Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5–8 September 2003, Göteborg University, Sweden.
Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 90, ISSN 0349–1021, pp. 87–90.

Some strategies in prolonging speech segments in spontaneous Japanese

Yasuharu Den

Faculty of Letters, Chiba University, Japan

Abstract

In this paper, we investigate segmental prolongation in a corpus of spontaneous Japanese monologues consisting of over 700,000 words. We examine effects on the rate of prolongation of various factors including speech types, the genders of speakers, word classes, word positions in the phrase and in the inter-pausal unit, and the presence of preceding fillers. Based on the empirical findings, we state some strategies in prolonging speech segments used by Japanese speakers.

1. Introduction

In spontaneous speech, speakers may prolong their speech segments anywhere in an utterance. They may prolong a filler placed at the beginning of an utterance or the initial phoneme immediately after starting a major constituent. Or, they may prolong a phoneme at a clause final position. Some of them serve as a signal to forthcoming problems in communication. For instance, Den [1] showed that when the first token involved in a word repetition is disrupted in the middle, the phoneme at the disruption point is considerably prolonged, which would inform listeners of speaker's difficulty in producing the rest of the constituent. Although several researchers focused on the phenomena with limited interests, prolongation in general has not been fully studied so far (for notable exceptions, see [2, 3, 4]).

This paper investigates segmental prolongation in the *Corpus of Spontaneous Japanese* (CSJ) [6], which is a huge-sized corpus (ca 740,000 words) of spontaneous monologues in Japanese. Watanabe and Den [7] have already reported their analysis on prolongation in the CSJ. They assumed that prolongation, as well as suspension and restart, is a device for speakers to inform listeners of foreseeable troubles in communication. They found that (i) speakers are most likely to prolong a filler before initiating a constituent, next most likely to prolong a vowel within the initial word of it, and less likely to prolong after an initial word, and that (ii) the more complex a constituent is, the more likely speakers are to prolong their speech segments in an initial commitment to it. They, however, analyzed only prolongations in utterance-initial noun phrases and did not look at prolongations at other places or in words of other syntactic classes. The current paper aims at investigating all occurrences of prolongation in the CSJ, without adopting particular hypotheses, to obtain basic facts about the phenomena.

2. Method

2.1. Corpus

We analyzed the *Corpus of Spontaneous Japanese* (CSJ, Monitor Version 2002) [6], that is being developed at the National Language Research Institute as a part of their five-year Spontaneous Speech Project (fiscal years of 1999–2003). It comprises speech, transcripts and morphological analyses of 134 academic presentations and 189 simulated public speech. The former is live recordings of researchers' presentations in meetings of several academic societies, while the latter is short speech spoken specifically for the purpose of the data collection by paid non-professional speakers mostly in recording studios. The speakers include both females and males (33 females and 101 males for academic presentations, and 120 females and 69 males for simulated public speech), and their ages range between early thirty and early eighty with the average and the median at mid sixty. Some speakers engaged in more than one session, but we disregard variation within speakers across sessions. The speech data amounts to 70 hours, and the morphological data to 740,000 words excluding fillers. Table 1 shows the summary statistics of the corpus for each combination of speech type and the gender of the speaker.

Table 1: Summary statistics of the CSJ.

	Academic		Simulated		Total
	Female	Male	Female	Male	
# of sessions	33	101	120	69	323
Duration	9.2hrs	26.6hrs	21.4hrs	12.9hrs	70.2hrs
# of words	98141	285640	221911	136199	741891
# of morae	199034	588032	427084	259335	1473485

2.2. Changes made to the corpus

In the transcripts, speech segments are divided into basic units according to the following criteria: a stretch of speech either (i) delimited by silent pauses longer than 200ms or (ii) ending with sentence final elements such as verbs in finite form and final particles [5]. Since there is no indication of whether or not the second criterion could also be applied when the first one was applied, we cannot precisely know whether or not a unit boundary coincides with a syntactic boundary. Thus, we decided to use, for uniformity, inter-pausal units (IPUs) determined by the first criterion only, discarding the boundaries at which only the second criterion was applied, i.e., the boundaries followed by silent pauses shorter than 200ms.

The transcripts also include the information about the boundaries of basic syntactic phrases, i.e., bunsetsu phrases, and occurrences of prolongation. Prolongations of vowels are marked by an <H> tag in the transcripts.[1] Although the assignment of <H> tags is based on the transcribers' intuition, we accepted all of them as instances of prolongations.

