Journal of Phonetics 76 (2019) 100912

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Prosodic encoding in Mandarin spontaneous speech: Evidence for clause-based advanced planning in language production

Alvin Cheng-Hsien Chen [a,*], Shu-Chuan Tseng [b]

[a] National Taiwan Normal University, 162, Section 1, Heping E. Rd., Taipei 106, Taiwan
[b] Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan

ARTICLE INFO

Article history:
Received 25 May 2018
Received in revised form 16 May 2019
Accepted 21 July 2019

Keywords:
Prosodic encoding
f0 shifting
Preplanning
Spontaneous speech
Clause
Syntax-phonology interface
Proposition

ABSTRACT

This study reports the cross-boundary f0 shifting of prosodic units (PU) in Mandarin conversational speech by analyzing the PU-initial and PU-final f0 heights as a function of its semantic structure. Initial and final f0 heights were defined as the f0 values extracted at the energy max of the first and the last syllable of the PU. The semantic structure of the PU was defined based on its co-extensiveness with a semantic unit in discourse (DU), i.e., a proposition, often encoded by a clause. Our analysis shows significant relationships between the cross-boundary f0 heights and the PU-DU co-extensiveness. PU-DU left alignment introduces a significant up-shifting effect on both initial and final f0 heights. This pitch resetting is effective across the whole PU, suggesting speakers' sensitivity to the initiation of propositions in production. On the other hand, PU-DU right alignment introduces a down-shifting effect on both f0 heights. The *regressive* f0-lowering observed in the PU-initial f0 heights (anticipatory effect based on the PU-terminal semantics) and the *progressive* f0-raising of the PU-final f0 heights (carried-over effect based on the PU-initial semantics) both shed light on the psycholinguistic importance of the proposition and support its central role in advanced planning in spontaneous speech production.

1. Introduction

Preplanning in speech production has been a central topic in linguistic and psycholinguistic research. Linguists are concerned with the concurrence of message planning and speech articulation (Ferreira & Swets, 2002; Keating & Shattuck-Hufnagel, 2002; Levelt, 1989; Wagner, Jescheniak, & Schriefers, 2010). It is posited that language production is incremental: planning of the upcoming messages and articulation of the current message may proceed simultaneously. These studies are structured around two central questions: (1) how far ahead do speakers plan before the onset of the articulation? (2) what is the base unit in this advanced planning (e.g., clauses or sub-clausal phrases)?

It is evident that speech is "not produced in a continuous, uninterrupted flow but in spurts" (Chafe, 1994, p. 57). These segments in speech production are assumed to "verbalize the information active in speaker's mind" at the onset of articulating the segment (Chafe, 1994, p. 63). Analyses of the relationship between the *prosodic encodings* and the *semantic contents* of these speech segments may provide a clue to the question of how much information has been active in the speaker's mind before articulation. This in turn will give us acoustic evidence for advanced planning in speech production.

For example, Lee, Brown-Schmidt, and Watson (2013) observed, when analyzing the duration of the lexical head noun of a noun phrase used for an entity description task, that lexical head nouns with more complex long-distance dependents in the upcoming phrase (e.g., relative clauses) tended to be longer, and their speech onset time significantly increased as well. This may suggest that during the articulation of the lexical head, its syntactic dependents, be it in a long- or short-distance dependency, may have been planned concurrently. Similarly, in analyzing the read speech of native German speakers, Fuchs, Petrone, Krivokapić, and Hoole (2013) also found that pause and inhalation duration was significantly longer when speakers knew they were about to read a longer sentence. Similar acoustic patterns have been reported in inhalation depths and breathing (Fuchs et al., 2013; Whalen & Kinsella-Shaw, 1997). Studies have also shown that the

* Corresponding author.
E-mail address: alvinchen@ntnu.edu.tw (A. C.-H. Chen).

https://doi.org/10.1016/j.wocn.2019.100912
0095-4470/

use of filled pauses before articulation serves as an efficient cue for a more complex upcoming message, which has been effectively utilized by both native and non-native listeners (Watanabe, Hirose, Den, & Minematsu, 2008). The length and types of the filled pauses may also be predictive of the strength of the discourse boundaries (Swerts, 1997) and complexity of the upcoming message (Clark & Tree, 2002).

Pitch-related preplanning has been studied extensively in controlled read speech by analyzing the pitch encodings of the speech utterances in relation to their grammatical complexity (Fuchs et al., 2013; Ladd & Johnson, 1987; Lee et al., 2013; Liberman & Pierrehumbert, 1984; Prieto, D'Imperio, Elordieta, Frota, & Vigário, 2006; Scholz & Chen, 2014; Shih, 2000; Wang & Xu, 2011). These experimental settings often utilize prompt utterances which are syntactically and semantically complete utterances (i.e., grammatical sentences), but vary in terms of grammatical complexity. Evidence for pitch-related preplanning relies on the determination of a significant relationship between the f0 shifting and the semantic/syntactic complexity of the utterances. For example, in manipulating the lengths of the sentences in a read-speech setting, research has found that the initial f0 height of the sentence increases as the sentence grows longer (Liberman & Pierrehumbert, 1984; Prieto et al., 2006; Shih, 2000; Thorson, 2007; Wang & Xu, 2011). More conservative local pitch-related preplanning has also been argued in Fuchs et al. (2013), where the initial f0 height varies only with the length of the first constituent in the sentence. In addition, longer sentences also tend to have less steep f0 slopes (Yuan & Liberman, 2014), and lower final f0 values (Shih, 2000), which could be taken as preliminary effects of phrasal preplanning. When speakers initiate the articulation of the sentences, their prosodic encoding may have been prearranged in anticipation of the complexity of their upcoming sentences (e.g., sentences with more words), thus providing acoustic support for pitch-related preplanning.

Previous research on pitch-related preplanning has mostly focused on read speech (cf. Asu, Lippus, Salveste, & Sahkai, 2016), and for Mandarin, only limited support has been reported in scripted speech (e.g., broadcast news speech in Yuan & Liberman, 2014). In these experimental settings, speakers are often given the whole utterance either in print or on the screen before their reading. Moreover, these utterances used for the reading experiments are often semantically-coherent full-fledged sentences. With an overview of the whole sentence to be articulated, speakers know the structural and semantic contents they are about to produce. These semantic contents may have been already active in the mind of the speaker before articulation. The pitch-related preplanning observed in read speech is largely based on the condition that speakers have learned the complete proposition (semantic content) they are expected to articulate. As the main objective in research on advanced planning is to find out how much information has been active in the speaker's mind *before* articulation, the prosodic preplanning found in these reading experiments may thus be rather different from the preplanning observed in on-line spontaneous speech production. While these studies have reported a strong connection between the pitch change (e.g., utterance-initial/final f0 heights) and the grammatical complexity of the utterances (e.g., sentences with more words, or more complex subjects), this relationship may be indirect or ambiguous for advanced planning. It is less clear whether this pitch-related preplanning is still evident and effective in a spontaneous context, where utterances are produced incrementally without an overview of the whole proposition before articulation.

More importantly, unlike read speech, research on prosody-grammar interface of conversational speech has found that our production units in conversational speech do not necessary match a semantically-coherent unit (cf. Croft, 1995; Ladd, 2008; Matsumoto, 2000; Park, 2002; Tao, 1996). Works on spontaneous speech have defined a functional speech segment in production, identifiable via various perceptual cues, such as pausing in timing, and/or changes in pitch, duration, rhythm, intensity, or voice quality. Most notably, Chafe (1994) refers to this perceptually-prominent speech segment as an *intonation unit* (IU). IU-based studies have shown that IUs in conversational discourse do not always correspond to the basic semantic unit in discourse, i.e., a proposition. A proposition refers to a single event or state, which is often structurally encoded by a clause. The mismatch between IU and the clause may lead to a question of whether the pitch-related preplanning in speech production proceeds on a propositional basis. It remains unclear whether the prosodic encodings of the IU are connected to the semantic contents articulated in the IU (e.g., whether the IU articulates a complete proposition, or a semantic fragment of the proposition). Relatedly, it would be interesting to examine to what extent the whole proposition may be active in the speaker's mind before articulating the IU.

This question is theoretically relevant in at least two important senses. First, the association between the prosodic encodings of the IU and the semantic completeness of the IU may be taken as acoustic evidence for the entrenchment of the clause schema in speaker's mental representation. While cross-linguistic studies have suggested that around 50–60% of the IUs in conversation each correspond to a clause (cf. Croft, 1995; Matsumoto, 2000; Park, 2002; Tao, 1996), they have also found that the online immediacy of turn-taking may add additional pressure to speakers' message planning and often result in a piecemeal presentation of a proposition, i.e., a series of IUs consisting of sub-clausal fragments (e.g., noun phrases, prepositional phrases). Many factors may be involved in on-line production and result in this discrepancy between prosody and syntax, including syntactic constituency (Selkirk, 1984), information structure (Chafe, 1994), interactional considerations (Ono & Thompson, 1995; Park, 2002), or simply performance arrangement (Ferreira, 2007). Alternatively, on-line speech production can be spontaneous to the extent that speakers may also integrate more than one proposition within an IU. These mismatches between IU and the clause often raise doubts to the functional relevance of the clause schema in spontaneous speech. However, if there is a strong connection between the prosodic encodings of the IUs and the co-extensiveness of the IU and the clause, these mismatches can be re-analyzed as supporting evidence for the importance of the clause schema in speech production, i.e., speaker's sensitivity to the propositional completeness of the IUs. Secondly, the relationship between the prosodic encodings and the semantic completeness of the IUs may shed light on the base unit in advanced planning. Research on IUs builds upon the hypothesis that the idea asserted in

one IU is already active in the "speaker's focus of consciousness" at its onset of articulation (Chafe, 1994, p. 63). A strong correlation between the prosodic encoding and the semantic coherence of the IU (e.g., whether the IU is a full-fledged proposition or not) may indicate the speaker's prosodic prearrangement of the IU on the basis of the upcoming semantic contents to be asserted in the IU.

In this study, we refer to the semantically-defined units as "discourse units (DU)", and the perceptually-defined units as "prosodic units (PU)". We decide to use PUs instead of Chafe's IUs for its more generic sense. Detailed operational definitions for DU and PU will be given in Section 3. Focusing on spontaneous speech of Taiwan Mandarin, this study investigates three main questions:

- RQ1: Do speakers show sensitivity to the semantic unit, DU, in speech production?
- RQ2: Do speakers show signs of preplanning a whole DU in speech production?
- RQ3: Do speakers show signs of preplanning a unit beyond a single DU?

In this study, the semantic contents (or semantic completeness) of the PU will be analyzed as the degrees to which the PU is a full-fledged proposition, i.e., the co-extensiveness of PU and DU: the PU-DU left- and right-alignments. In RQ1, we examine whether speakers show their sensitivity to a proposition (e.g., event or state), which has been considered a basic semantic unit in our daily interactions (Thompson & Couper-Kuhlen, 2005). As shown in Fig. 1, a DU may be presented with one canonical PU (i.e., $PU_{(D)}$ in Fig. 1), or in a piecemeal fashion with DU-initial (i.e., $PU_{(A)}$), DU-medial (i.e., $PU_{(B)}$) and/or DU-final PUs (i.e., $PU_{(C)}$). We examine whether the initial and final f0 heights of PUs are connected to the semantic completeness of the PU. If speakers are sensitive to the DUs, we will expect the *initial* prosodic encoding of the PU to correlate with the PU-DU *left* alignment (cf. Fig. 1a). We will expect the initial f0 height of PUs that initiate a proposition to differ significantly from that of those PUs starting in the middle of the proposition. We will also expect the *final* prosodic encoding of the PU to correlate with the PU-DU *right* alignment (cf. Fig. 1b). Finding the relationship between PU cross-boundary pitch variation and a discourse juncture is compatible with previous research on prosody-syntax interface as they suggest that speakers use different prosodic cues to demarcate PU boundaries within versus across semantic junctures.

