Conceptual Planning in Conversational Mandarin: Pitch Variation in Prosodic Phrasing

Alvin C.-H. Chen [1], *Shu-Chuan Tseng* [2]

[1] Taiwan International Graduate Program (TIGP), Academia Sinica, Taiwan
[2] Institute of Linguistics, Academia Sinica, Taiwan
alvinworks@gmail.com, tsengsc@gate.sinica.edu.tw

Abstract

The present study addresses the question of whether the grammatical configuration of prosodic units (i.e., their alignment with the clause unit) may contribute to systematic prosodic patterns. Specifically, we focus on the pitch variation on prosodic boundaries in conversational Mandarin. Results demonstrate that the grammatical configuration of the prosodic phrasing correlates with systematic pitch variation, on which implications for incremental production are drawn. The cross-boundary pitch variation signals not only whether speakers are going to finish their proposition by the end of the prosodic phrasing, but also how much information they have planned to package in the prosodic unit. A clause-based conceptual planning in conversational speech is empirically supported by our observation of the pitch variation on prosodic boundaries. Prosodic phrasing is found to emerge from the stream of conversational speech with a high degree of satisfying consistency reflecting our clause-based conceptualization in interaction.

Index Terms: pitch variation, prosodic phrasing, incremental production, conceptual planning

1. Introduction

Conversation is a primary means for our social interaction. One of its important characteristics is to exchange propositional contents and subjective perspectives, and during this transaction the interlocutors may process, regulate, or work out their intended message to make it proceed smoothly and efficiently [1]. It is generally acknowledged that discourse is structured [2-4]. The definition for the building blocks in conversational speech differs according to the discourse model at stake. Crucially, the interaction between prosodic phrasing and grammatical structure has given rise to various proposals to account for the mapping of their boundaries.

In conversational speech, the articulation results in a basic unit at the prosodic level — i.e., a prosodic unit. This has been referred to variously as *tone unit* [5], *intonation group* [6], *intonation phrase* [7], *intonational phrase* [8, 9], *intermediate phrase* [10], and *intonation unit* [11]. Specifically, cross-linguistic studies on intonation units have observed that an intonation unit often correlates with a clause unit in typologically unrelated languages [12], but this cross-linguistically established correlation is still not a perfect congruence of one-to-one correspondence [13]. The boundaries of grammatical, pragmatic, and prosodic units may not necessarily coincide. In face of this somewhat mixed character of the correlation between prosodic phrasing and the clause unit, our research aims to seek a more empirical account of the relation between prosody and grammar from a computational-acoustic perspective. More specifically, we address the question of whether the grammatical configuration of prosodic units (i.e., their alignment with the clause unit) may contribute to systematic pitch variation.

2. Prosodic phrasing

2.1. Prosodic unit

We adopted the annotation of the prosodic unit (hereafter PU) in the Mandarin Conversational Dialogue Corpus (MCDC) [14] as our starting point, totaling 3.5 hours with 16 different speakers. A PU is defined as a perceptually coherent prosodic constituent featuring possible pitch reset, final lengthening, occurrences of paralinguistic sounds, and/or alteration of speech rate [14]. Based on proper auditory cues, the boundaries of prosodic phrasing were annotated in each conversational turn. The operational criteria were essentially the same as those used in the intonation-unit framework [11, 15]. A satisfactory inter-transcriber agreement for PU annotation has been achieved [14] and a computational acoustic-based modeling for automatic PU boundary detection has also yielded promising results [16], thus rendering more psychological reality to the PU annotation.

2.2. Clause unit

A clause unit (hereafter CU) is often recognized as encoding a single proposition, which in turn is taken by linguists as a basic grammatical unit of information [17]. From a discourse-functional linguistic perspective, a workable definition for a CU is utterances with predicate and the center participants coming around it [18]. Based on proper operational criteria, the boundaries of CUs were annotated in each conversational turn. Issues on the annotation of the CU have been discussed in more detail in [19].

