Spring 1992

Stress in Japanese English: Evidence from native perceptual judgements

Brian D. Teaman
University of Pennsylvania

Stress in Japanese English: Evidence from native perceptual judgements

Stress in Japanese English: Evidence from native perceptual judgements

Brian D. Teaman
University of Pennsylvania
Department of Linguistics

In this pilot study looking at interlanguage prosody, normal and contrastively focused constructions in English were collected from four L1 English speakers and four L1 Japanese speakers. These productions were then played to six native English speakers to see how well they could identify the stress placement of the utterances. The judgements were used as a diagnostic tool to study the salient characteristics of problems in non-native stress productions. It was found that stress placement was easier to recognize in native speaker tokens, although it was not clear what features of stress were most important to the judges. Possible explanations and the directions they suggest for further study in second language prosody are given.

Introduction

The study of prosody is a relatively untouched aspect of second language acquisition (SLA) research. There have been few studies of prosodic development or descriptions of prosody in second language learning.[1] This lack of analysis continues, in spite of the suggestion that prosody might be more important than segmental effects in determining L2 comprehensibility (Gilbert, 1990). This is a very strong claim and will not be possible to test without first understanding interlanguage prosody in a clear way. One important aspect of prosody in English is stress. Not only does stress help identify and locate words, but the main stress of a phrase acts as a locus for intonational contours. This study will look at stress of two different types, normal and contrastive, in order to begin to understand how well stress variability is controlled by Japanese speakers of English.

This study was motivated by casual observations in the classroom that Japanese English speech often seems to be characterized by a relatively level pitch with high rise-falls on the nucleus of the intonational phrase. The intonational

contours of these speakers often seem inappropriately insistent. Although there seem to be certain characteristics of the pitch contour that determine the oddness of these types of sentences, sound properties of the intonational phrase play only a part; pragmatic differences also have an effect. An example of this is shown in (1). These intonational contours[2] seem perfectly normal on their own. The problem is their juxtaposition. To put two declarative intonations of roughly the same contour on this phrase seems inappropriate. The focusing of both "awareness" and "issues" suggests different pragmatic intentions if only one intonation peak were found on either element. Explaining characteristics of interlanguage intonation involves not only an understanding of its characteristic intonation and stress, but also a consideration of pragmatic concerns.

(1)

I will first look at the realization of sentence stress in different pragmatic contexts. The sentence "I have a red dog" has a standard declarative pitch contour as shown in (2a). In (2b) a contrastively focused counterpart is shown.

(2)

The way that contrasting elements are highlighted in standard English is to simply move the nucleus of the intonation contour from the normal position ("dog" in this case) to the contrasted element ("red"). Some background in stress and intonation in Japanese and English is necessary before discussing the current study and results.

Accent and Intonation in English and Japanese

While contrastive analysis and error analysis have fallen into disfavor as explanatory theories in SLA, the influence of transfer cannot be disregarded (loup, 1984). I argue that due to the complexity of accent and intonation coupled with the inherent variability

of interlanguage systems, a thorough understanding of the relevant L1 and L2 systems will not only be helpful, but it is necessary. The most logical beginning point for understanding characteristics of the interlanguage data is to first look at stress and intonation in the L1 and the target language.

Japanese and English prosody have been the subject of many recent studies which provide a thorough descriptive and theoretical basis with which to approach second language data.[3] Beckman (1986:5) distinguishes English as a prototypical stress-accent language and Japanese as a prototypical non-stress-accent language. Van der Hulst and Smith (1988) use the term "pitch-accent" to describe Japanese. Though the form differs for both languages, accent functions to mark certain syllables distinctively–distinguishing words like cóntent and contént in English, and distinguishing the Japanese words *háshi* 'chopsticks', *hashí* 'bridge,' and *hashi* 'edge.' Stress also functions to delimit words in a string just as the verb and noun are stressed in the English sentence 'he dránk a cóffee.' In Japanese, accent is not such a reliable indicator in marking words in a string, since it is possible to have extended stretches containing no pitch accent (compare the pitch-accented 8a, b and c with the unaccented counterpart in 13, noting the extended high-pitched, unaccented portion of 13). Japanese and English are similar in that they have lexical accent, which means that accent cannot be predicted, it must be specified for each word.[4]

Perceptually, there are three aspects to stress-accent in English: pitch, length and loudness. Fry (1958) tested these three parameters by eliciting judgements of synthesized tokens while carefully varying pitch, length and loudness. Pitch was found to be the most perceptually significant variable, followed by duration and then loudness. In spite of the significance of pitch, Lea (1977) found that an algorithm for recognizing stress worked best if length and volume were also taken into account.

