Brain Responses to Focus-Related Prosodic Mismatch in Japanese

Kiwako Ito[†] & Susan M. Garnsey[‡]

[†]Department of Linguistics, Ohio State University [‡]Department of Psychology, University of Illinois at Urbana-Champaign [†]ito@ling.ohio-state.edu, [‡]sgarnsey@psych.uiuc.edu

Abstract

WH-questions generally lead to intonational prominence on words carrying the inquired information (i.e. words under focus) in the answer. Past German studies have demonstrated that Event-Related-Potentials (ERPs) are sensitive measures of listeners' reaction to such focus-related prosody [1, 2]. According to [2], missing expected accents on focused words in short German dialogues lead to posterior negativity, whereas unexpected accents on non-focused words do not evoke any particular ERP component. These findings suggest that prosodic information may be processed differently for focused words. In order to test whether the ERP patterns reported in [2] reflect universal or language-specific brain responses to prosodic information, a similar auditory ERP study was conducted in Japanese, which has a very different prosodic structure from German. The Japanese ERP data confirmed a distinction between the responses to focused and non-focused words: lack of intonational prominence for expectedly focused words led to (1) posterior positivity for the subject; and (2) non-significant, but widely observable anterior negativity for the object. Similarly to the German data, unexpected prominence for non-focused words did not invoke ERP differences in Japanese. Despite the discrepancy in which ERP components were observed in response to the absence of expected prominence in German and Japanese, the present results suggest the general principles of prosodic processing that distinguish focused words from non-focused words across languages.

1. Introduction

In successful oral communication, listeners interpret speakers' messages by understanding what words have been uttered in what manner. Comprehending spoken sentences does not merely require recognition of words but also processing their prosodic properties. The pragmatic meaning, as opposed to the semantic meaning, of a sentence is often expressed by prosody rather than by word choice (e.g. Do you know who he is? can convey multiple different messages depending on how it is said). In languages such as English and German, words carrying pragmatically important information are often produced with a distinctive pitch movement or a pitch accent over the stressed syllable [3, 4]. For example, a word daughter will be accented if it appears in That's my daughter as an answer to Who is that girl?, whereas the same sentence would have a clear accent on my to answer Whose daughter is she? As these examples indicate, WH-questions generally lead to focus on particular words in their answers, or they evoke expectations about what should be intonationally prominent in the answer sentence.

As prosodic cues in running speech flow into listeners' ears quite rapidly along with segmental information, investigating the on-line processing of prosody requires a measure that allows fine time course analysis of listeners' responses to speech signals. Event-related potentials (ERPs) are a very useful measure of immediate responses to rapidly presented stimuli, and studies on on-line visual sentence processing have demonstrated that both syntactic and semantic processing are reflected in distinctive ERP patterns [5, 6].

ERPs have been used as an on-line measure of speech processing in only a few studies over the past decade. Some previous studies in German have reported interesting patterns in brain responses to certain prosodic cues or unexpected prosody in a given discourse context [1, 2]. Steinheuer, Alter, and Friederici [1] first reported a positive-going ERP waveform, which they called a closure positive shift (CPS), in response to intonational boundaries in their auditory stimuli (e.g. a sustained positive-going component was found after each prosodic boundary indicated by a square bracket in Peter verspricht] Anna zu entlasten] und das Büro zu putzen. 'Peter promised Anna to work and to clean the office'). This CPS component was still observed pauses were edited out, and a similar ERP pattern was also observed in another study using an artificial language with a group who had not mastered the language. Thus Steinhauer et al. suggest that CPS is induced by prosodic cues other than silence signaling a boundary, which may guide syntactic parsing but may not require syntactic knowledge. In a follow-up study, Hruska, Alter, Steinhauer, and Steube [2] also found a CPS right after a focused (i.e., accented) word.

