The following paper was presented at The 9th Workshop on Disfluency in Spontaneous Speech (DiSS 2019) held at ELTE Eötvös Loránd University in Budapest, Hungary on 12–13 September, 2019.

Title: Five pieces of evidence suggesting large lookahead in spontaneous monologue

Author(s): Kikuo Maekawa

Abstract: There is considerable disagreement among the researchers of speech production with respect to the range of lookahead or pre-planning. In this paper, five pieces of evidence suggesting the presence of relatively large lookahead in spontaneous monologues are presented, based on the analyses of the Corpus of Spontaneous Japanese. This evidence consistently suggests that the range of a lookahead is six to seven accentual phrases long, which corresponds on average to 3–4 seconds in the time domain.

DOI: https://doi.org/10.21862/diss-09-003-maekawa

Citation (JIPA): Maekawa, Kikuo. 2019. Five pieces of evidence suggesting large lookahead in spontaneous monologue. In: R. L. Rose & R. Eklund (eds.), *Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech*, 12–13 September, 2019, Budapest, Hungary, 7–10.

The complete proceedings for DiSS 2019 are available as follows.

ISBN: 978-963-489-063-8

DOI: https://doi.org/10.21862/diss-09

DiSS 2019 was sponsored by The Faculty of Humanities, ELTE Eötvös Loránd University and the International Speech Communication Association (ISCA).

Five pieces of evidence suggesting large lookahead in spontaneous monologue

Kikuo Maekawa
Division of Spoken Language, The National Institute for Japanese Language and Linguistics, Tokyo, Japan

Abstract

There is considerable disagreement among the researchers of speech production with respect to the range of lookahead or pre-planning. In this paper, five pieces of evidence suggesting the presence of relatively large lookahead in spontaneous monologues are presented, based on the analyses of the Corpus of Spontaneous Japanese. This evidence consistently suggests that the range of a lookahead is six to seven accentual phrases long, which corresponds on average to 3–4 seconds in the time domain.

Introduction

There is wide consensus among researchers of speech that human speech production involves some sort of 'lookahead' or 'pre-planning,' a process whereby a preverbal message is transformed into a verbal one prior to phonological encoding. There is, however, a lack of consensus with regard to the size of the lookahead; some say that the lookahead is, in principle, a single word (Levelt, 1988), while others suggest its domain is as wide as an intonation phrase (Keating & Shattuck-Hufnagel, 2002).

Part of the discrepancy stems from the confusion of linguistic levels. It is not surprising if the amount of lookahead observed at one level is considerably different from the lookahead observed in a different level; Levelt's study, cited above, is primarily concerned with the level of the lexicon, while Keating and Shattuck-Hufnagel (2002) are concerned with prosodic structure, which often spans a domain much wider than a word.

Another factor in this discrepancy, is the insufficiency in the quantitative analysis of spontaneous speech. It goes without saying, that research on lookahead needs to be based on the analysis of spontaneous speech. But, as Nooteboom (1995) points out, such a study is difficult to conduct, primarily due to the limitation in the size of spontaneous data available for analysis.

In this paper, evidence suggesting the presence of unexpectedly large lookahead in spontaneous monologue in Japanese will be presented; but before that, the results of previous studies that dealt with the issue of lookahead in spontaneous Japanese, either directly or indirectly, will be reviewed briefly.

Watanabe (2009) reported a positive correlation between the length of silent pauses and the grammatical complexity of the upcoming clauses; The more complex the upcoming clauses, the longer the pause.

Similarly, Koiso and Den (2015) found a direct causal relationship between four types of speech disfluencies and the complexity of the constituents that followed weak clause boundaries.

Other studies observed f0 and the length of utterance; they revealed an anticipatory f0 rise at the beginning of utterance (Ishimoto, Enomoto & Iida, 2011; Koiso & Ishimoto, 2012; and Maekawa, 2017).

Data

The analyses reported below are all concerned with the Core part of the Corpus of Spontaneous Japanese, or CSJ (Maekawa, 2003). The CSJ-Core is a speech corpus of 500,000 words (44 hours) spoken by 201 speakers of standard Japanese. It consists mainly of recordings of spontaneous monologues such as academic presentations and simulated public speaking by paid subjects of balanced in ages and genders. The CSJ was used in most of the studies reviewed in the previous section.

