Yasuharu Den*

Some phonological, syntactic, and cognitive factors behind phrase-final lengthening in spontaneous Japanese: A corpus-based study

Abstract: In this study, we investigated segment lengthening in spontaneous Japanese based on a quantitative analysis of a large-scale corpus, focusing on the following three locations at which lengthening frequently occurs: the final segments of (i) clause-initial preface tokens (fillers and conjunctions), (ii) clause-initial *wa*-marked topic phrases, and (iii) clause-final particles. Two cognitive factors, namely clause complexity and boundary depth, were precisely analyzed using statistical models that also accounted for several phonological and syntactic factors. The results showed that in addition to the reliably strong effects of some phonological factors such as the presence of a following pause and the presence of boundary pitch movement, the effects of two cognitive factors were also evident. The way in which lengthening is related to the cognitive factors, however, varies significantly by location and token type. Lengthening of clause-final particles was affected by boundary depth, while lengthening of the topic marker *wa* of clause-initial topic phrases was influenced by clause complexity. Lengthening of the filler *e* was affected by both factors. A significant interaction between the two factors was also observed for the filler *ano*. We discuss the implications of these results as well as agendas for improving the current analysis.

Keywords: lengthening, cognitive factors, phonological and syntactic factors, statistical models, corpus-based study

1 Introduction

1.1 Background

Spontaneous speech produced in a natural setting reveals remarkable characteristics that differ from those of laboratory speech. Among these characteristics are

*Corresponding author: Yasuharu Den,** Chiba University, 1–33 Yayoicho, Inage-ku, Chiba 263–8522, Japan, E-mail: den@chiba-u.jp

various sorts of disfluencies such as unwanted pauses, fillers (such as *uh* and *um*), inserted expressions (such as *I mean* and *you know*), lengthened segments, word fragments, self-repair, and repeated words. Numerous studies have been conducted to describe the distribution of these disfluencies and to elucidate factors underlying them (Maclay and Osgood 1959; Goldman-Eisler 1968; Beattie 1979; Chafe 1980; Levelt 1983, Levelt 1989; Blackmer and Mitton 1991; Clark 1994; Shriberg 1994; Fox et al. 1997; Clark and Wasow 1998; *inter alia*). Most disfluencies are considered evidence of speech planning. When speakers have problems with speech production, they may suspend their speech and insert pauses or fillers before continuing or modify words or phrases they have already produced.

In this study, through a quantitative analysis of a large-scale corpus, we investigate segment lengthening in spontaneous Japanese. Lengthening, also known as *prolongation*, is non-lexical stretching of speech segments. It can occur anywhere in an utterance. We focus on three locations where lengthening frequently occurs in Japanese: the final segments of (i) clause-final particles, (ii) clause-initial preface tokens (fillers and conjunctions), and (iii) clause-initial *wa*-marked topic phrases. The following example illustrates these three locations, where the symbol ":" indicates lengthening of a segment:

(1)   ...   *hakken-suru-n-desu-keredo-**mo:*** (1.1) ***e:***   (0.3) *saiaku-na-no-**wa:*** ...
           find-do-N-POL-but   um   worse-COP-N-TOP

           Clause ending   Preface token   Topic phrase
      '... find (something), but   (1.1) um (0.3) what's worse is that ...'

In Example (1), lengthening occurs at three different locations. The first instance is *mo:* at the final mora of the first clause.[1] Typical words appearing at this location are final and conjunctive particles as well as ending-form verbs, adjectives, and auxiliary verbs.[2] Among these, particles are the most frequent type

---

1 The mora is the basic phonological unit in Japanese and basically consists of a vowel possibly preceded by a consonant, e.g., /ka/, /mo/, and /i/. In addition, there are three types of special morae, which are perceived as bearing an equivalent rhythmic weight as a standard (C)V mora: a moraic obstruent (the second mora of /niQpoN/ 'Japan'), a moraic nasal (the final mora of /niQpoN/), and the second half of a long vowel (the second mora of /kyoHto/ 'Kyoto'). Including these special morae, there are 100 or more types of morae in Japanese.
2 Some conjunctive particles like *keredo* and *ga* can also be used as conjunctions when attached to the beginning of a new clause rather than to the end of a clause. The two usages are distinguished in the corpus used in this study based mainly on prosodic criteria, e.g., particles are, in general, not preceded by pauses while conjunctions are.

and relate to particular usages in spontaneous discourse. The second instance is *e:* at the beginning of the second clause. It is among a set of tokens typically used at the beginning of a new clause in spontaneous discourse – *preface tokens*. They include fillers (such as *e*, *eeto*, *ano*, and *ma*) and conjunctions (such as *de*, *zya*, and *sosite*). The third instance is *wa:* at the final mora of the first substantive phrase of the second clause. At this location, topic phrases, marked by the particle *wa*, often appear.[3] We focus on lengthening of the topic-marker *wa* in clause-initial topic phrases, and other kinds of clause-initial phrases are not considered, since *wa*-marked topic phrases, which are often propositionally uninformative, may be related to some functions in spontaneous discourse.

Factors behind lengthening have been studied from two perspectives, linguistic and cognitive. Linguistic factors have been discussed in phonology and phonetics as well as research on computerized speech synthesis; this research has been based mainly on speech materials collected in experimental settings. Cognitive factors, on the other hand, have been investigated by psycholinguists, corpus linguists, and computational linguists working with spontaneous discourse. We review some of these previous studies in the following two sections.

1.2 Linguistic factors behind lengthening

Numerous studies have been conducted on the syntactic and phonological factors affecting segmental duration in continuous speech.

In the syntactic domain, Klatt (1975) found that syllables are longer at the end of sentences than in other places within the sentence. Additionally, Oller (1973) found *final lengthening*: final syllables are longer than non-final syllables at various levels. Word-final syllables are longer than non-word-final ones, phrase-final syllables are longer than non-phrase-final ones, and sentence-final syllables are longer than non-sentence-final ones. In addition, syllables before pauses have been found to be considerably lengthened, a phenomenon often called *pre-pausal lengthening* (Oller 1973; Klatt 1975).

---

3 In Japanese, arguments of predicates can be omitted when recoverable from the context. In most cases, an omitted element constitutes a topic of the discourse, and could be realized by a topic phrase if it were overtly expressed (Kuno 1973). Since the rate of omitted arguments is quite high in spontaneous Japanese discourse – about 70% (Clancy 1980; Hinds 1983), the rate of overt topic phrases at the clause-initial position, although considerable, is relatively low.

In the phonological domain, Klatt (1975) found that, in general, stressed vowels are longer than unstressed vowels. He also found that emphatic or contrastive stress is accompanied by an increase in the duration of a word (Klatt 1976). In addition, lengthened final syllables often carry a final pitch movement such as rising (Cruttenden 1986). It has also been noted that segmental duration is affected by the surrounding phonological context. For instance, Sagisaka and Tohkura (1984) reported that in Japanese the duration of a vowel is lengthened/shortened when the preceding consonant has a short/long duration – a *compensation* effect. In contrast, Campbell and Isard (1991), and Campbell (1992) proposed a syllable-timing hypothesis, insisting that both a consonant and a vowel in a syllable, or a mora, have similar lengthening characteristics if viewed in terms of variance on the normalized scale. Thus, in Japanese, when the duration of a vowel increases, the preceding consonant is also expected to lengthen to the same degree. Some of these factors, among others, have been applied to computerized speech synthesis. Kaiki and Sagisaka (1992) successfully applied a multivariate statistical model using such variables as (i) the durations of the neighboring phonemes; (ii) the position of the phoneme in the word, the prosodic phrase, and the utterance; (iii) presence of a following pause; (iv) the syntactic category of the word; (v) the inherent duration of the phoneme; and (vi) the overall speech rate of the speaker.

One of the linguistic motivations for speakers to elongate speech segments is to mark syntactic and/or prosodic boundaries. Indeed, this seems to be the major reason for the final and pre-pausal lengthening observed in the data used in the above studies, which came from laboratory speech. If we turn our attention to spontaneous speech, however, we find other factors triggering segment lengthening.

1.3 Cognitive factors behind lengthening

In the past few decades, increasing attention has been paid to spontaneous speech beyond laboratory speech. Numerous studies have been conducted, using quantitative analyses of spontaneous speech from large-scale corpora, to examine disfluencies in spontaneous speech.

As for lengthening, Clark and Fox Tree (2002) examined the difference between two forms of fillers in English, *uh* and *um*, as well as the difference between their normal and prolonged versions. They showed that in both forms the prolonged version is more likely to be immediately followed by a pause than the normal version. Fox Tree and Clark (1997) also showed that the definite

article *the* pronounced as /ðiː/ with a non-reduced vowel is more likely to be immediately suspended than when it is normally pronounced as /ðə/. These studies suggest that lengthening, as well as other outcomes such as choice between different forms and pronunciations, is related to problems in speaking. When speakers have problems in speech production, they may signal an expected delay by lengthening a speech segment being produced.

Although these studies concerned themselves only with lengthening involved in particular tokens in English such as *uh*, *um*, and *the*, other studies have investigated the overall characteristics of lengthening in spontaneous discourse in various languages. Eklund (2001) studied lengthening in Swedish dialog and showed that the rate of lengthening varied as a function of segment type, phonological length, position in the word, word class (open vs. closed), and the domain dependency. Lee et al. (2004), analyzing spontaneous Mandarin conversations, showed that lengthening was often found in word-final, phrase-final, and utterance-medial positions. They also found that lengthening was particularly frequent in transitive verbs, adverbs, nouns, and particles. Den (2003) reported several strategies for lengthening speech segments based on a corpus analysis of spontaneous monologs in Japanese. Japanese speakers frequently prolonged utterance-initial, mono-moraic words such as *de* and *zya*, and sometimes elongated the final vowels of utterance-initial and phrase-final content words. They often prolonged the final vowels of phrase-final function words, which were typically immediately followed by pauses. Though exhaustive, these studies did not directly address the cognitive factors involved in the production of lengthened segments.

Clark and Wasow (1998) analyzed repeated articles in a large corpus of spontaneous conversations in English, directly addressing cognitive factors behind speech disfluencies. They showed that repeated *the* and *a* are more frequent when followed by complex noun phrases than when followed by simple noun phrases. This indicates that speakers are more likely to suspend speech when involved in planning more complex constituents – that is, when facing a heavier cognitive load in planning.

Following Clark and Wasow's (1998) influential study, Watanabe et al. (2006) investigated how the use of clause-initial fillers is affected by clause complexity, measured by the number of words in the clause, and by boundary depth, expressed by the type of the morpho-syntactic form of the preceding clause's ending, using a corpus of spontaneous monologs in Japanese. They showed that at shallower boundaries, the filler rate increases as clauses get longer. Den (2009) applied the same line of analysis to examine the effects of clause complexity and boundary depth on the rate of prolonged clause-initial words in Japanese monologs and found no reliable evidence for these effects. He also conducted a

detailed analysis of a particular token, the conjunction *de*, to examine whether or not its duration was affected by clause complexity. No complexity effect was found. Watanabe and Den (2010) extended this analysis to cover a broader range of clause-initial phrases such as fillers (*ano*, *e*, *eeto*, and *ma*), conjunctions (*de*), and topic phrases (nouns and pronouns marked by the topic particle *wa*). Among these items, only *e*, *eeto*, and topic phrases were found to be affected by clause complexity and/or boundary depth. More recently, Koiso and Den (2013) investigated how the complexity of clauses following shallow boundaries relates to acoustic and linguistic features such as the presence of a clause-initial filler, the presence of boundary pitch movement at the end of the preceding clause, and the duration of the last mora of the preceding clause. They showed that all these features are positively correlated with clause complexity: clauses get longer when they are prefaced by fillers, preceded by boundary pitch movements, or preceded by a stronger degree of final lengthening.

All of these studies shed new light on the research on segment lengthening in spontaneous speech. However, they pose the following problems. First, they lacked systematic control of lower-level factors, such as the inherent durations of phonemes or morae, and possible influences from surrounding phonological contexts. Second, some studies, e.g., Koiso and Den (2013), interpreted the cause-effect relationship between lengthening and cognitive factors differently from the other studies. In Koiso and Den's (2013) study, the degree of lengthening (the duration of the last mora of the clause) was a predictor (cause), while in other studies, it was a response variable (effect). Third, the relationship among lengthening at different locations was not investigated at all. The present study solves some of these problems.

