THE KINEMATIC PROPERTIES OF PROSODIC BOUNDARIES IN CONVERSATIONAL TURN-TAKING

Ruaridh Purse[1] and Jelena Krivokapić[1, 2]

[1]University of Michigan, [2]Haskins Laboratories
rupurse@umich.edu, jelenak@umich.edu

ABSTRACT

Studies on the kinematic properties of read speech have established an effect of lengthening on gestures at prosodic boundaries, which increases in strength according to the hierarchical position of the boundary. However, the effect of turn-position—whether a boundary occurs in the middle or at the end of a conversational turn—on these properties is underexplored. In an electromagnetic articulography semi-spontaneous speech study, pairs of participants (with one speaker in the magnetometer per dyad) asked each other questions to collaboratively solve a puzzle. We compare targetwords in phrase-medial position, phrase-final turn-medial position, and phrase-final turn-final position, in both questions and answers, exploring two competing hypotheses: (1) turn-final boundaries exhibit a stronger lengthening effect owing to a higher hierarchical position, and (2) lengthening is reduced in turn-final position since there is no imminent speech to plan, thus no additional lengthening is required to accommodate that planning time.

Keywords: Turn-taking, final lengthening, prosody, kinematics, planning

1. INTRODUCTION

Articulatory gestures at prosodic boundaries are longer than equivalent phrase-medial gestures [1,2,3,4]. This lengthening effect is larger at hierarchically higher prosodic boundaries (those delimiting higher prosodic units) and smaller and hierarchically lower boundaries [1,3,4]. Research on the scope of this lengthening effect finds that it extends from the phrase edge. Gestures and acoustic segments closest to the edge (e.g., the last gesture of a phrase-final word) are most strongly and reliably affected. Gestures and acoustic segments further away show less or no lengthening relative to phrase-medial gestures [5,6,7,8,9,10].

Studies describing the kinematic properties of prosodic boundaries in this way, robust as they are, have been almost exclusively limited to read speech monologues. While we do not expect read speech and spontaneous speech to be fundamentally different, some kinematic properties may differ depending on context: Specifically, final lengthening may differ depending on whether speakers need to plan an upcoming utterance. Indeed, longer or structurally more complex utterances are preceded by longer pauses than shorter and/or less complex utterances [11,12,13,14]. This has been related to the speakers' need to plan the upcoming utterance, with longer pauses allowing the speaker more time to plan upcoming speech. Since pauses are part of prosodic boundaries, and speakers plan speech continuously, it is likely that speech planning will also affect final lengthening.

The present study examines final lengthening in turn-taking. Despite the broad interest in the timing properties of turn-taking in the field of Conversation Analysis [15,15], the temporal properties of boundaries at turn-ends have not been examined extensively [cf. 17] and the fine-grained kinematic properties of prosodic boundaries at turn-ends remain unexplored. We propose two competing hypotheses about how phrase-final lengthening at turn-final prosodic boundaries compares to final lengthening at turn-medial boundaries if, indeed, these conditions are found to differ at all.

(1) Turn-final prosodic boundaries are hierarchically higher than turn-medial and therefore induce more lengthening in nearby gestures.

(2) Planning loads are lighter turn-finally than turn-medially and therefore turn-final gestures exhibit less planning-induced lengthening compared to phrase-final turn-medial gestures.

We also explore the turn-medial lengthening effect by comparing phrase-final turn-medial gestures to phrase-medial gestures (preceding a word boundary).

2. METHODS

2.1. Procedure

2.1.1. Participants

Six pairs of participants, native speakers of American English, participated in the study. They were naïve to the purposes of the study. In each dyad, synchronized acoustic and kinematic recordings were taken from one participant (participant A) using a Carstens AG501 Articulograph. The other participant (participant B) acted as an interlocutor.

2.1.2. Stimuli and experimental design

There are three conditions (phrase-medial, phrase-final turn-medial, phrase-final turn final) and one targetword, *mima* (pronounced ['mimə]).