In RQ2, we explore the role of the semantic unit, the proposition, in the look-ahead planning in speech production. While previous studies on read-speech have shown support of pitch-related preplanning for upcoming sentences that are longer, or structurally more complex, more fundamental to our understanding of language production is the connection between prosodic encoding and the semantic completeness of the PU in a spontaneous context. In this study, we ask whether speakers show signs of preplanning the whole proposition in conversation. If the preplanning in speech production proceeds on a propositional basis, we will expect the *initial* f0 height of PUs to correlate with the PU-DU *right* alignment (cf. Fig. 2a) and the *final* f0 height to correlate with the *left* alignment (cf. Fig. 2b). The former suggests that speakers provide anticipatory prosodic cues for the propositional completeness of the upcoming utterance at the onset of the articulation. The latter suggests that speakers are aware of the fresh-start of a proposition leaving a carried-over prosodic cue at the completion of the utterance.

In RQ3, this study will further explore the possibility of advanced-planning beyond a single DU. We will investigate whether speakers show signs of preplanning semantic contents beyond one proposition. In addition to the semantic completeness of PUs, PUs can be further divided into two types in terms of their semantic complexity. In this study, PUs integrating more than one proposition (DU) (e.g., a PU spanning at least two propositions) will be referred to as *complex* PUs, as schematically represented in Fig. 3. PUs examined in RQ1 and RQ2 are referred to as *simple* PUs (More comprehensive operational definitions will be provided in Section 3). If speakers are able to preplan an idea beyond a single proposition, we will expect the *initial* f0 height of complex PUs to correlate with the PU-DU *right* alignment (cf. Fig. 3a) and the *final* f0 height to correlate with the PU-DU *left* alignment (cf. Fig. 3b). If no strong correlations are found in complex PUs, this may suggest that conceptual planning in speech production is more likely to proceed at a semantically-limited scale.

2. Prosodic preplanning

2.1. Utterance-initial prosodic encoding

Language production is incremental. While speakers work on their utterances in a piecemeal fashion, their message planning and articulatory implementation often operate concurrently. It remains a controversial issue as to how far ahead speakers normally plan before the onset of the articulation and, in particular, whether speakers plan the articulation of an upcoming prosodic unit with respect to its position inside any wider-span linguistic units. The former issue has often been studied by examining the relationship between the prosodic encoding of the utterance and the utterance length (Asu et al., 2016; Liberman & Pierrehumbert, 1984; Prieto et al., 2006; Shih, 2000; Thorson, 2007; Wang & Xu, 2011; Yuan & Liberman, 2014) and the latter issue is often concerned with the distinct prosodic patterning contributed by different grammatical boundaries (Fuchs et al., 2013; Ladd & Johnson, 1987; Lee et al., 2013; Scholz & Chen, 2014; Wagner et al., 2010). Pitch shifting has been argued to provide acoustic evidence for pitch-related preplanning in speech production. In general, pitch patterning has been studied from at least three perspectives: the raising of the utterance initial f0, the lowering of the utterance final f0, and the slope of the f0 declination across the utterance.

The association between utterance-initial f0 heights and the utterance length has been taken as the most compelling evidence for pitch-related preplanning. Although previous studies have operationalized the utterance lengths on different bases, such as syllables (Fuchs et al., 2013; Prieto et al., 2006; Shih, 2000; Thorson, 2007; Wang & Xu, 2011), pitch-accents (Ladd & Johnson, 1987; Prieto et al., 2006; Thorson, 2007), words (Liberman & Pierrehumbert, 1984; Scholz & Chen, 2014) or time durations (Asu et al., 2016; Thorson, 2007; Yuan &

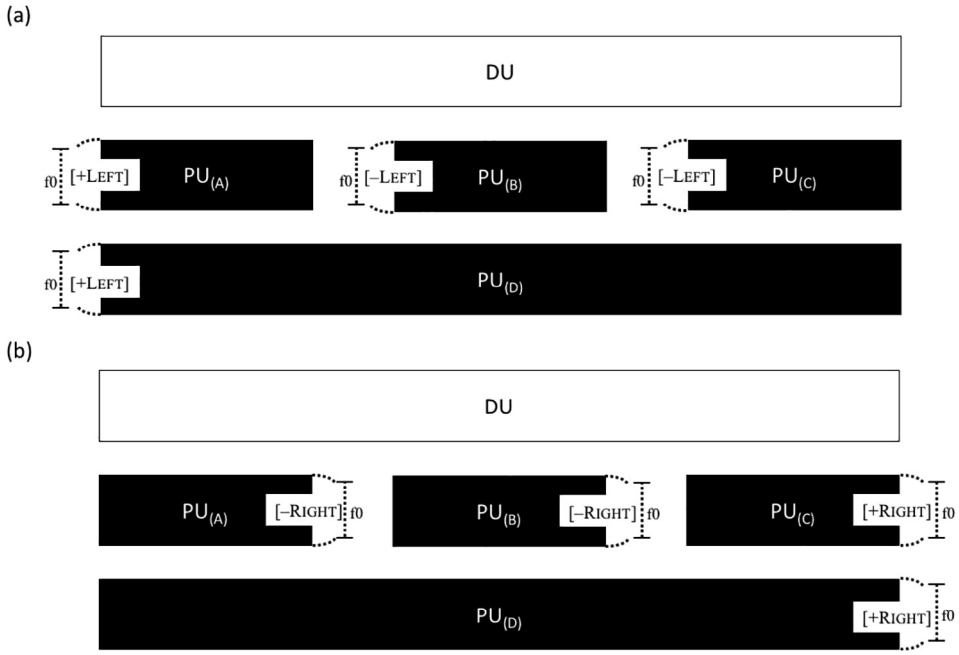


Fig. 1. A schematic representation of Research Question I. The dashed lines represent the prosodic encodings investigated in this study, i.e., PU-initial (a) and PU-final (b) f0 heights.

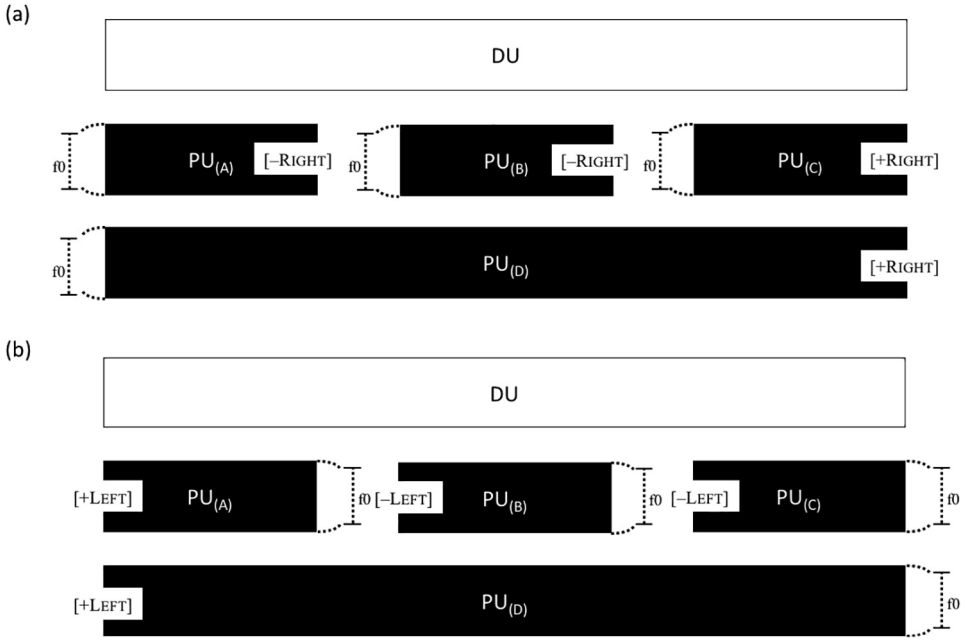


Fig. 2. A schematic representation of Research Question II. The dashed lines represent the prosodic encodings investigated in this study, i.e., PU-initial (a) and PU-final (b) f0 heights.

Liberman, 2014), a general tendency suggests that the initial f0 height at the onset of the utterance articulation tends to increase in longer utterances. Yet for a tone language like Mandarin, this f0 variation may be subject to tonal influences as the global utterance-initial/final pitch height may be conditioned on the lexical pitch excursion (Chen & Gussenhoven, 2008; Shih, 2000). In this respect, one of the pioneering studies on Mandarin pitch-related preplanning, Shih (2000), has observed an important relationship between utterance initial f0 values and utterance length (measured in number of syllables), by controlling for the lexical tone effect. Wang and Xu (2011) have reported that correlation between the utterance initial f0 and the utterance length is independent of the tonal influence of other prominence factors, such as topic and focus. Yuan and Liberman (2014) have also demonstrated a positive relationship between utterance initial f0 values and the utterance length in English and Mandarin broadcast news speech. Furthermore, the initial f0 height shifting in association to utterance length shows idiosyncratic variation even among native speakers (Prieto et al., 2006; Thorson, 2007). This

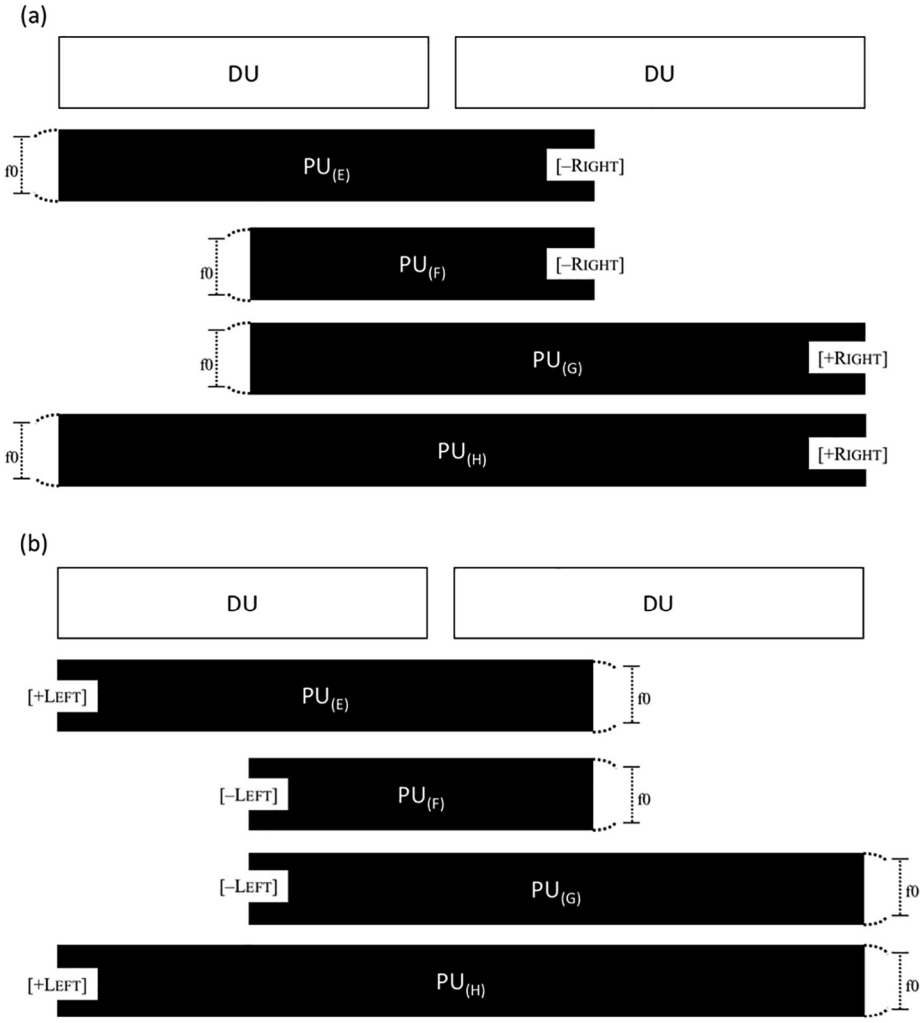


Fig. 3. A schematic representation of Research Question III with complex PUs. The dashed lines represent the prosodic encodings investigated in this study, i.e., PU-initial (a) and PU-final (b) f0 heights.

cross-speaker variation has motivated a mitigated conservative hypothesis of *soft preplanning* (Liberman & Pierrehumbert, 1984; Prieto et al., 2006), suggesting that speakers may strategically opt for the use of this pitch-related preplanning. All these findings point to a strong tendency that speakers have tonally marked an anticipatory cue at the onset of the articulation for the length of their upcoming utterance.

