TR-IT-0180

Final accent, dialects, and the perception of Japanese tones.

Natasha WARNER

1996.8

ABSTRACT

In this report, I discuss results confirming that final accented and unaccented words are different within the word, particularly for long (4-5 mora) words. A perception experiment showed that listeners do not perceive this difference more accurately with long words than with short, however, and that Tokyo speakers are more likely than Kansai speakers to perceive it. Finally, an experiment on perception of accent contrasts in different dialects showed that while few Tokyo listeners can identify Kansai accent distinctions, Kansai listeners are often better at perceiving Tokyo distinctions than their own. This indicates that Kansai listeners will also be critical of accent mistakes in speech synthesis.

I have been working on several topics within the area of Japanese prosody while here at ATR. This report will cover the following main topics:

I. Additional work on a previously conducted production experiment to test for the possible difference between final accented and unaccented phrases in 4-5 mora words/phrases;
II. A perception experiment to test whether listeners can hear the difference between final accented and unaccented phrases found in part one, and;
III. A perception experiment to determine how well listeners can distinguish and identify accent patterns in dialects other than their own.

I. Final accent versus no accent: Production experiment

Before coming to ATR, I conducted a production experiment in which four speakers produced sentences such as those in 1 in both regular and reiterant speech.

1. /imooto' da/ "It's my little sister."
   /hahaoya da/ "It's my mother."

These sentences consisted of a noun or /ano/ plus a noun, with either 4 or 5 moras, excluding /da/. Previous research testing final accented versus unaccented words used either one or two mora words, which I believe introduce a confound. Please see my paper on this experiment for details of experimental design, as well as for a discussion of the subject.

In this experiment, I found that the slope of the fundamental frequency contour between the phrasal peak and the accent peak differs for the two types of words: although f0 falls during this period for both accent types, it falls more for unaccented than for final accented words, in accordance with the prediction of Pierrehumbert and Beckman's (1988) theory.

After coming to ATR, I performed additional analyses on the data from this previous experiment, and collected additional data on the subject as well. I had previously analyzed only the change in f0 between the phrasal peak and the accent peak, but when the values of f0 at the phrasal peak and the accent peak were analyzed separately, I found that not only is the accent peak higher in final accented words than in unaccented words, the phrasal peak is higher in the final accented words as well. Thus, in Figure 1, point A, as well as point B, is higher in final accented than in unaccented words.

For this set of data, the difference at point A in final accented and unaccented words is approximately 6-7 Hz (average for all non-reiterant data). This difference is significant (F=16.32, F(1,3)=10.1 at p=.05) when the data is analyzed as a three factor within subjects design (with accent type, reiterant vs. original speech, and number of moras as the three factors). Reiteration is also significant, with point A higher for original than reiterant speech (since reiterant speech generally has less pitch variation overall). Number of moras is significant (F=47.9), with point A 8-9 Hz higher for 5 mora words than for 4 mora words. The reiteration by number of moras interaction is also significant, but when the simple effects are investigated, it appears to be very small.

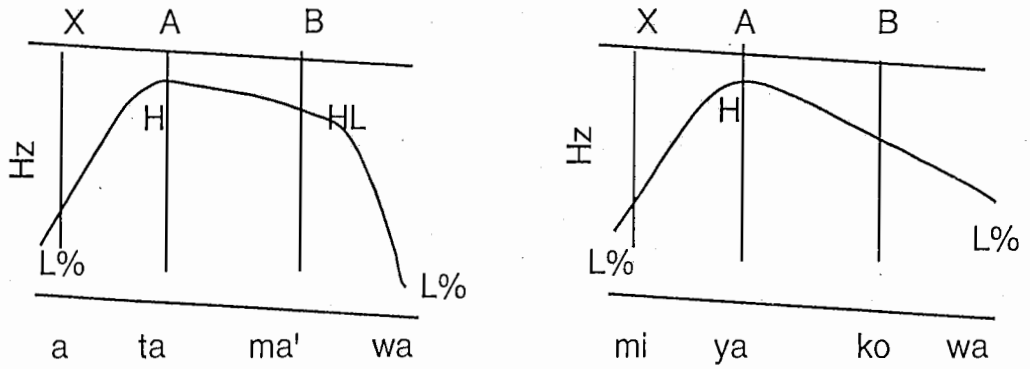

Figure 1

Data on f0 at point X in Figure 1 was then also collected. This point was the second or third f0 point produced by the ESPS pitch tracker, and chosen to be very near the beginning of the pitch track but after it becomes stable. I found that for the four speakers recorded before coming to ATR, the height of the beginning point does not vary depending on the accent type of the following word (F=0.27). There was also no effect for number of moras. As one would expect, the difference between point X and point A was greater for final accented than for unaccented words, since point A is higher, but point X is equal for the two types.