2.3. A computational-acoustic representation

In order to examine how prosodic phrasing works in conversational discourse, a transcription system needs to cope with two preprocessing tasks: unit annotation and prosodic transcription [19]. The purpose of the former procedure is to define the levels of the PU in the transcription convention and manually identify the boundaries of the PUs. The latter procedure concerns the way how a transcription system characterizes the structure of the PU. Our one-tier PU annotation [14] bears great resemblance to the intonation-unit framework [11] in that only one level of PU is annotated in the database. However, our PU annotation differs significantly from the more phonology-based conventions, where the prosodic hierarchy is manually labeled with an assumed set of prosodic tiers [10, 20]. Furthermore, instead of decomposing the global pitch contour into sequences of *a priori* contrastive pitch accents or boundary tones [20], we adopt a computational-acoustic approach to characterizing the manually labeled PUs with a comprehensive set of quantitative acoustic-prosodic measures [19].

We are not concerned with how grammatical structures are *aligned* with prosodic phrasing. Rather we are more interested in the *gradient variation* of the PU resulting from its grammatical configuration. Our rationale is that if the PU alignment with the CU boundary significantly leads to different prosodic patterns, then the correlation between prosodic phrasing and the basic grammatical unit holds. Instead of being counter-evidence, these mismatches may better be analyzed as compelling evidence that the CU indeed has its hand on the structure of the PU in a *gradient* way. In this paper, we will present our empirical results on the pitch variation contributed by the configuration of the PU-CU alignment.

2.4. Feature definition

We relied on `Praat`'s autocorrelation-based pitch tracking algorithm to extract raw pitch values of each PU, using gender-dependent pitch range (75-300 Hz for male, and 100-500 Hz for female). The raw pitch values were converted into semitones (w.r.t. 1Hz) for the computation of derived features. We computed 3 indexes to characterize the pitch variation of the PU: initial pitch reset, final pitch reset and pitch move. These indexes were based on reference pitch points derived from two types of stylization.

Since one of our major concerns was the global tendency of F0 declination for each PU, we first took the whole PU as our *global window* for stylization (**global stylization**). On the other hand, in order to capture the cross-boundary local variation of the fundamental frequency, we defined another *local window* for stylization, i.e., PU-final word (**local stylization**). We stylized the pitch values in the last word of the PU to model the cross-boundary local F0 variation.

For global stylization, we approximated the pitch values in the current PU with a first-order linear regression. For local stylization, we divided the pitch values of the target word in the PU-final position into halves and for each half we approached the pitch values with one linear regression line. Our local stylization was based on the acoustic modeling of the Japanese phrase-final pitch accent in [21], where a good correlation between the perceptual scores and the acoustic measures yielded by the bipartitioned linear fit was supported by perceptual experiments.

For each PU, four reference points were derived, two from the global stylization, one from the local stylization, and one based on the intensity:

- **F0_Pred_Initial_Global:** Predicted F0 value at the beginning of the PU from the global stylization
- **F0_Pred_Final_Global**: Predicted F0 value at the final of the PU from the global stylization
- **F0_Pred_Final_Local**: Predicted F0 value at the end of the PU from the second-half linearization of the local stylization
- **F0_dB_Initial**: Raw F0 value of the maximal dB value in the first word of the PU

For the current *i-th* PU, three acoustic indexes for its pitch variation were computed as follows:

- **PITCH MOVE** = $F0\_Pred\_Initial\_Global_i$ – $F0\_Pred\_Final\_Local_i$
- **INITIAL PITCH RESET** = $F0\_dB\_Initial_i$ – $F0\_Pred\_Final\_Global_{i-1}$
- **FINAL PITCH RESET** = $F0\_dB\_Initial_{i+1}$ – $F0\_Pred\_Final\_Global_i$

3. Method

3.1. Research questions

The present study addressed the question of how the left-edge, right-edge, and PU-internal CU boundaries may contribute to the pitch variation of the PUs in conversational Mandarin.