Current analyses of English intonation posit a connection between stress as discussed above and the intonational phrase. The most accentually prominent point–a stressed position–can then be used to "hang" one of a selected number of intonation contours. The metrical structure of words and/or phrases can be added to intontational contours by matching the "*" in the metrical string of (3) to the * of the intonational contour shown in (4). The "(L)H*+M" contour with an optional Low tone associated with the onset and a High fall to Mid-level could be added to these simple structures to yield the same basic elongated cáll.[5]

(3)

*	*
A bernathy	A li cia

(4) (L) H* +M (Calling contour)

In this case, the **H*** would be linked to the stressed syllable and the **M** tone would link with the coda syllables. In Abernathy, "A" would be linked with the high tone, while the other syllables would be associated with an **M** tone. Note that this word has no onset, so there is no material to align with the optional low. "Alicia," on the other hand, would get the optional **L** tone placed on the onset "A," while "li" would get a **H** tone and "cia," as the coda, would be associated with the **M**. Example (5) shows the resulting intonation contours that would result from the combination of the metrical structures represented in (3), with the contour of (4). Note that the difference between 5a and 5b is simply the existence of an onset to the nucleus in "Alicia" while "Abernathy" begins with an accented syllable.

(5)

a. A ber na thy b. A li cia

To understand English pitch as it is realized in phrases, it is important to keep separate the two different levels: stress assignment and the intonational contour. This understanding of the independence of stress and intonation as well as their interaction is crucial to understanding English prosody.

Japanese works differently. The shape of a pitch contour is determined by the pitch accent of individual words with phonological rules and focus constraints create a contour of the lexical pitch accents. While it is a simplification of the processes, it is easy to substantiate the claim that the pitch properties of individual words contribute to the overall intonation of phrases in ways that are not possible in English.

To demonstrate how the pitch properties of words determine different possible intonational phrases, I will refer to an example given by Kurasawa-Williams (1992). The basic phrase is shown in sentence (6).

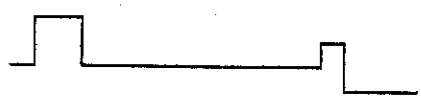
(6)

Yonémoto-san	no	ozyóo-tyan
Yonemoto-HON	GEN	woman-DIM
[Mr. Yonemoto's daughter]		

The Japanese phrase has pitch accents on two words in this phrase (indicated by the accent marks). Different kinds of focus allow for three different pitch patterns associated with (6). The questions 7a, 7b and 7c, force focusings that yield three different intonational phrases shown in 8a, 8b and 8c.

(7) a. Was it Mr. Yonemoto's son or daughter who called?
b. Whose daughter is coming?
c. Who is coming?

(8)



a. Yonémoto-san no ozyóotyan

b. Yonémoto-san no ozyóotyan

c. Yonémoto-san no ozyóotyan

English only allows two patterns for one intonational phrase. So, in English, 9a would be an answer to both 7a and 7b. While 9b would be an answer to 7c.

(9)

a. Mr. Yonemoto's daughter

b. Mr. Yonemoto's daughter

Therefore, Japanese allows up to three different possible intonational phrases for two pitch-accented words[6] while English allows only two possibilities for two stress-accented words in the same pragmatic context. This difference is due to the fact that

pitch is an inherent property of words in Japanese and rules of phonological focus as shown here depend crucially on the pitch-accent properties of the words in the sentence. In English, there is only the option of moving the nucleus of the pitch contour from "daughter" to "Yonemoto."