Kubozono (1993) also found that the phrasal peak is already higher if an accent is present later in the word than if no accent is present, but the words for which he found this were only three moras long, so the phrasal peak and accent peak were in adjacent moras. My data extends his finding even to five mora words, where the phrasal peak and accent peak are separated by three moras. This difference in the phrasal peak is not predicted by Pierrehumbert and Beckman's (1988) theory, which goes strictly from left to right in determining f0. However, Pierrehumbert and Beckman (1988:162)

mention that although they do not use any phonetic rules referring to rightward phonological context, they "would not be surprised to find that such rules exist, forcing a slight modification. . . ."

The Fujisaki model can easily account for point A being higher in final accented than in unaccented words, since it posits a larger accent command when there is an accent than when there is none. However, the Fujisaki model does not appear to account for the difference in slope between points A and B. Professor Fujisaki suggests (personal communication) that the relatively higher values for point B in final accented words (producing the difference in slope) could mean that there is a new, slightly larger, accent command beginning with the noun itself in the /ano/+noun sequences. This requires the assumption that the degree of separation between two words is scalar, not a binary decision of new accentual phrase or no new accentual phrase, as in Pierrehumbert and Beckman's theory. However, I believe the difference in slope between point A and B is present not only when the "word" consists of /ano/+noun, but also in the cases of a single word of four to five moras (as those in 1). Such cases could not possibly have a new accent command part way through the word.

Thus, either theory requires a modification in order to account for both the higher f0 at point A (the phrasal peak) and the lesser degree of fall between point A and point B (phrasal peak and accent peak) in final accented words. The fact that f0 at point X, the beginning of the phrase (as close to the boundary low of Pierrehumbert and Beckman's theory as possible) is the same for both accent types is in accordance with both theories.

For the perception experiment discussed in Part II below, I collected additional data on final accented and unaccented words. Different sentences were used, since minimal pairs were necessary. All sentences consisted of /ano/ plus a one or two mora noun, followed by /da/. The list of sentences appear in 2. 15-16 were excluded, because the speaker reads both as final accented.

2.

1.	/ano hana' da/	'flower'	2.	/ano hana da/	'nose'
3.	/ano hasi' da/	'bridge'	4.	/ano hasi da/	'edge'
5.	/ano hati' da/	'bowl'	6.	/ano hati da/	'bee'
7.	/ano kaki' da/	'fence'	8.	/ano kaki da/	'persimmon'
9.	/ano e' da/	'picture'	10.	/ano e da/	'handle'
11.	/ano ki' da/	'tree'	12.	/ano ki da/	'feeling'
13.	/ano ha' da/	'tooth'	14.	/ano ha da/	'leaf'
15.	/ano hi' da/	'fire'	16.	/ano hi da/	'day'
17.	/ano na' da/	'vegetables'	18.	/ano na da/	'name'

Labels were also placed in this data, at the same points as used for the data from the previous experiment, and analyzed in a similar way. This speaker also has the slope difference and the difference at the phrasal peak (greater fall between point A and B for unaccented phrases than for final accented, higher f0 at point A for final accented than for unaccented). In this speaker's data, the average f0 change between A and B in final accented phrases is a 3 Hz rise, while for unaccented there is a 12 Hz fall. This difference is significant ($F=10.79$, $F(1,12)=4.75$) when this speaker's data alone is analyzed as a between subjects design with accent type and number of moras as factors. The measurements of point A show that point A is 7 Hz higher for final accented than for unaccented phrases, and 5 Hz higher for 4 mora phrases than for 3 mora phrases. The effect of accent type on point A is significant ($F=9.35$), but the effect of number of moras is not.

However, this speaker also has a significant difference at the beginning of the phrase (the boundary low tone): f0 at point X is significantly higher ($F=8.19$, $F(1,12)=4.75$) for final accented phrases than for unaccented phrases. There is no effect for number of moras at this point. It is not yet clear why this difference appears for this speaker, but not for the four recorded in the previous experiment, and the position of the labels for point X for the different speakers should be reexamined. If further research shows this difference to be consistent for more than this one speaker, it would be very difficult for any of the theories of Japanese pitch accent to account for. A difference already at the beginning of an utterance depending on the existence of an accent several moras later would imply that phrases with an accent anywhere in them are simply different from phrases with no accent in them, throughout the utterance. However, this difference is not yet clear enough to warrant such conclusions.

This additional data has implications for Chatr, especially in those points where the data differs from Pierrehumbert and Beckman's predictions. Specifically, since Pierrehumbert and Beckman do not predict a difference in height of the phrasal peak, the finding that the phrasal peak is significantly higher in final accented phrases even when the accent comes considerably after the phrasal peak could be used to improve Chatr's f0 predictions. If this were to be implemented, the target for a phrasal peak should be set at least 6 Hz higher when there is an accent somewhere in the phrase than when there is not.

