

The Intonational Phrase as a planning unit: Some phonetic evidence

Mei-Ying Ki, Jason Bishop

City University of New York

During speech production, speakers engage in multiple stages of encoding/planning a spoken utterance in advance of actual articulation. It remains unclear what unit or units speakers utilize at the early stage of planning coarse phonological information, although it has been claimed that speakers plan in chunks at least as large as Intonational Phrases [1], rather than smaller, word-sized units [2]. Yet prosodic structure consists of multiple levels: in English, for example, relevant units include the intermediate phrase (ip) and the Intonational Phrase (IP) [3], although research on planning scope has often relied on phonetic units such as the utterance/interpause intervals (IPI) rather than phonological constituents like the ip or IP [4].

In the present study, we use production data collected in the context of a previous study to estimate the extent of lookahead speakers typically engage in, with the hypothesis that planning is generally limited to a single IP. Connected speech by a large group of speakers was used to test two things: pre-phrasal pause durations and the height of utterance-initial f_0 peaks, both of which are indicative of speakers' planning, as both are sensitive to the amount of upcoming speech material [5, 6; see also 7]. We therefore examine whether these cues best correlate with the amount of speech material in an upcoming ip, IP, or IPI. Data come from a corpus of read speech produced by 144 monolingual English speakers (74 females and 70 males) from New York City, all reading the same 163-word passage. Mixed-effects linear regression was used to predict durations of silent pauses and the height of utterance-initial f_0 peaks as a function of the length and prosodic structure of upcoming speech.

Silent pause duration: We distinguish between two types of IPIs: simple IPIs (consisting of a single IP) and complex IPIs (consisting of multiple IPs). For each type, we calculate the correlation between the length of an upcoming IPI (measured in syllables) and the duration of the preceding silent pause. If speakers plan the entire IPI, the correlations are expected to be similar for both simple and complex IPIs. If, however, speakers plan only a single IP instead, a stronger correlation is expected for simple IPIs, as the planning process of additional IPs in complex IPIs would not be reflected in the preceding pause and would therefore weaken the correlation. The results support the latter pattern, indicating that speakers' planning is likely limited to the upcoming IP rather than the entire IPI (Fig. 1).

Utterance-initial f_0 : This part of analysis used a subset of the corpus (63 speakers; 34 females and 29 males). For each IPI, the first IP was identified, followed by the identification of the first ip within that IP. The f_0 peak within this ip was automatically identified using Parselmouth [8], a Python library that works with Praat [9]. Specifically, f_0 information was extracted using Praat's raw autocorrelation algorithm, after which the f_0 peak within the ip was detected. All automatically identified f_0 peaks were then manually checked and adjusted in Praat to eliminate errors arising from factors such as f_0 tracking errors, creaky voice, and microprosody. The adjusted f_0 peaks for each ip were subsequently re-extracted using Parselmouth. Following the logic of the analysis of silent pause duration, we compare the relationship between utterance-initial f_0 and upcoming length for simple versus complex IPIs. Consistent with IP-limited planning, a stronger correlation is observed with simple IPI lengths rather than with complex IPI lengths (Fig. 2).

In summary, results from both preceding silent pause duration and utterance-initial f_0 suggest that speakers' planning scope is likely by default limited to the upcoming Intonational Phrase (although speakers may flexibly adjust the length of the IP [10]). We also discuss individual differences that may modulate this typical amount of lookahead.

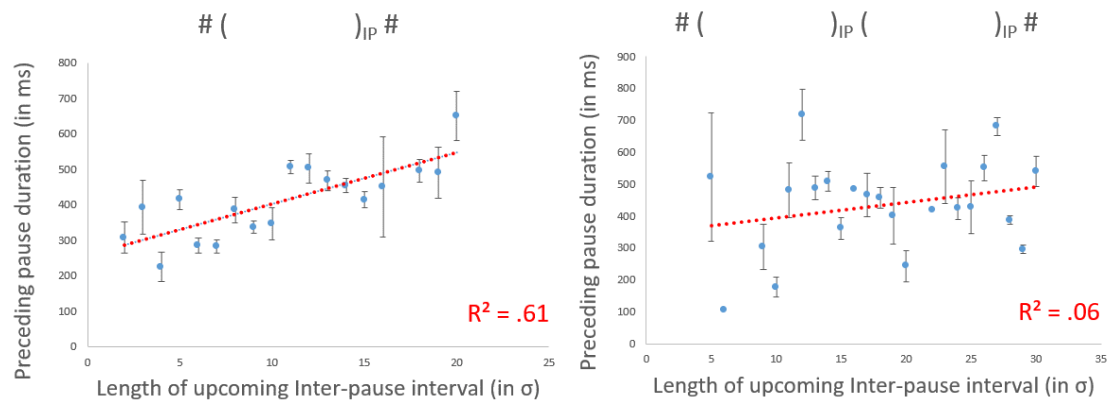


Fig. 1. Correlations between preceding silent pause duration and the length of an upcoming IPI. Left: Before simple IPIs (those with one single IP); Right: Before complex IPIs (those with multiple IPs).

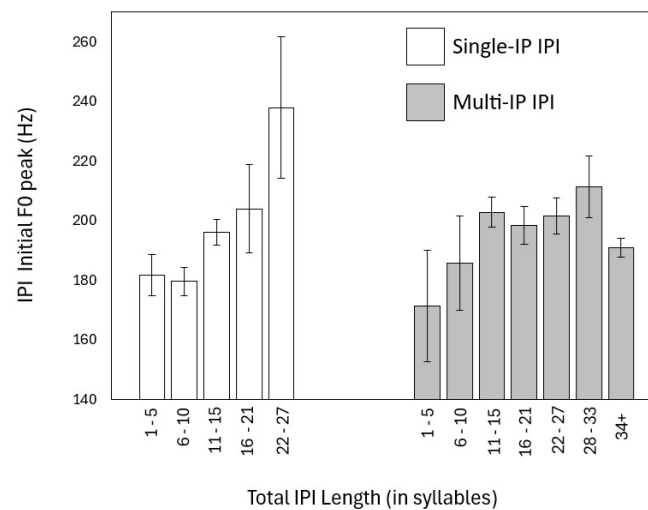


Fig. 2. The relationship between utterance-initial f0 and the length of an upcoming IPI. Left: Simple IPIs (those with one single IP); Right: Complex IPIs (those with multiple IPs).

References

- [1] Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 101, pp. 112–156.
- [2] Beckman, M. and J. Pierrehumbert (1986). Intonational structure in Japanese and English, *Phonology Yearbook*, III, 15-70.
- [3] Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38.
- [4] Whalen, D., & Kinsella-Shaw, J. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica*, 54 (3-4), 138–152.
- [5] Krivokapić, J. (2012). Prosodic planning in speech production. In Susanne Fuchs, Melanie Weihrich, Daniel Pape, Pascal Perrier (eds.) *Speech planning and dynamics*. Peter Lang, pp. 157–190.
- [6] Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin broadcast news speech. *Speech Communication*, 65, 67-74.
- [7] Kim, S. E., & Tilsen, S. (2024). Planning for the future and reacting to the present: Proactive and reactive F0 adjustments in speech. *Journal of Phonetics*, 104, 101322.
- [8] Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1–15.
- [9] Boersma, P., & Weenink, D. (2026). Praat: doing phonetics by computer [Computer program]. Version 6.4.58, retrieved 25 January 2026 from <https://praat.org>
- [10] Bishop, J., & Intlekofer, D. (2020). Lower working memory capacity is associated with shorter prosodic phrases: Implications for speech production planning. *Proceedings of Speech Prosody 10*, 191–195.