changes in pitch context. It was predicted earlier that the vertically aligned pitch accent distinction might be altered to account for perceived relationships with relevant earlier pitch events; the facts that (1) no such effects were observed, either when one (H tone) context pitch event (*ame* tests) or an entire prompting contour (*shitai* tests) was altered, and that (2) no effect of any kind was found depending on whether context-providing contours were present or absent, help to show the extent to which tendencies observed in production are influential in perception. Technically, it could be either that the cues provided by context pitch are insufficiently important to effect perceptual changes or simply that the H and L pitch cues present locally within a (fully voiced) accented word are salient enough to override at least the types of conflicting context information examined in this study. However, considering the fact that at least qualitative context effects have been observed for pitch accent perception where less local information was available (eg Matsui 1993) and the general tendency for sensitivity to uniformly represented, clearly audible cues in the speech signal, it seems likely that the latter of these is the case. That is, of the two types of information available in making the perceptual distinction—previous context intonation and local H-L relationships—the latter, more immediate one is in this case preferred, where possible, when the two are in contradiction.

The results of this experiment, then, are in general consistent with—but demonstrate the incompleteness of—descriptions of intonational phenomena such as those of Ladd (1996), in which only phonologically contrastive tones are assumed to autonomously maintain linguistic status. While precise F0 scaling of tones may not have the same type of linguistic function as the tones themselves, it appears clear that the phonetic realization of pitch range *is* predictable, to a larger extent than seems to be assumed, by contextual cues. Present data are by far insufficient to propose a precise system of functional relationships, but it seems in order to assume at this point that some such formulation could result from further study.

<u>Appendix A</u>

Sentences used as verbal prompts for production experiment

1) kore-wa nán desuka (for sentences 6–9)
'what is this (a picture of)?'

2) yónde kudasai (for sentences 1-5, 10-14)
'please read'

Appendix B

Phrases/ sentence pairs for production

1) okawari vs. okáwari
'change' / 'another helping'

2) kakeru vs. kakéru
'to lack' / 'to gallop'

3) kaeru vs. káeru
'frog' / 'to return'

4) ame vs. áme
'candy' / 'rain'

5) shitai desu vs. shitái desu
'dead body' / 'I want to'

6) akai hashí desu vs. akai háshi desu
'it is a red bridge' / 'they are red chopsticks'

7) tomeru hito desu vs. toméru hito desu
'he is a person who stops' / 'he is a rich person'

8) hakushi desu vs. hákushi desu
'it is blank paper' / 'she is a doctor'

9) róndon-no ame desu vs. róndon-no áme desu
'it is candy from London' / 'it is London rain'

10) kátte vs. katte
'keep as a pet' / 'buy', 'reap'

11) kátta vs. katta
'kept as a pet' / 'bought', 'reaped'

12) útta vs. utta
'hit' / 'sold'

13) shimátta vs. shimatta
'closed' / 'put in order'

14) shimátteiru vs. shimatteiru
'closed' / 'putting in order'

NOTES

[1] The experiments outlined here represent part of a larger study described in Wade (2001).

[2] With the exception of the second mora of a long (VV, VN) nucleus and the limitation that a word may have at most one accent.

[3] As in Matsui (1993; as reported by Kitahara, 2001), where with a continuum of accent-following L values in an accent-opposed minimal pair in which the accented vowel is voiceless, listeners' perception differed according to preceding context: subjects were somewhat more likely to hear the unaccented form when there was a pitch accent earlier in the phrase. Presumably, as Kitahara observes, the difference was due to subjects' accounting for their expectation of a downstepped pitch range in the accented context.

[4] Of course, these levels also vary with location, particularly if declination is assumed; however, any such effect would be small in the stimuli used here (provided that the H* tones analyzed are not preceded by other accents and thus subject to downstep; for this reason, only the initial H* tone is considered in utterances with multiple accents) compared to the effect of position in an unaccented phrase.

[5] Specifically, only H* and H peaks from responses to sentences 1-5 were used as prompt 1 data. Responses to sentences 10-14 had geminate stops immediately following the target syllables; as a result, H peaks were somewhat higher and less regular in the much shortened vowels and, while useful in creating pilot experiments due to their simplified pitch contours, were not considered in production analysis. Similarly, sentences with voiceless vowels on the second mora were excluded from phrasal H analysis because of their lack of an observable peak, and H* accents after the third syllable were excluded as a few subjects placed additional phrasal boundaries before the accents.