The CSJ-Core is annotated richly with respect to word segmentation, POS classification, clause boundary labeling, segmental and prosodic labeling by means of the X-JToBI scheme (Maekawa et al., 2002), and the bunsetsu-based dependency structure, among others.

The relational database (RDB) version of the corpus (Koiso et al., 2014) was used in the following analyses.

Analysis

Anticipatory shortening in AP

Anticipatory shortening implies a shortening of stressed syllable duration as a function of the number of other syllables within a word (Nooteboom & Slis, 1972; Bishop & Kim, 2018). This effect has not been reported for Japanese, which is a mora-timed language (Mora is the unit of phonological length in Japanese. In most, but not all cases, it coincides with the syllable).

Figure 1 shows the relationship between the length of accentual phrases (AP), as measured by the number

https://doi.org/10.21862/diss-09-003-maekawa

of constituent mora (abscissa), and the mean speaking rate (ordinate, unit is [mora/sec]). AP is the basic constituent unit of Japanese intonation, which is marked by f0 rise in the beginning and various boundary pitch movements in the end (see Pierrehumbert & Beckman, 1988).

The mean speaking rate of each AP length in the abscissa, is represented by a red circle with the standard error. The blue line is the LOESS non-parametric regression curve, and the shaded area around the curve is the 95% confidence interval. The LOESS curve was computed using the `ggplot2` library (Ver. 3.1.0) of the `R` language (Ver. 3.5.1) with the smoothing parameter (`'span'`) set to 0.9. Figure 1 is the case that pooled all 67,923 APs (excluding APs longer than 16 morae). The same tendency is observed in all AP locations (1st, 2nd, …, Nth) in an utterance. This figure shows clearly the presence of lookahead at the level of AP.

AP-internal anticipatory f0 rise

Figure 2 shows AP-internal anticipatory f0 rise. The method of presentation is basically the same as in Figure 1. The abscissa stands for the length of AP and the ordinate stands for the *z*-normalized logarithm of f0. The ordinate value plotted here is the difference between the AP-initial low tone (%L), and the phrasal high tone (H−). See Pierrehumbert and Beckman (1988) for the tonal structure of AP in Japanese. The longer the AP, the larger the phrase-initial f0 rise.

Utterance-internal anticipatory f0 rise

Figure 3 shows the anticipatory f0 rise as observed in the beginning of utterance. The abscissa and ordinates stand, respectively, for the length of utterance measured in terms of the number of constituent APs, and the *z*-normalized logarithm of
the mean f0 of the first AP of utterance. There is a clear positive correlation between the utterance length and the mean f0 of the first AP, in the range between 2–6 APs in the abscissa, before the curve reaches the plateau. Note that utterances that consist of a single AP are omitted from the analysis. Note also, that unlike the Figures 1 and 2 that dealt with the lookahead within an AP, Figure 3 shows the evidence of lookahead in much larger prosodic domain.

Duration of the silent pause

As noted above, Watanabe (2009) found a correlation between the silent pause length and utterance length. Here, similarly, Figure 4 plots the relationship between the length of the upcoming utterance (number of constituent AP, abscissa), and the duration ([sec.]) of the silent pause preceding the utterance. Note that exceedingly long silent pauses (those longer than 3 seconds) were omitted from the analysis, because long silent pauses are often caused by troubles that are external to speech, for example,

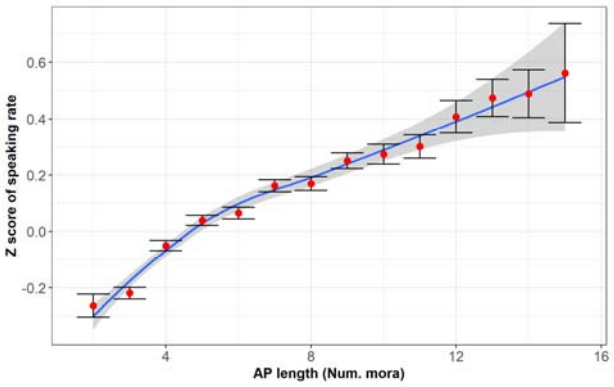

Figure 1. Anticipatory shortening within an AP.