II. Perception experiment: Final accented versus unaccented phrases

Although the data discussed above show consistently that speakers produce a difference between final accented and unaccented words within the word (without reference to the following mora, such as /da/), this does not necessarily mean that listeners can hear this difference. It is relatively clear that the sharp fall in f0 after an accented mora is the main perceptual cue for Japanese pitch accent (Hasegawa & Hata 1992:87, 88), but if that cue is missing, can listeners make use of secondary cues to distinguish the final accented and unaccented types?

To test this question, I compiled the list of sentences in two, above. All consist of /ano/ plus a one or two mora noun, and /da/. It would be preferable to use single words of 4-5 moras, but no minimal pairs for final accent versus no accent longer than 2 moras could be found. (In the case of verbs which are minimal pairs for accent, and can take the suffix /-kata/, it is possible to find 4 mora minimal pairs, such as /kaikata'/ and /kaikata/ ('way to feed' and 'way to buy,' respectively). Several such pairs exist. However, the speaker recorded for this experiment has the other possible accent placement for the accented verbs, /kaika'ta/, so these words could not be used.) I attempted to make the phrases longer by using /kooiu/ 'this kind of' instead of /ano/ before the noun. However, there was generally a dip in the f0 contour between /kooiu/ and the noun, so that one could not be sure they formed a single accentual phrase. Therefore, the resulting phrases (excluding /da/) were only three to four moras long, but they do still show both the differences found for the previous experiment, as discussed above. In addition to the final accented and unaccented phrases, several sentences which are minimal pairs for other accent distinctions in Tokyo Japanese, shown in 3, were also recorded. These distinctions should be clear even within the word.

3.

19. /ano ha'si da/	'chopsticks'			
20. /ano isi' da/	'stone'	21. /ano i'si da/	'intention'
22. /ano ka'nzi da/	'manager'	23. /ano kanzi da/	'character'
24. /ano ke'eki da/	'cake'	25. /ano keeki da/	'times'
26. /ano ka'mi da/	'god'	27. /ano kami' da/	'paper'
28. /ano ka'ta da/	'shoulder'	29. /ano kata' da/	'shape'

The sentences in both 2 and 3, as well as two times as many unrelated short sentences, were written on individual cards, and read 12 times each by a female native speaker of Tokyo Japanese (Ms. Shimoda). The recording was made under high quality recording conditions. The speaker is educated about Japanese pitch accent. The sentences were arranged in random order (except that the members of each pair were not near each other), with each of

these sentences separated from the next by at least two unrelated short sentences of a different syntactic pattern from the sentences of interest, in order to distract the reader from the accent patterns of the experimental sentences. Since the two members of a pair were frequently separated by 20 or more other sentences, it is very unlikely that the speaker could have remembered what one member of a pair sounded like when she read the other member of the pair.

Ten tokens each of sentences 1-18 (excluding 15-16 because of the unexpected accent placement), and two each of sentences 19-29, were sampled at 16,000 Hz and trimmed to leave 150 msec. of silence from the original recording before the beginning of the utterance, and cut off approximately 18 msec. before the end of the vowel of the last mora before /da/. (All path names are included at the end of the paper.) It was necessary to remove the final portion of the vowel, because f0 often starts to fall shortly before the end of the accented mora (although the majority of the fall occurs in the following mora). 18 msec. proved to be enough to remove all or almost all of the f0 fall in final accented phrases, while leaving enough of the vowel to be clearly perceptible even in the case of very short vowels, such as in /ano-kaki da/. The speaker read both sentences 15 and 16 as final accented, so these sentences were excluded.

The trimmed phrases (with a uniform length of silence at the beginning, and less /da/ and the final portion of the preceding vowel) were arranged in a random order (except that the members of a minimal pair never occurred next to each other). Each token of sentences 20-29 was used approximately 8 times each, in order to have as many sentences where final accent/noaccent was not the distinction as sentences in which it was. I felt that the purpose of the test would be too obvious if only the sentences of interest (final accent/no accent pairs) were used. Also, if some listeners did not make use of accent information at all, the words in 20-29 (shown in 3) would make that clear. Sentence 19 (/ano ha'si/, 'those chopsticks') was excluded so that the answer sheet could consistently have two choices for each stimulus.

For the sentences of primary interest (1-18), 10 different tokens were used, only once each, in case a difference or lack of difference between the two accent types was specific to a particular token. However, one token each of half the phrases in 1-18 was repeated one additional time, but not counted in the score. This was to provide an unbalanced design, so that listeners would not base decisions on the number of tokens to which they had already given a particular answer. The resulting 328 tokens were separated into numbered blocks of ten, and a tape was made with a two second pause between each sound file.