Now that a basic sketch of Japanese and English accent and intonation has been discussed, a few general statements concerning the differences between these two languages can be delineated.

(10)

accentless words: Japanese has accentless words, English does not.

stress accent: Accent in English involves pitch, loudness and length while Japanese uses only pitch (Beckman 1986).

pitch accent: In English, the intonational phrase is made by applying a pitch contour to a metrically marked phrase while in Japanese the pitch properties of words contribute more to the realization of pitch.

How then, do these differences pose problems for Japanese speakers learning English? The existence of accentless words would seem to be a greater problem for English speakers learning Japanese than for Japanese speakers learning English. For Japanese speakers, English represents fewer possible stress types since English words are free be accented on any syllable, but must be accented on at least one, while Japanese words are free to have no accent at all.

Stress accent would seem to be a more serious problem for Japanese learners of English, since this division of labor is quite foreign to these learners. A Japanese learner of English is faced with a language that relates any one of a number of intonational contours to the same metrically marked string, whereas Japanese has constraints on pitch contours that are derivable only by understanding the underlying pitch properties of words.

The third aspect, a different phonetic realization of accent, seems as though it would be extremely difficult to acquire. English and Japanese both have abstract accent but this accent is realized quite differently. As Lea (1977) has shown, native speakers of English manipulate the three parameters of pitch accent in complicated ways to mark stressed words in context. Not only are the relationships between pitch and intensity potentially complicated, duration alone seems almost intractably complex. Van Santen and Olive (1990) have attempted to derive an algorithm which would model duration in English vowels. They look at several factors that contribute to

vowel length such as vowel type, consonantal context and position in a phrase and find extremely complex interactions between the different variables. That the L1 learner has the capacity for reproducing these variants is amazing in itself.

The study

The research that will be described in this section will first look at how native speakers perceive native and non-native stress. These perceptions will be used as a basis for a discussion of the interlanguage English productions. There were two major phases to the pilot study: a data-collection phase and a judgement phase. The data-collection phase involved an activity used to generate natural speech with different types of normal and contrastive stress. The judgement phase used the data generated in the first part to elicit native-speaker judgements of English stress-accent. The judgements were used as a diagnostic tool to determine the salient characteristics of problems in non-native stress productions.

Data Collection

Subjects

Volunteers were solicited from students of English as a second language and Japanese as a second language at the University of Pennsylvania. Four native Japanese speakers (NJS) and four native English speakers (NES) participated. All participants were college graduates and had some abilities in both languages. The researcher, who was trained in the ACTFL oral proficiency rating system, estimated that the level of the NJSs ranged from "intermediate-mid" to "advanced-high." Two of the NJSs had been in the U.S. for more than two years and two had been in the U.S. for less than a year. Impressionistically, the two with the most time in the U.S. were more proficient than the other two.

Materials and Procedure

Materials were prepared in order to elicit noun phrases with pitch-accents located in normal and non-normal (i.e. exceptionally focused) positions. Pictures were used instead of written material so that natural speech could be elicited. The task allowed speakers to use language to solve a problem, not just language for language's sake. Reading sentences would have failed to control adequately for communicative goals.

Two pairs of picture cards were prepared. Each card had a 4x4 grid with 10 pictures and 6 blank spaces. The blank spaces were added to make it more difficult for the participants to go through the exercise in a "list-like" repetition of phrases, with none of the variation in stress desired. Participants were placed in pairs and asked to orally compare the pictures on their card with the pictures on their partner's card. Pairs consisting of all possible combinations were used. There were two NES-NES pairs, two NJS-NJS pairs, and four NES-NJS pairs. After an item was compared, each participant was asked to determine whether there were 0, 1 or 2 differences between the picture on their own card and the picture on their partner's card. Each participant recorded the number of differences on a separate piece of paper. For example, if the first person had a picture of a red dog as shown in 11, the partner could have any number of differences in their picture such as those represented by sentences 12a-12d.

(11) I have a red dog.