For word fragments, which are marked by a <D> tag in the transcripts, <H> tags are never used; that is, for those sounds, long vowels are transcribed using standard orthography. Since they may include instances of prolonged word fragments involved in word repetitions [1], we checked all occurrences of them and substituted long vowels with <H> tags when the intended word was reliably recovered and prolongation could be supposed there.

Fillers such as "eto" and "ano" are treated as genuin words in the morphological analyses of the corpus, but we changed this treatment. We attached fillers to the succeeding words

[1] Prolongations of consonants are not marked, but they are rare in Japanese.

regarding their presence as a property of the succeeding words. Thus, fillers per se were never counted as words. For instance, when an IPU begins with "*Ee Nihon de-wa*" (um in Japan), "*Nihon*" (Japan) is considered as the initial word with a property of being preceded by a filler.

2.3. Classification

The positions of morae in the word were classified into 'Single', 'Initial', 'Medial', or 'Final'. The class 'Single' was used for words consisting of a single mora, i.e., mono-moraic words, and the other classes were used for words consisting of more than one mora. Similarly, the positions of words in the phrase and in the IPU were classified into 'Single', 'Initial', 'Medial', or 'Final'. Words were classified into content words or function words, in a traditional grammatical sense, or word fragments. They were also classified according to the presence of the preceding disfluent items: `Fillers' when preceded by one or more fillers, 'Fillers + Pause' when preceded by fillers and intervening or following silent pauses, and 'None' when preceded by no fillers.

3. Results

3.1. PR rates

Table 2 shows the rate of prolongations in the CSJ.

Table 2: The rates of prolongations.

	Academic		Simulated		
	Female	Male	Female	Male	Total
# of PRs	335	1542	4313	2202	8392
% PRs/words	0.34%	0.54%	1.94%	1.62%	1.13%
% PRs/morae	0.17%	0.26%	1.01%	0.85%	0.57%

The overall PR rate of 1.13% per word is comparable to the PR rate of 1.27% reported for Swedish [3]. There is, however, an obvious speech type difference. The PR rates are much greater in simulated public speech than in academic presentations both for female and male speakers. This is mainly because academic presentations are pre-planned and sometimes rehearsed, and thus more trouble-free than simulated public speech, which is usually improvisational.

3.2. PR position in the word

Table 3 shows a breakdown of prolongations according to their positions in the word.

Table 3: The positions of prolongations in the word.

	Academic		Simulated		
	Female	Male	Female	Male	Total
# of Single PRs	114	577	2738	1095	4524
% per word	0.12%	0.20%	1.23%	0.80%	0.61%
% per mora	0.06%	0.10%	0.64%	0.42%	0.31%
# of Initial PRs	17	63	159	99	338
% per word	0.02%	0.02%	0.07%	0.07%	0.05%
% per mora	0.01%	0.01%	0.04%	0.04%	0.02%
# of Medial PRs	13	45	97	63	218
% per word	0.01%	0.02%	0.04%	0.05%	0.03%
% per mora	0.01%	0.01%	0.02%	0.02%	0.01%
# of Final PRs	191	857	1319	945	3312
% per word	0.19%	0.30%	0.59%	0.69%	0.45%
% per mora	0.10%	0.15%	0.31%	0.36%	0.22%

As can be seen, about a half of the prolongations occur in mono-moraic words. Word initial and medial prolongations are rare in general, and the ratio among initial, medial, and final positions is approximately 10–5–85, excluding mono-moraic words. This ratio is quite different from the 30–20–50 ratio reported for Swedish [2].

These tendencies do not depend on speech type or the gender of the speaker, although the prolongations of mono-moraic words are relatively infrequent in academic presentations compared to simulated public speech.

3.3. Word classes

Table 4 shows the rates of prolongations relative to word classes.

Table 4: The rates of prolongations relative to word classes.

	Content	Function	Fragment
# of words	369189	361986	10716
# of morae	1003479	455971	14035
# of PRs	4116	4211	65
% PRs/words	1.11%	1.16%	0.61%
% PRs/morae	0.41%	0.92%	0.46%

The distribution is more or less 50-50 between the content and the function word classes. The PR rate per word is also nearly the same between these two classes. The PR rate per mora, however, is much greater in the function word class than in the content word class. This is because function words are in general short, in Japanese typically consisting of a single mora (277829 out of 361986 words, or 76.8%) such as grammatical particle *ga*, conjunctive particle *te*, sentence final particle *ne*, and copula *da*. The PR rate of word fragments is less than the PR rate of complete words, but it is still considerable.