1.4 The purpose of the study

In this study, we investigate segment lengthening in spontaneous Japanese through a quantitative analysis of a large-scale corpus. More specifically, we examine phonological and syntactic factors as well as two cognitive factors, namely clause complexity and boundary depth, that may affect segmental durations at the three locations depicted in Example (1), i.e., the final segments of (i) clause-initial preface tokens (fillers and conjunctions), (ii) clause-initial *wa*-marked topic phrases, and (iii) clause-final particles. Clause complexity is considered to reflect the amount of cognitive load required for planning the clause or utterance. Boundary depth, on the other hand, is thought to indicate qualitative differences involved in the planning process. Planning of local syntax/semantics may proceed at shallower boundaries, while planning of

global discourse structure may be involved at deeper boundaries. We demonstrate how lengthening in spontaneous Japanese is governed not only by phonological and syntactic factors but also by cognitive factors, more precisely, by the planning of syntax/semantics and discourse.

2 Method

2.1 Methodological concerns

Spontaneous speech data is messy. It is always questionable whether results obtained from such messy data are as reliable as those obtained from controlled, experimental data. In order for corpus-based research on spontaneous speech to be dependable, the following prerequisites should be taken into account.

–   The amount and the quality of the data. If the size of a corpus used for quantitative analysis is very small, the data may be distorted due to sampling biases. It is ideal to use a corpus that is representative of the speech under study, but in practice, use of a large-scale corpus in a particular domain is a sub-optimal solution. Fortunately, we have such a corpus in Japanese, the *Corpus of Spontaneous Japanese* (CSJ) (Maekawa 2003). Corpus-based analysis of spontaneous speech also makes heavy use of linguistic annotations such as phoneme boundaries, the parts of speech of words, and the intonation of utterances. If the accuracy of the annotation is not high, the results may not be trustworthy. The CSJ contains a subset of the data, called the *Core*, which comes with rich annotations carefully checked by hand (see Section 2.2).

–   Control of confounding factors. There is a risk of confounding variables when using uncontrolled data. Such variables can interfere with other variables of primary interest. In the present study, phonological and syntactic factors may affect lengthening and interfere with cognitive factors. There are two ways to address confounding variables. The first is to keep constant the values of variables to be controlled. For instance, if we want to study lengthening of fillers and control the influence of their forms and positions in the utterance, we could restrict our data to tokens occurring at the clause-initial position and apply a separate analysis for each form, provided there is enough data for each analysis. The second way is to introduce the variables to be controlled into statistical models. For instance, if we want to study lengthening of clause-final segments and control the

influence of the boundary pitch movements accompanying those segments, we could use the presence of boundary pitch movement as a variable in the statistical model (see Section 2.4).

– Adequate statistical modeling. Even if we have a large corpus of good quality and a way to properly control confounding factors, we also need adequate statistical models for the nature of the data. For instance, to eliminate the influence of differences in inherent segmental durations and speech rates of speakers, the duration data should be normalized by using the mean and the standard deviation calculated for each segment type and for each speaker. The hierarchical nature of corpus data should also be addressed. That is, the entire sample from a corpus is an amalgam of subsets, each taken from a population of utterances produced by a single speaker, who, in turn, is taken from a population of speakers. For such hierarchically organized data, mixed-effects models are becoming a standard tool in psycholinguistics and corpus linguistics (Baayen 2008; Baayen et al. 2008; see Sections 2.4 and 2.6).

2.2 Corpus and annotation

In this study, a subset of the *Corpus of Spontaneous Japanese* (CSJ) (Maekawa 2003) was used. The CSJ is a large-scale corpus of spontaneous Japanese, consisting mainly of monologs. Its *core* includes hand-corrected annotations of various sorts, including clause units, *bunsetsu* phrases, words, phonetic segments, and prosodic information. The CSJ-Core contains monologs from two different recording sources: the academic presentation speech (APS: 70 monologs) and the simulated public speech (SPS: 107 monologs). APS consists of live recordings of academic presentations from meetings of scholars in engineering, the humanities, and the social sciences, while SPS consists of casual 10- to 12-minutes narratives on everyday topics given by laypeople in front of small, friendly audiences. Of these two types, SPS is closer to natural discourse in everyday situations. Thus, the SPS part of the CSJ-Core was used in the present study. There were 54 female and 53 male speakers, ranging in age from their early 20s to their late 60s.

Clause units were basically syntactic clauses ending with either of the following two types of morpho-syntactic forms:

– **Absolute Boundaries (AB)**, which correspond to sentence boundaries in the usual sense.

(2) *Tokyo-ni   iki-masu*
     Tokyo-DAT go-POL
     '(I) will go to Tokyo.'

– **Strong Boundaries (SB)**, which correspond to the boundaries of elemental clauses in compound sentences, typically marked by conjunctive particles such as *ga*, *kedo*, and *keredo*, expressing coordination. In spontaneous Japanese discourse, two or more clauses are connected by these boundaries, which results in a long stretch of chained utterances (Iwasaki and Ono 2001).

(3) *Tokyo-ni   iki-masu-ga*
     Tokyo-DAT go-POL-but
     '(I) will go to Tokyo, but'

Clause units include other types of clauses and phrases that are not accompanied by morpho-syntactic ending-forms but are better viewed as independent utterances. The boundary types of these clause units are classified as follows:

– **Weak Boundaries (WB)**, which correspond to the boundaries of utterances in the form of subordinate clauses that are used without main clauses. The boundary is typically marked by conjunctive particles such as *node, kara,* and *tara,* expressing causality, reason, condition, or other meanings.

(4) *Tokyo-ni   iku-node*
     Tokyo-DAT go-because
     'Because (I) will go to Tokyo.'

– **Non-Clausal Boundaries (NCB)**, which correspond to the boundaries of utterances lacking a main verb, such as one-word or one-phrase utterances.

(5) *Tokyo-ni*
     Tokyo-DAT
     'To Tokyo.'

The above four types of clause boundaries form a scale for degree of boundary depth, from non-clausal boundaries (the shallowest) to absolute boundaries (the deepest). We use these clause-boundary types to measure the cognitive factor of boundary depth.

Clause units are segmented into *bunsetsu* phrases, which consist of one content word possibly followed by one or more function words. Both content

and function words are assigned morphological information such as part of speech, conjugation type and form, and the dictionary form.[4] Below the word level, the starting and the ending times of phonetic segments are precisely identified, which enables us to calculate the durations of units at various levels such as phoneme, mora, word, and clause. Some of these will be used as response variables and predictors in the analyses presented below. When a boundary is uncertain, it is marked as such.

In addition to these syntactic, morphological, and segmental annotations, the CSJ-Core is also annotated for prosody using the X–JToBI scheme (Maekawa et al. 2002). Among the labels of X–JToBI, we focus on the final boundary tones of accentual phrases. The right edge of an accentual phrase is always character-ized by a low tone (L%), which can be followed by an additional tone such as a simple rising tone (H%), a rising-falling tone (HL%), a rising tone with a sustained low (LH%), or a rising-falling-rising tone (HLH%). These extra move-ments at the right edge of accentual phrases are called *boundary pitch move-ments*. The presence of a boundary pitch movement may affect the duration of the segment bearing that tone, since realization of a complex trajectory in a pitch contour generally requires extra time.

All the above-mentioned annotations are compiled into a relational data-base (Koiso et al. 2014) and can be easily accessed via SQL queries. Table 1 shows the summary statistics of the data used in this study.

Table 1: Summary statistics of the data.

Total running time	20 hours
No. of discourses	107
No. of clause units	9,781
No. of *bunetsu* phrases	97,906
No. of long-unit words	198,353

2.3 Unit of analysis

The phonological unit chosen as a domain of lengthening (e.g., phoneme, mora, or syllable) is a crucial decision. For the following reasons, the mora was used as a basic unit of analysis in the present study:

---

4 Two kinds of words with different granularities are identified in the CSJ: short-unit and long-unit words. The former roughly corresponds to headwords in a dictionary, while the latter includes compound and complex words such as "National Institute for Japanese Language and Linguistics." In this study, mainly long-unit words were utilized.

- It is believed that the mora plays a major role in the phonological processing of the Japanese language (Kubozono 1989; Otake et al. 1993; Cutler and Otake 1994; *inter alia*). It is, thus, natural to assume that lengthening operates on the mora, rather than on other units, as its basic processing domain.
- Empirical evidence suggests that both a vowel and consonant in a mora are simultaneously elongated. Figure 1 shows scatter plots between durations of vowels and those of their preceding consonants in our data, with data for different vowel types and their positions in the utterance shown separately.
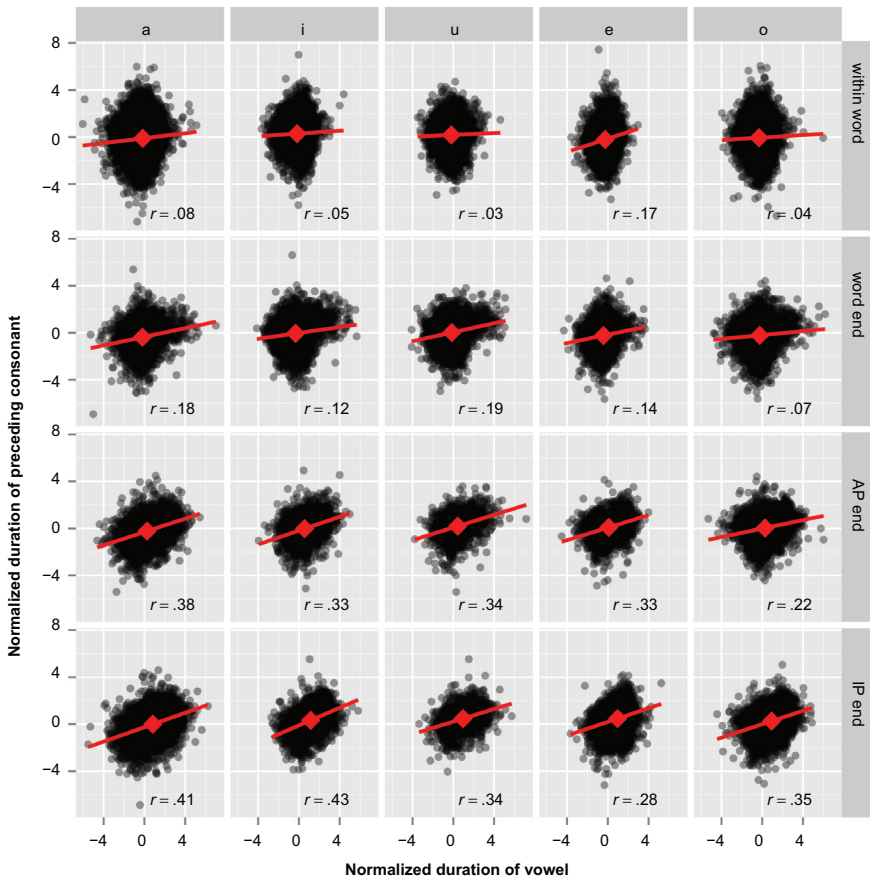
Figure 1: Scatter plots between the normalized durations of vowels and their preceding consonants, separately shown with respect to each vowel type and each position in the utterance. Solid lines represent regression lines, and diamonds the centers of distributions. *r*s indicate correlation coefficients.

The duration values were transformed into z-scores for each phoneme type and each speaker, using a similar formula to Equation (6) below. There was no evidence that vowel duration compensated for consonant duration. Moreover, at the end of accentual phrases (APs) and intonation phrases (IPs), which were most common in the three locations we focused on, the durations of both vowels and consonants were longer on average than their normal durations, and they were positively correlated. For instance, the mean z-scores of the vowel /o/ and its preceding consonant at the IP end are 0.91 and 0.26, respectively, and their correlation coefficient is $r = 0.35$ (the bottom-right panel in Figure 1). Therefore, lengthening seems to operate on a whole mora rather than on a vowel alone.

– In standard Japanese, high vowels, /i/ and /u/, in their pre-pausal positions, are often devoiced following a voiceless consonant. It has been noted that in spontaneous Japanese, these high vowels are also devoiced in other environments, although at a lower rate, and that devoicing of non-high vowels is not uncommon (Maekawa and Kikuchi 2005). When vowels are devoiced, their starting boundaries are quite likely to be uncertain (over 90% of the time), which prevents us from using cases involving de-voiced vowels in analysis of durations. Using the mora as a unit of analysis enables us to cover these cases.