Ten listeners participated in the perception experiment. The listeners were chosen to be native speakers of the Tokyo dialect, or of a dialect which is considered Tokyo-type for pitch accent (Okayama, Nagoya, Nagano). Because they were all employees of ATR-ITL, most have at least some knowledge of Japanese pitch accent. (Even many of those whose current research has no relationship to accent have previously worked on synthesis of prosody, or learned about Japanese pitch accent in classes.) Nine listeners were male, one female. The female listener was also the reader for the sentences, but they were presented in a different order from her reading.

They were provided with an answer sheet, on which the two possible spellings (usually two different Chinese characters) for each word were written. They were instructed to circle the word they heard on the tape, and told about the format of the answer sheet. They were warned that the members of a pair would not appear the same number of times. They were given a list of the words to appear in the experiment (1-14, 17-18, 20-29), in random order, and asked to read the words silently to themselves, and then to mark whether each word was one they used often, sometimes, or rarely. Because minimal pairs for this contrast are not very common, it is very difficult to find minimal pairs in which both words are of approximately equal frequency. Even in the case of a pair where both words are common, such as /hana'/ 'nose' and /hana/ 'flower' (the best known such pair), adding /ano/ before the noun makes one less likely. 'That flower' is much more likely than 'that nose.' Several of the words used are also quite uncommon (/e/, 'handle,' or /kaki'/, 'fence,' for example). Because of the word frequency problem, the listeners were trained by asking them to read the words to themselves, so that each listener had at least thought about each of the words and its pronunciation immediately before taking the test. Asking them to mark frequency gave them further reason to think about each word, and also gave an estimate of relative frequency specific to each listener.

After the listeners took the test, their answers were scored as a confusion matrix, using number of sentences perceived as final accented which were actually produced as such, number perceived as final accented which were actually produced as unaccented, number perceived as unaccented which were actually produced as final accented, and number perceived as unaccented which were actually produced as unaccented. If a listener was able to hear the difference (at least some of the time), then among the sentences to which s/he responded as final accented, more will have been produced as final accented than were produced as unaccented.

After the experiment, the listeners were interviewed individually about their dialect background and their degree of knowledge about Japanese pitch accent. These two factors were then converted into points on a scale of 1-5. For dialect background, a 5 is a listener who has spent his/her life entirely in the immediate Tokyo area (except a few years while working at ATR), and lower numbers are assigned to listeners who spent several years of their childhood in non-Tokyo-type accent areas, or who come from areas which have Tokyo-type accent patterns, but otherwise different dialects, such as Okayama, Nagoya, or Nagano. For degree of knowledge about Japanese pitch accent, a 5 is assigned to listeners who have a strong background in Japanese pitch accent and who knew the purpose of the experiment, a 4 to listeners who either are currently doing research on Japanese pitch accent, but did not become aware of which patterns were being tested during the experiment, or whose current research is not on pitch accent, but who realized during the test that final accent was at issue. 3 or 2 was assigned to listeners who said they knew some or a little about pitch accent, and a 1 to the single listener who said he knows nothing about the pitch accent system. It would obviously be preferable to do this experiment with naive listeners, but this was not possible at this point.

The listeners were also asked to read the list of final accented and unaccented words used in the experiment, adding /da/ after the words, to find out whether they read the words with the same accent marking given by the dictionary (and used by the reader for the experiment). This was quantified as the number of words out of the 16 used for which the listener agreed with the accent marking on the word list. Some listeners were quite unsure of whether a given word was final accented or unaccented (they could use either possible pronunciation when adding /da/ to the word), so this also gives a measure of the degree of certainty the listener has about the accent type of the words.

The results of the test, chi-square values for each subject, and the information gained from the interviews, are as follows:

subject	Tokyo-ness	know-ledge	agree-ment	A resp. as A	U resp. as A	A. resp. as U	U resp. as U	chi square
1	5	5	16	71	59	9	21	5.90*
2	4	5	16	53	32	27	48	11.06*
3	2	4	12	48	47	32	33	0.026
4	3	3	15	56	46	24	34	2.7
5	5	2	14	49	32	31	48	7.22*
6	5	3	9	63	60	17	20	0.32
7	5	4	15	44	24	36	56	10.22*
8	4	4	9	31	41	49	39	2.02
9	3	1	16	43	43	37	37	0
10	4	4	14	46	44	34	36	0.10

* Values over 3.841 are significant at $p<.05$, with df=1.

5 of the listeners heard final accented tokens as final accented more often than they heard unaccented tokens as final accented, with a relatively large difference. Of these, 4 have a significant difference using a chi-square test. 3 additional listeners have very small differences in this same direction, 1 listener heard final accented tokens as such exactly as often as he heard unaccented tokens as final accented, and one listener has a relatively large, but non-significant, difference in the opposite direction. Most of the listeners are biased toward final accented, as also reported in Vance (1996). When results from all of the listeners together are evaluated using a within-subjects ANOVA (with one factor, actual accent type of the token, and two responses from each listener, those in the columns "A responded as A" and "U responded as A" above), the effect of accent type on response is significant ($F=5.71$, $F(1,9)=5.12$ at $p=.05$). These results are similar to those in previous research using words of one to two moras, as in Sugito (1982) and Vance (1996), where it was also found that approximately half of the listeners could perceive a difference between final accented and unaccented words without a following syllable. Thus, although use of longer words (or phrases) makes the difference between final accent and accentless types more clear in production, it does not seem to improve perception.