3.4. Word classes and PR position

Table 5 shows a breakdown of Table 4 according to PR positions.

Table 5: The positions of prolongations in the word relative to word classes.

	Content	Function	Fragment
# of Single PRs	1078	3400	46
% per word	0.29%	0.94%	0.43%
# of Initial PRs	292	42	4
% per word	0.08%	0.01%	0.04%
# of Medial PRs	215	3	0
% per word	0.06%	0.00%	0.00%
# of Final PRs	2531	766	15
% per word	0.69%	0.21%	0.14%

As was mentioned above, the majority of the function words are mono-moraic words. The PR rate for this class (Function-Single) is high. Other classes with a high PR rate are the final positions of content words (Content-Final) and word fragments resulting in a single mora (Fragment-Single). The Fragment-Single class is particularly important, since it, together with the Fragment-Final class, comprises prolongations at the disruption point of word cut-off. Interestingly, in the two-thirds (41 out of 61) of the instances of these classes, the prolonged word fragment was followed by a word whose initial part phonetically matches or similar to (e.g., *su* vs. *so*)[1] the fragment. These can be seen as instances of prolonged word fragments involved in word repetitions [1].

[1] The phonetic transcripts in the CSJ are written in Katakana, and there is no way to describe word cut-off at a consonant. But, *si* and *su* at the disruption point are likely to be describing cut-off at a consonant [s], since high vowels like [i] and [u] are usually devoiced between a voiceless consonant and a silence. If this is the case, a word fragment transcribed as *su* or *si* phonetically matches the initial *so* or *sa* of the following word like "*sore* (it)" or "*san* (three)".

3.5. Word positions in the phrase

Now we turn to the analysis of syntactic factors. First, we calculated PR rates considering word positions in the phrase. The PR rates relative to word positions in the phrase for the content word class and for the function word class are shown in Tables 6 and 7, respectively.

Table 6: The rates of prolongations relative to word positions in the phrase for the content word class.

	Single-W	Initial-W	Medial-W	Final-W
# of words	59974	217603	75103	16509
# of morae	162080	631722	172578	37098
# of PRs	2298	1448	255	115
% PRs/words	3.83%	0.67%	0.34%	0.70%
% PRs/morae	1.42%	0.23%	0.15%	0.31%

Table 7: The rates of prolongations relative to word positions in the phrase for the function word class.

	Single-W	Initial-W	Medial-W	Final-W
# of words	3189	8276	141759	208762
# of morae	3587	11437	203814	237132
# of PRs	89	27	878	3217
% PRs/words	2.79%	0.33%	0.62%	1.54%
% PRs/morae	2.48%	0.24%	0.43%	1.36%

The PR rate is particularly high in the Single-W class, which is a class of words solely comprising a phrase. The PR rate is relatively high in the phrase-final word class. There tendencies apply to both content and function word classes.

Next, we make breakdowns of these tables according to PR positions in the word. The results are shown in Tables 8 and 9, respectively.

Table 8: The positions of prolongations in the word relative to word positions in the phrase for content word class.

	Single-W	Initial-W	Medial-W	Final-W
# of Single PRs	981	53	31	13
% per word	1.64%	0.02%	0.04%	0.08%
# of Initial PRs	191	88	8	5
% per word	0.32%	0.04%	0.01%	0.03%
# of Medial PRs	89	115	8	3
% per word	0.15%	0.05%	0.01%	0.02%
# of Final PRs	1037	1192	208	94
% per word	1.73%	0.55%	0.28%	0.57%

Table 9: The positions of prolongations in the word relative to word positions in the phrase for function word class.

	Single-W	Initial-W	Medial-W	Final-W
# of Single PRs	78	19	722	2581
% per word	2.45%	0.23%	0.51%	1.24%
# of Initial PRs	0	2	24	16
% per word	0.00%	0.02%	0.02%	0.01%
# of Medial PRs	0	0	2	1
% per word	0.00%	0.00%	0.00%	0.00%
# of Final PRs	11	6	130	619
% per word	0.34%	0.07%	0.09%	0.30%

In general, for mono-moraic words solely comprising a phrase (Single-W-Single), the PR rate is very high. For content words, the PR rate for the word-final position is also high when the word solely comprises a phrase (Single-W-Final), and relatively high when the word is at a phrase boundary (Initial-W-Final/Final-W-Final). Function words consisting of a single mora are frequently prolonged when it appears at a phrase boundary (Final-W-Single). The PR rates for non-word-final positions are very low.