2.4 Variables for statistical analysis

2.4.1 Response variables

In each of the subsequent analyses, the duration of the final mora at a particular location was used as a response variable. The locations examined were (i) clause-initial preface tokens (fillers and conjunctions), (ii) clause-initial *wa-* marked topic phrases, and (iii) clause-final particles, as shown in Example (1).

The duration data were transformed into z-scores, using the mean and the standard deviation calculated for each mora type and for each speaker, to eliminate the influence of variance in the inherent segmental durations and in the speech rates of speakers. More precisely, a raw duration, d_{ij}, of ith speaker's jth token was converted into a z-score, z_{ij}, using the following formula (cf. Campbell and Isard 1991):[5]

---

5 Campbell and Isard (1991) used duration values on a linear scale instead of log-transformed values. The distribution of durations in our data was right-skewed, however, and use of log-transformed values seemed more adequate.

(6)

$$z_{ij} = \frac{\log d_{ij} - \mu_{i,\, type(j)}}{\sigma_{i,\, type(j)}}$$

where *type(j)* is the mora type of the *j*th token, and μ_{it} and σ_{it} are the mean and the standard deviation of (log-transformed) raw durations for the speaker *i* and the mora type *t*, respectively.

Morae with uncertain starting or ending boundaries were excluded from the above calculation as well as from the subsequent analyses.

2.4.2 Predictors concerning cognitive factors

In this study, we examined two cognitive factors related to lengthening: clause complexity and boundary depth. Clause complexity is considered to reflect the amount of cognitive load required to plan an upcoming clause. Boundary depth, on the other hand, is thought to indicate qualitative differences in the planning process.

Clause complexity can be estimated by the number of words, syntactic nodes, or phrasal nodes in the clause. These numbers are found to be highly correlated (Wasow 1997). In the present study, the duration of the clause unit, which is also highly correlated with these numbers, was used. The duration of the clause unit was obtained from the region (the gray region in Figure 2) excluding initial preface tokens, initial topic phrases, and disfluencies (word fragments and fillers) immediately following topic phrases. This region of the clause unit, called the *body*, conveys the substantial content of the clause, and its duration is considered to reflect cognitive load in planning the meaning and the structure of the clause. The duration values were log-transformed to better approximate the normal distribution.
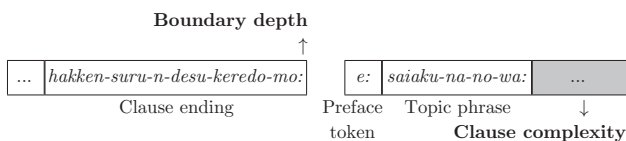
Figure 2: Predictors of the model. Clause complexity is calculated from the duration of the gray region (the body) in the second clause unit, while boundary depth is determined by the end-form of the first clause unit.

Boundary depth, on the other hand, was measured by the clause-boundary types described in Section 2.2 (Figure 2). The following order on the scale of the boundary depth was assumed:

(7)　　Absolute > Strong > Weak > Non-Clausal

Planning of local syntax/semantics may proceed at shallower (weak and non-clausal) boundaries, while planning of global discourse structure may occur at deeper (absolute and strong) boundaries.

The two cognitive variables are not independent of each other. The duration of the clause body varied depending on the clause-boundary type – the mean durations were AB: 4.50 s, SB: 4.99 s, WB: 4.73 s, and NCB: 3.96 s for the overall data, corrected for per-speaker variance. This suggests that an interaction between the two predictors may be considered in statistical analysis.

2.4.3 Predictors for phonological and syntactic factors

In addition to cognitive factors, we also examined phonological and syntactic factors that may interfere with cognitive factors. The following predictors concerning phonological and syntactic factors were used in some analyses; e.g., the type of the vowel in the target mora was used in the analysis of clause-final particles, but was not used in the analysis of *wa*-marked topic phrases, since in the latter case, the vowel type is always fixed to /a/.

– Type of the vowel in the mora. Although the influence of inherent duration was nullified by normalization using Equation (6), lengthening tendencies may vary by mora type. Since the number of mora types in Japanese is quite large (over 100), and vowels play a larger role in lengthened morae, the type of the vowel in the target mora, rather than the type of the mora, was used as a predictor. The presence of vowel devoicing at the target mora was also considered. When the vowel in the target mora is devoiced, the preceding consonant may lengthen or shorten instead. A special vowel type /V/ 'devoiced' was used instead of a separate variable for devoicing. That is, when the target mora involved a devoiced vowel, its vowel type was labeled as /V/. When the target mora did not involve a devoiced vowel, its vowel type was labeled, in a usual way, as /a/, /i/, /u/, /e/, /o/, or /N/.[6]

– Type of the vowel in the preceding mora. The surrounding phonological context has an effect on segmental duration. Since we are concerned with phrase-final positions, only the influence of the preceding mora was

---

6 /N/ is not a vowel but is included in the list, since it is the only part within this special mora that can lengthen or shorten. The other two special morae, /Q/ and /H/, never appear at the position being analyzed, but may appear at other positions (see "Type of the vowel in the preceding mora" below).

considered. The type of the vowel in the preceding mora was labeled in the same way as the type of the vowel in the target mora, except that the former had one more category, /H/ 'the second half of a long vowel.'

– Syntactic category of the word. The syntactic category of the word was found to affect segmental durations (Kaiki and Sagisaka 1992). In the analysis of clause-final particles, the syntactic sub-category was considered, which was either of the following five categories: final particles (*ne*, *yo*, etc.), conjunctive particles (*te*, *ga*, *kedo*, etc.), case particles (*ga*, *ni*, etc.) including quotational particles (*to*), adverbial particles (*mo*, *toka*, etc.), and topic particles (*wa*). (For other locations, word forms, and hence their syntactic categories, were fixed.)

– Number of morae in the word. The moraic length of the word may also influence the duration of the last mora of that word. Since most clause-final particles are mono-moraic or bi-moraic, only three categories, '1,' '2,' and '3 + ,' were distinguished. (For other locations, word forms, and hence their moraic lengths, were fixed.)

– Syntactic category of the head word in the phrase. In the analysis of *wa*-marked topic phrases, the syntactic category of the head word, noun or pronoun, was used as a predictor. Since the distribution of these syntactic categories varied by the clause-boundary type of the preceding clause (the rates of pronouns were AB: 44.4%, SB: 37.2%, WB: 39.9%, and NCB: 44.0%, corrected for per-speaker variance), an interaction between the syntactic category of the head word and clause-boundary type was also considered.

– Presence of a following pause. Pre-pausal lengthening is one of the major sources of prolonged segments. There were a considerable number of instances in which phrase boundaries being analyzed were not followed by a pause; in particular at boundaries other than clause-final ones, "no-pause" was the majority. Thus, this binary variable was one of the most important predictors of the analysis. Since the probability of the presence of a following pause varied by clause-boundary type (e.g., for pauses following clause-final particles, the probabilities were AB: 94.3%, SB: 91.4%, WB: 91.5%, and NCB: 94.4%, corrected for per-speaker variance), an interaction between the presence of a following pause and clause-boundary type was also considered.

– Presence of boundary pitch movement. This is an important predictor as well, since morae bearing complex pitch movements tend to be long. The probability of the presence of boundary pitch movement also varied by clause-boundary type (e.g., for boundary pitch movements in clause-final particles, the probabilities were AB: 78.0%, SB: 79.4%, WB: 58.7%, and

NCB: 54.2%, corrected for per-speaker variance), and thus, an interaction between the presence of boundary pitch movement and clause-boundary type was considered.

One may note that an important phonological factor is missing in the above list – the duration of the preceding mora. The influence of the surrounding phonological context is an important factor to be considered. In the above list, the type of the vowel in the preceding mora is included as a predictor, but the duration of the preceding mora is not included. Although it may be important, the use of the duration of the preceding mora as a predictor could lead to wrong conclusions: If the duration of the preceding mora is itself affected by cognitive factors, the true effect of these factors on the target mora might be hindered by an indirect effect of the preceding mora on the target mora. We will discuss this issue in more detail in Section 4.3.

2.5 Other predictors

When speakers have several alternative locations for lengthening, their choice may be influenced by how many alternatives they have. For instance, when a filler precedes a topic phrase, the speaker may choose to prolong the filler instead of the topic phrase. Thus, the probability of lengthened topic phrases may be lower when preface tokens precede them than when they do not. To see such dependence between different lengthening locations, the presence of a preface token and the presence of a topic phrase were used as predictors.

2.6 Statistical models

To identify factors affecting lengthening in the three locations, a linear model with the two cognitive predictors as well as phonological and syntactic predictors was separately applied to a subset of the data for each location. Since the data were heterogeneously clustered by speakers, linear mixed-effects models, rather than ordinary linear models, were applied, as recommended in recent psycholinguistic and corpus-linguistic research (Baayen 2008; Baayen et al. 2008). Model parameters were obtained by restricted maximum likelihood (REML) estimation using the *lme4* package (Bates et al. 2014) of the R language (R Core Team 2014). Significance was tested via *p*-values computed by likelihood-ratio tests using the *drop1* function in the *lme4* package as well as by least-squares means for factors using the *lsmeans* package (Lenth 2015).

A random intercept for speakers was used throughout the analyses; a random slope of the clause duration was not used, since it was not significant in any analysis. In the analysis of clause-final particles, a random intercept for word forms was also considered another random effect. In fact, there was large variance in the distribution of final mora duration of clause-final particles according to word form. Taking this variance into account enhances the confidence of the analytic results.

3 Results

The following three sections report on the results for lengthening in each of the three targeted locations. In each section, the data screening process and predictors used are first described, and then statistical results are provided.

3.1 Lengthening of clause-initial preface tokens

3.1.1 Data screening

There were 8,982 pairs of clause units, like the one in Figure 2, in which neither the first nor the second clause unit was a false start or a disrupted utterance. Of these, 5,522 cases (61.5%) contained preface tokens at the initial position of the second clause unit.

To ensure precise analysis of segmental durations and control over phonological and syntactic conditions, cases meeting either of the following conditions were excluded from the analysis:

– The clause-initial preface token was followed by another preface token, i.e., in the case of multiple occurrence.
– The last mora of the preface token had an uncertain boundary or was pronounced in a non-canonical way.

This procedure left us 2,823 cases, 51.1% of the entire preface token data. Table 2 shows the top 10 items in these cases. The first four items, i.e., the conjunction *de* and the fillers *e*, *ma*, and *ano*, constituted 72% of the data, and were used for the subsequent analyses (*de* after AB: 367, SB: 141, NCB: 19; *e* after AB: 253, SB: 230, WB: 84, NCB: 42; *ma* after AB: 150, SB: 289, WB: 121, NCB: 31; *ano* after AB: 66, SB: 180, WB: 48, NCB: 22). Since fillers in different forms behave differently (Watanabe 2009), and conjunctions have different characteristics from fillers (Den 2009), separate models were applied for the forms of preface tokens.

Table 2: Top 10 items of preface tokens and their frequency counts.

Token	Frequency
e (filler)	609
ma (filler)	591
de (conj.)	527
ano (filler)	316
eeto (filler)	98
nde (conj.)	70
n (filler)	64
demo (conj.)	59
a (filler)	48
sorede (conj.)	43

3.1.2 Predictors

All models included the following predictors in addition to the cognitive predictors (clause complexity [the duration of the second clause's body] and boundary depth [the clause-boundary type of the first clause unit]):

– Presence of a pause immediately following the preface token
– Presence of a *wa*-marked topic phrase following the preface token

The presence of boundary pitch movement was not used, since in the preface token data, boundary pitch movements were rarely assigned (only three instances in the filler *ma* data). The other predictors described in Section 2.4.3 were not used either, since their values were fixed in the current analyses (e.g., the type of the vowel in the target mora was always /e/, /a/, or /o/, depending on the four forms of preface tokens being analyzed).

For each of the four preface-token items, a linear mixed-effects model with an interaction between clause complexity and boundary depth as well as an interaction between the presence of a following pause and clause complexity was first applied. Then, a likelihood-ratio test comparing the model with the interaction terms to the model excluding each of the interaction terms was conducted. Only when the test revealed significance did the interaction remain in the model.