3.2. Grouping factors

Based on the configuration of the PU-CU alignment, we classified PUs according to their alignment with CUs and the number of PU-internal CU boundaries. Three grouping factors for PUs were highlighted:

- **LEFT**: their left-alignment with CUs (*Y* for being left-aligned and *N* for not being left-aligned with CU left-ledge boundaries)
- **RIGHT**: their right-alignment with CUs (*Y* for being right-aligned and *N* for not being right-aligned with CU right-edge boundaries)
- **INTCU**: whether they integrate more than one CU boundary (For a PU that integrates more than one CU, we define it as a *complex* PU; for a PU that is a sub-clausal prosodic fragment, not integrating additional CU boundaries, we define it as a *simple* PU.)

3.3. Statistical evaluation

We conducted three linear mixed effect analyses with the three pitch-related features as our dependent variables and the 3 grouping factors as our predictors (i.e., fixed effects). We included a random subject intercept as the random effect in the linear models. To test the significance of the fixed effects, a bootstrap approach was adopted to find more accurate *p*-values for the likelihood ratio test, using Markov chain Monte Carlo sampling. The statistical analysis was done in `R`, using `languageR` package to compute the *p*-values of the fixed effects – RIGHT, LEFT, INTCU, and all pairwise interactions (α-level = 0.01).

4. Results

4.1. Descriptive statistics

Alignment with PU	With internal PU	No internal PU
Both	1296 (28.53%)	1505 (33.13%)
Left	322 (7.09%)	427 (9.40%)
None	94 (2.07%)	104 (2.29%)
Right	295 (6.49%)	500 (11.01%)
Total	2007 (44.18%)	2536 (55.82%)

Table 4-1: Classification table of the CUs according to their alignment with PUs and the number of internal PU boundary

The frequency distribution suggests that 61.66% (the BOTH row) of the CUs are coextensive with PUs and about 80% of the them (the BOTH row and the RIGHT row) are provided with a prosodic phrasing at the right-edge CU boundary. Over 95% of the CUs are aligned with PUs in at least one clausal boundary — either at the clause-initial or clause-final boundaries. The general tendency is that speakers predominantly signal the beginning and ending of their intended CUs via prosodic phrasing in spontaneous speech production. In other words, speakers tend to package semantically coherent propositions in terms of PU boundaries.

4.2. Experimental results

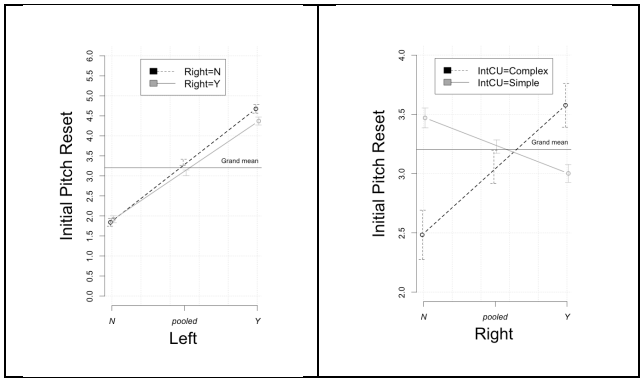

Figure 1: *Interactions plots for Initial Pitch Reset. The left panel shows the Left×Right interaction; the right panel shows the Right×IntCU interaction.*

For INITIAL PITCH RESET, we observed one significant main effect — LEFT (β = 2.94, p < 0.01), and two significant interactions — LEFT × RIGHT (β = -0.57, p < 0.01) and RIGHT × INTCU (β = 1.09, p < 0.01). The LEFT main effect suggests that if the PU starts at the onset of the CU, its initial pitch reset from the previous PU will be significantly larger. Figure 1 summarized the two interactions on INITIAL PITCH RESET. As illustrated by the left-panel plot in Figure 1, the LEFT × RIGHT interaction suggests that given a left-aligned PU (LEFT = Y), if it is also right-aligned (RIGHT = Y), its initial pitch reset is lower than if it is non-right-aligned (RIGHT = N). A schematic illustration of the LEFT × RIGHT is given in Figure 2, where the two contours represent two hypothetical intonation contours of PUs. The upper-panel contour is an example of a left-aligned PU which is not right-aligned; the lower-panel contour is an example of a left-aligned PU which is also right-aligned. The vertical dashed lines represent the hypothetical CU boundaries. When starting a CU with a PU (a left-aligned PU), speakers seem to adjust their degree of initial pitch reset according to their intention whether to finish the current CU by the end of this PU. If at the beginning of a CU they have planned to finish the proposition by the end of the current PU, they would demonstrate a smaller degree of initial pitch reset. However, if they have not been fully prepared to finish the CU within this PU, they would show a higher degree of initial pitch reset.