The information from the interviews can help determine why some listeners hear the secondary cues differentiating final accented and unaccented tokens, while others do not. The difference between "A responded as A" and "U responded as A" was found, and the correlation between this difference and each of the factors investigated in the interviews was evaluated. (The difference in number of final accented responses to the two accent types was used instead of the chi-square value, for instance, because listener 8 has a difference in the opposite direction from all other listeners, but this cannot

be reflected in the chi-square value, since it must be positive.) The correlation between degree of difference in responses and degree of the listener's agreement with the accent markings in the list is the highest, with a correlation coefficient of 0.648. The correlation between this difference and the "Tokyo-ness" of the listener's dialect background is 0.419, and the correlation between the difference and the amount of knowledge the listener has about Japanese pitch accent is only 0.209. Thus, it appears that listeners who pronounce the words with the same accent pattern as they were hearing, and who are more sure of the accent type of the words, are better able to hear the difference between the two types. Secondarily, listeners who have a more strongly Tokyo background are better able to hear the difference. Knowledge about the pitch accent system does not give listeners much help in hearing the difference. It would still be better to run this experiment with naive listeners, but this does substantiate the results from these non-naive listeners.

This experiment was run again with six listeners from the Kansai area as the subjects. None of them are educated about Japanese pitch accent. One lived in Tokyo until age three, but all others have lived their entire lives in Osaka, Kyoto, Nara, or Wakayama. One listener was able to identify the final accented and unaccented phrases with greater than chance accuracy (chi square=7.37, values over 3.841 are significant at p=.05). This is listener K8 from the experiment described below, the results of which show clearly that she has exceptional skill in making accent distinctions. One additional listener produced a significant difference in the opposite direction (of the stimuli to which he answered final accented, the majority were recorded as unaccented; chi square value of 4.258), but his responses seem to be conditioned primarily by word frequency. The remaining four listeners had no significant difference in either direction, and appear to be guessing randomly.

III. Perception of different dialects

It is well known that there is a great deal of variation at many levels within Japanese pitch accent. A large proportion of the population speaks dialects with the Kansai-type accent system, or one of the dialects with no pitch accent distinction at all. Many other dialects have the Tokyo-type accent system, but modify that system in various ways. Even among speakers born and raised in Tokyo, there is considerable variation in placement of accent for a given word. Many speakers were not raised in only one dialect area, but were exposed to more than one of the three major dialect groups (Kansai type, Tokyo type, accentless). Finally, all speakers are exposed to the standardized dialect (based on Tokyo accent) through television, and

also to some extent to the Kansai dialect through its use in certain television shows. Thus, when two speakers converse, the speech they are hearing is rarely the strict Tokyo accent system used in studies such as those above. Still, there is no problem with mutual intelligibility.

When a speaker of one dialect, for example a Kansai type dialect, listens to a speaker of a dialect with a different accent system, such as the Tokyo dialect, there are at least two possibilities for what use the listener can make of accent information. If the listener has at least a passive knowledge of the other dialect's accent system, s/he might make use of the accent information in perceiving the speech, even though the accent information is different from what the listener produces. Alternatively, the listener might use the accent information only to identify the speaker as being from Tokyo, and not use it in understanding the speech. If listeners are not able to use accent information from other dialects (other than to identify the speaker as being from a different dialect area), then it might be less important to produce correct accent placement in synthesized Tokyo Japanese--especially when the synthesis is intended for a Kansai market. However, if listeners can use accent information which is different from their own, they will perceive mistakes in synthesis as mistakes, and not just as a different dialect.

To find out whether listeners were able to make use of accent information from a dialect different from their own, I designed an experiment in which Kansai type and Tokyo type listeners were asked to identify words which are minimal pairs for accent read by a speaker of the opposite dialect. The word lists for this experiment appear in 4 and 5.