3.6. Word positions in the inter-pausal unit

We conducted an analysis similar to the previous section considering word positions in the IPU. The PR rates relative to word positions in the IPU for the content word class and for the function word class are shown in Tables 10 and 11, respectively.

Table 10: The rates of prolongations relative to word positions in the inter-pausal unit for the content word class.

	Single-W	Initial-W	Medial-W	Final-W
# of words	4753	70360	282122	11954
# of morae	11921	197310	760698	33550
# of PRs	544	1404	1791	377
% PRs/words	11.45%	2.00%	0.63%	3.15%
% PRs/morae	4.56%	0.71%	0.24%	1.12%

Table 11: The rates of prolongations relative to word positions in the inter-pausal unit for the function word class.

	Single-W	Initial-W	Medial-W	Final-W
# of words	669	6851	289275	65191
# of morae	787	8849	365886	80448
# of PRs	32	66	2023	2090
% PRs/words	4.78%	0.96%	0.70%	3.21%
% PRs/morae	4.07%	0.75%	0.55%	2.60%

Breakdowns of these tables according to PR positions in the word are shown in Tables 12 and 13, respectively.

Table 12: The positions of prolongations in the word relative to word positions in the inter-pausal unit for content word class.

	Single-W	Initial-W	Medial-W	Final-W
# of Single PRs	312	603	137	26
% per word	6.56%	0.86%	0.05%	0.22%
# of Initial PRs	9	86	178	19
% per word	0.19%	0.12%	0.06%	0.16%
# of Medial PRs	6	60	138	11
% per word	0.13%	0.09%	0.05%	0.09%
# of Final PRs	217	655	1338	321
% per word	4.57%	0.93%	0.47%	2.69%

Table 13: The positions of prolongations in the word relative to word positions in the inter-pausal unit for function word class.

	Single-W	Initial-W	Medial-W	Final-W
# of Single PRs	26	58	1728	1588
% per word	3.89%	0.85%	0.60%	2.44%
# of Initial PRs	0	2	25	15
% per word	0.00%	0.03%	0.01%	0.02%
# of Medial PRs	0	0	2	1
% per word	0.00%	0.00%	0.00%	0.00%
# of Final PRs	6	6	268	486
% per word	0.90%	0.09%	0.09%	0.75%

We can observe tendencies similar to those observed in Tables 6–9. Besides these tendencies, we can see in Table 12 that the PR rates for mono-moraic words and for the word-final position are relatively high when the word is a content word and appears at the initial position in the IPU (Initial-W-Single/Initial-W-Final).

In Table 12, we can see that for mono-moraic words solely comprising an IPU (Single-W-Single), the PR rate is particularly high. Since we employ IPUs, not syntactic units, for the unit of analysis, this can be interpreted in the following way: When a mono-moraic word at the initial position in an utterance is prolonged, it is usually followed by a pause (hence, by an IPU boundary). A typical example is the prolongation of discourse markers such as *de* and *zya*. For

example, in the CSJ, there are 288 instances of prolonged *de* followed by a pause longer than 200ms.

3.7. Preceding fillers

Finally, we examine the effect of fillers on PR rates. The PR rates for words with one or more preceding fillers, with preceding fillers plus intervening or following silent pauses, and with no preceding fillers, indicated by `Fillers', `Fillers + Pause', and 'None' classes, respectively, are shown in Table 14, and its breakdown according to RP positions in the word is shown in Table 15.

Table 14: The rates of prolongations relative to preceding disfluent element classes.

	None	Fillers	Fillers+Pause
# of words	695966	37490	8435
# of morae	1339380	109685	24420
# of PRs	7701	537	154
% PRs/words	1.11%	1.43%	1.83%
% PRs/morae	0.57%	0.49%	0.63%

Table 15: The positions of prolongations in the word relative to preceding disfluent item classes.