As illustrated by the right-panel plot in Figure 1, the RIGHT × INTCU interaction suggests that given a PU integrating more than one CU (INTCU = Complex), if it is also right-aligned (RIGHT = Y), its initial pitch reset is larger than if it is non-right-aligned (RIGHT = N). A schematic illustration of the RIGHT × INTCU is given in Figure 3. The upper-panel contour is an example of a complex PU which is right-aligned with the CU boundary; the lower-panel contour is an example of a simple PU which is right-aligned with the CU boundary. The vertical dashed lines represent the hypothetical CU boundaries. If speakers integrate more than one CU in their current PU and have planned to finish this clause-complex sequence by the end of the current PU, they will show a significantly larger degree of initial pitch reset. In contrast, if speakers have reached to the mid part of a CU and start the current PU to finish the remaining parts of the CU by the end of the PU, they will show a significantly smaller degree of initial pitch reset. Therefore, the degree of the PU-initial pitch reset signals not only whether speakers are going to finish the CU by the end of the prosodic phrasing, but also how much information (measured in number of CUs) they have planned to package in the PU.

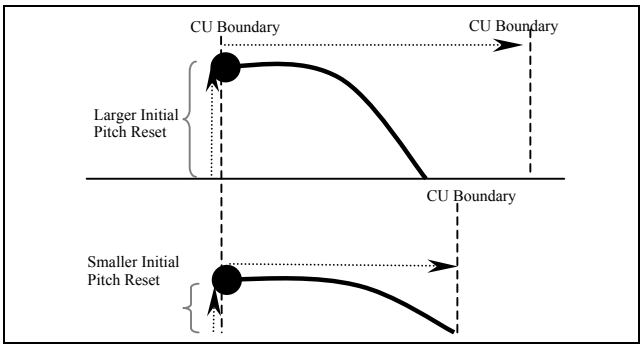

Figure 2: *Schematic diagram of Left × Right.*

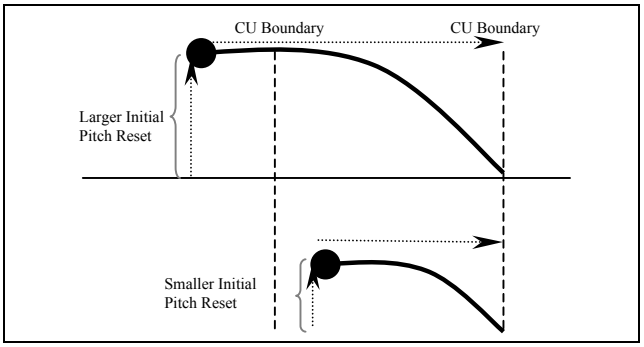

Figure 3: *Schematic diagram of Right × IntCU*

For FINAL PITCH RESET, we observed one significant main effect — RIGHT (β = 2.57, p < 0.01). For PITCH MOVE, we observed one significant main effect — RIGHT (β = 1.39, p < 0.01). Both of these measures suggest that the grammatical configuration of the PU-CU alignment on the right-edge boundaries may contribute to a significantly larger degree of final pitch reset as well as F0 declination within the prosodic phrasing.