Previous studies have also reported evidence for more narrow-scoped pitch-related preplanning by showing the time-dependency of utterance-initial f0 heights on the local grammatical configuration (e.g., the length of a sub-clausal constituent) of the utterance (Fuchs et al., 2013; Ladd & Johnson, 1987; Scholz & Chen, 2014). These two different views of pitch-related preplanning have been referred to as *global* vs. *local* pitch-related preplanning (Scholz & Chen, 2014). Studies in favor of a local view have reported that the initial f0 heights correlate with the lengths of the sub-constituents in the utterance (e.g., the subject or the object NPs). In a qualitative analysis of the read speech from two English native speakers, Ladd and Johnson (1987) presented a case where the initial f0 height varied as a function not of the whole utterance length, but of the length of the first major constituent of the sentence. Scholz and Chen (2014) analyzed the sentence-initial f0 heights in Wenzhou Chinese in relation to the lengths of the sub-clausal constituents, i.e., subject and object NPs. They varied the length of the subjects and objects in terms of word numbers and examined how the length of the sub-constituents correlated with the initial f0 shifting of the whole utterance. According to their results, the sentence-initial f0 heights varied both as a function of the length of the subject and of the object, with the former effect found to be stronger. Their analyses suggested that while utterance length had an effect on the initial f0 shifting, part of the correlation could be attributed to local pitch-related shifting from the subject, which in turn could be taken as a mitigated argument for weak global f0 preplanning. Analyzing German read speech varying in length of the first constituent, Fuchs et al. (2013) also reported a similar observation: that the initial f0 height of the utterance was associated with the length of the first constituent in the utterance, but not with the overall utterance length. Be it global or local pitch-related preplanning, it is clear

that the utterance initial prosodic encoding is indicative of durational properties of the upcoming discourse in speech production. It is unclear, however, to what extent the local grammatical configuration of the utterance (e.g., the sub-constituents) may contribute to this pitch-related preplanning.

2.2. Utterance-final prosodic encoding and F0 declination

In connection to the relationship between initial f0 height and utterance length, the evidence for pitch-related preplanning based on the final f0 height may be more ambiguous. Shih (2000) observed that the final f0 mean in a controlled sentence with all lexical high-tone syllables (with a shared low and rising tone starting frame, *Lao3Wang2*) tended to be lower when the sentence was longer. However, this relationship became weaker in sentences with clear prominence focus. Studies on read speech in general have reported a tendency that the utterance-final f0 height decreased in longer phrases but the duration-dependency of the final f0 predictably followed an exponential decay from the initial f0 and asymptoted to a speaker-based baseline (Liberman & Pierrehumbert, 1984), or that the utterance-final f0 was possibly constant for utterances of different lengths (Yuan & Liberman, 2014). Analyzing the f0 heights of a reading of different berry names, Liberman and Pierrehumbert (1984) found that the last f0 height was lower than the corresponding non-final f0 heights at the same serial position. For example, the last f0 height of the three-item berry name reading consistently fell below the third f0 height of the four-item or five-item berry names. However, no statistically important relationship was reported between utterance length and the utterance-final f0 height in Liberman and Pierrehumbert (1984). Using homophones to construct utterances with different levels of internal grammatical junctures (i.e., lexical, phrasal, and clausal boundaries), Wang, Xu, and Ding (2018) further examined how different boundary types were prosodically marked in relation to different pragmatic contexts (i.e., focus and newness). Their results suggested that while all boundary types were consistently marked by different cross-boundary durational patterns, only the strongest grammatical boundary (i.e., clausal boundary) showed significant effects on f0 changes, i.e., pre-boundary f0 lowering and post-boundary f0 raising. Also, they found that while the pre-boundary f0 lowering was evident in both maximum and minimum f0, the post-boundary f0 raising was only observed in minimum f0. On the other hand, working on scripted broadcast speech in Mandarin and English, Yuan and Liberman (2014) found no correlation between the length of the inter-pause unit (IPU) and the final f0 height, which was measured as the f0 valleys in the final 500 ms of the IPU identified using the convex-hull algorithm. This would render initial f0 shifting to be more indicative evidence for pitch-related preplanning, and the final f0 shifting representing "invariant characteristics of a speaker's voice" (Liberman & Pierrehumbert, 1984).

Relevant to the initial and final pitch shifting is the degree of the f0 declination of the utterances in speech production. F0 declination has been a central pitch-related variation discussed in research on pitch encodings (Beckman, Hirschberg, & Shattuck-Hufnagel, 2006; Collier, 1975; Liberman & Pierrehumbert, 1984; Maeda, 1976; Shih, 2000; 't Hart, 1979; Yuan & Liberman, 2014). Many factors have been suggested to contribute to this general pitch decline in speech production (see a summary in Shih, 2000), including physiological constraints (Collier, 1975), down-stepping effect (Liberman & Pierrehumbert, 1984), final lowering effect of grammatical boundaries (Beckman et al., 2006), or sentence types (Thorsen, 1980). Of particular relevance to pitch-related preplanning is the relationship between f0 declination and the utterance length. A strong association between pitch range arrangement (e.g., the slope of the f0 declination) and the utterance length may provide support for preplanning. Studies have shown that shorter utterances usually have steeper f0 slopes, which are evident not only for the f0 peaks but also f0 valleys of the utterances (Maeda, 1976; 't Hart, 1979). However, it has been argued that this correlation may arise from a down-stepping effect where the f0 peaks of the high pitch accents, due to the influences of utterance-medial low lexical tones, may follow a predictable exponential decay, i.e., one f0 peak being a constant fraction of the previous one (Liberman & Pierrehumbert, 1984). Shih (2000) observed that the f0 means of high tone words showed a significant exponential declination as well, with a faster dropping rate in the early segment of the utterance. She argued that the declination would asymptote to a predicable speaker's low-bound baseline. However, she did not identify any important relationship between the f0 slope and the utterance length.

2.3. From read speech to spontaneous speech

As most studies on f0 declination have only analyzed controlled read speech, an investigation of the f0 declination in spontaneous speech may provide a more conclusive generalization. Pitch declination has been shown to vary greatly in different speaking styles (Swerts, Strangert, & Heldner, 1996). In read speech, the f0 declination is suggested to be steeper with more time-dependency (Swerts et al., 1996). Yuan and Liberman (2014) conducted a cross-linguistic comparative study on the f0 declinations in English and Mandarin broadcast news speech. In their study, the slopes of the f0 declination in the inter-pause units (IPUs) in broadcast news speech were strongly associated with the IPU length, even after removing the initial raising and final lowering effects. Shorter IPUs showed steeper declinations in both English and Mandarin broadcast speech, and this tendency was significantly more prominent in Mandarin. In their analysis, the utterance final f0 heights, however, did not vary as a function of the utterance length. Their results may reiterate the claim that prosodic encoding in speech production is a linguistically meaningful behavior (Liberman & Pierrehumbert, 1984; Yuan & Liberman, 2014).

Working on spontaneous speech of Estonian, Asu et al. (2016) analyzed the f0 declination slopes of the intonational phrases (IPs) in relation to the lengths of IPs. They observed that IP length correlated not only with the f0 declination slope but also the final f0 heights as well. Their results suggested that shorter IPs showed steeper f0 slopes than long IPs, but longer IPs tended to have lower phrase-final f0 heights, suggesting a larger pitch range pre-arranged in longer IPs in

spontaneous speech. The different findings on the duration-dependency of the final f0 heights in Yuan and Liberman (2014) and Asu et al. (2016) may be connected to the fact that read speech often comes with a larger pitch range (Swerts et al., 1996). It is thus more likely for a speaker to reach their individual low bound in read speech, no matter the utterance is long or short. However, in more spontaneous speech, the online message planning often leads to more fragmented utterances (i.e., IPs), possibly with disfluencies/interruptions. Therefore, the final f0 height of a prosodic unit in spontaneous speech may be more duration-dependent. Speakers may be more likely to drop to their pitch low-bound baseline in a longer prosodic unit. We posit that the more spontaneous the speaking style is, the more duration-dependent the utterance-final f0 height may be.

As already noted, previous studies on pitch-related preplanning have mostly analyzed the f0 shifting in read speech while spontaneous speech has been rather understudied. Furthermore, in controlled read-speech analyses, the prompt utterances are often a syntactically complete clause/sentence and different studies analyze the prosodic encoding of the utterance (e.g., initial/final f0 shifting, f0 slope) in relation to (1) the length of the whole sentence (Liberman & Pierrehumbert, 1984; Prieto et al., 2006; Shih, 2000; Thorson, 2007; Wang & Xu, 2011), (2) the length of the sub-constituent of the sentence (Fuchs et al., 2013; Ladd & Johnson, 1987; Scholz & Chen, 2014), or (3) the utterance types (Shih, 2000; Swerts et al., 1996; Thorsen, 1980). In these controlled settings, speakers often have an overview of the sentences to be articulated, which in turn makes the semantic contents of these sentences already active in the mind of the speaker before articulation. The pitch-related preplanning observed under the condition that speakers have learned the complete propositions they are expected to articulate may not be optimal. In a normal spontaneous speech setting, it remains an empirical question of whether speakers plan their talk on a propositional basis. It is still unclear to what extent a full-fledged proposition is active in the mind of the speaker before articulating each prosodic unit. As in spontaneous speech a prosodic unit does not necessarily correspond to a proposition (Park, 2002; Tao, 1996), it would be interesting to see the role of semantic completeness of the prosodic units and its effect on the prosodic encodings of the prosodic units in spontaneous speech. Therefore, different from previous studies, the present study will examine the initial and final f0 shifting of the prosodic units in relation to their semantic configuration. In particular, the semantic configuration of a prosodic unit in conversational discourse will be defined based on their degrees of alignment with a basic semantic unit in discourse, the proposition.