(12)

a.	I have a BLUE dog	1 difference (adjective)
b.	I have a red CAT	1 difference (noun)
c.	I have a red dog (too)	0 differences
d.	I have a BLUE CAT	2 differences (adjective and noun)

These are only sample sentences; participants were free to complete the exercise using any phrasing they chose. A professional quality cassette recorder and stereo microphone were used for recording.[7]

Native Speaker Judgements

Subjects

Six graduate students in Linguistics volunteered to make judgements. These subjects were asked to participate because it was thought that they would be more consistent in making judgements of the sort required. Researchers have found that it is difficult to get reliable results on stress judgements.[8]

Materials and Procedures

The recordings of the data were digitized and tokens were segmented from the digitized speech. Some "Adj + Noun" phrases elicited were not used because they

overlapped other speech or were barely audible. The final number of tokens segmented for this experiment was 153 including 73 native and 80 non-native segments. These tokens were segmented within the intonational phrase in which they appeared such as: "I have a red dog," "next, a red dog," or "red dog." Judges listened to the tokens through headphones as a computer program randomly chose one of the 153 tokens. They were asked to make one of three judgements for each token: whether stress was on the noun, on the adjective, or even. The judges were allowed to listen to the token up to five times. It was thought, that by allowing for repetitions, the judges would make more reliable judgements. However, they were encouraged to make their judgments in as few listens as possible. If each subject used only the number of repetitions actually needed to make a judgement, this number of repetitions could be used to provide a measure of the difficulty of judging the stress in that token. Information was stored on number of listens as well as on which stress type was judged.

Results

Repetitions

The number of repetitions needed varied from one to three; therefore, the maximum of five listens was not needed. In Table 1, the mean number of repetetions needed to make a judgement are listed. The numbers clearly demonstrate that judges needed to listen to non-native tokens more than native tokens. The first column, the mean number of listens, shows that there were between 1.0 to 2.0 mean listens. There were many tokens that only needed to be listened to once. For the native tokens, 40% were not repeated by any of the judges, while only 24% of the non-native tokens were unrepeated. These repetitions might reflect the difficulty of perceiving stress, however, other factors might also be involved--such as difficulty in overall comprehension due to segmental effects.

Table 1 Number of tokens needed to make a judgement

Mean	native tokens	non-native tokens
1.0	30	19
1.2	25	22
1.3	09	19
1. 0	09	12
1.6	0	04
1.8	0	03
2.0	0	01
TOTAL	73	80

Agreement among judges

Amount of agreement among the judges is a better measure of difficulty in judging stress, because it directly addresses the question of stress and not just overall difficulty in comprehension of the token. Agreement was judged "high" if at least four out of the six judges agreed on the token as either noun-stressed, adjecive-stressed, or even-stressed. The results are shown in Table 2.

Table 2 High agreement tokens

native		non-native	
54/73	74%	44/80	55%

Native speech samples elicited a much higher agreement among the judges. It is clear that there is a definite L1 effect. Table 3 rank orders the eight speakers by language in respect to the number of high agreement tokens. In this example, American English speakers are represented by NES1, NES2, NES3 and NES4 and Japanese speakers by NJS1, NJS2, NJS3 and NJS4.

Table 3 High agreement tokens: By individual speakers

Subject	High Agreement/Total	Percent
NES1	16/18	89
NES2	13/17	76
NES3	13/19	68
NES4	12/19	63
NJS1	12/19	63
NJS2	12/20	60
NJS3	11/19	58
NJS4	9/22	41

There was a much higher tendency for agreement on the NES tokens than the NJS tokens. The native speaker "NES1" had 18 total tokens played to judges and 16 tokens were agreed on by five or more of the judges. NJS4 only produced 9 of 22 high agreement tokens. It is clear that there is an L1 effect, since the non-native speakers were never above 63%. However, there was a lot of variability in judgements of native speech; considering native vs. non-native speech does not totally account for this variability.

Table 4 below gives more information by showing not only the high tokens, but also the non-high tokens.