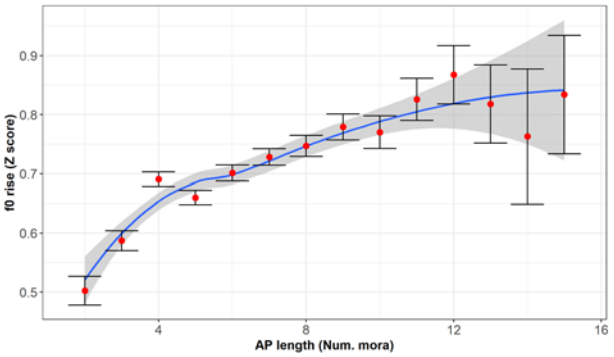

Figure 2. Anticipatory f0 rise within an AP.

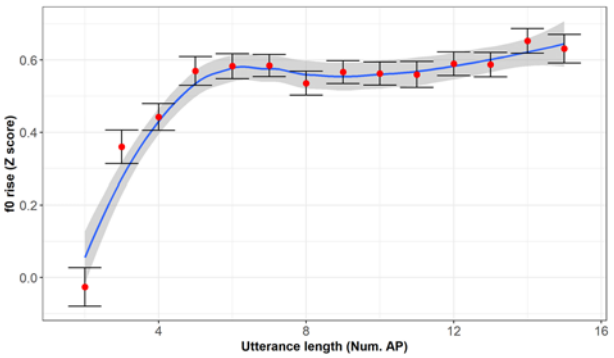

Figure 3. Anticipatory f0 rise within an utterance.

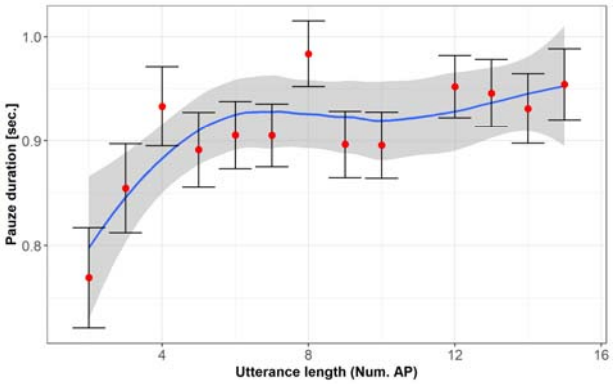

Figure 4. The length of upcoming utterance and the duration [sec.] of utterance-initial silent pause.

difficulties in handling presentation material such as the overhead projector and microphone. Here, the pause duration increases within the range of 1–5 APs, before it reaches the plateau. Note also that this figure has a relatively large confidence interval.

Dependency distance and the f0

Lastly, the relation with syntactic complexity is examined using the dependency structure. In traditional Japanese grammar and Japanese natural language processing, syntactic structure is often represented by means of 'bunsetsu', which is a surface syntactic unit that usually consists of one content word followed by one or more function words.

In Figure 5, the upper part shows the bunsetsu dependency relationship in the example of "kinoo Taro-to Hanako-ga Kyoto-made it-ta" ("Yesterday Taro and Hanako went to Kyoto"), where the time adverbial "kinoo" at the beginning of the sentence modifies the last bunsetsu, which is the predicate and the fourth bunsetsu from the time adverbial. On the other hand, the second bunsetsu modifies the third one. In these cases, the dependency distance is counted as 3 and 0 respectively; the general rule is that if a modified bunsetsu is *N* phrases apart from the modifying bunsetsu, the distance is *N*−1.

The lower part of the figure shows the dependency relation among APs. In Japanese, it is often the case that more than one bunsetsu merged into a single AP. In this example, the second and third bunsetsu on the one hand, and the fourth and fifth on the other, are merged. The dependency distance between these APs is counted in the same manner as in the case of bunsetsu (see Figure 5).

Figure 6 shows the relationship between the dependency distance of a given AP and the mean difference in the *z*-normalized logarithm f0, between the AP in question and the phrase that follows immediately.

This figure shows, that the behavior of f0 is different depending on the presence of discontinuity in the dependency; when there isn't discontinuity (i.e. the distance is zero), the f0 difference is negative, implying that the f0 of the following AP is lower than the AP in question; on the other hand, when there is large discontinuity (e.g. the distance is larger than two), the difference is positive, meaning that the following AP has higher f0 than the AP in question.