3. Methods

3.1. Data

The Sinica Phone-aligned Chinese Conversational Speech Database (SPCCSD) was used in this study. The SPCCSD consists of face-to-face free conversations of totaling around 3.5 h from 16 different speakers: 7 males (age range: 23–43) and 9 females (age range: 16–46). It is available via the

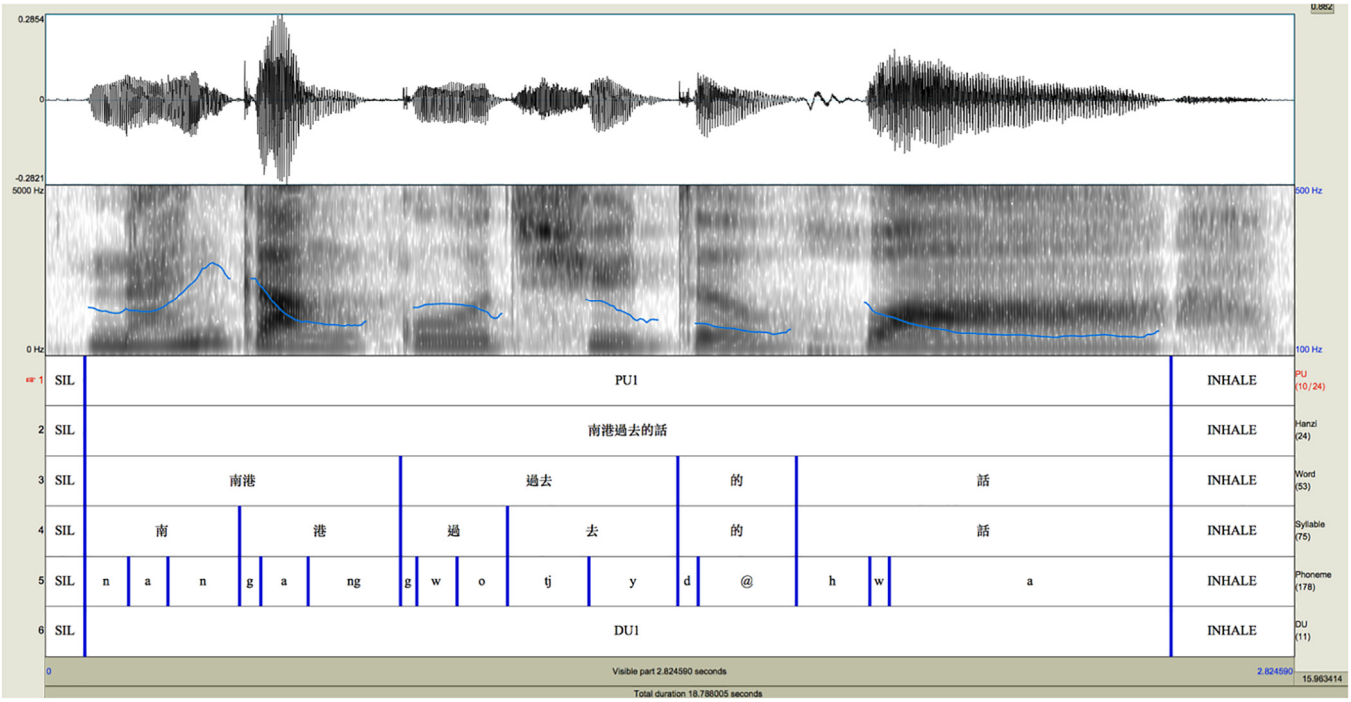


Fig. 4. Example of a time-aligned Praat TextGrid transcript from the SPCCSD. The SPCCSD provides the human annotations of syllables and phone segments for each speaker turn, which are time-aligned with each sound file, i.e., the *Syllable* and *Phoneme* tiers in the TextGrid. The symbols used in the *Phoneme* tiers were based on the conventions used in the HTK speech recognition toolkit. The additional tiers shown in the graph, i.e., *PU*, *Hanzi*, and *Word* were further enriched in Liu and Tseng (2009); the DU tier was first annotated in Prévot et al. (2015) and further refined in this study.

Association for Computational Linguistics and Chinese Language Processing (ACLCLP) (http://www.aclclp.org.tw/corp.php). The audio files of the SPCCSD have been manually time-aligned with the transcripts to the segmental level, stored in Praat's TextGrid format, as shown in Fig. 4.

3.2. Prosodic units and discourse units

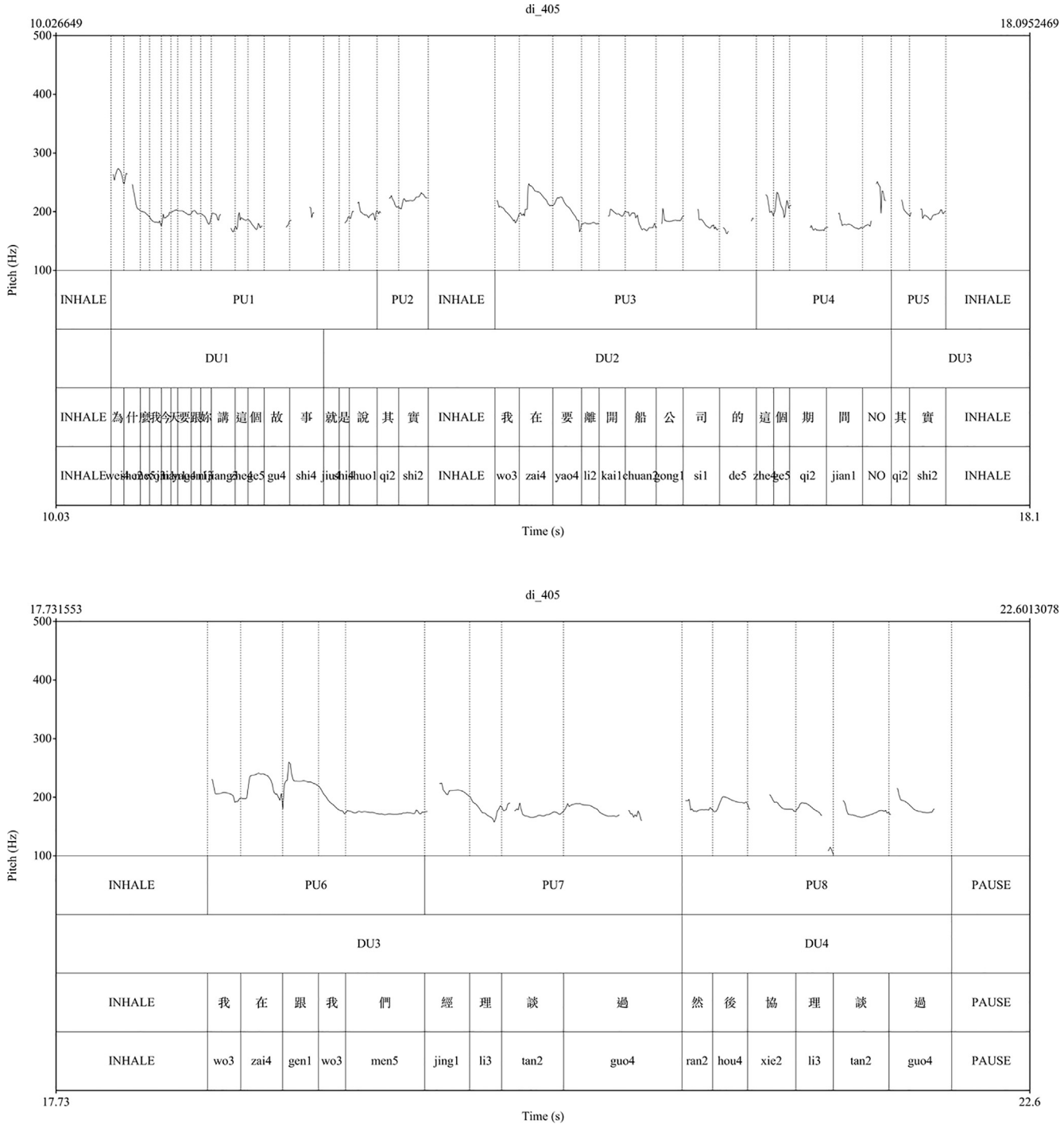
The prosodic units (PU) analyzed in this study were taken from Liu and Tseng (2009). Following Chafe's definitions of intonation units, Liu and Tseng enriched the prosodic annotation of the SPCCSD by manually segmenting each speaker turn into non-overlapping intonation units by auditory means. As the word "intonation" was ambiguous for non-stressed languages without prominent pitch-accents such as Taiwan Mandarin, these speech units were referred to as *prosodic units* (PU) in this project. Liu and Tseng defined a PU as a stretch of speech by a single speaker uttered with a "perceptually coherent prosodic constituent" (p. 150). The perception of coherence was based on two types of cues: the pitch variation and the timing patterns of the utterance. The former featured a perception of an up-ward shift in pitch relative to the speaker's previous utterance; the latter included a perception of a prosodic lengthening of a (final) syllable, acceleration in tempo on initial syllables, or a noticeable disjuncture or disruption of utterances (e.g., pauses, inhalation, laughters). The annotators marked the PU boundaries based on at least one of these perceptual cues (i.e., preferring to use the convergence of these cues). Not all perceptual cues were equally important in determining the PU boundaries. Noticeable pauses were often more central than the others. PUs may differ in types and levels as they demonstrate varying degrees of canonical prosodic auditory cues.

This perceptually-defined speech unit was evident in spoken discourse and various names have been proposed to refer to units of a similar, though not necessarily identical, construct. It is generally acknowledged that the prosodic structure in spoken discourse forms a hierarchy (Beckman et al., 2006; Selkirk, 1986; Tseng, Pin, Lee, Wang, & Chen, 2005). These hierarchy-dominant prosodic frameworks often require a more explicit marking of the levels/types for the prosodic boundaries above our PU level. It should be stressed that the PU annotation adopted in this study does not imply that we are claiming these PU boundaries are all of the same boundary type/level. As the present study is dealing with spontaneous speech, which is more likely to demonstrate greater prosodic variability than read speech, we believe that identifying a baseline prosodic boundary that is perceptually prominent to the annotator is more practical and workable. Classifying these perceptual boundaries into different types/strengths on top of the boundary identification would introduce additional concerns of inter-transcriber reliability (Syrdal & McGory, 2000). Most importantly, the former (identifying perceptually prominent prosodic boundaries) is what conversational participants routinely engage in in daily interactions (Mo, Cole, & Lee, 2008) while the latter (grouping the prosodic boundaries into levels/types) is a rather professional task requiring considerable amount of training (cf. Beckman et al., 2006). To put it differently, our analysis values the importance of PUs at this baseline level in the study of speech preplanning in spontaneous speech production.

The perceptual cues used in the PU segmentation in Liu and Tseng (2009) bear great resemblance to the criteria used for identifying prosodic boundaries in other theoretical frameworks. The main difference lies in the fact that the annotators in our project were not required to further determine the types and levels of the PU boundaries in the manual annotation. A comparison of the definitions of our PUs and those of the other prosodic boundaries defined in other hierarchy-dominant theories suggest that our PUs may correspond to the prosodic level of *intermediate phrase* in the autosegmental-metrical framework (Beckman et al., 2006). As our PU was largely based on Chafe's IU, Chafe has also pointed out that "evidently it is the intermediate phrase that corresponds to the intonation unit here" (Chafe, 1994, p. 57). While this study did not explicitly mark the boundary types/levels of the PUs, we examined the variability of their prosodic encodings in a different way. Our working assumption is that if the prosodic encodings of these PUs are connected to the grammatical configurations of the PUs (e.g., whether these PUs are structurally or semantically complete), we may be able to present more acoustic support for speakers' sensitivity to a particular grammatical or semantic template, and at the same time group these PUs into different types/levels. Therefore, the boundary types and/or levels of the PUs in this project are not considered to be a *categorical* feature of a prosodic boundary, but a *continuous* prosodic property of each PU.