5. Discussion

Results demonstrated that an underlying clause schema is embedded in these pitch-related patterns of the PUs in conversational speech. The PU left-alignment with a CU is anticipated in the degree of INITIAL PITCH RESET and the PU right-alignment with a CU correlates with its FINAL PITCH RESET and PITCH MOVE. The interaction effects on the PU-initial pitch reset may be germane in a larger context of incremental speech production [22, 23]. Incremental production puts forward the idea that speakers may perform the conceptual planning during the articulation process at the same time. When a speaker is formulating the morpho-phonological encoding and articulating, s/he is capable of conceptually planning the upcoming words at the same time. We suggest that the "lookahead" conceptual planning is somewhat anticipated in the acoustic-prosodic measures of prosodic phrasing. Following [23], we wish to go one step further and claim that our incremental production may proceed on a *clausal* basis.

When a PU starts at the onset of a CU, the right-alignment of the PU with the CU is prosodically anticipated in the initial pitch reset. A larger initial pitch reset may project a more distant right-edge CU boundary, namely, not within the

current PU domain (i.e., the upper-panel plot in Figure 2). A smaller initial pitch reset may project a more imminent right-edge CU boundary, namely, at the end of the current PU (i.e., the lower-panel plot in Figure 2). When speakers reset their pitch contour for a new prosodic phrasing, they have embedded anticipatory prosodic cues for how far they intended to go at the onset of the prosodic contour.

The interaction of RIGHT × INTCU on the initial pitch reset reveals another iconic relationship between the initial pitch reset and the semantic complexity carried by the PU. A complex PU is defined as integrating more than one PU-internal CU boundary, thus rendering it more complex than a simple PU in terms of the syntactic configurations and propositional loadings. A right-aligned complex PU refers to a PU that integrates a clause complex (either starting at the onset of the clause complex or not) and ends at the end of the clause complex. Its semantic complexity is found to be reflected in its larger degree of initial pitch reset, as illustrated by the upper-panel plot in Figure 3. A right-aligned simple PU refers to a PU that integrates at most one complete CU, or a part of the CU, and ends at the end of the CU. Its semantic complexity also correlates with its smaller degree of initial pitch reset, as illustrated by the lower-panel plot in Figure 3.

It seems that at the onset of the PU, a roughly-sketched propositional format for the CU has been "active" in the mind of the speaker, and the PU-initial pitch reset suggests that the propositional contents of the CU to be verbalized in the PU indeed are active for the speaker at this onset point. Based on our computational-acoustic analysis, we contend that speakers should have generated a primitive propositional outset for the intended message in their conceptual preparation at the *onset* of the PUs. This would explain why they produce a PU whose initial pitch reset is significantly indicative of not only the upcoming propositional terminal (right-edge CU boundaries at the end of the PU), but also an upcoming sequence of CUs (internal CU boundaries within the PU domain). While we agree on the tenets of incremental production in its broad sense that speakers are capable of planning upcoming portions of an utterance as they are articulating, we believe that such an incremental production may proceed on a clausal (propositional) basis. PUs emerge from the stream of speech with a high degree of satisfying consistency reflecting our CU-based conceptualization in conversational Mandarin. The correlation between the pitch variation and CU boundaries suggests that a proposition-based CU should play a functional role in the prosodic phrasing of conversational Mandarin.

6. Conclusions

If we take prosodic phrasing as a linguistic window into the process of the speakers' conceptual preparation before grammatical encoding, the systematic prosodic structures in PUs contributed by CU boundaries indeed indicate that the conceptual planning seems to proceed on a *clausal* basis. The grammatical configuration of the PUs correlates with systematic pitch variation in speech production. Speakers seem to provide anticipatory cues in their prosodic phrasing for the onset and the finality of their intended propositions. Furthermore, our acoustic characterization of the cross-boundary pitch variation highlights the crucial role of pitch reset in creating more "conversational space" [24].

7. Acknowledgements

This study was funded by the Fellowships for Doctoral Candidates in the Humanities and Social Sciences, Academia Sinica, Taiwan grant to the first author, and the National Science Council, Taiwan grant (NSC 100-2410-H-001-093) to the second author. The authors would like to thank Yi-Fen Liu for helpful comments and technical support.