3.1.3 Statistical results

Figure 3 shows scatter plots between the (normalized) duration of (the last mora of) the preface token and the (log) duration of the clause body, i.e., clause
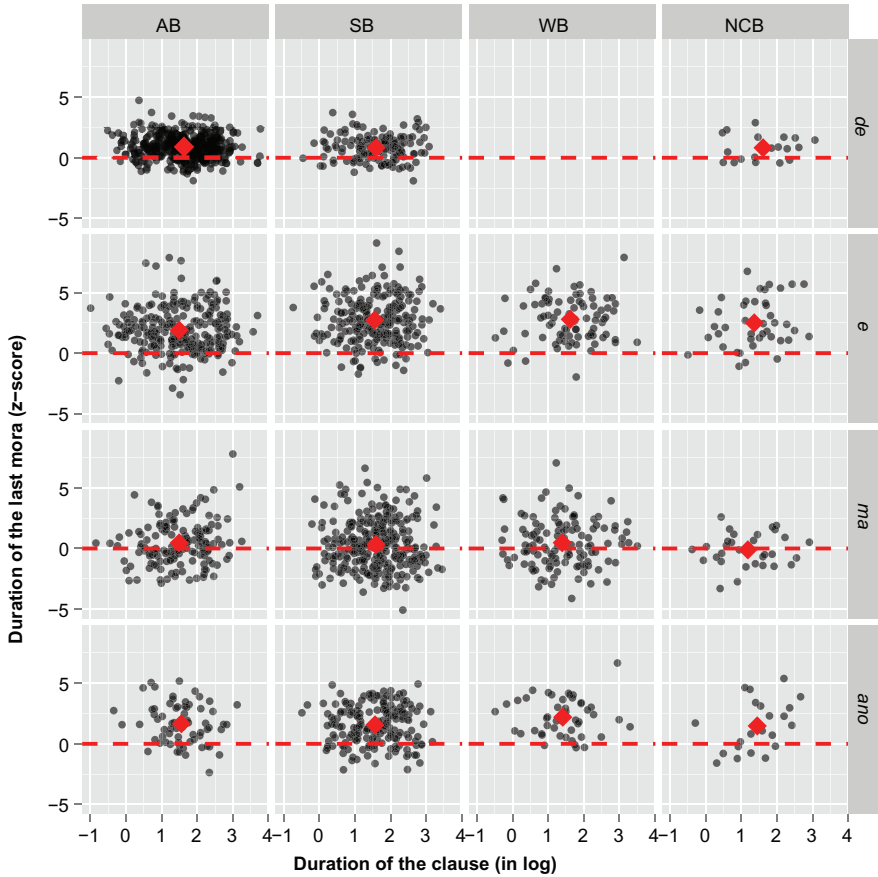
Figure 3: The (normalized) duration of the last mora of the preface token vs. the (log) duration of the clause body (clause complexity) relative to the clause-boundary type of the preceding clause unit (boundary depth), separately shown with respect to each preface-token item, *de* (top row), *e* (second row), *ma* (third row), and *ano* (bottom row). Dashed lines indicate the baseline, i.e., zero on the normalized scale, and diamonds indicate the centers of distributions.

complexity, relative to the clause-boundary type of the preceding clause unit, i.e., boundary depth. The top row gives the results for the conjunction *de*, the second row for the filler *e*, the third row for the filler *ma*, and the bottom row for the filler *ano*. In general, the duration of (the last mora of) the preface token shifted positively on the normalized scale, meaning that it lengthened from its normal duration. This tendency, however, was not so clear for the filler *ma*.

Likelihood-ratio tests revealed a significant interaction between clause complexity and boundary depth only for the filler *ano* (*de*: $\chi^2(2) = 0.45$, $p = 0.80$; *e*: $\chi^2(3) = 5.07$, $p = 0.17$; *ma*: $\chi^2(3) = 3.25$, $p = 0.35$; *ano*: $\chi^2(3) = 11.41$, $p < 0.01$). Furthermore, no significant interaction between the presence of a following pause and boundary depth was found for any item (*de*: $\chi^2(2) = 2.48$, $p = 0.29$; *e*: $\chi^2(3) = 1.30$, $p = 0.73$; *ma*: $\chi^2(3) = 2.00$, $p = 0.57$; *ano*: $\chi^2(3) = 7.75$, $p = 0.052$). Thus, for *ano*, a model with an interaction between clause complexity and boundary depth was applied, while for the remaining items, models without interaction terms were applied.

3.1.3.1 The conjunction *de*

Figure 4 shows the effect displays for the four predictors of the model fitted to the data for the conjunction *de*, created by using the *effects* package (Fox et al. 2014) of the R language.[7] An effect display generally illustrates fitted values under the model computed by absorbing the lower-order effects marginal to a high-order term and by averaging over other terms in the model (Fox 2003). In the current case, since no interaction term is included in the model, the effect display of each predictor is computed by fixing the values of other predictors at typical values, i.e., their means.

No clear tendency was observed for either of the two cognitive factors, i.e., the (log) duration of the clause body (the top-left panel) and the clause-boundary type of the preceding clause unit (the top-right panel). The other two predictors, however, revealed reliable effects on the (normalized) duration of the mora /de/: In the presence of a following pause, the duration tended to be longer (the bottom-left panel); and in the presence of a following topic phrase, the duration tended to be shorter (the bottom-right panel).

Investigation of the estimated model parameters, as well as the analysis of deviance table, indicate these points more clearly (Table 3).[8] Neither the effect of clause duration nor the effect of clause-boundary type was significant

---

7 Some of the displays are superimposed by kernel density plots of the observed data. Kernel density plots are similar to histograms, but have properties such as smoothness or continuity by using a suitable kernel. In Figure 4, kernel density plots are vertically placed on both sides of the estimated mean and its 95% confidence interval.

8 The left half of Table 3 shows the summary statistics of the model (estimated coefficients, standard errors, and *t*-values), and the right half shows the analysis of deviance table (χ^2 statistics, degrees of freedom, and *p*-values for fixed effects). Note that for a categorical variable with more than two levels, a treatment contrast is used; the treatment contrast compares each level of the categorical variable to a base reference level, e.g., SB vs. AB of the clause boundary type. In the analysis of deviance table, the statistics for a categorical variable are displayed on the first of the rows for such comparisons.
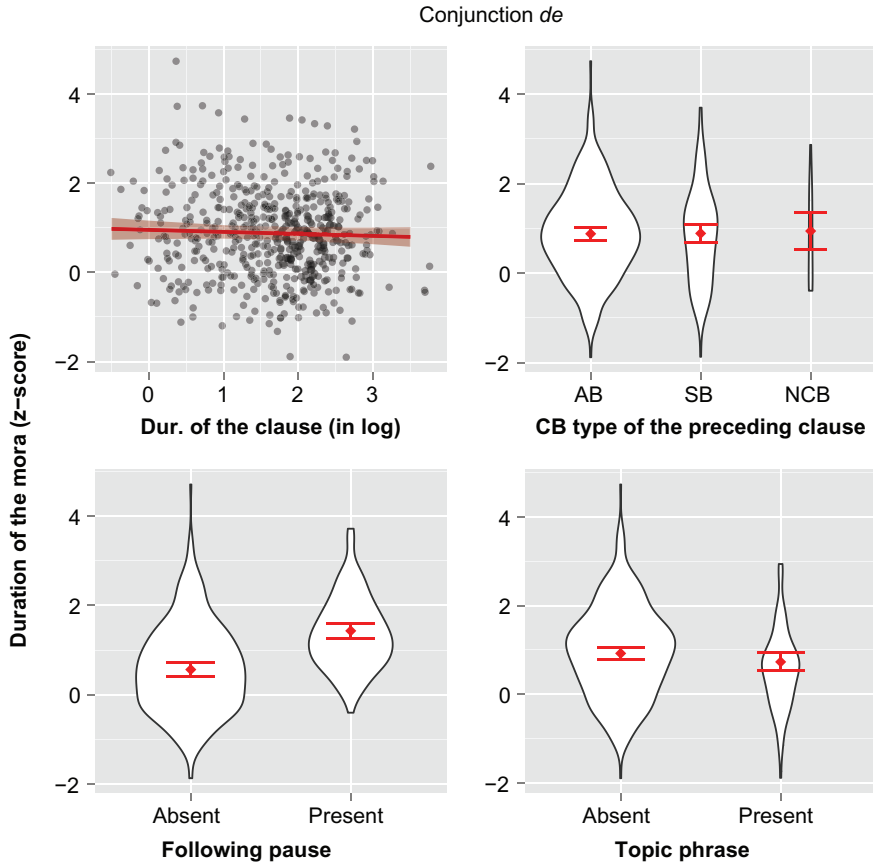
Figure 4: Effect displays for the four predictors of the *de* model. The top-left panel shows the estimated slope coefficient for the (log) duration of the clause body, with its 95% confidence interval, which was computed by averaging over other terms in the model, superimposed by a scatter plot of the observed data. The other three panels show the estimated coefficients for categorical predictors, the clause-boundary type of the preceding clause unit (top-right), the presence of a following pause (bottom-left), and the presence of a topic phrase (bottom-right), with their 95% confidence intervals, which were computed by averaging over other terms in the model, superimposed by kernel density plots of the observed data.

according to a likelihood-ratio test (Clause Duration: $\chi^2(1) = 0.90$, $p = 0.34$; CB Type: $\chi^2(2) = 0.09$, $p = 0.96$). The effects of the following pause and the topic phrase, on the other hand, were both significant (Following Pause: $\chi^2(1) = 101.07$, $p < 0.001$; Topic Phrase: $\chi^2(1) = 4.29$, $p < 0.04$). The estimated coefficient for the former was positive, 0.867, meaning that *de* was longer when immediately followed by a pause than when it was not. In contrast, the estimated coefficient for the latter was

Table 3: Estimated parameters of the *de* model and its analysis of deviance table. σ_S indicates the standard deviation of the random intercept for speakers, and σ the residual standard deviation.

	Coef.	SE	t-value	χ^2	df	p-value
Intercept	0.671	0.116	5.79			
Clause duration	−0.045	0.048	−0.94	0.90	1	0.34
CB type$_{SB \text{ vs. } AB}$	0.008	0.095	0.09	0.09	2	0.96
CB type$_{NCB \text{ vs. } AB}$	0.062	0.207	0.30			
Following pause	0.867	0.082	10.56	101.07	1	<0.001
Topic phrase	−0.192	0.093	−2.07	4.29	1	<0.04
						($\sigma_S = 0.49$, $\sigma = 0.81$)

negative, −0.192, meaning that *de* was lengthened less when followed by a *wa*-marked topic phrase than without such a phrase.

3.1.3.2 The filler *e*

Figure 5 shows the effect displays for the two cognitive factors of the model fitted to the data for the filler *e*.[9] The effects of the two predictors may not be obvious in the plots, but investigation of the estimated model parameters,
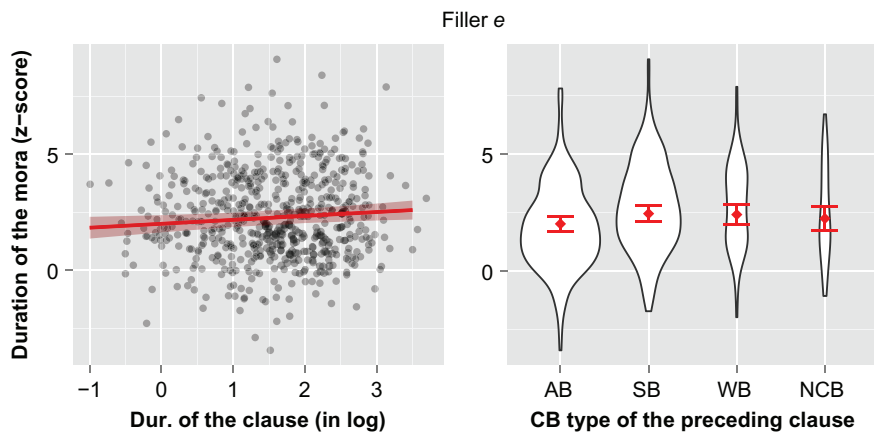
Figure 5: Effect displays for the two cognitive predictors of the *e* model. These panels show similar plots to those in the upper part of Figure 4.

---

9 Hereafter, the effect displays for dichotomous predictors, such as the presence of a following pause, are not shown, because these effects can be easily understood.

Table 4: Estimated parameters of the *e* model and its analysis of deviance table.