A satisfactory inter-transcriber agreement over 80% for PU annotation was reported by Liu and Tseng, and a computational acoustic-based modeling for automatic PU boundary detection has also yielded promising results (Liu, Tseng, Jang, & Chen, 2010), thus affording considerable practical validity and reliability for the prosodic coherence of these units. In (1), we provide an example of PU segmentation in the SPCCSD, with the f0 plots on the top-panel and the linguistic annotations on the bottom panel. For clarity of expression, only four tiers of annotations were presented here, including PU (Tier 1), DU (Tier 2), Syllables in Characters (Tier 3), and Syllables in Pinyin (Tier 4). PUs were numbered by a unique index number within each example. In (1), several PUs were determined based on a preceding noticeable disjuncture in speech (e.g., inhalation), such as PU1, PU3, and PU6. Also, several PUs were determined based on a cross-boundary durational pattern of acceleration-deceleration (e.g., lengthening of the final syllables in the previous segment and the shortening of the initial syllables in the subsequent segment), such as PU1/2, PU3/4, PU4/5. Several PUs were identified on the basis of a clear initial pitch reset from the previous PU, such as PU3/4, PU4/5, PU6/7, PU7/8. It should be noted that all these PU annotations were solely based on the annotator's perception of relevant cues from the audio signals. It was evident that the identification of PUs was supported by the convergence of multiple perceptual cues.

(1) Examples of PU segmentation[1]



Translations: “Why would I want to share this story with you? / In fact, at the time when I was about to leave the ship company, / I actually discussed (that) with our manager. / And (I) talked with the associate manager.”

[1] All the examples presented in this study follow this simplified textgrid format. The first tier includes the annotations of PU; the second tier the DU; the third tier the Chinese characters of each syllable; the last tier the pinyin of each syllable. Under each example we provide the idiomatic English translation of each DU. The DU boundary in the English translation is preserved and marked by a forward slash “/” for readability. A longer utterance is sometimes split into two textgrids for clarity of expression.

This project continued to enrich the semantic annotations of the SPCCSD by manually segmenting each speaker turn into non-overlapping *discourse units* (DU) based on semantic criteria. We used a more general label for this semantic unit, i.e., discourse unit (DU), in the hope of acknowledging the

spontaneous characteristics and functional importance of other structural equivalents of similar pragmatic functions in conversational speech, such as reactive tokens (e.g., *dui a* 'yeah right') and formulaic expressions (*jiu4shi4 zhe4yang4* "it's like this", *zhe4yang4zhi5* "that's it"). A DU was defined as a part of an utterance with a predicate and the key arguments of the predicate, which described a proposition, e.g., a state or event (Croft, 1995; Givón, 1984; Huang & Chui, 1997; Matsumoto, 2000; Park, 2002; Tao, 1996; Thompson & Couper-Kuhlen, 2005; Thompson & Hopper, 2001). This semantic unit is often structurally encoded by a clause in languages (Givón, 1984, p. 239; Thompson & Couper-Kuhlen, 2005), serving as a practical and useful analytic unit for research on conversational speech (Huang & Chui, 1997; Lehmann, 1988; Ono & Thompson, 1996; Tao, 1996; Thompson & Couper-Kuhlen, 2005) and cross-linguistic comparative analyses on spontaneous speech segmentation (Iwasaki & Tao, 1993; Prévot, Tseng, Peshkov, & Chen, 2015).

The main predicate was the semantic core of the proposition. Its semantics informed the number of participants involved in the proposition, which in turn gave clues to the boundaries of a DU. Predicates included the following categories: one-participant predicates (2a), two-participant predicates (2b), three-participant predicates (2c), copular (2d), prepositional predicates (2e), and complement-taking predicates (2f).

(2) Types of Propositions (DU)

a. *ta1 ye3shi4 hen3 nu3li4 de5 zai4 gong1zuo4* (di_269)
"he's also working very hard."

b. *suo3yi3 ji1ben3shang4 yuang3uang1 ni3 hui4 zhi2she4 dao4 ni3 yan3jing1* (di_044)
"So basically the high beam would directly shoot at your eyes."

c. *wo3men5 dou1 jiao4 ta1 wu2 ye2ye5* (di_213)
"We all call him Granpa Wu."

d. *qi2shi2 shi4 hen3 jian4quan2 de5 yi2 ge5 guan3* (di_017)
"(It's) actually a very well-designed shop."

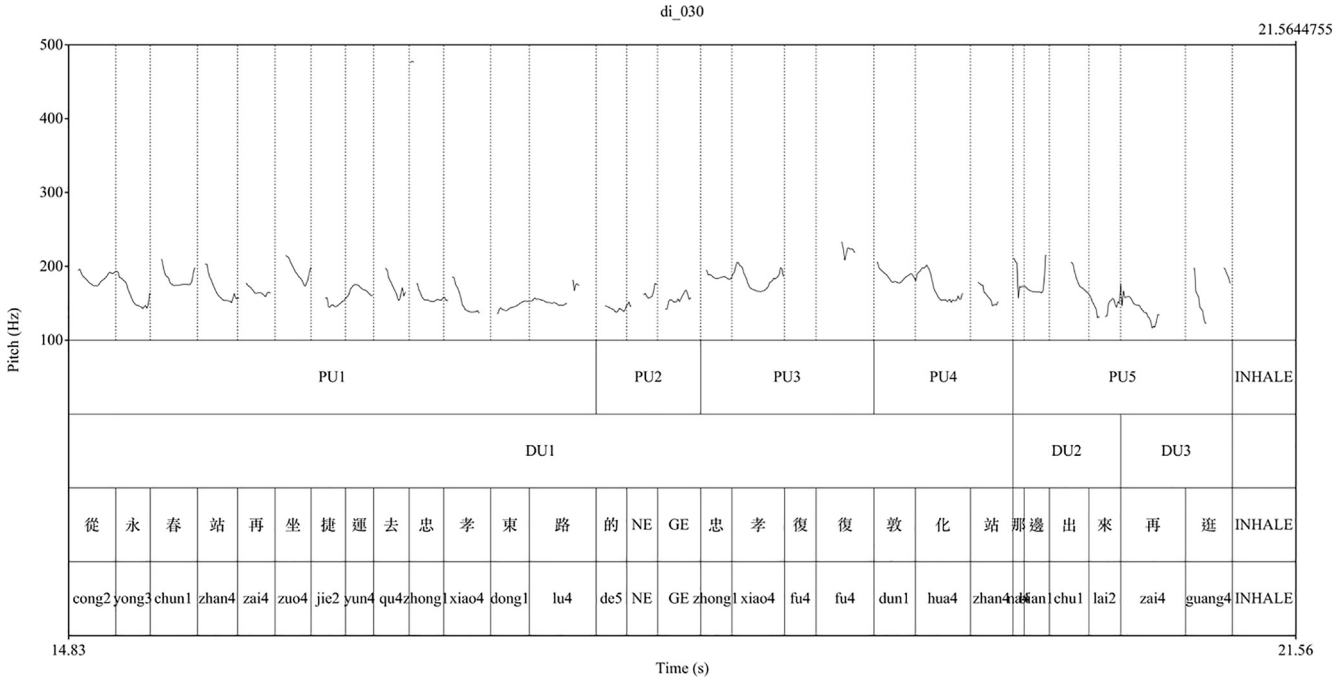
e. *ran2hou4 yi4 tai2 xiao3xiao3 mo2tuo1che1 zai4 pang2bian1* (di_677)
"And a very small scooter was on the side."

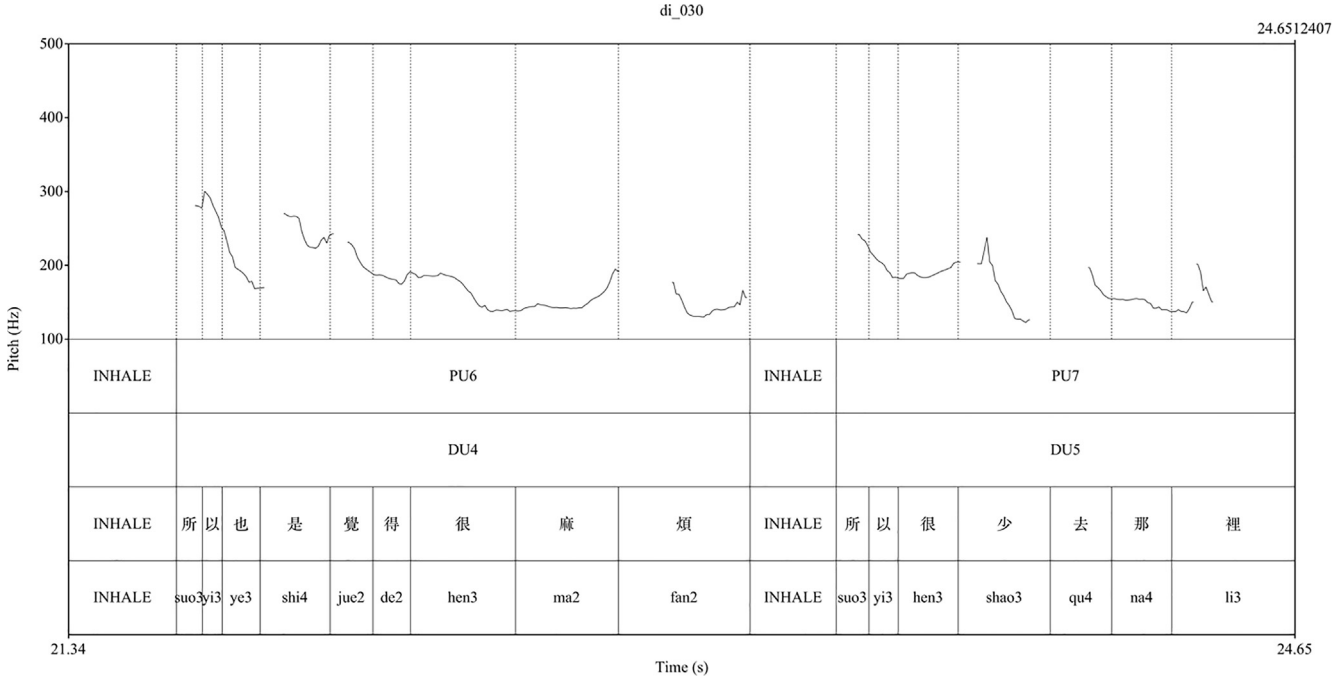
f. *wo3 jue2de5 hui4 bi3jiao4 hao3wan2* (di_013)
"I think it would be more fun."