Note that the f0 rise in the case of discontinuous dependency cannot be explained as a simple resetting of downstep (Pierrehumbert & Beckman, 1988). Here, the f0 difference between the two APs is not positive when the distance is one, while most theories of downstep (Kubozono, 1993 among others) predict that the resetting occurs whenever there is discontinuity. In any case, the correlation between the distance in dependency, and the f0 behavior can be observed in the range between 0 and 5 or 6 in the abscissa. It is equivalent to saying that the maximum range of lookahead is 6 or 7.

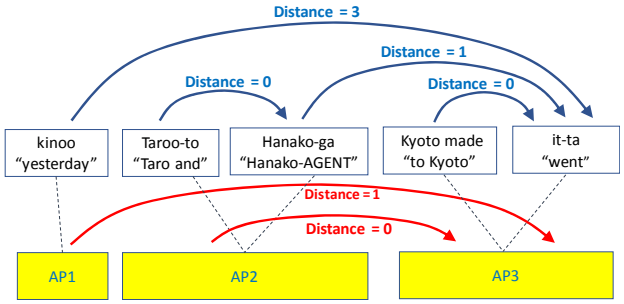

Figure 5. Schematic representation of the dependency distance among bunsetsu and accentual phrases.

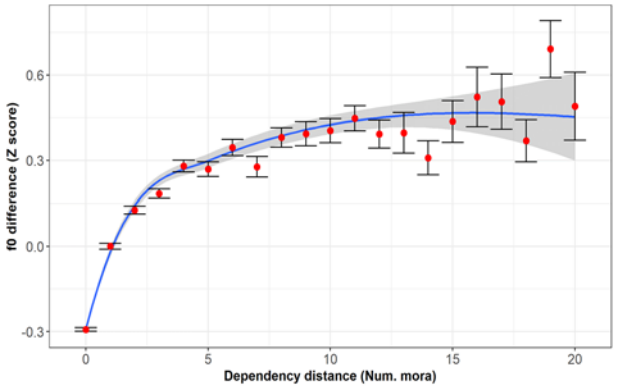

Figure 6. Dependency distance of an AP and the mean f0 difference between the AP and the immediately following AP.

Discussions and conclusion

The results reported in the previous section are the evidence of the presence of lookahead in spontaneous monologue of Japanese. They include evidence for lookahead both inside and outside of AP. Inside an AP, the relationship between the AP length and the mora shortening, or AP-initial f0 rise, was almost linear without any plateau effect; this suggests that there is no clear limit in the size of lookahead within an AP.

On the other hand, there were limits in the size of lookahead observed at the level of utterance; all the figures presented above the curves suggesting the presence of lookahead showed clear plateau effects. More importantly, the maximum size of lookahead estimated independently, coincided almost with the values between 5 and 7 APs.

Given that the mean duration and mean mora length of AP in the CSJ-Core are 0.56 sec and 4.91 morae respectively, the size of lookahead suggested above corresponds to about 2.8–4.0 sec and 25–35

morae. To which linguistic unit, then, does this rather large domain of utterance-level lookahead correspond?

The most obvious possibility is the intonation phrase. In Japanese, an intonation phrase, aka intermediate phrase, is a prosodic domain consisting of several APs, and it is presumed to be the domain of downstep (Pierrehumbert & Beckman, 1988). Several studies concluded that the domain of lookahead in English is the intonation phrase (Keating & Shattuck-Hufnagel, 2002; Bishop & Kim, 2018).

In Japanese, however, this hypothesis is difficult to support. One negative evidence is that, in the case of the CSJ-Core, the mean length of the intonation phrase, which is defined as the prosodic domain as demarcated by the boundary indices (BI) 3 on both edges in the X-JToBI annotation, is much shorter than the size of lookahead estimated in the current study.

In the CSJ-Core, nearly one-third of the intonation phrases coincide with AP (i.e. intonation phrase consists of a single AP), and the mean length of the intonation phrases is about 1.70 AP. Even if we remove the cases of a single AP phrases from the computation, the mean length is no longer than 2.64.

The second negative evidence comes from results in Figure 6. This figure showed that the difference of f0 between the AP in question and the immediately following AP is, on average, close to zero when the dependency distance is 1, and it keeps increasing until it reaches the plateau of about 0.2 at around the distance of 6.