	Coef.	SE	*t*-value	χ^2	df	*p*-value
Intercept	1.372	0.211	6.52			
Clause duration	0.169	0.072	2.35	5.52	1	< 0.02
CB type$_{SB\ vs.\ AB}$	0.437	0.141	3.10	10.51	3	< 0.02
CB type$_{WB\ vs.\ AB}$	0.397	0.195	2.04			
CB type$_{NCB\ vs.\ AB}$	0.230	0.249	0.92			
Following pause	1.693	0.146	11.61	122.75	1	< 0.001
Topic phrase	−0.166	0.154	−1.08	1.17	1	0.28
						($\sigma_S = 1.19$, $\sigma = 1.40$)

shown in Table 4, revealed significant effects for both clause duration and clause-boundary type (Clause Duration: $\chi^2(1) = 5.52$, $p < 0.02$; CB Type: $\chi^2(3) = 10.51$, $p < 0.02$). Pairwise comparisons of the estimated coefficients for the four levels of the clause-boundary type were performed by using the *lsmeans* package (Lenth 2015) of the R language, with *p*-values adjusted for the multiplicity of the levels involved by the Tukey method. A significant difference was found only between the absolute and the strong boundaries (adjusted $p < 0.02$), showing that after strong boundaries, *e* tended to be more lengthened than after absolute boundaries.

For the other two predictors, a strong effect of a following pause was replicated, but topic phrase did not reveal a significant effect (Following Pause: $\chi^2(1) = 122.75$, $p < 0.001$; Topic Phrase: $\chi^2(1) = 1.17$, $p = 0.28$).

3.1.3.3 The filler *ma*

Figure 6 shows the effect displays for the two cognitive factors of the model fitted to the data for the filler *ma*, and Table 5 shows the estimated model parameters and the analysis of deviance table. There was a tendency for the duration of *ma* to increase slightly as a function of the clause duration, which was marginally significant ($\chi^2(1) = 3.64$, $p < 0.06$); *ma* was slightly more lengthened when clause duration became longer. In addition, the effect of the following pause was, again, highly significant ($\chi^2(1) = 140.33$, $p < 0.001$). The effects of the other two predictors were not significant (CB Type: $\chi^2(3) = 5.14$, $p = 0.16$; Topic Phrase: $\chi^2(1) = 0.15$, $p = 0.69$).

3.1.3.4 The filler *ano*

In contrast to the three preceding cases, the model fitted to the data for the filler *ano* included an interaction between clause complexity and boundary depth, as
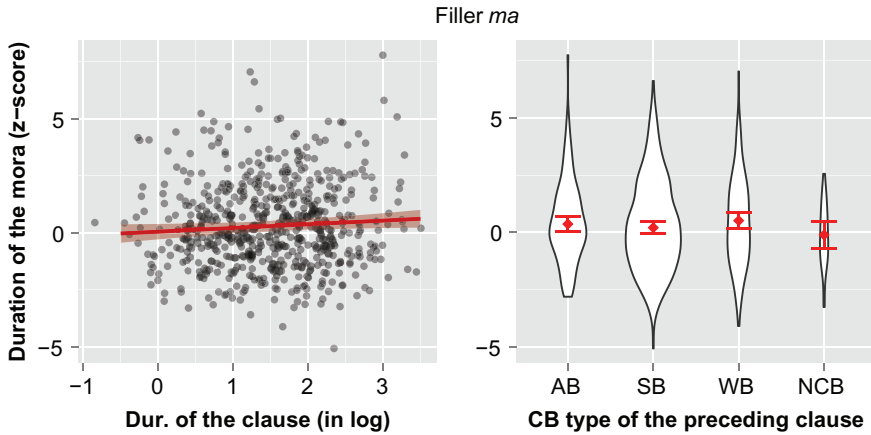
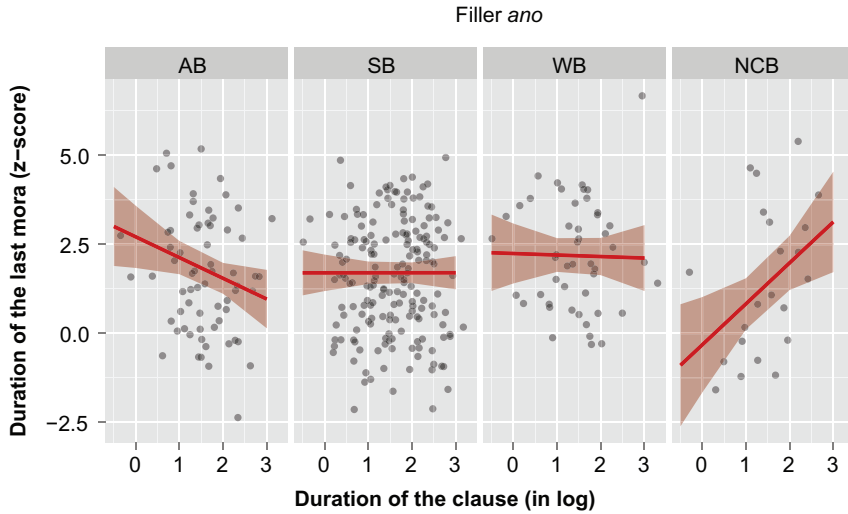Figure 6: Effect displays for the two cognitive predictors of the *ma* model.

Table 5: Estimated parameters of the *ma* model and its analysis of deviance table.

	Coef.	SE	t-value	χ^2	df	p-value
Intercept	−0.340	0.211	−1.61			
Clause duration	0.162	0.085	1.91	3.64	1	0.057
CB type$_{SB \text{ vs. } AB}$	−0.157	0.170	−0.92	5.14	3	0.16
CB type$_{WB \text{ vs. } AB}$	0.150	0.206	0.73			
CB type$_{NCB \text{ vs. } AB}$	−0.475	0.324	−1.47			
Following pause	2.111	0.169	12.52	140.33	1	< 0.001
Topic phrase	0.071	0.184	0.39	0.15	1	0.69
				($\sigma_S = 0.73$, $\sigma = 1.55$)		

described at the beginning of this section. The interaction is evident in the effect display shown in Figure 7: after absolute boundaries, the duration of the last mora /no/ of the filler *ano* revealed a clear tendency to decrease as clause duration increased, while after non-clausal boundaries, this duration showed a clear increasing trend. Table 6 shows the estimated model parameters and the analysis of deviance table. These show that the interaction between clause complexity and boundary depth, as well as the effect of the following pause, are significant, as shown by likelihood-ratio tests (Interaction: $\chi^2(3) = 12.78$, $p < 0.006$; Following Pause: $\chi^2(1) = 76.51$, $p < 0.001$). The estimated slope coefficient for the clause duration was computed for each clause-boundary type by using the *lstrends* function in the *lsmeans* package. It was reliably estimated to be negative when *ano* appeared after absolute boundaries (slope = −0.583, $p < 0.03$) and to be positive

Figure 7: Effect display for the interaction between the two cognitive predictors of the *ano* model. The figure shows the estimated slope coefficients for the (log) duration of the clause body, with their 95% confidence intervals, relative to the clause-boundary type of the preceding clause unit, which were computed by absorbing the main effects of these two variables and averaging over other terms in the model, superimposed by scatter plots of the observed data.

Table 6: Estimated parameters of the *ano* model and its analysis of deviance table.

	Coef.	SE	t-value	χ^2	df	p-value
Intercept	2.113	0.444	4.76			
ClauseDur: CBType$_{SB \text{ vs. } AB}$	0.584	0.290	2.01	12.78	3	< 0.006
ClauseDur: CBType$_{WB \text{ vs. } AB}$	0.541	0.366	1.48			
ClauseDur: CBType$_{NCB \text{ vs. } AB}$	1.735	0.492	3.53			
Clause duraton	−0.583	0.257	−2.27			
CB type$_{SB \text{ vs. } AB}$	−1.012	0.502	−2.02			
CB type$_{WB \text{ vs. } AB}$	−0.468	0.608	−0.77			
CB type$_{NCB \text{ vs. } AB}$	−3.034	0.811	−3.74			
Following pause	1.539	0.168	9.16	76.51	1	< 0.001
Topic phrase	0.025	0.245	0.10	0.01	1	0.92
				$(\sigma_S = 0.51, \sigma = 1.36)$		

when *ano* appeared after non-clausal boundaries (slope = 1.151, $p < 0.007$); after the other two types of boundaries, the estimated coefficients were not found to be significantly different from zero (SB: slope = 0.001, $p = 1.00$; WB: slope = −0.043, $p = 0.87$).

3.2 Lengthening of clause-initial *wa*-marked topic phrases

3.2.1 Data screening

Of 8,982 pairs of consecutive clause units, 1,544 cases (17.2%) contained *wa*-marked topic phrases at the initial position of the second clause unit, possibly preceded by preface tokens. In searching for the topic phrase data, those containing relative clauses were not considered, since these clauses are likely to constitute the substantial content of the clause and hence become more like a part of the clause body than time-gaining elements for the planning of the clause.

To ensure precise analysis of segmental durations and control over phonological and syntactic conditions, cases meeting either of the following conditions were excluded from the analysis:

–   The topic phrase was composed of more than one *bunsetsu* phrase, i.e., in the case of compound or complex phrases.
–   The syntactic category of the head word (*wa*-marked word) was not a noun or pronoun.
–   The last mora, /wa/, of the topic phrase had an uncertain boundary or was pronounced in a non-canonical way.
–   The last mora, /wa/, of the topic phrase was not tightly connected to the preceding mora – that is, a pause occurred between them.

This procedure left us 852 cases (AB: 418, SB: 295, WB: 86, NCB: 53), 55.2% of the entire topic phrase data.

3.2.2 Predictors

In the present analysis, the following five predictors were used in addition to the two cognitive predictors:

–   Type of the vowel in the mora preceding the last mora /wa/
–   Syntactic category of the head word of the topic phrase: noun or pronoun
–   Presence of a pause immediately following the topic phrase
–   Presence of boundary pitch movement at the right edge (within the mora /wa/) of the topic phrase
–   Presence of a preface token preceding the topic phrase

Note that the other predictors described in Section 2.4.3 had constant values and hence were not used.

3.2.3 Statistical results

Figure 8 shows scatter plots between the (normalized) duration of the last mora, /wa/, of the topic phrase and the (log) duration of the clause body (clause complexity), relative to the clause-boundary type of the preceding clause unit (boundary depth). In general, the duration of *wa* in the topic phrase shifted positively on the normalized scale, lengthening from its normal duration.
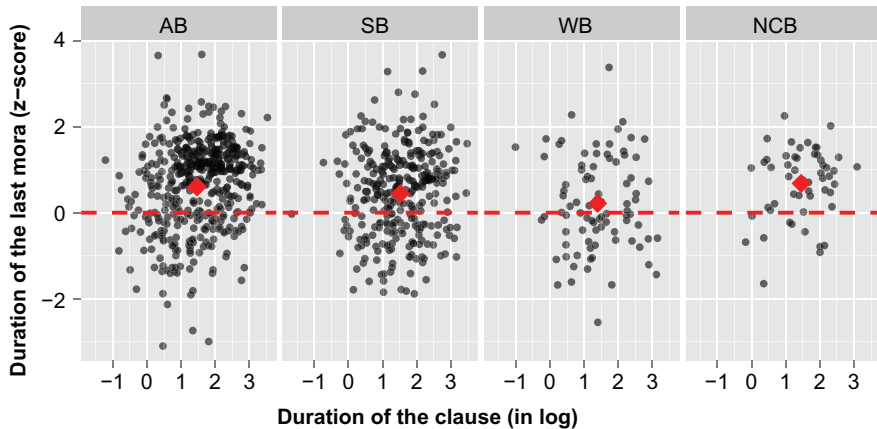
Figure 8: The (normalized) duration of the last mora, /wa/, of the topic phrase vs. the (log) duration of the clause body (clause complexity) relative to the clause-boundary type of the preceding clause unit (boundary depth). Dashed lines indicate the baseline (zero on the normalized scale) and diamonds indicate the centers of distributions.