4. Words read by Tokyo speaker, heard by Kansai listeners

	word	gloss	tones in Tokyo	tones in Kansai
1.	hasi	chopsticks	HL	LH
2.	hasi	bridge	LH	HL
3.	kata	shoulder	HL	LH
4.	kata	shape	LH	HL
5.	ame	rain	HL	LD
6.	ame	candy	LH	HH
7.	isi	thought	HL	HL
8.	isi	stone	LH	HL
9.	kami	god	HL	HL
10.	kami	paper	LH	HL
11.	kati	value	HL	HL
12.	kati	winning	LH	HL

5. Words read by Kansai speaker, heard by Tokyo listeners

	word	gloss	tones in Tokyo	tones in Kansai
1.	hasi	chopsticks	HL	LH
2.	hasi	bridge	LH	HL
3.	kata	shoulder	HL	LH
4.	kata	shape	LH	HL
5.	ima	now	HL	LH
6.	ima	living room	LH	HL
7.	ito	string	HL	LH
8.	ito	intention	HL	HL
9.	kaku	to write	HL	LH
10.	kaku	each	HL	HL
11.	katu	to win	HL	LH
12.	katu	cutlet	HL	HL
13.	aki	autumn	HL	LH
14.	aki	empty	LH	HH
15.	kanji	manager	HLL	HLL
16.	kanji	Chn. character	LHH	LLH
17.	keeki	cake	HLL	HLL
18.	keeki	times	LHH	HHH
19.	e	picture	H	A
20.	e	handle	L	H
21.	hi	fire	H	A
22.	hi	day	L	D
23.	haru	spring	HL	LH
24.	haru	to hang	HL	HH

All decisions about which tones a given word has in either dialect for this experiment are based on Professor Sugito's Tokyo-Osaka accent dictionary. As much as possible, Kansai words were chosen to have all six Osaka speakers surveyed for the dictionary agree on the accent type of the word. After compiling the word lists, they were checked with the two readers for the experiment, to make sure the readers actually had the accent types as marked. No frame sentences were used, because any frame sentence would be semantically more likely for one word than the other, and would allow for semantic influence in listeners' choices.

The words to be read by the Tokyo speaker, and listened to by the Kansai listeners, are all of course minimal pairs for accent in the Tokyo dialect.

They are divided into two groups, a group in which the same pairs are also minimal pairs in the Kansai dialect (although with different tones than in Tokyo), and a group in which each pair is truly homophonous in the Kansai dialect. Words 1-4 have the opposite tones in Tokyo from what they have in Kansai (HL in Tokyo, LH in Kansai for the same word, and vice versa). For words 7-12, one of the words in the pair has the same tones in both dialects, while the other word has opposite tones.

The words to be read by the Kansai speaker, and listened to by the Tokyo listeners, are divided into four categories. All, of course, are minimal pairs in the Kansai dialect. Two groups are also minimal pairs in the Tokyo dialect, while two groups are completely homophonous in the Tokyo dialect. The further division is based on whether the distinction present in the Kansai dialect is a distinction which Tokyo Japanese also uses, or not. This is because the Kansai dialect has more types of tonal distinctions than the Tokyo dialect. (Kansai can have a given word start either high or low, in addition to having an accent at some point in the word, and can also have rising or falling tones when two tones fall on the last mora. Tokyo dialect, however, can only have one accent during a word, and does not have the contour tones.) Words 1-6 again have opposite tones in the two dialects. Words 7-12 have one word in each pair with the same tones in both dialects, and the other word with opposite tones. In words 15-18, one word has the same tones in both dialects, and the other word, while formally assigned different tones, sounds quite similar in both dialects (for example /keeki/ 'times,' which has HHH in Kansai, and the similar sounding LHH in Tokyo). The remaining words are neither similar nor exact opposites in the two dialects.

A female native speaker of the Kansai dialect (from the Kyoto-Nara area) read the words on the Kansai speaker list 12 times each, in random order (except that the two members of each pair were kept separated). She was instructed to read in the Kansai dialect. As will be discussed later, many Kansai speakers, when asked to read words in isolation, produce some of them with Tokyo accent patterns, but the reader for this experiment is readily able to produce Kansai speech even for words in isolation. The Tokyo speaker, instructed to speak in the Tokyo dialect, read the words on her list 15 times each, in random order. She was the reader for both this and the previous perception experiment.

Ten tokens of each word, for each list, were sampled at 16000 Hz and trimmed to leave approximately 200 msec. of silence before and after the word. These files were arranged in random order (except that the members of a pair were kept apart). As in the previous experiment, 10 different

tokens of each word were used so as to minimize any effects specific to one production of a word, but one token was repeated an extra time for half the words, and not counted in the score, to provide an unbalanced experiment. For the Kansai listeners, the test started with a short test on listening to Kansai speech. For this portion, two tokens each of words 5-9 of the Kansai reader's list were used, for a total of 12 words. The purpose of this brief test on the listeners' own dialects was to provide a control, in case there are listeners who cannot identify words in isolation based on accent even in their own dialect. The Kansai listeners were then presented with the longer test (126 stimuli) of words in the Tokyo dialect.

The arrangement of stimuli was the same as in the previous experiment, except that no noise was added to the tokens. The answer sheet clearly showed which dialect would be presented in which section of the test. The listeners were instructed about which dialect they would hear for which section. They were told that when listening to Kansai speech, they should circle what the word they heard would be in the Kansai dialect, and when listening to Tokyo speech, they should circle what the word would be in the Tokyo dialect. They were, as in the previous experiment, told that the number of tokens was unbalanced. The test for the Tokyo listeners was the reverse: first, 12 tokens of Tokyo speech were presented (words 5-10 on the Tokyo reader's list), and then the 252 token Kansai test. Instructions were the same.