[6] For greater methodological detail here and in the production study, see Wade (2001).

[7] In this and other test groups, sample sizes depended partially on which tests remote subjects opted to take and therefore often differ between individual tests; this does not affect analysis of the results as such groups were regarded as independent samples.

REFERENCES

Beckman, M. and J. Pierrehumbert (1986). "Intonational structure in Japanese and English." *Phonology Yearbook* 3, 255-309.

Connine, C., C. Clifton, and A. Cutler (1987). "Effects of lexical stress on phonetic categorization." *Phonetica* 44, 133-46.

Cutler, A. and T. Otake (1999). "Pitch accent in Spoken word recognition in Japanese." *Journal of the Acoustical Society of America* 105, 1877-88.

Fox, R.A. and J. Unkefer (1985). "The effect of lexical status on the perception of tone." *Journal of Chinese Linguistics* 13, 69-90.

Gussenhoven, C. (1999). "Discreteness and gradience in intonational contrasts." *Language and Speech* 42(2-3), 283-305.

Hasegawa,Y. and K. Hata (1992). "Fundamental frequency as an acoustic cue to accent perception." *Language and Speech* 35, 87-98.

Kawai, G. (1999). *Spoken Language Processing Applied to Nonnative Pronunciation Learning*. Doctoral dissertation, University of Tokyo.

Kitahara, M. and S. Amano (2000). "Functional load of pitch accent based on word familiarity." *ASJ fall meeting presentation.*

Kitahara, M. (2001). *Category Structure and Function of Pitch Accent in Tokyo Japanese*. Doctoral dissertation, University of Indiana.

Ladd, D. (1996). *Intonational Phonology*. Cambridge University Press, Cambridge.

Ladd, D. (1997). "The perception of intonational emphasis: continuous or categorical?" *Journal of Phonetics* 25, 313-42.

Liberman, M. and J. B. Pierrehumbert (1984). "Intonational invariance under changes in pitch range and length." in M. Aronoff and R.T. Oehrle, eds., *Language Sound Structure: Studies in Phonology Presented to Morris Halle*, 157-233. MIT Press, Cambridge.

Matsui, M. (1993). "Museihaku joo no aksentu kaku no chikaku no tsuite." [Perceptual study of the accent in the devoiced mora] paper presented at Kinki Onsei Gengo Kenkyuukai, Osaka, Japan.

Moore, C. and A. Jongman (1997). "Speaker normalization in the perception of Mandarin Chinese tones." *Journal of the Acoustical Society of America* 102, 1864-77.

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

Pierrehumbert, J. and J. Hirschberg (1990). "The meaning of intonational contours in the interpretation of discourse." in P. Cohen, J. Morgan, and M. Pollack, eds. *Intentions in Communication*. MIT Press, Cambridge.

Poser, W.J. (1984). *The Phonetics and Phonology of Tone and Intonation in Japanese*. Doctoral Dissertation, MIT.

Remjisen, B. and J. van Heuven (1999). "Linguistic versus paralinguistic status of prosodic contrasts: the case of high and low pitch in Dutch." 4th HIL Phonology Conference, HIL/Leiden University.

Sugito, M (1982). *Nihongo Akusento-no Kenkyuu* [Research on Japanese accent]. Sanseido, Tokyo.

Varden, J.K. (1998). *On High Vowel Devoicing in Standard Modern Japanese: Implications for Current Phonological Theory*. Doctoral dissertation, University of Washington.

Venditti, J., K. Maeda, and J. van Santen (1998). "Modeling Japanese boundary pitch movements for speech synthesis." in *Proceedings of the 3rd ESCA Workshop on Speech Synthesis*, 317-22. Jenolan Caves, Australia.

Venditti, J. (forthcoming). "The J_ToBI model of Japanese intonation." In S. Jun, ed. *Prosodic Typology and Transcription: a Unified Approach* (Collection of papers from the ICPhS 1999 satellite workshop on "Intonation: Models and ToBI Labeling.") San Francisco.

Wade, T. (2001) *Evaluating possible linguistic function of vertical scaling in speech intonation: evidence from Japanese tone*. Master's thesis, University of Kansas.