8. References

[1] D. C. O'Connell, and S. Kowal, *Communicating with one another: Toward a psychology of spontaneous spoken discourse*, Berlin: Springer Verlag, 2008.

[2] J. Miller, and R. Weinert, *Spontaneous spoken language: Syntax and discourse*, Oxford: Oxford University Press, 1998.

[3] W. C. Mann, and S. A. Thompson, "Rhetorical structure theory: A theory of text organization," *Text,* vol. 8, no. 3, pp. 243-281, 1988.

[4] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language,* vol. 50, no. 4, pp. 696-735, 1974.

[5] D. Crystal, *Prosodic systems and intonation in English*, Cambridge: Cambridge University Press, 1969.

[6] A. Cruttenden, *Intonation*, 2nd edn ed., Cambridge: Cambridge University Press, 1997.

[7] J. Pierrehumbert, "The phonology and phonetics of English intonation," dissertation, MIT, Cambridge, MA, 1980.

[8] M. Nespor, and I. Vogel, *Prosodic phonology*, Berlin: Walter de Gruyter, 1986.

[9] E. Selkirk, *Phonology and syntax: The relation between sound and structure*, Cambridge: MIT Press, 1984.

[10] K. Silverman, M. E. Beckman, J. Pitrelli *et al.*, "ToBI: A standard for labeling English prosody," in ICSLP-92, Vol. 2, 1992, pp. 867-870.

[11] W. Chafe, *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*, Chicago: University of Chicago Press, 1994.

[12] J. S.-Y. Park, "Cognitive and interactional motivations for the intonation unit," *Studies in Language,* vol. 26, no. 3, pp. 637-680, 2002.

[13] J. Cole, Y. Mo, and S. Baek, "The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech," *Language and Cognitive Processes,* vol. 25, no. 7, pp. 1141-1177, 2010.

[14] Y.-F. Liu, and S.-C. Tseng, "Linguistic patterns detected through a prosodic segmentation in spontaneous Taiwan Mandarin speech," *Linguistic patterns in spontaneous speech*, S.-C. Tseng, ed., pp. 147-166, Taipei: Institute of Linguistics, Academia Sinica, 2009.

[15] H. Tao, *Units in Mandarin conversation: Prosody, discourse, and grammar*, Amsterdam: John Benjamins, 1996.

[16] Y.-F. Liu, S.-C. Tseng, J.-S. R. Jang *et al.*, "Coping imbalanced prosodic unit boundary detection with linguistically-motivated prosodic features," in INTERSPEECH 2010, Makuhari, Japan, 2010, pp. 1417-1420.

[17] T. Givón, *Syntax: A functional and typological introduction*, Amsterdam: John Benjamins, 1984.

[18] S. A. Thompson, and E. Couper-Kuhlen, "The clause as a locus of grammar and interaction," *Discourse Studies,* vol. 7, no. 4-5, pp. 481-506, 2005.

[19] A. C.-H. Chen, "Prosodic phrasing in Mandarin spontaneous speech: A computational-acoustic perspective," dissertation, National Taiwan University, Taipei, 2011.

[20] M. E. Beckman, J. Hirschberg, and S. Shattuck-Hufnagel, "The original ToBI system and the evolution of the ToBI framework," *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun, ed., pp. 9-54, Oxford: Oxford University Press, 2006.

[21] C. T. Ishi, P. Mokhtari, and N. Campbell, "Perceptually-related acoustic-prosodic features of phrase finals in spontaneous speech," in Proceeding of Eurospeech 2003, 2003, pp. 405-408.

[22] F. Ferreira, and B. Swets, "How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums," *Journal of Memory and Language,* vol. 46, no. 1, pp. 57-84, 2002.

[23] W. J. M. Levelt, "Producing spoken language: A blueprint of the speaker," *The neurocognition of language*, C. M. Brown and P. Hagoort, eds., pp. 83-122, Oxford: Oxford University Press, 1999.

[24] D. Schiffrin, *Discourse markers*, Cambridge: Cambridge University Press, 1987.