The Kansai listeners were 8 native speakers of Japanese, all of whom have lived their entire lives in Osaka, Kyoto, Nara, or Wakayama. In most cases, their parents are also from the Kansai area. The Tokyo listeners were 8 of the 10 listeners used in the previous experiment. The amount of exposure they have had to the Kansai dialect varies, but none is able to speak it.

The results, reported as percent correct for the listener's own dialect and for each of the groups of 6 words for the other dialect, are as follows. K designates a Kansai listener, and T a Tokyo listener.

listener	own dialect	1-6	7-12	13-18	19-24
K1	75	51.7	73.3		
K2	33.3	83.3	88.3		
K3	58.3	70	96.7		
K4	83.3	81.7	93.3		
K5	58.3	18.3	16.7		
K6	75	46.7	83.3		
K7	41.7	86.7	86.7		
K8	100	100	100		
T1	100	88.3	75	28.3	55
T2	100	0	83.3	83.3	41.7
T3	100	48.3	96.7	95	46.7
T4	100	6.7	63.3	95	58.3
T5	100	0	25	78.3	43.3
T6	100	6.7	66.7	78.3	38.3
T7	100	10	98.3	100	78.3
T8	100	91.7	71.7	41.7	58.3

Although the groups of words were chosen depending on their status as minimal pairs in one dialect or the other, they also conveniently divide the words into groups based on the relationship of the accent contrast between the two dialects, in most cases. Thus, as discussed above, words 1-6 have the opposite tones in the two dialects for both tests. Words 7-12 have one member of each pair identical in both dialects, while the other member has different tones in the two dialects. Thus, if a listener adopts the strategy that the other dialect has the opposite tones from his/her own, s/he should do well on the first group, but badly on the second.

If a listener, however, decides to evaluate the stimuli based on what tones they have in his/her own dialect, s/he would be guessing randomly on words 7-12, because both members of a pair have exactly the same tones in his/her own dialect, but would score near zero on the first group, where the tones are consistently opposite in the two dialects. For example, a Tokyo listener hearing word 8 (/ito/ 'intention,' HL in Kansai) from the Kansai reader would would not know whether it was "string" or "intention" based on his own dialect, since both words have HL in the Tokyo dialect. When this listener heard word 7 (/ito/ 'string,' LH in Kansai), it would not match either of the words in his own dialect. For these words (7-12), since both members of each pair have identical tones in the listener's own dialect, interference from the first dialect might be less than for the word in 1-6, so listeners will get

no help from their own dialect, but might find it easier to learn the tones of the other dialect.

For the Tokyo listeners, the third group (13-18) should be relatively easy, since two of the three pairs have similar tones in both dialects. In the fourth group, however, none of the words have similar tones to the Tokyo dialect, but neither do they have opposite tones, so using information from the Tokyo dialect would not be helpful, but the Tokyo dialect would still provide interference.

There is a great deal of individual variation in the data from this test. Different listeners seem to have adopted different strategies, and with differing degrees of success. There is also some variation within some of the groups of words, so that the percentages above do not tell the whole story. However, some patterns do emerge. When one word in the pair has the same tones in both dialects (the cases where the two words are homophonous in the listener's own dialect), most listeners of both dialects are able to identify both members of the pair with better than chance accuracy, as shown by the relatively high scores for words 7-12. This implies that listeners have been better able to learn distinctions in the other dialect when interference from their own dialect is weak. Many listeners do worse on the member of the pair (in 7-12) which has the same tones as in their own dialect than they do on the member of the pair which has different tones. This might reflect the "use the opposite of my own dialect" strategy as a secondary influence after lack of interference.

If listeners evaluate a word based on the tones it has in their own dialect, they should do very badly on the words in group 1, where the tones are opposite in the two dialects. This method should give scores much less than chance, as is the case with several of the Tokyo listeners. However, several of the listeners who have high scores on the group 2 words also have relatively high scores for group 1. It is possible that these listeners have learned, lexically, which words have opposite tones from their own dialect and which do not. This pattern, in which a listener can distinguish words of the other dialect at more than chance frequency regardless of whether the words have the opposite tones to their own dialect or not, is more common among Kansai listeners than Tokyo listeners, indicating that Kansai listeners have either had more exposure to the Tokyo dialect than vice versa, or have had more motivation to learn it, since it is the "standard." (One should remember that all of the Tokyo listeners currently live in Kansai, so their exposure to the Kansai dialect is more than minimal. Kansai listeners' exposure to the Tokyo dialect comes primarily from television, and to some degree from working at ATR, but all of the Kansai listeners spend

most of their time working with other Kansai speakers.)