The targetwords were elicited using a modified game of tangrams, with each participant solving up to 10 tangrams. For each trial, both participants were shown the same two tangram puzzles, with the outlines of their component shapes visible. For one of these tangrams, its component shapes were colored and labelled; labels for triangles contained the targetword *mima* and labels for quadrangles contained the targetword *nana*. The shapes making up the other visible tangram were left blank. Each targetword was embedded in a phrase consisting of the definite article, a color word, and a targetword, for example "the Lava Nana", "the Big Mocha Mima"). Participants were instructed how to produce the targetwords (['mimə]) and ['nɑnə]). To ensure that the targetwords do not have a pitch accent, they were also instructed to place emphasis on the color words.

The labelled tangram seen by participant A matched the blank tangram seen by participant B and vice versa, such that each participant had access to information about the location of shapes in their interlocutor's uncompleted tangram (Fig. 1). Each participant was prompted to find the location of the labelled shapes in their blank tangram by asking their interlocutor questions.

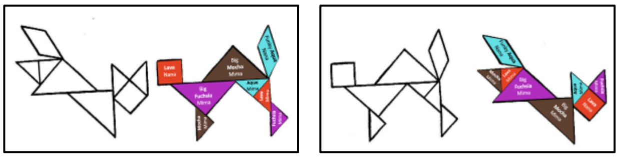

Figure 1: Example tangrams seen by participant A (left) and participant B (right) for a given trial.

Participants took turns asking and answering questions about the position of one named shape relative to another. On each turn, a participant could choose to ask one question or two questions, ensuring that sentences occur in both turn-medial and turn-final condition. To ensure that the targetwords occurred in all boundary conditions (phrase-medial, phrase-final turn-medial, phrase-final turn final), participants were asked to produce the questions and statements in a specific format. For example, for the two-question sequence, the template was: "Is the Aqua Mima to the right of the Big Mocha Mima? And is the Lava Nana right above the Fuchsia Mima?" (with the first *mima* being phrase-medial, the second *mima* being phrase-final turn-medial, and the third *mima* being phrase-final turn-medial). The answers had the parallel format (Table 1).

One question	Is the **Aqua** Mima$_1$ to the right of the Big **Mocha** Mima$_3$?
Two questions	Is the **Aqua** Mima$_1$ to the right of the Big **Mocha** Mima$_2$? And is the **Lava** Nana above the **Fuchsia** Mima$_3$?
One-sentence answer	Yes, the **Aqua** Mima$_1$ is to the right of the Big **Mocha** Mima$_3$.
Two-sentence answer	Yes, the **Aqua** Mima$_1$ is to the right of the Big **Mocha** Mima$_2$. But no, the **Lava** Nana is not above the **Fuchsia** Mima$_3$.

Table 1: Template for sentences. Bolding indicates targeted pitch accents. The indices at the targetwords indicate the boundary condition: targetword$_1$ indicates phrase-medial, targetword$_2$ indicates phrase-final turn-medial, and targetword$_3$ indicates phrase-final turn-final.

2.2. Analysis

2.1.1. Data labelling

Targetword tokens were labelled for their phrase position, turn position, and whether they were used as part of a question or an answer. Targetwords in the middle of a question or answer were labelled "phrase-medial" and act as a control condition. Targetwords at the end of a whole question or answer turn were labelled both "phrase-final" and "turn-final", and targetwords at the end of a question or answer in a two-question or two-answer turn were labelled both "phrase-final" and "turn-medial".

Kinematic data were semi-automatically labelled in MATLAB using the *findgest* function in Mview (custom software written by Mark Tiede at Haskins Laboratories, New Haven, CT). Using velocity criteria, we label gesture onset, peak velocity of the closing movement, nucleus onset, maximum constriction (velocity minimum), nucleus offset, peak velocity of the constriction opening movement, and gesture offset. For the targetword *mima*, the lip aperture trajectory was used, for *nana*, the vertical tongue tip movement was used. The gestures for both consonants were labelled (Fig. 2).