As Mandarin does not have an overt tense-marking on the main predicate of the clause, the zero-anaphora (i.e., unexpressed participants) in discourse may sometimes introduce additional challenges in determining the main predicate out of a series of verbal lexical units in a continuous speech. Given two consecutive verbal units in a row, we used semantic criteria to determine whether each verbal unit corresponded to an independent DU or was analyzed as a complement of another matrix predicate. We adopted the principle of semantic integration proposed in cognitive semantics (Givón, 1993). Givón (1993) identified three types of predicates that may take verbal units as complements, i.e., modality, manipulative, and perception-cognition-utterance (PCU) verbs. These complement-taking predicates are often semantically bonded with their verbal complements in terms of the co-temporality, implicativity, and shared participants of the events encoded by these two verbs, and therefore are more likely to be structurally encoded with their complements as one clause. Based on large-scaled typological analyses, Givón found that modality and manipulative predicates were more often syntactically integrated with their verbal complements as one clause, while PCU predicates were more likely to be encoded as an independent clause from its verbal complement. In this project, we considered all these three types of complement-taking predicates in the same way.

Our DU annotation followed three heuristics: (a) every verbal unit in the speaker turn was a candidate for the main predicate of a proposition (e.g., *chu1lai2* 'get off' of DU2 and *guang4* 'go shopping' of DU3 in (3)), unless (b) it was the complement of the modality, manipulative, and PCU verbs (e.g., *ma2fan2* 'troublesome' of DU4 in (3)), or (c) it was embedded in a relative clause (e.g., *li2kai1* 'leave' of DU2 in (1)). These decisions ensured a transparent connection between the "participants" and the core predicate in each proposition. Unlike relative and nominal clauses, adverbial dependent clauses were annotated as an independent DU due to their looser semantic connection to the proposition encoded in the matrix clause as well as their distinct discourse-pragmatic functions (Ford, 1993; Wang, 2002). For example, the adverbial clause introduced by *yin1wei4* 'because' in DU4 of (4) was identified as an independent DU. One-seventh of the whole dataset was annotated by two trained linguists and the inter-labeler agreement for DU annotation measured using kappa statistic was 0.86. The remaining dataset was then annotated by one labeler.

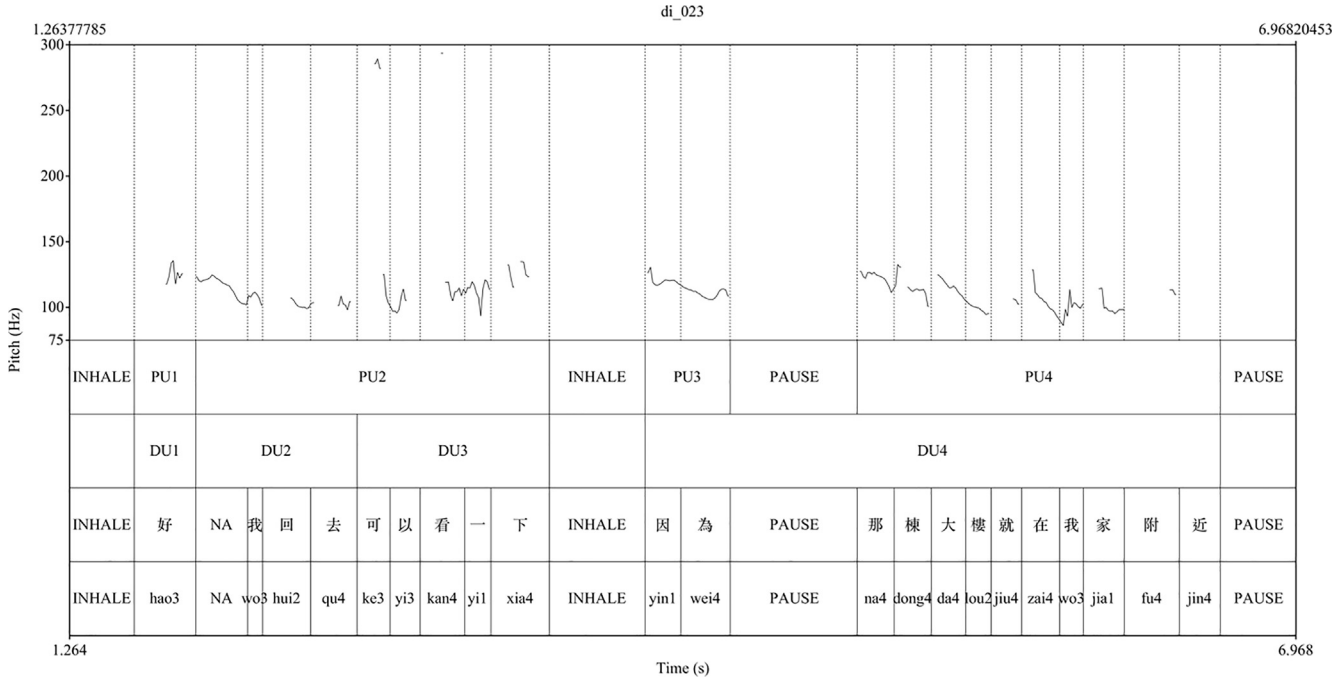
(3) DI_030





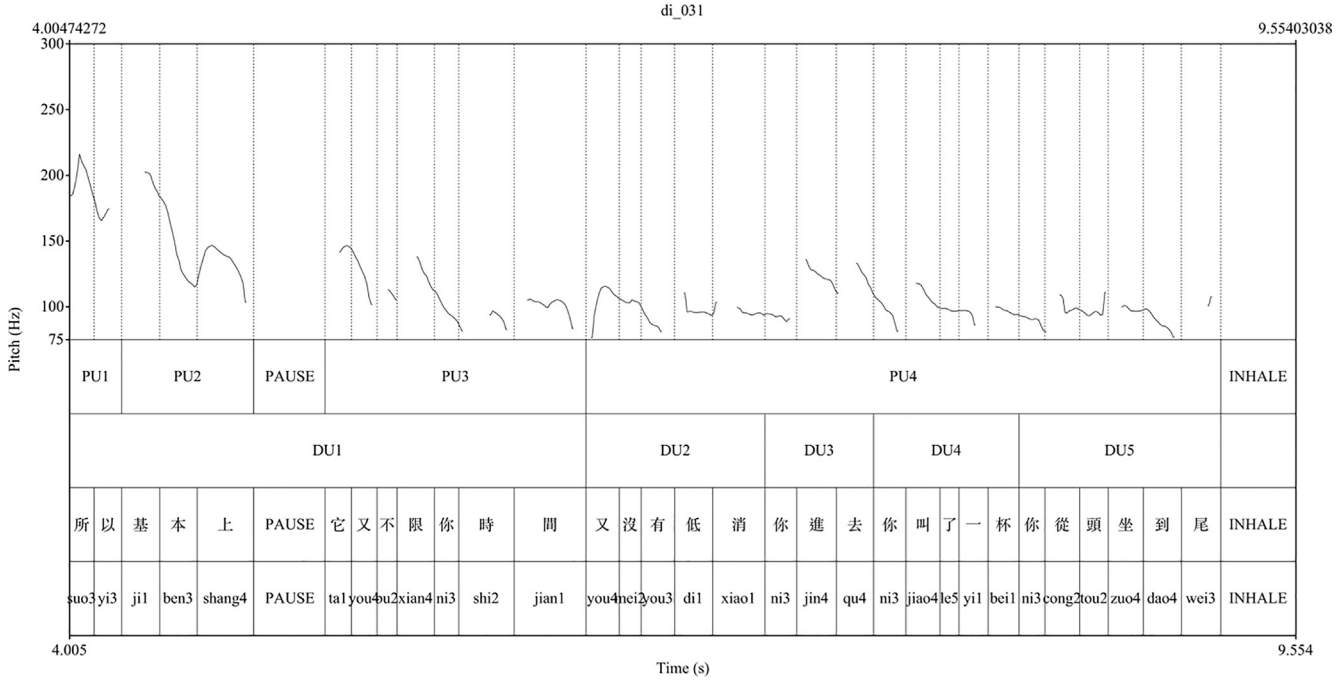
Translation: “(We) took the MRT from Yong-Chun Station to Zhong-Xiao Dun-Hua Station on Zhong-Xiao East Road, / got off from there, / and then shopped around. / So (I) think it's inconvenient as well. / So (I) seldom go there.”

(4) DI_023



Translation: “Ok. / And when I get back, / I’ll take a look / because the building is right next to my place.”

(5) DI_031



Translation: “So, basically, there’s no time limit, / and they don’t have a minimum charge. / If you go there, / you can order a drink, / and sit from the beginning till the end.”

With the semantic segmentation of DUs in each speaker turn, we were able to analyze each PU in terms of their semantic completeness. First of all, we categorized PUs into two classes, i.e., simple and complex PUs, according to their semantic complexity. A complex PU referred to one which spanned more than one DU in the prosodic contour (e.g., PU4 in (5)) while a simple PU referred to one which presented a single proposition or propositional fragments. Secondly, both simple and complex PUs were further categorized according to their co-extensiveness with the DU, namely, whether the PU and DU were co-extensive on the left (LEFT) and RIGHT (R) boundaries. Take simple PUs for example. Speakers in spontaneous speech may articulate a DU in one PU, which is coextensive with the whole proposition on both sides (e.g., PU6 and PU7 in (3)). On the other hand, speakers may sometimes present the DU incrementally with a series of smaller propositionally-fragmented PUs, consisting of a *DU-initial PU* (e.g., PU1 in (5)), *DU-final PU* (e.g. PU3 in (5)), and/or *DU-internal PU* (e.g. PU2 in (5)). Therefore, the semantic contents of each PU were evaluated based on its co-extensiveness with the DU. The varying degrees of PU-DU correspondences formed the basis of our hypothesis-driven factors on the cross-boundary prosodic encodings of PUs.

Two types of PUs in the SPCCSD were excluded from the current analysis. First, we excluded fully disfluent PUs which presented a partial proposition that was not complete and abandoned by the speakers. However, if the intended DU presented or mediated by the disfluent PUs was completed in the speaker's subsequent speech, these disfluent PUs were still included in the current analysis (e.g., PU3 in (3)). Secondly, PUs that were too short to extract reliable f0 values were also removed from the analysis. In the original annotations of PUs provided in Liu and Tseng (2009), there were 8563 prosodic units identified in the SPCCSD. After removing fully-disfluent PUs in the abandoned utterances (477 tokens) and short PUs (47 tokens), there remained 8039 PUs in our final dataset.

3.3. F0 values extractions

F0 extraction has been performed by researchers over different domains, e.g., the first pitch accent (Ladd & Johnson, 1987; Liberman & Pierrehumbert, 1984; Prieto et al., 2006), syllable (Shih, 2000), lexical tone (Wang & Xu, 2011), word (Fuchs et al., 2013; Scholz & Chen, 2014), or a relative proportion of the utterance (e.g., the first 25% of the phrase as in Asu et al. (2016)). Moreover, while most studies have adopted the observed f0 values as the measure, others may utilize more sophisticated interpolation and stylization to ensure a more reliable measure of f0 values. This study defined the PU-initial and PU-final f0 values as follows.

First, we extracted raw pitch values of each PU, using Praat's autocorrelation-based pitch tracking algorithm, with gender-dependent pitch ranges (75–300 Hz for male, and 100–500 Hz for female). Raw pitch values were then converted into semitones (with 50 Hz as the reference). Secondly, we identified the f0 value at the energy max (i.e., at the maximal dB value) of the first and last syllable of the PU based on the dB values. All syllable boundaries in the SPCCSD were manually annotated and cross-checked, thus providing reliable bases for our f0 extraction. As suggested by Grabe, Kochanski, and Coleman (2007, p. 287), this energy-controlled extraction can ensure the integrity and reliability of the observed f0 values and reduce the chance of f0 tracking errors. F0 values based on energy max were also found useful in Ladd and Johnson (1987).