One very noticeable pattern in this data is that Tokyo listeners can identify words of their own dialect at 100% accuracy, for all listeners, but Kansai listeners cannot do the same for their own dialect. In fact, many Kansai listeners can identify the Tokyo words more accurately than the Kansai words. Subsequent elicitation from the Kansai listeners showed that some listeners have exactly the same tones as listed on the word list (Kansai tones) and used by the reader for the experiment, but some listeners, when asked to read words in isolation, produce several of them with Tokyo accent patterns. The patterns they produced in these cases are not possible Kansai pronunciations (not used by any of the six speakers used for Sugito's Osaka accent dictionary). They are also not due to regional differences (such as Kyoto versus Osaka), since there are both Kyoto and Osaka speakers who produce strictly Kansai tones, and also who produce some words with Tokyo tones. Both Kansai listeners who produce consistently Kansai patterns and Kansai listeners who produce several words with Tokyo patterns scored badly on the Kansai portion of the test, so the low scores are not due to individual differences in accent placement. I believe that the low scores of Kansai listeners on words of their own dialect reflect the relative status of the two dialects, and the fact that while school teachers in the Kansai area usually speak to their classes in the Kansai dialect, the Kansai dialect is probably not taught about as much as the Tokyo dialect is. This may be similar to differences between Mandarin and Cantonese in how well speakers are able to compare tones.

These results do have some implications for speech synthesis. If it were the case that Kansai listeners could not make use of accent information in the Tokyo dialect, it might not be as important to synthesize correct Tokyo accent patterns, since this would mean that a large portion of the population would not be listening to those accent patterns anyway. However, this experiment shows that not only are Tokyo listeners extremely sensitive to their own accent patterns, Kansai listeners are almost as sensitive to Tokyo accent as well. Therefore, mistakes in accent synthesis will not be forgiven by speakers of either dialect group. These results also explain why, when there are mistakes in synthesis of accent, Tokyo listeners sometimes perceive the synthesis as sounding like Kansai dialect, but Kansai listeners hear it as neither Kansai nor Tokyo dialect (from anecdotal evidence involving Chatr). Since Tokyo listeners are not as good at perceiving Kansai distinctions as vice versa, when Tokyo listeners hear something unusual, they can simply assume it is Kansai dialect, but when Kansai listeners hear something unusual, they know it cannot be either dialect.

Pathnames: (all directories are in /as52/nwarner/)

/percep/data/

Files for perception experiment on final accent and no accent. Naming conventons are x.y.d for sampled data, where x is the sentence number from the list above, and y is the token number. x.y.cut.d are the trimmed files, x.y.cut.noise500 have noise added at the end, x.y.f0 are the f0 files, and x.y.lab are label files for points x, a, and b. Files beginning with "output" contain the numbers extracted from the f0 files.

/percep/lxdir/

EGG signals for corresponding to the data files. Filenames are x.y.lx.d. This data is unfiltered.

/dialects/kansai/

Files of Kansai speaker's data for cross-dialect experiment. x.y.d are sampled data files, where x is sentence number corresponding to the list in 4 above, and y is token number. x.y.nodc are the same files with the dc component removed. x.y.nodc.d are the files trimmed to leave approximately the same amount of silence around each utterance. playlist.kan creates the tape for the Kansai reader part of the experiment.

/dialects/tokyo/

Same as /kansai/, but for the Tokyo reader.

/dialects/playlist.both

Creates a tape with both the Tokyo and the Kansai reader's parts of the experiment.

/producexp/data/

Sampled data files created in Berkeley, for the original final accent versus no accent experiment. (These are the ones for Hirai-san to try fitting to if he wants, especially those which are single words.) Sampled data filenames are (by example) 15a1.sd, where 15 is the sentence number (see my paper on this for the word list), a is the speaker (a-f, but a and c excluded from data for reiterant speech), and 1 is the token number. 15a1.sd.out are the f0 files, and 15a1.lab are the label files, containing only points A and B (with corresponding C and D for the reiterant speech).

/producexp/newlabels/

Label files for the above data, with point X (and corresponding Y for reiterant speech) added by Ohta-san. Filenames are also 15a1.lab (same as the original label files).

/producexp/stat/

Files containing numbers extracted from the files above.

References

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

Pierrehumbert, J. B., & Beckman, M. E. (1988) *Japanese tone structure. Linguistic Inquiry, Monograph 15.* Cambridge: MIT Press.

Sugito, M. (1982) *Nihongo akusento no kenkyuu* [Research on Japanese accent]. Tokyo: Sanseido.

Sugito, M. (1995) *Osaka Tokyo akusento onsei jiten* [Osaka Tokyo accent phonetic dictionary]. CD ROM. Maruzen.

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

Kubozono, Haruo. (1993) The organization of Japanese prosody. Tokyo, Kurosio.