From these measures we calculate the following variables for each consonant gesture:

- Closing movement: from gesture onset to nucleus offset
- Opening movement: from nucleus offset to gesture offset

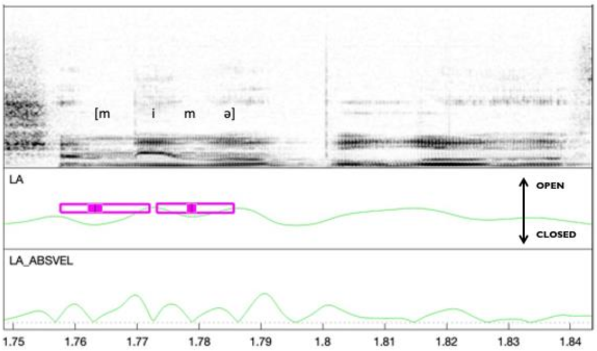

Figure 2: Sample token with the word *mima*, in the context of "Is the big fuchsia mima below…" The rectangles show labelled vocal tract gestures. The whole rectangle is the gesture, the filled part the gesture nucleus, the dashed line the time of maximum constriction. LA: the lip aperture trajectory (for the consonant [m]).

To ascertain that targetwords were produced in the intended boundary position, pauses after targetwords were also labelled according to acoustic cues. Phrase-medial targetwords followed by a pause longer than 100ms were excluded from the analysis [19] as these could not be ascertained to not be phrase-final targetwords. Targetwords preceded by pauses were also excluded from the analysis, as these might be phrase-initial, and therefore subject to phrase-initial lengthening. Targetwords produced with disfluencies or a pitch accent (as noted by a trained ToBI labeller) were also excluded.

2.1.2. Statistical analysis

Analysis is limited to instances of the targetword *mima* in the four speakers analysed to date, owing to the scarcity of instances of the targetowrd *nana* following exclusions.

	P. Medial		T. Medial		T. Final	
	Q.	A.	Q.	A.	Q.	A.
M1A	22	40	10	16	19	23
F3A	67	53	11	13	37	23
F4A	35	62	16	21	24	29
F5A	25	45	11	19	15	15

Table 2: Per-speaker token counts by phrase/turn position and sentence type (question/answer).

Separate mixed-effects linear regression models were built predicting variance in duration for each of the four extracted movements: C1 closing, C1 opening, C2 closing, and C2 opening. Each model contained an interaction between phrase position and sentence type (sum-coded) and a random intercept for speaker. Additional linear regressions were used to model individual speaker behavior.

3. RESULTS

Aggregating across speakers and sentence types (Fig. 3), the closing movements are longer on average in phrase-final turn-medial position than in phrase-medial position for both C1 [7.1ms, p<.05] and C2 [3.9ms, p<.05]. However, C2 opening movements—which lie closest to the prosodic boundary—are not significantly longer in phrase-final turn-medial position than in phrase-medial position. This pattern also holds for two out of four speakers individually (M1A, F4A), while the other two speakers do not exhibit significant final lengthening on any measures.

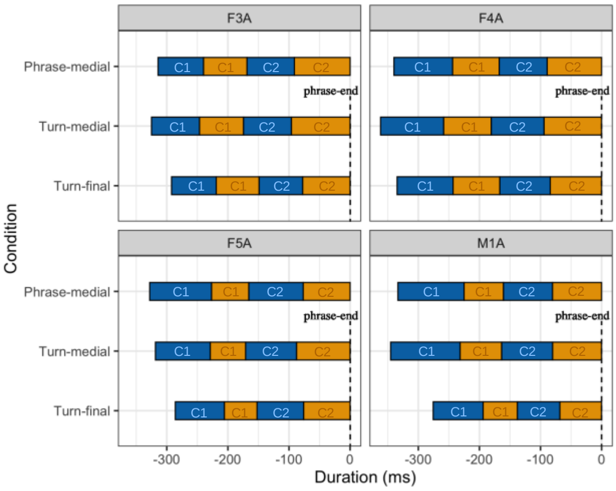

Figure 3: Average duration of C1 and C2 closing (blue) and opening (orange) movements.