To ensure the reliability of this energy-based f0 extraction, we replicated the method of f0 extraction used in Wang and Xu (2011), and extracted the PU-initial/final f0 values on the basis of the High/Low tones in the first/last word of the PU (i.e., the maximum f0 of the High tone in the first word of the PU was extracted as the PU-initial f0; the minimum f0 of the Low tone in the last word of the PU was extracted as the PU-final f0). Comparisons of our energy-based f0 values with these tone-based f0 values suggested high correlations (PU-initial: $r = 0.93$, $n = 8037$, $p < 0.01$; PU-final: $r = 0.89$, $n = 8037$, $p < 0.01$), and the tone-based f0 values yielded consistent statistical results as to be reported later in this study. Furthermore, we also found that the energy-based f0 values were strongly correlated with the maximum and minimum f0 values of the PU-initial and PU-final syllables. These correlations suggest that (a) the f0 change contributed by the lexical tone may be rather limited in spontaneous speech, and (b) the pitch contour of the lexical tones may be greatly reduced in spontaneous speech. However, several studies have suggested that different local tonal contours may influence the f0 extraction (Chen & Gussenhoven, 2008; Wang & Xu, 2011). To properly address this tonal effect, we included the lexical tones of the syllables from which the energy-controlled f0 values were extracted and the lexical tones of the neighboring syllables in our later statistical models as control factors (cf. Section 3.3).

It should be noted that we did not include the metric of the f0 slope in our analysis. Slope metrics are usually a mathematical construct approximating the general trend of the f0 downward movement in a particular prosodic domain, rendering this metric less transparent and computationally more complex. It is posited that the variation of initial and final f0 heights should suffice for the purpose of determining a relationship between pitch-related patterning and message preplanning. In this study, an *anticipatory* initial f0 shifting associated with the semantics of the upcoming utterance (e.g., whether the utterance is a semantically complete proposition) may provide support for advanced planning; a *carried-over* final f0 shifting associated with the semantics of the articulated utterance may also give a clue to the semantic base unit on which the prosodic structure has been pre-arranged.

3.4. Statistical analysis

The present study examined to what extent the PU-DU alignment on each side of the boundary was related to the shifting of the PU-initial and PU-final f0 heights in spontaneous speech production. To determine the importance of the relationships, linear mixed-effect analyses were performed on the PU-initial and PU-final f0 heights, using R and *lme4* (Bates, Mächler, Bolker, & Walker, 2014). To answer our RQ1 and RQ2, our first analysis examined the relationship between prosodic encodings and semantic contents of all simple PUs by taking PU-initial and PU-final f0 heights as dependent variables on the one hand, and the two alignments, LEFT

Table 1
Distributions of simple PUs according to PU-DU alignment.

	Left	Right	*N*	%	Average Syllable Number (SD)	Average Length in Seconds (SD)
Simple PU	–	–	1421	0.21	4.06(3.02)	0.78(0.48)
	–	+	1672	0.24	6.90(3.70)	1.21(0.57)
	+	–	1701	0.25	4.19(3.11)	0.73(0.48)
	+	+	2127	0.31	6.94(4.40)	1.16(0.69)
Total			6921	100%		

Notes. For Left and Right, + = alignment; – = mismatch.

Table 2
Distributions of complex PUs according to PU-DU alignment.

	Left	Right	*N*	%	Average Syllable Number (SD)	Average Length in Seconds (SD)
Complex PU	–	–	221	0.20	11.23(4.75)	1.74(0.74)
	–	+	282	0.25	14.24(5.54)	2.23(0.86)
	+	–	246	0.22	10.17(4.85)	1.51(0.72)
	+	+	369	0.33	13.13(5.91)	2.01(0.93)
Total			1118	100%		

Notes. For Left and Right, + = alignment; – = mismatch.

and Right, and their interaction, Left:Right, as the fixed effects on the other. Both models were built with speaker-dependent and gender-dependent random intercepts[2]. To answer our RQ3, we did the same analyses on cross-boundary f0 heights for all complex PUs.

As previous studies have suggested a range of important factors that may influence the utterance-initial/final f0, we included these important factors as control fixed effects in our mixed-effect models. All the following factors were particularly included in the model to ensure that the cross-boundary f0 heights analyzed in this study did not result from these confounding impacts.

(6) Control Factors in the Mixed-effect Models

a. SylTone: the lexical tone of the syllable from which the f0 at the energy max was extracted (categorical, 6 levels).
b. SylToneNeighbor: the lexical tone of the syllable next to the syllable from which the energy-based f0 was extracted. In the PU-initial f0 analysis, this would be the lexical tone of the second syllable of the PU; in the PU-final f0 analysis, this would be the lexical tone of the penultimate syllable of the PU (categorical, 6 levels).
c. PULength: the length of the PU measured in seconds (continuous).
d. PUOnsetTime: the time difference between the onset of the PU and the onset of the speaker turn in which the PU was articulated, normalized by the length of the turn (continuous).
e. NewTopic: the change of the new topic in discourse (categorical, 2 levels).

We specify the operational definitions and data types of all the control factors included in the mixed-effect models in (6). SylTone and SylToneNeighbor addressed the concerns that the extracted PU-initial/final f0 heights were conditioned on the local pitch contour introduced by (a) the lexical tones from which the f0 values were extracted, and (b) the tonal co-articulation of the neighboring syllables. These factors included five lexical tones and one undefined tonal level for discourse particles. PULength was a control factor to accommodate the strong correlation between the f0 heights and utterance length observed in previous studies (cf. Section 2.1). PUOnsetTime addressed the general tendency of f0 declination in a continuous stretch of speech by considering the onset of each PU relative to the onset of the speaker turn where the PU was articulated.

Prior research has also suggested that the information status has great influence on the f0 variation (Smith, 2004; Wang & Xu, 2011; Wichmann, 2000; Zellers, Post, & D'Imperio, 2009). Relevant to our cross-boundary f0 heights is the topic structure of the DUs. As every DU boundary was potentially a topic transition in discourse, the cross-boundary f0 shifting in our analysis may be partially contributed by a DU which introduced a new topic in discourse. To control the influence of the topic structure, we marked the nature of topic transition for each DU boundary based on the labeling scheme of Nakajima and Allen (1993). Each DU boundary was labeled according to whether the following DU was New (e.g., a topic shift or speaker change) or Given (e.g., a topic elaboration or continuation). About one-fifth of the dataset were labelled by two annotators for inter-labeler agreement check (Kappa = 0.85). Therefore, we included NewTopic as a control fixed effect in our model to address the pitch variation introduced by new-topic DUs, i.e., regressing the f0 shifting contributed by new topics in discourse.

The present study analyzed the significance of the PU-DU alignment effects (Left, Right, and Left:Right) on the f0 shifting of PUs, by controlling effects of utterance length (PULength), tonal influences (SylTone, SylToneNeighbor), pitch declination (PUOnsetTime) and topic structure (NewTopic). All reported models had low collinearity [vif's < 2.4]. Following the model-building principle suggested in Field, Miles, and Field (2012, p. 619), we first defined default baseline models with all control factors included as fixed parts of the mixed-effect model, i.e., (7a) and (7c). We added our hypothesis-driven fixed effects (i.e., Left, Right, and Left:Right) incrementally to the baseline models and evaluated their significance through model comparisons using chi-square likelihood ratio tests. After we determined the structure of the fixed-effect parameters, we tested

[2] In the random part of the model, the gender variable (GENDER) was nested within the speaker variable (SPID) by setting the random factor as ~1|SPID/GENDER in R. The statistical results did not differ significantly when GENDER was treated as a fixed part of the formula. The random intercepts were used to control the varying f0 baselines for different speakers and genders.

Table 3
Model parameters of simple PUs.[1]

	PU-DU Co-extensiveness	β	SE	df	*t* value	*r*
PU-INITIAL F0	LEFT	0.3412	0.0769	6887	4.4348***	0.05
	RIGHT	−1.2533	0.0757	6887	−16.5609***	0.20
PU-FINAL F0	LEFT	0.5179	0.0784	6887	6.6083***	0.08
	RIGHT	−1.5990	0.0878	6887	−18.2086***	0.21

Notes: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

[1] Field et al. (2012, p. 640) suggest that a mixed-design model should compute an effect size which summarizes a focused effect. Therefore, the effect size (*r*) reported in this study used the method provided in Field et al.: $r = \sqrt{\frac{t^2}{t^2+df}}$.

the need of the inclusion of speaker-dependent random *slopes*.[3] Speaker-dependent *intercepts* were always preloaded as a fixed part of the baseline model. The model comparisons suggested that the high-order interaction effect, LEFT:RIGHT, and random slopes did not contribute significantly to the improvement in the goodness-of-fit as these saturated models showed larger values of AIC or BIC (i.e., an indicator of overfitting). Therefore, our final models included LEFT and RIGHT as the fixed effects and speaker- and gender-dependent random intercepts as the random effects. The two statistical mixed-effect models are schematically represented in (7b) and (7d):

(7) Formulas for Linear Mixed Effect Models

a. PUINITIALF0 BASELINE = SYLTONE + SYLTONENEIGHBOR + PULENGTH + PUONSETTIME + NEWTOPIC + (1/SPEAKER/GENDER)
b. PUINITIALF0 = LEFT + RIGHT + SYLTONE + SYLTONENEIGHBOR + PULENGTH + PUONSETTIME + NEWTOPIC + (1/SPEAKER/GENDER)
c. PUFINALF0 BASELINE = SYLTONE + SYLTONENEIGHBOR + PULENGTH + PUONSETTIME + NEWTOPIC + (1/SPEAKER/GENDER)
d. PUFINALF0 = LEFT + RIGHT + SYLTONE + SYLTONENEIGHBOR + PULENGTH + PUONSETTIME + NEWTOPIC + (1/SPEAKER/GENDER)

4. Results

Tables 1 and 2 show the distribution of the PUs according to PU-DU alignment, LEFT and RIGHT, as well as the average length of each PU type. Similar distributions were observed in both simple and complex PUs. About one-third of the PUs are coextensive with DUs on both sides (i.e., [+LEFT] and [+RIGHT]) and about 55% of the PUs are co-extensive with DUs on the right-end boundaries (i.e., [+RIGHT]). More than 80% of the PUs are aligned with the DU boundaries on at least one side (i.e., either [+LEFT] or [+RIGHT]). The average length of the 8039 PUs analyzed in our analysis was 6.6 syllables (SD = 4.8) with average duration of 1.11 s (SD = 0.73) per PU.

4.1. Simple PUs

Our first analysis investigated the effects of PU-DU co-extensiveness on the f0 shifting of simple PUs. The effects of LEFT and RIGHT on PU-initial f0 shifting were significant (see Table 3): speakers showed higher initial f0 values in left-aligned PUs (Fig. 5), compared to non-left-aligned ones ($\beta = 0.34$, $t = 4.43$, df = 6887, $p < 0.01$, $r = 0.05$), and lower initial f0 in right-aligned PUs (Fig. 6), compared to non-right-aligned ones ($\beta = -1.25$, $t = -16.56$, df = 6887, $p < 0.01$, $r = 0.20$). No significant interaction effect was found on PU-initial f0 shifting.