Likelihood-ratio tests compared the model with interactions between the clause-boundary type and other four predictors – clause duration, the syntactic category of the head word, the presence of a following pause, and the presence of boundary pitch movement (BPM) – with the four models, each excluding one of these interaction terms. No interaction terms revealed significant results (Clause Duration: $\chi^2(3) = 1.23$, $p = 0.75$; Syn. Head: $\chi^2(3) = 3.40$, $p = 0.33$; Following Pause: $\chi^2(3) = 2.71$, $p = 0.44$; BPM: $\chi^2(3) = 3.74$, $p = 0.29$). Thus, a model without interaction terms was applied.

Figure 9 shows the effect displays for the two cognitive factors of the model fitted to the *wa* data. An increasing trend in the duration of *wa* according to an increase of clause duration may be clear. On the other hand, the clause-boundary type of the preceding clause seemed to have no effect. Table 7 shows the estimated model parameters and the analysis of deviance table. The
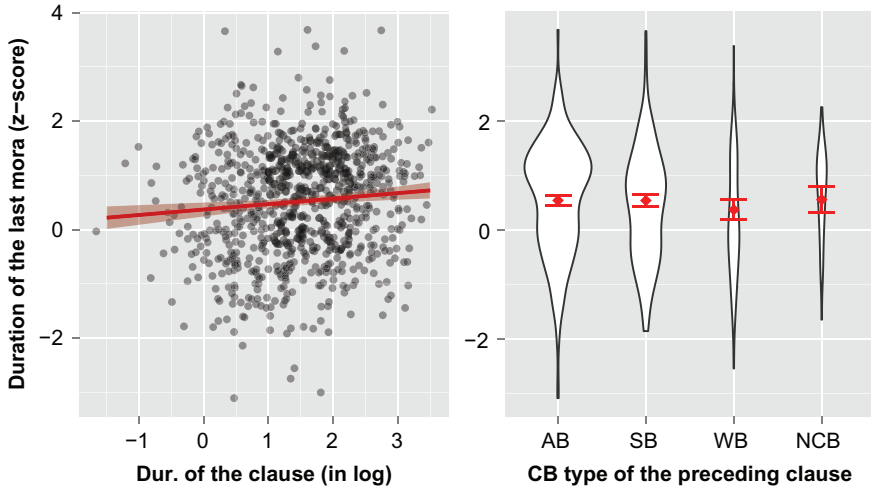
Figure 9: Effect displays for the two cognitive predictors of the *wa* model.

Table 7: Estimated parameters of the *wa* model and its analysis of deviance table.

	Coef.	SE	t-value	χ^2	df	p-value
Intercept	−0.183	0.134	−1.36			
Clause duration	0.101	0.033	3.02	9.16	1	< 0.003
CB type$_{SB\ vs.\ AB}$	−0.001	0.066	−0.02	3.19	3	0.36
CB type$_{WB\ vs.\ AB}$	−0.171	0.103	−1.67			
CB type$_{NCB\ vs.\ AB}$	0.015	0.125	0.12			
Prev. vowel type $_{/i/\ vs.\ /a/}$	0.053	0.121	0.44	4.65	7	0.70
Prev. vowel type $_{/u/\ vs.\ /a/}$	0.010	0.143	0.07			
Prev. vowel type $_{/e/\ vs.\ /a/}$	0.052	0.122	0.43			
Prev. vowel type $_{/o/\ vs.\ /a/}$	−0.092	0.124	−0.75			
Prev. vowel type $_{/V/\ vs.\ /a/}$	−0.236	0.291	−0.81			
Prev. vowel type $_{/N/\ vs.\ /a/}$	−0.027	0.156	−0.17			
Prev. vowel type $_{/H/\ vs.\ /a/}$	0.056	0.161	0.35			
Syn. head$_{Pronoun\ vs.\ Noun}$	−0.070	0.067	−1.04	1.10	1	0.29
Following pause	1.070	0.061	17.54	265.93	1	< 0.001
BPM	0.365	0.074	4.93	24.03	1	< 0.001
Preface token	0.096	0.062	1.55	2.46	1	0.12

$(\sigma_S = 0.19,\ \sigma = 0.81)$

tendency for the duration of *wa* to increase as a function of clause duration was, in fact, significant ($\chi^2(1) = 9.16$, $p < 0.003$): *wa* was more lengthened when clause duration was longer. In line with the results for preface tokens, a strong effect of the following pause was evident ($\chi^2(1) = 265.93$, $p < 0.001$); the duration of *wa* tended to be longer in the presence of a following pause. In addition, the effect of the boundary pitch movement was also highly significant ($\chi^2(1) = 24.03$, $p < 0.001$), with boundary pitch movement enhancing the lengthening of *wa*.

3.3 Lengthening of clause-final particles

3.3.1 Data screening

Among 8,982 pairs of consecutive clause units, particles appeared 5,289 times (58.9%) in the final position of the first clause unit. Auxiliary verbs also frequently appeared in this position, 3,225 cases (35.9%); these might also be considered for the present analysis. We, however, focus on clause-final particles only, and do not analyze clause-final auxiliary verbs for the following two reasons:

– The distribution of the clause-boundary type in the auxiliary verb data was quite uneven, which makes statistical analysis difficult. Absolute boundaries occupied 90% of the entire data, and non-clausal boundaries occurred only 35 times.
– In the auxiliary verb data, 46.6% of final vowels were devoiced. This was mainly due to particular words, *desu* and *masu*. Although morae with devoiced vowels in these words could also be lengthened, the underlying mechanism would be different from those involved in other cases analyzed in this study.

As in the previous analyses, cases meeting the following conditions were excluded from the analysis:

– The last mora of the clause-final particle had an uncertain boundary or was pronounced in a non-canonical way.
– The last mora of the clause-final particle was not tightly connected to the preceding mora – that is, a pause occurred between them.

Of 5,289 cases of clause-final particles, 4,653 cases, or 88.0% (AB: 891, SB: 2919, WB: 597, NCB: 246), were selected by the above procedure and retained for the subsequent analysis.

3.3.2 Predictors

In the present analysis, the following eight predictors were used in addition to
the two cognitive predictors:
1. Type of the vowel in the final mora of the clause-final particle
2. Type of the vowel in the mora preceding the final mora
3. Syntactic sub-category of the clause-final particle
4. Number of morae in the clause-final particle
5. Presence of a pause immediately following the clause-final particle
6. Presence of boundary pitch movement at the right edge of the clause-final
    particle
7. Presence of a preface token in the following clause unit
8. Presence of a topic phrase in the following clause unit

In addition to a random intercept for speakers, another random intercept
for word forms was included in the models to capture the large variance in final-
mora duration for clause-final particles according to word form.

3.3.3 Statistical results

Figure 10 shows scatter plots between the (normalized) duration of the last mora
of the clause-final particles and the (log) duration of the following clause's body
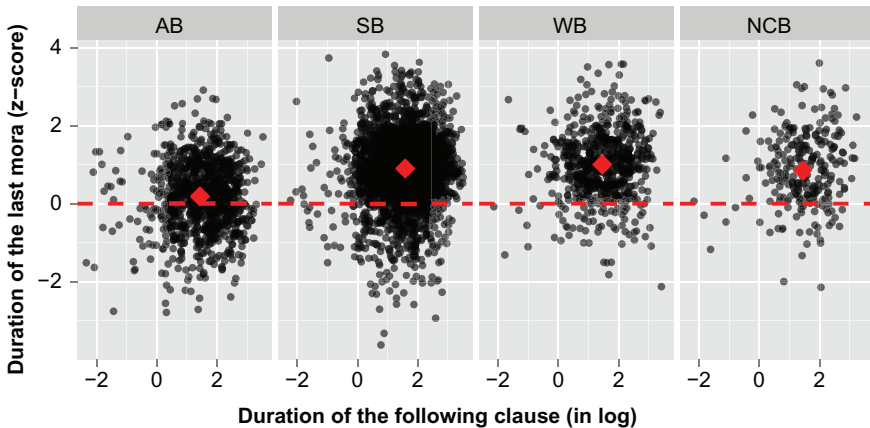

Figure 10: The (normalized) duration of the last mora of the clause-final particle vs. the (log)
duration of the following clause's body (clause complexity) relative to the clause-boundary
type of the current clause unit (boundary depth). Dashed lines indicate the baseline (zero on
the normalized scale), and diamonds indicate the centers of distributions.

(clause complexity), relative to the clause-boundary type (boundary depth). In general, the duration of the last mora of the clause-final particle shifted positively on the normalized scale, meaning it lengthened from its normal duration. The tendency, however, was weaker at absolute boundaries.

Likelihood-ratio tests compared the model with interactions between the clause-boundary type and other three predictors – clause duration, the presence of a following pause, and the presence of boundary pitch movement – with the three models, each excluding one of these interaction terms. The first two interaction terms revealed no significant results (Clause Duration: $\chi^2(3) = 2.19$, $p = 0.53$; Following Pause: $\chi^2(3) = 2.91$, $p = 0.41$), but the last term revealed a significant result (BPM: $\chi^2(3) = 31.06$, $p < 0.001$). Thus, a model with an interaction between the presence of boundary pitch movement and the clause-boundary type was used.

Figure 11 shows the effect displays for some of the predictors of the model fitted to the clause-final particle data. The estimated model parameters are also shown in Table 8. In this model, a significant interaction held between the presence of boundary pitch movement and clause-boundary type. Although lengthening of the last mora was always facilitated by boundary pitch movement, regardless of clause-boundary type, differences in the estimated coefficients among the four clause-boundary types varied between when boundary pitch movement was present and when it was absent. Only when the clause-final particle had boundary pitch movement was the duration of the final mora less lengthened at absolute and strong boundaries than at weak and non-clausal boundaries (AB vs. WB: adjusted $p < 0.001$; AB vs. NCB: adjusted $p < 0.001$; SB vs. WB: adjusted $p < 0.004$; SB vs. NCB: adjusted $p < 0.004$). When the clause-final particle did not have boundary pitch movement, no significant differences were found among the four clause-boundary types.

A strong effect of the following pause was confirmed along with an effect of the presence of a preface token (Following Pause: $\chi^2(1) = 189.39$, $p < 0.001$; Preface Token: $\chi^2(1) = 5.31$, $p < 0.03$), but no significant effects were observed for clause duration, syntactic sub-category of the particle, number of morae in the particle, or presence of a topic phrase. Both the vowel types of the last and the preceding morae had strong effects. The results of the multiple comparisons for the vowel type of the last mora, however, were very complicated. In general, clause-final morae ending with /i/ were more lengthened and those involving vowel devoicing were less lengthened (/i/ vs. /e/: adjusted $p < 0.02$; /i/ vs. /V/: adjusted $p < 0.001$; /a/ vs. /V/: adjusted $p < 0.003$; /o/ vs. /V/: adjusted $p < 0.001$). For the vowel type of the preceding mora, significant differences were found only between /o/ and /H/ and between /u/ and /H/ (/u/ vs. /H/: adjusted $p < 0.04$; /o/ vs. /H/: adjusted $p < 0.03$); clause-final morae following /H/ were more lengthened than those following /u/ or /o/.
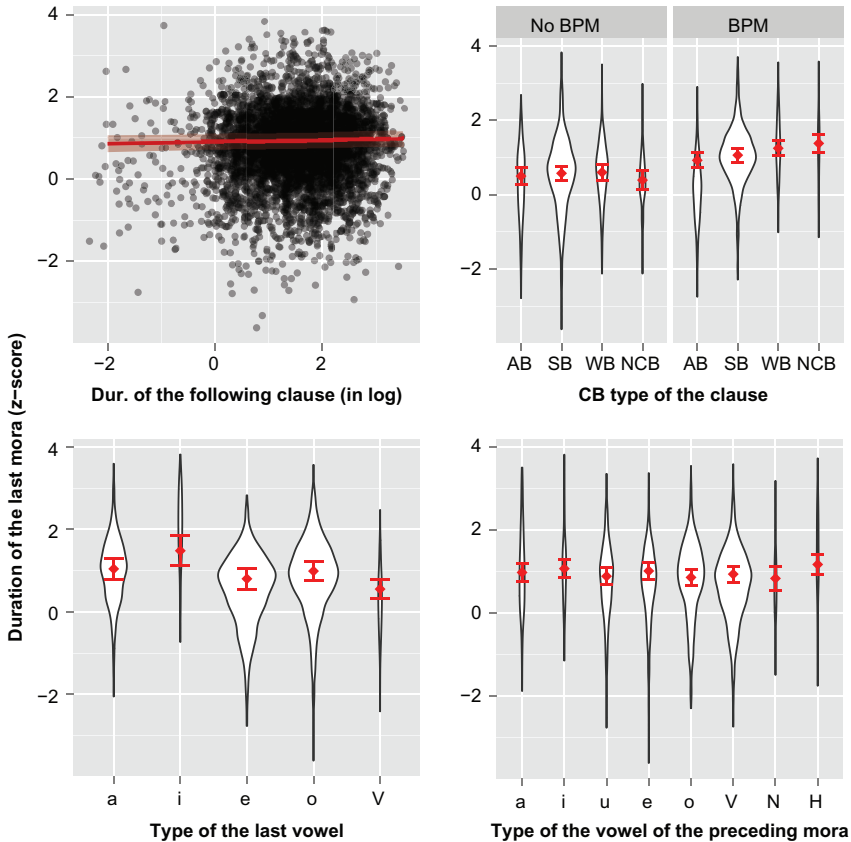
Figure 11: Effect displays for some predictors of the clause-final particle model. Upper-left: clause duration; upper-right: the interaction between clause-boundary type and the presence of boundary pitch movement; bottom-left: the vowel type of the target mora; bottom-right: the vowel type of the preceding mora.