Since a positive value in the f0 distance implies the resetting of downstep (Kubozono, 1993), and thereby the presence of an intonation phrase boundary, the figure suggests that there are many instances where intonation boundaries are included within the domain of lookahead.

At the present stage of this study, it is difficult to conclude exactly what the domain of lookahead corresponds to, but it seems plausible that the domain is larger than the intonation phrase. It might be a type of clause; or it might be that a simple correspondence between the domain of lookahead and linguistic or prosodic structure does not exist. It would not be surprising if the domain of lookahead differs considerably from one speaker to another, reflecting the difference in their working memories.

Acknowledgments

This work is supported by the Kakenhi grants (17H02339 and 19X21641) and the budget of the Center for Corpus Development, NINJAL. The author thanks his colleagues in NINJAL for their comments on earlier version of the paper.

References

Bishop, J. & B. Kim. 2018. Anticipatory shortening: Articulation rate, phrase length, and lookahead in speech production. In: K. Klessa, J. Bachan, A. Wagner, M. Karpiński & D. Śledziński (eds.), *Proceedings of Speech Prosody*, 13–16 June 2018, Poznań, Poland, 13–17. https://doi.org/10.21437/SpeechProsody.2018-48

Ishimoto, Y., M. Enomoto & H. Iida. 2011. Projectability of transition-relevance places using prosodic features in Japanese spontaneous conversation. In: P. Cosi, R. De Mori, G. Di Fabbrizio & R. Pieraccini (eds.), *Proceeding of Interspeech,* 27–31 August, Florence, Italy, 2061–2064.

Keating, P. & S. Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101: 112–156.

Koiso, H. & Y. Den. 2015. Causal analysis of acoustic and linguistic factors related to speech planning in Japanese monologs. In: *DiSS 2015, Proceedings of the 7th Workshop on Disfluency in Spontaneous Speech*, 8-9 August 2015, University of Edinburgh, Scotland, UK.

Koiso, H. & Y. Ishimoto. 2012. Nihongo hanashikotoba koopasuo mochiita hatsuwano inritsutekitokuchoono bunseki: Intoneeshonkuwo kirikuchito shite [Prosodic Features of Utterances in the Corpus of Spontaneous Japanese: Intonational Phrase–Based Approach]. In: *Proceedings of the 1st Corpus Japanese Linguistics Workshop*, 56 March 2012, Tokyo, Japan, 167–176.

Koiso, H., Y. Den, K. Nishikawa & K. Maekawa. 2014. Design and development of an RDB version of the Corpus of Spontaneous Japanese. In: N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (eds.), *Proceedings of LREC 2014*, 26 May 2014, Rejkjavik, Iceland, 311–315.

Kubozono, H. 1993. *The Organization of Japanese Prosody*. Tokyo: Kuroshio Publishing.

Levelt, W. J. M. 1988. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Maekawa, K. 2003. Corpus of Spontaneous Japanese: Its Design and Evaluation. In: *Proceedings of SSPR 2003*, 13–16 April 2003, Tokyo, Japan, 7–12.

Maekawa, K. 2017. A new model of final lowering in spontaneous monologue. In: *Proceedings of Interspeech 2017*, 20–24 August 2017. Stockholm, Sweden, 1233–1237. https://doi.org/10.21437/Interspeech.2017-175

Maekawa, K., H. Kikuchi, Y. Igarashi & J. Venditti. 2002. X-JToBI: An extended J_ToBI for spontaneous speech. In: *Proceedings of ICSLP 2002*, 16–20 September 2002, Denver, CO, 1545–1548.

Nooteboom, S. G. 1995. Limited lookahead in speech production. In: F. Bell-Berti & L. R. Raphael (eds.), *Producing speech: Contemporary issues—for Katherine Safford Harris*, NY: AIP Press, 3–18.

Nooteboom, S. G. & I. H. Slis. 1972. The Phonetic Feature of Vowel Length in Dutch. *Language and Speech*, 15(4): 301–316. https://doi.org/10.1177/002383097201500401

Pierrehumbert, J. & M. Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Watanabe, M. 2009. *Features and Roles of Filled Pauses in Speech Communication*. Tokyo: Hituzi.