Contrary to our first hypothesis, we also do not find more pronounced lengthening in turn-final position compared to phrase-final turn-medial position. In fact, in aggregated data, phrase-final turn-medial movements are longer compared to turn-final movements across the board: for C1 closing [14.7ms, p<.001], C1 opening [5ms, p<.001], C2 closing [8ms, p<.001] and C2 opening [10.6ms, p<.001]. Every individual speaker also produced at least two of these four movements significantly shorter [p<.05] in turn-final position, with C2 closing and opening at least trending towards significantly shorter turn-finally for all four individuals [p<.08].

3.1 Kinematic properties of different sentence types

Comparing the kinematic properties of prosodic boundaries in questions and answers separately (Fig. 4), we find significant duration differences in all measures between sentence types for phrase-medial targetwords in the aggregate data. Responses show shorter C1 closing [5.4ms, p<.01], C1 opening [4.7ms p<.001], and C2 closing [4.2ms p<.001]. However, phrase-medial targetwords in responses also show a longer C2 opening [6.3ms p<.001] than in questions. The only interaction effect that emerges between sentence type and phrase/turn position shows that there is no such effect of longer C2 opening in questions in turn-final position.

Looking at individuals, it emerges that one speaker (F5A) *does* produce a C2 opening that is significantly longer in phrase-final, turn-medial position than phrase-medial position, even though this effect is not evident in the aggregate data, but only in questions [11.5ms, p<.001].

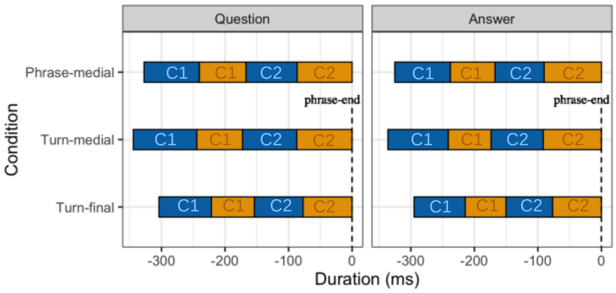

Figure 4: Average aggregate duration of C1 and C2 closing (blue) and opening (orange) movements.

The effect whereby turn-final gestures are shorter than turn-medial phrase-final gestures holds significantly for most measurements across all individuals in both sentence types.

4. DISCUSSION

4.1. Reduced planning load turn-finally

Across the board, turn-final gestures are reliably shorter than phrase-final turn-medial gestures. Thus our hypothesis that a greater lengthening effect may be found here due to a hierarchically stronger prosodic boundary was not borne out. This finding also contrasts with previous studies associating turn-ends with a lengthening effect [15, 19].

The results support our second hypothesis that the effect of final lengthening would be reduced in turn-final position, relative to turn-medial prosodic boundaries, because of a difference in planning load. In turn-medial positions, additional lengthening may occur in order to afford a speaker more time to plan upcoming speech in the same turn. In turn-final position, there is no need for this additional lengthening since there is no more speech left to plan in that turn.

4.2. Simultaneity of lengthening effects

The effect of planning may also explain the pattern of results found for the comparison between phrase-medial and phrase-final turn-medial gestures. That is, the highest planning load may be found in phrase-medial positions, which are relatively early in a sentence, as speakers continuously formulate their questions and answers as they produce them. Thus, the need for more planning time might lead to lengthening. Even though phrase-medial tokens with following pauses greater than 100ms were excluded, their original presence in the dataset supports that planning loads were high in phrase-medial position [see also 20,21]. As such, the effects of structural prosodic lengthening may be obscured by planning-induced lengthening in phrase-medial position, which reduces the difference in gesture duration that might otherwise be found between phrase-medial and phrase-final movement.