4 Discussion

4.1 Phonological and syntactic factors

Without exception, phrase-final morae in the three locations examined in this study were longer than their normal durations. In the statistical analyses, the effects of phonological factors were very robust. Throughout the analyses of the three locations, the effects of the following pause and the boundary pitch movement were always significant. Lengthening of phrase-final morae was

Table 8: Estimated parameters of the clause-final particle model and its analysis of deviance table. σ_S indicates the standard deviation of the random intercept for speakers, σ_W the standard deviation of the random intercept for word forms, and σ the residual standard deviation.

	Coef.	SE	t-value	χ^2	df	p-value
Intercept	0.164	0.180	0.92			
Clause duration	0.020	0.013	1.56	2.46	1	0.12
BPM:CB type$_{SB\ vs.\ AB}$	0.063	0.079	0.81	29.98	3	< 0.001
BPM:CB type$_{WB\ vs.\ AB}$	0.224	0.091	2.48			
BPM:CB type$_{NCB\ vs.\ AB}$	0.563	0.120	4.71			
CB type$_{SB\ vs.\ AB}$	0.075	0.083	0.91			
CB type$_{WB\ vs.\ AB}$	0.100	0.078	1.28			
CB type$_{NCB\ vs.\ AB}$	−0.108	0.111	−0.97			
BPM	0.424	0.072	5.88			
Vowel type$_{/i/\ vs.\ /a/}$	0.440	0.203	2.17	46.61	4	< 0.001
Vowel type$_{/e/\ vs.\ /a/}$	−0.240	0.153	−1.57			
Vowel type$_{/o/\ vs.\ /a/}$	−0.053	0.138	−0.39			
Vowel type$_{/V/\ vs.\ /a/}$	−0.490	0.131	−3.74			
Prev. vowel type $_{/i/\ vs.\ /a/}$	0.092	0.079	1.16	19.72	7	< 0.007
Prev. vowel type $_{/u/\ vs.\ /a/}$	−0.091	0.063	−1.44			
Prev. vowel type $_{/e/\ vs.\ /a/}$	0.036	0.081	0.44			
Prev. vowel type $_{/o/\ vs.\ /a/}$	−0.119	0.072	−1.67			
Prev. vowel type $_{/V/\ vs.\ /a/}$	−0.041	0.060	−0.68			
Prev. vowel type $_{/N/\ vs.\ /a/}$	−0.144	0.131	−1.10			
Prev. vowel type $_{/H/\ vs.\ /a/}$	0.191	0.093	2.05			
Sub-cat.$_{Final\ vs.\ Conj.}$	−0.419	0.224	−1.87	5.31	4	0.26
Sub-cat.$_{Adv.\ vs.\ Conj.}$	0.088	0.209	0.42			
Sub-cat.$_{Case\ vs.\ Conj.}$	−0.120	0.190	−0.63			
Sub-cat.$_{Topic\ vs.\ Conj.}$	−0.206	0.426	−0.48			
Mora length$_{2\ vs.\ 1}$	−0.115	0.179	−0.65	0.84	2	0.66
Mora length$_{3+\ vs.\ 1}$	−0.120	0.207	−0.58			
Following pause	0.517	0.037	13.89	189.39	1	< 0.001
Preface token	0.055	0.024	2.32	5.31	1	< 0.03
Topic phrase	−0.026	0.030	−0.87	0.76	1	0.38

($\sigma_S = 0.16$, $\sigma_W = 0.37$, $\sigma = 0.81$)

facilitated when followed by a pause or when bearing extra movement in the pitch contour. This is not surprising given that pre-pausal lengthening is a widely-known phenomenon and that a complex pitch contour occupies a larger space than a simple falling contour. Furthermore, the presence of boundary pitch movement interacted, in a significant way, with the clause-boundary type, a cognitive factor, which will be discussed in the next section.

In the analysis of clause-final particles, lengthening was also affected by the vowel types of the last and the preceding morae. Morae involving vowel devoicing were less lengthened, while morae ending with /i/ were more lengthened at the end of clause-final particles. Note that our response variable, the duration of the final mora, was normalized per mora type in order to eliminate the influence of the inherent duration of the mora. The degree of lengthening was nevertheless uneven among different vowel types. A closer look at word forms involved in these data revealed that the effect of the vowel type might be attributed to a few word forms that were either extremely or minimally prolonged.

Table 9 shows some frequent word forms and their statistics in the clause-final particle data. The numbers in the third column indicate the frequency count and its percentage within each vowel type. In most cases, the majority of cases involved one or two word forms. For instance, the conjunctive particle *si* accounted for 72.4% of the clause-final particles ending with /i/. More importantly, it was extremely lengthened: the mean duration, on the normalized scale, of the vowel /i/ in *si* was 2.358, as depicted in the fourth column in Table 9, which was very large. By contrast, the quotational particle *to* accounted for 59.9% of the clause-final particles whose last vowels were devoiced[10] and was minimally lengthened, the sample mean being only 0.348. Therefore, a large

Table 9: Frequent word forms and their statistics relative to the last vowel's type in the clause-final particle data. The third column indicates the frequency count and its percentage within each vowel type. The fourth column represents the mean (normalized) duration of the last mora, not adjusted for other variables. The fifth column represents the least-squares mean of the (normalized) duration for each vowel type, and the last column the estimated deviation in the random effect for each word form.

Vowel type	Word form	Freq. (%)	Sample mean	LS mean	Deviation
/a/	*ga* (conj.)	523 (63.9%)	0.940	0.663	−0.255
/i/	*si* (conj.)	123 (72.4%)	2.358	1.103	0.790
/e/	*te* (conj.)	859 (46.1%)	0.929	0.422	0.018
	ne (final)	825 (44.2%)	0.045	0.422	−0.441
/o/	*mo* (adverbial)	803 (48.3%)	0.893	0.610	−0.239
	kedo (conj.)	444 (26.7%)	0.623	0.610	−0.365
/V/	*to* (quotational)	82 (59.9%)	0.348	0.173	0.090

---

10 *to* does not end with a high vowel, but devoicing of non-high vowels in spontaneous Japanese is common (Maekawa and Kikuchi 2005).

value in the least-squares mean for the vowel /i/, 1.103, and a small value in the least-squares mean for the devoiced vowel, 0.173, could be due to these influential word forms.[11]

One important point to note is that the final vowel of *si* is a high vowel in the devoicing context but is nonetheless voiced. In fact, there were only 6 instances of *si* with its vowel devoiced. One factor behind the voicing of /i/ in *si* might be the high rate of boundary pitch movements within this mora (65.1%). But regardless of presence/absence of boundary pitch movements, this vowel was almost always voiced and lengthened. That is, for some reason, the conjunctive particle *si* was frequently lengthened, and for it to be prolonged, the last high vowel was voiced.

The results for the vowel type of the preceding mora are much harder to interpret. Since there are many clause-final particles that are mono-moraic, some preceding morae belong to the previous words rather than the final words. This makes it difficult to conduct a similar analysis to that applied to the vowel type of the final mora. Further investigation is left for future study.

4.2 Cognitive factors

The two cognitive factors of interest, clause complexity and boundary depth, related to lengthening in a complex way. There were five distinct relationships:
1. For the filler *e*, both clause complexity and boundary depth had significant effects; *e* was longer when clause duration was longer, and also longer after strong boundaries than absolute boundaries.
2. For the topic marker *wa* in clause-initial topic phrases, only clause complexity had a significant effect. *Wa* was longer when the clause duration was longer. The same tendency was also observed for the filler *ma*, although the effect was only marginal.
3. For clause-final particles, boundary depth had a significant effect only when the particles were accompanied by boundary pitch movements. In general, the lengthening of morae with boundary pitch movements was further facilitated by shallower boundary depths.

---

11 One might wonder why the extreme lengthening of *si* is reflected in the fixed effect of the vowel type rather than the deviation from it (i.e., random effect of the word form). Since a random effect is expressed in terms of a zero-centered normal distribution with an estimable standard deviation, no deviation can be too large or too small. In fact, the deviation for *si*, 0.790 (in the last column of Table 9), is already too big compared to the estimated standard deviation of this random effect, $\sigma_W = 0.37$ (in Table 8).

4. For the filler *ano*, the interaction between clause duration and boundary depth was significant: after absolute boundaries, lengthening of *ano* was inhibited by increased clause duration, while it was facilitated after non-clausal boundaries.
5. For the conjunction *de*, neither of the two cognitive factors had a significant effect.

It seems that the difference between the first two cases was attributable to differences in the positions in which these tokens occurred. For the fillers *e* and *ma*, *e* tended to come earlier than *ma*; in 134 cases in which *e* and *ma* appeared at the same time as clause-initial preface tokens, *e* came earlier than *ma* 92.5% of the time. As for *wa*-marked topic phrases, they come later than preface tokens. Therefore, boundary depth plays a role in earlier positions, while clause complexity is relevant for later positions. The third case also seems consistent with this; clause-final particles come earlier than preface tokens, and hence boundary depth plays a role. This is quite natural if we assume that boundary depth indicates a qualitative difference in the planning process involved and that clause complexity reflects the amount of cognitive load for planning the remainder of the utterance. Speakers may engage in discourse-level and abstract-level planning at an earlier stage, and then, at a later stage, proceed to the planning of more local and concrete syntax and semantics of the utterance/clause.

As for the effect of clause complexity, the increasing trends in the durations of *e*, *ma*, and *wa* were relatively small. An increase of 1.0 on the log scale, i.e., 2.7 times on the linear scale, in the duration of the clause body yields an increase of only 0.169 for *e*, 0.162 for *ma*, and 0.101 for *wa* in the normalized duration of the mora. (Compare these with an increase of 1.693/ 2.111/1.070 with the presence of a following pause in these data.) Although these numbers are small, they are reliable. It is interesting that the lengthening of the topic-marker *wa* is affected by clause complexity. In spontaneous Japanese discourse, about 70% of arguments are omitted (Clancy 1980; Hinds 1983). Thus, for the topic-marker *wa* to be utilized to gain time for planning, the topic phrase should be overtly expressed in the first place. Den and Nakagawa (2013) showed that the duration of the clause body tended to be longer in the presence of overt topic phrases that could have been omitted in that context, based on an analysis of dialog data in the CSJ.[12] When speakers

---

12 In their data, topic phrases in utterances responding to the other speaker's initiating utterances were overtly expressed 37.4% of the time, which is relatively high given that topic phrases can be basically omitted in Japanese.

need extra time for planning, they sometimes overtly express omissible topic phrases and lengthen the last word *wa*. This is an unusual lengthening strategy never observed in English, nor in other languages that disallow the omission of arguments.

Moving on to the effect of the boundary depth, the degree of lengthening of the filler *e* and clause-final particles varied by clause-boundary type, although in the latter case the effect was observed only in the presence of boundary pitch movement. The interaction between clause-boundary type and boundary pitch movement suggests that boundary pitch movements may serve as a vehicle for lengthening. As described in Section 2.4.3, the rates of boundary pitch movements were lower at weak and non-clausal boundaries (WB: 58.7%, NCB: 54.2%) than at absolute and strong boundaries (AB: 78.0%, SB: 79.4%). Thus, using boundary pitch movement at shallower boundaries, where its use is not canonical, may be part of a strategy to make room for lengthening to operate to a significant degree.