In addition to the CPS, [2] reports interesting differences in ERP patterns between two prosodic mismatch conditions: missing accent on a focused word vs. unexpected accent on a non-focused word. In this study, subjects heard short WHquestion + answer pairs in which the accent was placed either on the word under focus or on the non-focused word in the answer sentence. For example, two renditions of Peter verspricht Anna zu arbeiten und das Büro zu putzen were recorded with an accent either on Anna or on arbeiten. The two types of answers were cross-matched with two types of WH-questions such as (a) Wem verspricht Peterzu arbeiten und das Büro zu putzen? 'Whom does Peter promises to work and to clean the office?' and (b) Was verspricht Peter Anna zu tun? 'What did Peter promise Anna to do?' which led to focus on Anna and arbeiten (and das Büro zu putzen), respectively. When the expected focused word did not have an accent (e.g., unaccented Anna after question (a) or unaccented arbeiten after question (b)), a negative-going waveform was observed after the unaccented word most prominently at posterior sites, as compared to properly matched question-answer pairs. However, when a repeated word had an unexpected accent (e.g. ANNA after (b) or ARBEITEN after (a)), no difference was found between these mismatch cases and properly matched question-answer pairs. These findings suggest that prosodic information is processed in a different manner for words that are and are not under focus. In order to test whether the ERP responses described above for prosodic mismatch for the focused words are language-specific phenomena or instead reflect general, universal principles underlying the processing of focus-related prosodic information, brain responses to similar pragmatics-prosody mismatches should be investigated across languages with different prosodic structures. Japanese is an ideal language to examine this issue, as its intonational expression of focus is very distinctive from that in German. Japanese has accented and unaccented words, and the basic tonal shape of each word is lexically specified. When a word is under focus, its pitch range is expanded or reset, and the tonal movement of the following words is remarkably compressed, which often leads to a grouping of post-focal words with the preceding focused word into the same prosodic unit [7, 8]. Thus, the focal status of a word is expressed in Japanese by tonal scaling, instead of the presence or absence of a pitch accent. The present study addresses the following two questions: (1) Does a lack of pitch range expansion for a focused word lead to the same ERP patterns that were induced by a lack of accent in German?; (2) How would Japanese listeners react to unexpected pitch expansion for non-focused words? If the data in the current study show a distinctive ERP component for focused words lacking expected intonational prominence but no difference for non-focused words with unexpected intonational prominence, as was found for German, a general principle of prosodic processing that distinguishes focused words from non-focused words would be suggested. Otherwise, the differences in reactions to context-prosodic mismatches may well be explained by language-specific prosodic properties.

2. Experiment

Subjects heard short WH-question + answer pairs and made a decision about whether or not the just-heard dialogue made sense as conversation for each trial while their EEGs are recorded.

2.1. Materials

Sixty-four sets of target question-answer pairs were prepared by cross-matching two types of WH-questions and four types of answers differentiating the location of focus-related pitch range expansion. Half of the questions had the WH-word (dare-ga 'who-NOM' or nani-ga 'what-NOM') in the subject position, and the other half contained WH-word (dare-o 'who-ACC' or nani-o 'what-ACC) in the object position. The answer sentences were produced with pitch range expansion either on the subject or on the object noun. An example question-answer set is shown in Table 1. The nominative or topical case makers (-ga or -wa) were matched between the question and the answer in each pair, so that the prosodic match pairs were distinguished from prosodic mismatch pairs only by the intonational properties of the answer sentences. Another 128 question-answer pairs were prepared as distracters. Half of the distracter pairs had WH-questions focusing either on the verb or the adverb (e.g. when and where), half of which had pitch range expansion at the expected focus position while the other half had prominence at some other unexpected location. The rest of the distracter pairs had WH-words focusing on the subject, object, verb, or the adverb but these questions were paired with sentences that did not properly answer them (e.g. Where does Midori practice the piano? – Midori was in the gym between 6 and 8pm.)

Table 1: *Example stimuli set.* (An apostrophe indicates a lexical pitch accent on the preceding mora, and capitalization indicates prosodic prominence)

Subject	Da're-oa kaoi'-o nakushita'-no?		
Wh-Q	Who lost the key?		
Match	MA'SAYA-ga kagi'-o nakushita'-N-da-yo.		
(A1)	MASAYA lost the key.		
Mismatch	Ma'saya-ga KAGI'-o nakushita'-N-da-yo.		
(A2)	Masaya lost the KEY.		
Object	Ma'saya-wa na'ni-o nakushita'-no?		
Object Wh-Q	Ma'saya-wa na'ni-o nakushita'-no? What did Masaya lose?		
Object Wh-Q Mismatch	Ma'saya-wa na'ni-o nakushita'-no? What did Masaya lose? MA'SAYA-wa kagi'-o nakushita'-N-da-yo.		
Object Wh-Q Mismatch (A3)	Ma'saya-wa na'ni-o nakushita'-no? What did Masaya lose? MA'SAYA-wa kagi'-o nakushita'-N-da-yo. MASAYA lost the key.		
Object Wh-Q Mismatch (A3) Match	Ma'saya-wa na'ni-o nakushita'-no? What did Masaya lose? MA'SAYA-wa kagi'-o nakushita'-N-da-yo. MASAYA lost the key. Ma'saya-wa KAGI'-o nakushita'-N-da-yo.		