The notion that planning-induced lengthening may obscure structural lengthening is also relevant for the turn-final position. We cannot rule out that an underlying effect of structural prosodic lengthening is present in turn-final position, only that it is not large enough to be evident under the current conditions. Put differently, the sources of increased or reduced effects of final lengthening in turn-final position that we illustrated in our original two hypotheses are not necessarily mutually exclusive. It may be true that turn-final phrase boundaries, since they are hierarchically higher, induce a stronger effect of structural lengthening. However, this effect will only be evident when the difference in speech planning load between turn-medial and turn-final position is not so large as to induce an even greater reduction in planning-induced lengthening turn-finally.

5. CONCLUSIONS

Gestures at turn-final prosodic boundaries are produced with shorter durations than gestures before turn-medial prosodic boundaries. This phenomenon is explained by the fact that a turn-final position is associated with a lighter planning load than turn-medial positions.

Heavy speech planning loads in phrase-medial position may also obscure phrase-final turn-medial lengthening relative to these tokens. Properly accounting for the role of performance factors like planning is crucial as we attempt to understand kinematic properties and the representation of prosodic boundaries.

6. ACKNOWLEDGMENTS

Thank you to Wyatt Barnes, Ana-Patricia Collazo, Benjamin Hermen Paige Marshner, Lucas White and Brennan Yaghmour for data labelling, and to attendees of ASA183, and the members of the University of Michigan Phondi and Prosody groups for their feedback. The study is supported by the Associate Professor Fund of the University of Michigan.

7. REFERENCES

[1] Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. JPhon, 26(2), 173-199.
[2] Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. JPhon, 29 (2), 155–190.
[3] Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. JASA, 101(6), 3728-3740.
[4] Krivokapic, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. JPhon, 40 (3), 430–442.
[5] Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. JPhon, 31(2), 149-180.
[6] Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects
[7] Oller, K. D. (1973). The effect of position in utterance on speech segment duration in English. JASA 54, 1235—1247.
[8] Berkovits, R. (1993). Progressive utterance-final lengthening in syllables with final fricatives. Lang. Speech 26, 89—98.
[9] Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. Journal of Phonetics, 44, 62-82.
[10] Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. Journal of phonetics, 55, 149-181.
[11] Ferreira, F. (1991). "Effects of length and syntactic complexity on initiation times for prepared utterances," J. Mem. Lang. 30, 210–233.
[12] Ferreira, F., and Swets, B. (2002). "How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums," J. Mem. Lang. 46, 57–84.
[13] Krivokapic´, J. (2012). "Prosodic planning in speech production," Speech planning dynamics in Speech planning and dynamics (Peter Lang, Frankfurt, Berlin, Bern, Bruxelles, New York, Oxford, Vienna), pp. 157–190.
[14] Watson, D., and Gibson, E. (2004). "The relationship between intonational phrasing and syntactic structure in language production," Lang. Cogn. Proc. 19, 713–755.
[15] Duncan, S., Jr. (1972). Some Signals and Rules for Taking Speaking Turns in Conversations. J. Personal. Soc. Psychol. 23, 283–292.
[16] Cowley, S. J. (1998). Of Timing, Turn-Taking, and Conversations. J. Psycholing. Research 27, 541–571.
[17] Rühlemann, C., & Gries, S. T. (2020). Speakers advance-project turn completion by slowing down: A multifactorial corpus analysis. Journal of Phonetics, 80,100976.
[18] Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with" articulatory" pauses. Language and Speech, 26(3), 203-214.
[19] Local, J., & Walker, G. (2012). How phonetic features project more talk. JIPA, 42 (3), 255–280.
[20] Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. J. Mem. and Lang., 60(1), 92-111.
[21] Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. JASA, 113(2), 1001-1024.