There is another point to account for related to the results for boundary depth: the least lengthening occurred at absolute boundaries, which had been assumed to be the deepest boundary in this study. There could be two reasons for this. First, there may still be a confounding variable affected by boundary depth but missing in our statistical models. One candidate for such a confounding variable is the duration of pauses immediately following the first clause unit. If speakers use silent pauses at this location for planning, lengthening before or after these pauses would be reduced when they are long. It is, in fact, the case that the durations of pauses were longer after absolute boundaries than after other types of boundaries, the median duration being 0.757 s after absolute boundaries compared with 0.477, 0.466, and 0.572 s after strong, weak, and non-clausal boundaries, respectively. Second, the hierarchy assumed in (7) may not be precise. As described in Section 2.2, weak and non-clausal boundaries occur when clauses are not accompanied by morpho-syntactic ending-forms but are better viewed as independent utterances. There are several criteria for identifying weak and non-clausal boundaries, some of which are related to pragmatics and discourse. Therefore, it might be incorrect to assume that absolute boundaries are always deeper than weak and non-clausal boundaries.

The interaction between the two cognitive factors was significant only for the filler *ano*. There were clear trends for the duration of the second mora of *ano* to decrease after absolute boundaries and increase after non-clausal boundaries as a function of the clause duration. These trends were, however, obtained

based only on 66 and 22 cases of these clause-boundary types, respectively,[13] and it would be unwise to conclude that these trends are due to characteristics peculiar to these clause-boundary types.

Finally, lengthening of the conjunction *de* showed no effect of the cognitive factors. The irrelevance of the clause-initial *de*'s duration to the two cognitive factors has also been reported in previous studies (Den 2009; Watanabe and Den 2010). Since *de* is often used as a discourse marker, it would be rather natural to infer that when this preface token occurs, a certain kind of discourse planning is always taking place, regardless of the clause-boundary type, so clause complexity is irrelevant.

4.3 Predictors not considered

As described in Section 2.4.3, the influence of the duration of the preceding mora was not considered in the present study. When the (normalized) duration of the preceding mora was included in each of the statistical models developed in Sections 3.1–3.3, it was found to always have a significant effect on the duration of the final mora (the bottom row in Table 10): the final mora tended to be longer when its preceding mora was longer. More importantly, in the presence of this predictor, the effects of the two cognitive factors weakened or disappeared (the other rows in Table 10).

Table 10: Summary of the (part of the) results for models with the duration of the preceding mora included. "+" indicates that the variable on that row significantly facilitates lengthening. An inequality represents significant differences in least-squares means among the levels of the factor on that row. n.s. = not significant. * = effect was only marginal.

	ano	*wa*	Particle
ClauseDur:CBType	*NCB: +	n.s.	n.s.
Clause duration		n.s.	n.s.
BPM:CBType			BPM: AB, SB < WB, NCB
CB type		n.s.	
Prec. mora duration	+	+	+

---

13 The use of *ano* in the clause-initial position may be atypical, as these small numbers indicate (cf. 253 and 42 cases for the filler *e*). This also suggests that we may not have a concrete conclusion at this stage.

So, are the effects found for the cognitive factors an artifact of other factors? This cannot be true, since the duration of the preceding mora itself was affected by the cognitive factors. When the (normalized) duration of the preceding mora, instead of the final mora, was used as a response variable of the statistical model, we obtained similar results to those with the final mora's duration as a response variable, in terms of the relationship between the response variable and the cognitive factors (Table 11). That is, for the filler *ano*, the interaction between the two cognitive factors was (marginally) significant; for *wa*-marked topic phrases, only clause complexity had a significant effect; and for clause-final particles, the interaction between the boundary pitch movement and boundary depth was (marginally) significant.

Table 11: Summary of the (part of the) results for models with the duration of the preceding mora as a response variable. "−" indicates that the variable on that row significantly inhibits lengthening. * indicates the effect was only marginal.

	ano	*wa*	Particle
ClauseDur:CBType	*AB: −	n.s.	n.s.
Clause duration		+	n.s.
BPM:CBType			*BPM: AB < SB, WB, NCB
CB type		n.s.	

These results suggest that lengthening operates not only on the last mora but also on the penultimate mora, and that the cognitive factors influence both morae. In this situation, if we use one of the two duration variables as a response variable and the other as a predictor, the effect of the cognitive factors on the response variable might be obscured by the predictor, leading to a wrong conclusion. This is why the duration of the preceding mora was not used as a predictor in our statistical models.

The same discussion may also be applied to the duration of the following pause. In the present study, the *presence* of a following pause, not its *duration*, was used as a predictor. This was sufficient to capture the pre-pausal lengthening effect. It might not be sufficient, however, if we want to see the influence of the following pause on lengthening of the target mora in terms of cognitive factors. As discussed in the previous section, the duration of pauses might be affected by cognitive factors. Thus, lengthening and pause durations would be related to each other. This could be a motivation to introduce the duration of the following pause into the statistical model as a predictor. However, as discussed above, introducing a variable with an indirect correlation with the response variable through the cognitive factors might lead to a faulty conclusion.

Currently, we have no good way to address these variables. Development of proper models that take these variables into account is left for future study.

4.4 Interrelationship among the three locations

Finally, this section addresses the interrelationships among the three locations investigated in this study. Our statistical models included the presence of a preface token and a topic phrase as predictors in order to examine the dependencies between different lengthening locations. We assumed that speakers may choose one location over the others when multiple locations are available for lengthening. The results showed that this assumption was only partly verified. The effect of the presence of another lengthening location was not significant except for two cases, the conjunction *de* and clause-final particles. Even when the effect was significant, the direction of the effect was not consistent: lengthening of *de* was *inhibited* when followed by a topic phrase, and lengthening of clause-final particles was *facilitated* when followed by a preface token.

Along with failure to obtain meaningful results, our way of investigating the interrelationships among the three locations was insufficient, since it could not directly model the dependency between the durations of the target morae at two locations. For the purpose of proper modeling of such a dependency, models able to handle two or more response variables at the same time will be required. This is another direction in improving our models in future research.

References

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/packa ge = lme4 (accessed 28 March 2015).

Beattie, Geoffrey. 1979. Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. *Linguistics* 17. 61–78.

Blackmer, Elizabeth R. & Janet L. Mitton. 1991. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition* 39. 173–194.

Campbell, Nick. 1992. Segmental elasticity and timing in Japanese speech. In Yoichi Tohkura, Erik Vatikiotis-Bateson & Yoshinori Sagisaka (eds.), *Speech perception, production and linguistic structure*, 403–418. Tokyo: Ohmsha.

Campbell, W. Nick & Stephen D. Isard. 1991. Segment durations in a syllable frame. *Journal of Phonetics* 19(1). 37–47.

Chafe, Wallace. 1980. The deployment of consciousness in the production of a narrative. In Wallace Chafe (ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, 9–50. Norwood, NJ: Ablex.

Clancy, Patricia M. 1980. Referential choice in English and Japanese narrative discourse. In Wallace Chafe (ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, 127–202. Norwood: Ablex.

Clark, Herbert H. 1994. Managing problems in speaking. *Speech Communication* 15. 243–250.

Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84. 73–111.

Clark, Herbert H. & Thomas Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37. 201–242.

Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.

Cutler, Anne & Takashi Otake. 1994. Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language* 33. 824–844.

Den, Yasuharu. 2003. Some strategies in prolonging speech segments in spontaneous Japanese. In *Proceedings of ISCA Research Workshop on Disfluency in Spontaneous Speech,* 87–90. Göteborg, Sweden.

Den, Yasuharu. 2009. Prolongation of clause-initial mono-word phrases in Japanese. In Shu-Chuan Tseng (ed.), *Linguistic patterns in spontaneous speech*, 167–192. Taipei: Institute of Linguistics, Academica Sinica.

Den, Yasuharu & Natsuko Nakagawa. 2013. Anti-zero-pronominalization: When Japanese speakers overtly express omissible topic phrases. In *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech*, 25–28. Stockholm, Sweden.

Eklund, Robert. 2001. Prolongations: A dark horse in the disfluency stable. In *Proceedings of ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*, 5–8. Edinburgh, UK.

Fox, John. 2003. Effect displays in R for generalized linear models. *Journal of Statistical Software* 8(15). 1–27.

Fox, John, Sanford Weisberg, Michael Friendly & Jangman Hong. 2014. Effects: Effect displays for linear, generalized linear, and other models. R package version 3.0-3, http://CRAN.R-project.org/package = effects (accessed 28 March 2015).

Fox Tree, Jean E. & Herbert H. Clark. 1997. Pronouncing "the" as "thee" to signal problems in speaking. *Cognition* 62. 151–167.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Hinds, John. 1983. Topic continuity in Japanese. In Talmy Givón (ed.), *Topic continuity in discourse*, 43–93. Amsterdam: John Benjamins.

Iwasaki, Shoichi & Tsuyoshi Ono. 2001. 'Sentence' in spontaneous spoken Japanese discourse. In Joan L. Bybee & Michael Noonan (eds.), *Complex sentences in grammar and discourse*, 175–202. Amsterdam: Benjamins.

Kaiki, Nobuyoshi & Yoshinori Sagisaka. 1992. The control of segmental duration in speech synthesis using statistical methods. In Yoichi Tohkura, Erik Vatikiotis-Bateson & Yoshinori Sagisaka (eds.), *Speech perception, production and linguistic structure*, 391–402. Tokyo: Ohmsha.

Klatt, Dennis H. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3. 129–140.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59. 1208–1221.

Koiso, Hanae & Yasuharu Den. 2013. *Acoustic and linguistic features related to speech planning appearing at weak clause boundaries in Japanese monologs*. In *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech*, 37–40. Stockholm, Sweden.

Koiso, Hanae, Yasuharu Den, Ken'ya Nishikawa & Kikuo Maekawa. 2014. Design and development of an RDB version of the Corpus of Spontaneous Japanese. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, 1471–1476. Reykjavik, Iceland.

Kubozono, Haruo. 1989. The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech* 32. 249–278.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.

Lee, Tzu-Lun, Ya-Fang He, Yun-Ju Huang, Shu-Chuan Tseng & Robert Eklund. 2004. Prolongation in spontaneous Mandarin. In *Proceedings of the 8th International Conference on Spoken Language Processing*, 2181–2184. Jeju Island, Korea.

Lenth, Russell V. 2015. lsmeans: Least-squares means. R package version 2.16, http://CRAN.R-project.org/package=lsmeans (accessed 28 March 2015).

Levelt, Willem J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14. 41–104.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Maclay, Howard & Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15. 19–44.

Maekawa, Kikuo. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. In *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition*, 7–12. Tokyo.

Maekawa, Kikuo & Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara (eds.), *Voicing in Japanese*, 205–228. Berlin: Mouton de Gruyter.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi & Jennifer J. Venditti. 2002. X–JToBI: An extended J_ToBI for spontaneous speech. In *Proceedings of the 7th International Conference on Spoken Language Processing*, 1545–1548. Denver, CO.

Oller, D. Kimbrough. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54. 1235–1247.

Otake, Takashi, Giyoo Hatano, Anne Cutler & Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32. 358–378.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (accessed 28 March 2015).

Sagisaka, Yoshinori & Yoh'ichi Tohkura. 1984. Phoneme duration control for speech synthesis by rule (in Japanese). *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* (Japanese edition) J67-A. 629–638.

Shriberg, Elizabeth E. 1994. *Preliminaries to a theory of speech disfluencies*. Berkeley: University of California Ph.D. thesis.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9. 81–105.

Watanabe, Michiko. 2009. *Features and roles of filled pauses in speech communication: A corpus-based study of spontaneous speech*. Tokyo: Hituzi Syobo.

Watanabe, Michiko & Yasuharu Den. 2010. Utterance-initial elements in Japanese: A comparison among fillers, conjunctions, and topic phrases. In *Proceedings of the DiSS-LPSS Joint Workshop 2010 (the 5th Workshop on Disfluency in Spontaneous Speech and the 2nd International Symposium on Linguistic Patterns in Spontaneous Speech)*, 31–34. Tokyo.

Watanabe, Michiko, Keikichi Hirose, Yasuharu Den, Shusaku Miwa & Nobuaki Minematsu. 2006. Factors influencing ratios of filled pauses at clause boundaries in Japanese. In *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, 253–256. Athens, Greece.