Table 4 Agreement among judges for different tokens (by L1)

Judges agreeing	English L1 subjects		Japanese L1 subjects	
No.	No.	%	No.	%
6	19	26	10	13
5	19	26	18	23
4	16	22	16	20
3	17	23	34	43
2	02	03	02	03

Although Table 2 shows clear differences in the perception of native and non-native speech, Table 4 shows it more clearly. The number of tokens where only three judges agreed was 23% for the natives and 43% for the non-native tokens.

Discussion

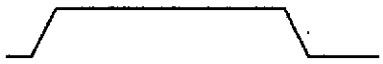
There are some general comments to be made regarding the results discussed above. First, Japanese learners are capable of producing stress forms that are categorically perceived by native listeners. All tokens used for the judgement test were not categorically perceived for NESs or NJSs. Secondly, the NJS subjects who participated in this study showed what might be developmental effects. The more advanced learners produced forms that were more often agreed upon by the native judges. This potential for development, or at least differential competence in stress would indicate that this is a variable worthy of further study.

There are clear indications of differences in how tokens produced by different speakers were perceived, but nothing has yet been said about what the cause for these differences in stress judgements might be. Why did the NJSs perform at a much lower level overall, compared to the NESs? In order to explore the possible sources for differences, I will first begin by discussing the three problem areas discussed above in 10. There seemed to be no evidence that learners interpreted English words as being accentless. This possibility should not be ruled out, however, since this was a small group of learners, none of whom were in the earliest stages of learning English. Learners of lower proficiency levels, different learners, or these same learners under different circumstances might do so. I have observed in casual situations that sentences are sometimes ended with no fall which would seem to indicate that the speaker was behaving as if there were no accent on the final word, where there should have been.

The second difference listed in (10), stress accent, probably has some effect, and might be a variable that is not manipulated properly. All of the syllable nuclei were measured for all of the tokens. It seemed that the NJSs varied little in the length of the syllable nucleus whereas the NESs lengthened stressed syllables to a greater extent. The database was not controlled for in terms of syllable shape to an extent great enough to warrant broad conclusions about length. An experiment that controls more for syllable shape will be necessary to explore this difference more fully.

The final difference listed in (10), pitch accent, seemed to have a great effect. Although it is not possible to formalize the demonstrated effects at this point, it seems clear that certain phonological rules of "pitch-accent" are operating on the English productions of these speakers. (8c) shows two pitch accented words in Japanese in a neutral context (no special focus on either term). If we look at the same phrase but with accent only on *ozyóo-tyan,* the result is a hat-shaped intonational pattern (13). This type of phrasing was observed on many tokens.

(13)



a. Hiraoka-san no ozyóotyan

In summary, each of the three major differences related in (10) seem to yield some interesting possibilities, but more work is needed to understand the characteristics of interlanguage stress productions by Japanese speakers of English. Further studies will be needed with larger numbers controlling for different levels of proficiency. For each of the different hypotheses, a data-set will need to be constructed that will explore the specific variables in a more complete way.

---

[1] A few of the more significant studies of second language prosody have been Backman (1977), Sethi (1982), Cruz-Ferreira (1980; 1987; 1989), and Berkovits (1980).

[2] The contours used in this paper are stylized, they do not represent certain phonetic factors such as downdrift and segmental effects since they are meant to represent only the phonologically significant aspects of the intonational contour.

[3] Some of the representative examples of approaches to English are Liberman (1975) and Liberman and Pierrehumbert (1984). For Japanese, see Poser (1984), Pierrehumbert and Beckman (1988), and Haraguchi (1988). Beckman and Pierrehumbert (1986) and Beckman (1986) compare Japanese and English.

[4] Here, I mean only to contrast English and Japanese with other languages where accent is much more predictable, and with very few exceptions. Autosegmental analyses have been offered for English in Halle and Vergnaud (1989) among many others. For Japanese, a comparable analysis has been given by Haraguchi (1991). These analyses show that accent is much more predictable than seems to be possible on a superficial level, however, the fact remains that the system is much less predictable with Japanese where there are many minimal pairs like the *háshi -hashí - hashi* paradigm as shown above. English has fewer examples, with réefer- refér, díffer-defér, and pérvert-pervért being differentiated solely by stress for many speakers (Flege & Ocke-Schwen, 1989).