	None	Fillers	Fillers+Pause
# of Single PRs	4474	40	10
% per word	0.64%	0.11%	0.12%
# of Initial PRs	311	18	9
% per word	0.04%	0.05%	0.11%
# of Medial PRs	185	26	7
% per word	0.03%	0.07%	0.08%
# of Final PRs	2731	453	128
% per word	0.39%	1.21%	1.52%

Table 15 shows that the PR rate becomes higher when the word is preceded by more disfluent items. This tendency is conspicuous in the word-final prolongation as in Table 15. It, however, is not observed for mono-moraic words. When these words are preceded by fillers, the PR rate remarkably decreases. This would suggest that the prolongation of mono-moraic words and the production of fillers are complementary.

4. Discussion

Based upon the empirical findings shown in the previous section, we can now state several strategies in prolonging speech segments used by Japanese speakers.

1. From the results of Sections 3.6 and 3.7: Japanese speakers frequently prolong utterance initial, mono-moraic words. These are typically discourse markers such as *de* and *zya*, and distribute complementarily with fillers. This usage might have the same function as fillers.
2. From the results of Section 3.6: Japanese speakers sometimes prolong the final vowels of utterance initial content words. These also include discourse markers, but nouns, including demonstrative nouns, are another typical example of this pattern. We conjecture that these nouns serve a topic of an utterance, since demonstrative nouns are often used as anaphoric expressions. This usage might be related to the information structure of the utterance.
3. From the results of Sections 3.5 and 3.6: Japanese speakers sometimes prolong the final vowels of phrase-final content words. These are mainly common nouns, but details are unclear.
4. From the results of Sections 3.5 and 3.6: Japanese speakers often prolong the final vowels of phrase-final function words, especially immediately before a silent pause. Typical examples include conjunctive particle *te*, politeness marker *masu*, copula *da*, and topic particle *mo*. These would be accounted for from a phonological point of view, i.e., as instances of pre-pausal lengthening.
5. From the results of Section 3.4: Japanese speaker sometimes prolong the vowels, and probably consonants like fricatives, at the disruption point of word fragments. In many cases, the disrupted word is immediately restarted from the beginning, resulting in a word repetition involving prolongation of the word fragments. This could be a signal for speakers to inform listeners of difficulty in producing the rest of the constituent.
6. From the results of Section 3.2: Japanese speakers rarely prolong vowels in the middle of a word. This might be a morphological constraint which is typical of Japanese.

5. Conclusion

In this paper, we have investigated segmental prolongation in the *Corpus of Spontaneous Japanese*, and stated, based upon the empirical findings, some strategies in prolonging speech segments used by Japanese speakers. We are planning to get into details of phonological aspects of prolongation in Japanese and to construct an integrated model of the phenomena taking into account phonological, morphological, and syntactic factors as well as discourse factors.

6. Acknowledgements

This research is funded by the CREST/ESP Project of Japan Science and Technology Corporation.

7. References

[1] Den, Yasuharu. 2001. Are word repetitions really intended by the speaker? *Proc. ISCA tutorial and research workshop on Disfluency in Spontaneous Speech*, Edinburgh, UK, pp. 25–28.

[2] Eklund, Robert. 2000. Crosslinguistic disfluency modeling: A comparative analysis of Swedish and Tok Pisin human–human ATIS dialogues. *Proc. ICSLP'00*, Beijing, vol. 2, pp. 991–994.

[3] Eklund, Robert. 2002. Prolongations: A dark horse in the disfluency stable. *Proc. ISCA tutorial and research workshop on Disfluency in Spontaneous Speech*, Edinburgh, UK, pp. 5–8.

[4] Eklund, Robert & Elizabeth E. Shriberg. 1998. Crosslinguistic disfluency modeling: A comparative analysis of Swedish and American English human–human and human–machine dialogues. *Proc. ICSLP'98*, Sydney, pp. 2631–2634.

[5] Koiso, Hanae. 2001. Transcription criteria for the *Corpus of Spontaneous Japanese*. *Proc. Spontaneous Speech Science and Technology Workshop*, Tokyo, pp. 13–20.

[6] Maekawa, Kikuo. 2002. Compilation of the *Corpus of Spontaneous Japanese*: A status report. *Proc. 2nd Spontaneous Speech Science and Technology Workshop*, Tokyo, pp. 7–10.

[7] Watanabe, Michiko & Yasuharu Den. 2003. When and why do speakers prolong their speech segments? *Proc. 1st JST/CREST International Workshop on Expressive Speech Processing*, Kobe, Japan, pp. 71–74.