2.2. Acoustic Properties of Target Materials

To ensure that the expected prosodic differences across the four types of answers were present, duration and F0 peak of the subject and objects were measured for each target answer sentence. Table 2 summarizes the mean across 64 target items. In general, focused words had higher F0 peak and longer duration than unfocused words, but the focus-related differences in pitch range and duration were larger for the object position than in the subject position.

Table 2: Acoustic measurement of critical	words in target
answer sentences	

	Subject	Subject	Object	Object
	Duration	F0 peak	Duration	F0 peak
	(ms)	(Hz)	(ms)	(Hz)
A1	325	365	253	222
A2	309	264	282	360
A3	325	363	221	223
A4	314	267	315	370

When the pitch range was expanded for the subject, the tonal movement for the following object was remarkably compressed (Figure 1), whereas when the object was under focus, the pitch range of preceding subject remained at a intermediate level (Figure 2), preserving the tonal contour for its lexical pitch accent.



Figure 1: F0 track of Subject-focus Answer A1: MA'SAYA-ga kagi'-o nakushita'-N-dayo



Figure 2: F0 track of Object-focus Answer A2: ma'saya-ga KAGI'-o nakushita'-N-dayo

2.3. Participants

Forty two native speakers of Japanese were recruited at the University of Illinois at Urbana-Champaign. Each received a small amount of compensation for participating.

2.4. EEG data Collection

EEG signals were continuously recorded at 200 Hz/16 bit from 24 cap-mounted silver/silver-chloride electrodes. The left mastoid was used as the reference during recording, but the data were later re-referenced to a combination of the Left and Right mastoids. The average impedance measured across the 42 subjects was 3.7 k Ω . The average EEG waveforms were computed for each critical region (i.e., subject & object) referenced to a 100-ms prestimulus baseline.

2.5. Procedures

Subjects were seated in front of a computer monitor in a sound-attenuated booth. They were told to listen carefully to each question-answer pair presented through a set of speakers and make a decision about whether the just-presented dialogue made a good conversation (i.e., whether the answer sentence straightforwardly replied to the preceding Wh-question). A cross '+' was presented in the middle of the screen for the subject to fixate on during the presentation of each question-answer pair. At the offset of the answer sentence, a sign '***' appeared to prompt a response (Left mouse button = made sense vs. Right mouse button = nonsensical) to each pair. Each subject heard four blocks of trials, with each block consisting of 16 target question-answer pairs intermixed with 32 filler pairs. Each session lasted about 1.5-2 hours.

3. Results

Twelve subjects' data were discarded due to either excessive eye movements or experimenter errors in the amplifier calibration process. The average EEGs across thirty subjects were submitted to ANOVAs for both subject and the object positions.

3.1. Missing Pitch Range Expansion for Focused Words

The ERP results manifest clear brain responses to the lack of expected intonational prominence for focused words in Japanese. Figure 3 shows the the ERP waveforms for the two Subject-focus WH-Q conditions (A1 vs. A2).



Figure 3: ERP waveforms for Subject-focus WH-Q-answer pairs (A1 vs. A2)

The ANOVAs showed that when the subject was missing the expected intonational prominence (A2), the waveform starting at the subject was more positive than the properly prominent subject (A1) at four posterior sites (CZ, PZ, CP3, T5, p<.05). Electrode site PZ is shown in Figure 3, illustrating that the largest difference between the two conditions appeared between 250-ms and 500-ms.

At the object position, however, missing intonational prominence did not lead to the same positivity at posterior sites. Instead, there was a trend toward *anterior negativity* across frontal electrodes. Figure 4 shows the two ERP waveforms for the Object-focus WH-Q question-answer pairs (A3 vs. A4) at a left frontal electrode, FT7. Compared to a properly prominent object (A4), a non-prominent object (A3) led to relatively more negative waveforms. A similar marginal effect was observed bilaterally at F3, F7 (left), and at F8 (right).



Figure 4: ERP waveforms for Object-focus WH-Q-answer pairs (A3 vs. A4)

3.2. Unexpected Pitch Range Expansion for Non-Focused Words

In congruence with the previous German study [2], the Japanese data showed no distinctive ERP component evoked by unexpected intonational prominence for non-focused words. At both subject and object positions, the pitch range expansion over a non-focused word (e.g. *MA'SAYA-wa* in A3 and *KAGI'-o* in A2) did not evoke any specific brainwave reactions, as illustrated in Figure 5 for. posterior electrode site PZ.