[5] The basic line of this argument comes from Liberman (1975) however there is some influence in the notation used by Pierrehumbert (1980). For the purposes of this paper, I have deviated from both of their systems in order to simplify the argument; it does not weaken the argument.

[6] For all of the possible permutations of accented and unaccented nouns and adjectives in Japanese, 9 different pitch contours are shown in Kurasawa-Williams (1992) and Teaman (1992).

[7] Central air conditioning caused an audible hum on the tape which made the quality of the tapes of less than optimal quality, though they were good enough to proceed with further analysis.

[8] (See Lea [1977] for an extensive discussion of methodology in stress judgements).

References

Backman, N. (1977). Intonation problems of eight Spanish-speaking adults learning English. Unpublished doctoral dissertation, Boston University.

Beckman, M. E. (1986). Stress and non-stress accent. Dordrecht, Holland: Foris Publications.

Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational Structure in Japanese and English. Phonology Yearbook, 3, 255-310.

Berkovits, R. (1980). Perception of intonation in native and non-native speakers of English. Language and Speech, 23 (3), 271-280.

Cruz-Ferreira, M. (1980). Non-native comprehension of intonation patterns in Portuguese and in English. Unpublished Ph.D. thesis, University of Manchester.

Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. In J. Allen & J. Leather. Sound patterns in second language acquisition. Dordrecht, Holland: Foris.

Cruz-Ferreira, M. (1989). A test for non-native comprehension of intonation in English. IRAL, February, 1989.

Flege, E. J. & Ocke-Schwen, B. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. Studies in Second Language Acquisition, 11: 35-62.

Fry, D. B. (1958). Experiments in the perception of stress. Language and Speech, 1:2, 126-152.

Gilbert, J. B. (1990). Pronunciation: What should we be teaching? Unpublished manuscript.

Haraguchi, S. (1988). Pitch accent and intonation in Japanese. In H. van der Hulst & N. Smith (Eds.), Autosegmental studies on pitch accent. Dordrecht, Holland: Foris Publications.

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

Hyman, L. (Ed.). (1977). Studies in Stress and Accent (Southern California Occasional Papers in Linguistics, 4). Los Angeles: Dept. of Linguistics, University of Southern California.

Ioup, G. (1984). Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition. Language Learning, 34(2): 1-17.

James, A. R. (1988). The acquisition of a second language phonology. Tübingen: Gunter Narr Verlag.

Kurasawa-Williams, N. (1992). Semantically-motivated intonational phrases in Japanese. Paper presented at the Linguistic Society of America Annual Meeting, Philadelphia, PA.

Lea, W. A. (1977). Acoustic correlates of stress and juncture. In L. Hyman (Ed.), Studies in stress and accent (Southern California Occasional Papers in Linguistics, 4: 83-120).

Liberman, M. (1975). The Intonational System of English. Doctoral dissertation, MIT, Cambridge. (Subsequently published by Garland Publishing , Inc. 1979.)

Liberman, M. & Pierrehumbert, J. B. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R.T. Oehrle (Eds.). Language sound structure. Cambridge: MIT Press.

Pierrehumbert, J. & Beckman, M. (1988). Japanese tone structure. Cambridge: MIT Press.

Poser, W. J. (1984). The phonetics and phonology of tone and intonation in Japanese. Doctoral dissertation, MIT, Cambridge. (To be published by Foris Publications.)

Sethi, J. 1982. Intonation in educated Panjabi-speakers' English. CIEFL Bulletin, 18 (1-2) 69-85.

Teaman, B. D. (1992). Phonetics, phonology and pragmatics: Determining the role of phonological focus in interlanguage production. Paper given at the Second Language Research Forum, Michigan State University, April 3-5.

van Santen, Jan P.H. and Olive, J. P. (1990). The analysis of contextual effects on segmental duration. Computer Speech and Language, 4, 359-390.

van der Hulst, H. & Smith, N. (1988). Autosegmental studies on pitch accent. Dordrecht, Holland: Foris Publications.