Figure 5: ERP waveforms for Object-focus WH-Qanswer pairs (A3 vs. A4)

4. Discussion

The Japanese ERP data reported here confirmed the distinction between two different types of pragmatics-prosody mismatch conditions, which was originally reported for German in [2]: When a WH-question led to focus on a specific word but the focused word was presented without intonational prominence, it evoked a distinctive ERP pattern, whereas when non-focused words bore unexpected pitch range expansion, no differences in brain responses were observed as compared to the match condition in which nonfocused words were appropriately produced without such prosodic prominence. The absence of an observable response to unexpected, excessive pitch range expansion was rather surprising at first, given the robust acoustic prominence in the stimuli. Obviously, ERPs do not reflect merely automatic, context-independent reactions to prominent auditory stimuli. The absence of ERP responses to unexpected prosodic prominence instead suggests rapid evaluation of prosodic information against the discourse context.

The present data are by no means sufficient to draw conclusive generalizations about prosodic processing, especially because the ERP patterns observed for missing intonational prominence in Japanese were different (posterior positivity) from those reported for German (posterior negativity [2]). However, the present data may indeed indicate a universal mechanism of processing focused words, which could have surfaced as different ERP patterns due to differences in the experimental tasks across studies, which will be described in more detail below. One possible account for the ERPs evoked by unexpected prosodic prominence is that the degree of prosodic prominence for the focused word is somehow easier to evaluate than the prominence of nonfocused word. Upon listening to a WH-question, listeners may build an expectation about the grammatical category and the prosodic properties of the word that provides the inquiredabout information. As a word that fits in the expected grammatical category is recognized in the answer, its prosodic information may be evaluated immediately against the expected degree of prominence that may be driven from the speaker's overall pitch range. On the other hand, listeners may not establish clear expectation about the prosodic status of non-focused words or the words that are assumed to convey background information in WH-questions (e.g. topical subject Ma'saya-wa in the Object-focus WH-Q and object kagi-o in the Subject-focus WH-Q), as these pieces of background information are often pronominalized or omitted in answer to such questions. Furthermore, a piece of background information sometimes does bear intonational prominence, especially if the first speaker seems to be making an incorrect presupposition and the second speaker corrects it (e.g. Who lost the key? - Nobody lost the KEY, it is the keychain that is missing.) Thus, either intonational prominence for nonfocused words may not be so unexpected, generally, or the prosodic status of non-focused words are less clearly established in listener's mind, or both.

The discrepancy in the kind of ERP response to missing expected accents between the previous German data and the present Japanese data might be attributable to differences in the experimental task between the two studies. In [2], subjects heard all four versions of question-answer combinations for each of 96 stimuli items across 8 blocks, and they were specifically told to judge the appropriateness of the prosodic patterns of the answer sentences. In contrast, in the present study, subjects heard only one of the prosodic versions of a particular question-answer pair, and targets were intermixed among twice as many distracters, which included semantically nonsensical question-answer pairs. The target prosodic mismatch trials were intended to be judged as semantically correct question-answer pairs, as opposed to the semantically nonsensical distractor question-answer pairs. Our subjects no doubt became aware that prosodic mismatch was an important feature of some of the stimuli, but they were still probably less focused on specifically attending to prosody than the subjects in the German studies, which may help explain the different patterns of ERP responses observed in the two studies.

5. Conclusions

The present study demonstrates some advantages of using ERP measures to investigate the on-line processing of prosodic information. However, discrepancies in the nature of the ERP effects observed in a previous German study and the present study suggest the importance of comparable experimental designs and tasks in cross-linguistic investigation. The authors are currently conducting another ERP study in English with a design strictly parallel to the present study. The data from these cross-linguistic studies are expected to provide a clearer picture of the rapid processing of focus-related prosody.

6. References

- Steinhauer, K.; Alter, K.; Friederici, A. D., 1999. Brain responses indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191-196.
- [2] Hruska, C.; Alter, K.; Steinhauer, K.; Steube, A., 2001. Misleading dialogues: Human's brain reaction to prosodic information. Presented at ORAlité et GEstualité. Aix en Provence, France.
- [3] Cruttenden, A., 1986. *Intonation*. Avon: Cambridge University Press.
- [4] Selkirk, E. O., 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT press.
- [5] Kutas, M.; van Petten, C., 1994. In *Handbook of Psycholinguistics*, Gernsbacher, M. A. (ed.). Academic: San Diego. 83-143.
- [6] Garnsey, S. M.; Tanenhaus M. K.; Chapman, R. M., 1989. Evoked potentials and the study of sentence comprehension', *Journal of Psycholinguistic Research: Special Issue on Sentence Processing*, 18, 51-60.
- [7] Pierrehumbert, J.; Beckman, M., 1988. *Japanese tone structure*. Cambridge: MIT Press.
- [8] Ito, K., 2002. The interaction of focus and lexical pitch accent in speech production and dialogue comprehension: Evidence from Japanese and Basque. PhD dissertation. University of Illinois at Urbana-Champaign.