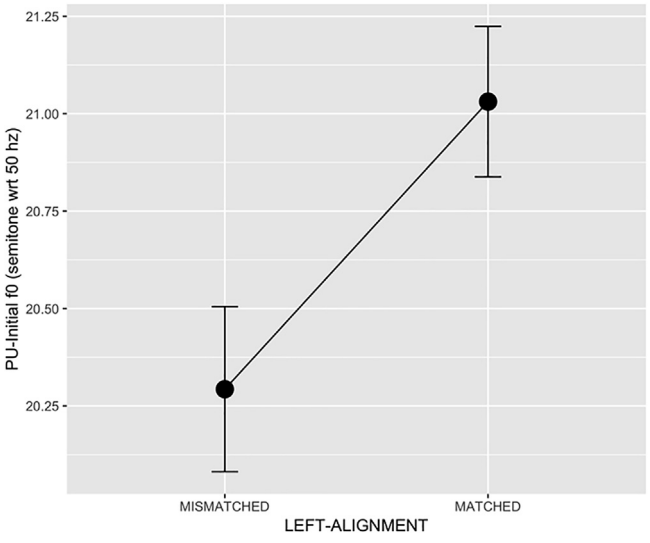


Fig. 5. Mean values of initial f0 heights of simple PUs in relation to LEFT with 95% confidence intervals. The bars represent the mean values of each sub-group, with confidence intervals delineated by the whiskers.

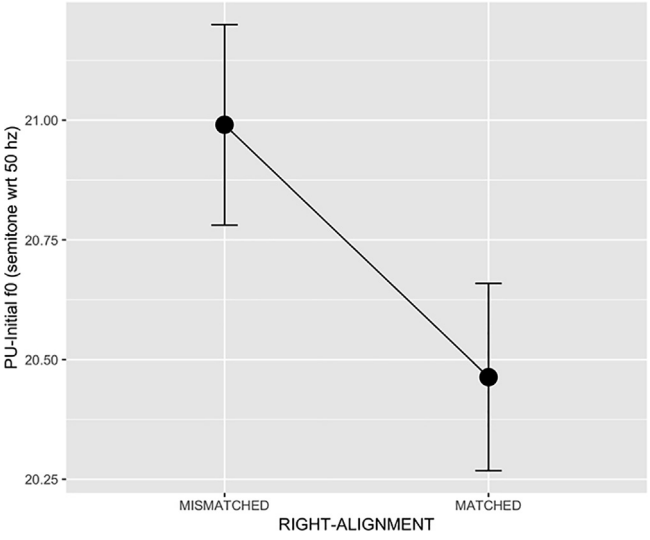


Fig. 6. Mean values of initial f0 heights of simple PUs in relation to RIGHT with 95% confidence intervals. The bars represent the mean values of each sub-group, with confidence intervals delineated by the whiskers.

The effects of LEFT and RIGHT on PU-final f0 shifting were also significant (see Table 3): speakers showed higher final f0 values in left-aligned PUs (Fig. 7), compared to non-left-

[3] For random slopes, we tested whether the effects of LEFT and RIGHT varied across different speakers.

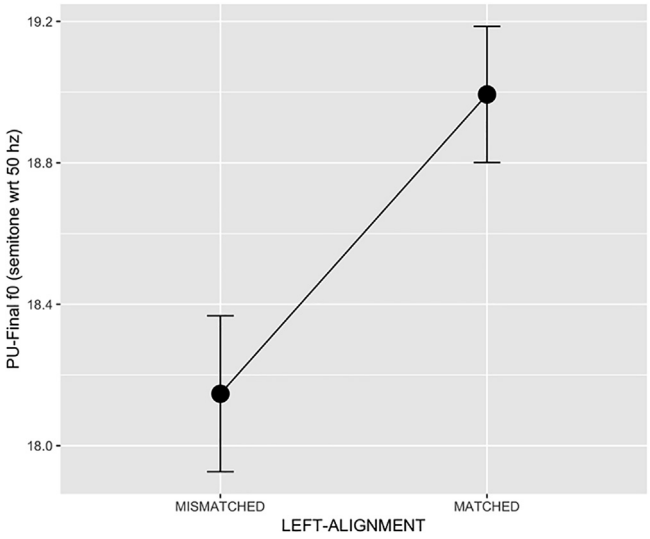


Fig. 7. Mean values of final f0 heights of simple PUs in relation to LEFT with 95% confidence intervals. The bars represent the mean values of each sub-group, with confidence intervals delineated by the whiskers.

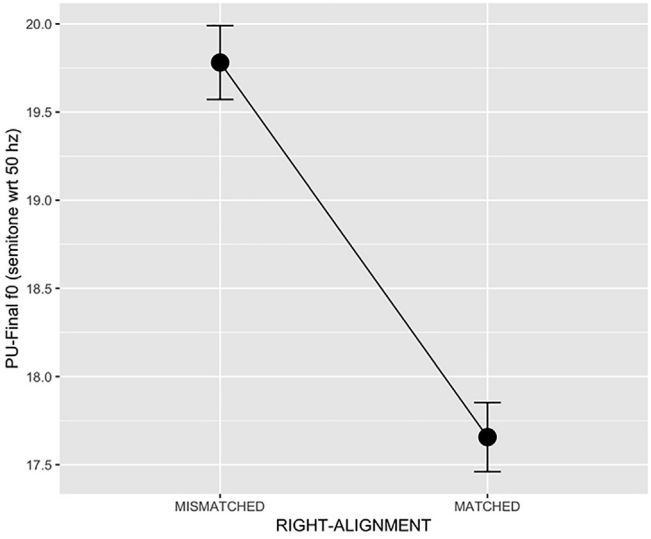


Fig. 8. Mean values of final f0 heights of simple PUs in relation to RIGHT with 95% confidence intervals. The bars represent the mean values of each sub-group, with confidence intervals delineated by the whiskers.

aligned ones ($\beta = 0.51$, $t = 6.60$, df = 6887, $p < 0.01$, $r = 0.08$), and lower final f0 in right-aligned PUs (Fig. 8), compared with non-right-aligned ones ($\beta = -1.58$, $t = -18.20$, df = 6887, $p < 0.01$, $r = 0.21$). No significant interaction effect was found on PU-final f0 shifting.

In summary, our results show that (a) speakers are sensitive to the boundaries of the semantic unit, i.e., DU, as the PU-DU alignment consistently introduces distinctive f0 shifting on both sides of the PU, and that (b) speakers' initial f0 height at the *onset* of the PU is connected to the semantic co-extensiveness at the *terminal* of the PU, and that (c) speakers' final f0 height at the *terminal* of the PU is connected to the semantic co-extensiveness at the *onset* of the PU.

4.2. Complex PUs

Our second analysis investigated the effects of PU-DU co-extensiveness on the f0 shifting of complex PUs. For PU-initial f0 shifting, no significant effects were found. For PU-final f0 shifting, we found a significant RIGHT effect ($\beta = -2.02$, $t = -10.35$, df = 1086, $p < 0.01$, $r = 0.30$): speakers showed lower final f0 values in right-aligned complex PUs, compared to non-right-aligned ones (see Table 4).

In sum, our second analysis shows that when producing a complex PU, i.e., a speech segment verbalizing more than one proposition, speakers show little sensitivity to the DU boundaries, and only limited evidence can be found in the PU-final f0 shifting. When speakers combine more than one proposition in one PU, they show lower f0 values if the complex PU terminates a DU.

5. Discussion

This study investigated PU-initial and PU-final f0 heights in relation to the PU-DU alignment on boundaries of both sides (LEFT and RIGHT) in Mandarin spontaneous speech. Cross-boundary f0 heights were defined as the f0 values at the energy max of the first and last syllables of the PU. In this section, we discuss the results of the two analyses in relation to our three research questions.

5.1. Sensitivity to the proposition

Our results provide an affirmative answer to our first research question: speakers consistently show sensitivity to the boundaries of the propositions in the cross-boundary prosodic encodings of the PUs. The pitch shifting contributed by the PU-DU alignment is effective in addition to the control factors (i.e., utterance length, tonal influences, pitch declination and topic structure). In particular, PU-initial f0 height is positively correlated with LEFT (Fig. 5), and PU-final f0 height is negatively correlated with RIGHT (Fig. 8). The former relationship indicates that speakers are aware of the initiation of a proposition at the onset of the PU articulation, which is reflected in the higher initial f0 height in speech production.

Table 4
Model parameters of complex PUs.

	PU-DU Co-extensiveness	β	SE	df	*t* value	*r*
PU-INITIAL F0	LEFT	0.0598	0.1865	1086	0.3204	0.01
	RIGHT	−0.0672	0.1732	1086	−0.3879	0.01
PU-FINAL F0	LEFT	0.4375	0.1790	1086	2.4448	0.07
	RIGHT	−2.0292	0.1961	1086	−10.3502***	0.30

Notes: ***$p < 0.001$.

The latter relationship indicates that speakers are aware of the termination of a proposition in the prearrangement of the prosodic contour, which is prosodically marked by the lower final f0 height in speech production.

These findings extend prior research on prosody-syntax interface in several ways. First, unlike IU-based studies, the co-extensiveness of PUs and DUs is re-analyzed in our study as a graded acoustic relationship, rather than a one-to-one mapping. Our analysis suggests that the mismatch of PU-DU boundaries is consistently marked by varying prosodic encodings, indicating speakers' sensitivity to this basic semantic unit in speech production. Second, our analysis goes in line with previous research showing that speakers are sensitive to discourse boundaries/junctures at multiple grain sizes (Fon, Johnson, & Chen, 2011). This study further contributes that speakers consistently show prosodic sensitivity to a more fundamental discourse unit, i.e., the proposition, in their prosodic arrangement in speech production. As a proposition is often structurally encoded by a clause, our finding may provide more substantive acoustic support for the effectiveness of the clause schema in speech production, suggesting the importance of this integral building block in language processing.

However, the connection between PU-initial f0 shifting and PU-initial semantics, or that between PU-final f0 shifting and PU-final semantics, only provides limited support for advanced planning. These findings so far only suggest that speakers show sensitivity to the initiation and termination of the propositions by their cross-boundary prosodic encodings. It is less unclear whether the whole proposition could have been preplanned before the onset of the articulation. The next section discusses this important issue by connecting our findings to our second question.

5.2. Proposition-based advanced planning

Our results provide an affirmative answer to our second question: speakers show signs of preplanning a whole proposition in speech production. In our first analysis of simple PUs, we have shown that PU-initial f0 shifting is connected to RIGHT, i.e., whether the PU is reaching the end of a proposition. Speakers consistently show lower initial f0 heights if the PU is to terminate a proposition (Fig. 6). Relatedly, we have demonstrated that PU-final f0 shifting is connected to LEFT, i.e., whether the PU initiates a proposition. Speakers consistently show higher final f0 heights if the PU has initiated a proposition (Fig. 7).

The former anticipatory relationship is interesting in the sense that utterance-initial f0 shifting is rarely discussed in association with the semantic configuration of the utterance-terminal positions in previous studies on read speech. Instead, previous studies have mostly reported the utterance-initial pitch heights to be correlated with the length of either the whole utterance (Liberman & Pierrehumbert, 1984; Prieto et al., 2006; Shih, 2000; Wang & Xu, 2011; Yuan & Liberman, 2014) or a more local sub-constituent, e.g., utterance-initial constituents (Fuchs et al., 2013; Ladd & Johnson, 1987; Scholz & Chen, 2014). Moreover, these experimental analyses on controlled speech often take semantically complete utterances as analytic units. On top of the random effects and other control factors, our analysis further suggests that the advanced planning may involve not only the complexity of the upcoming utterance (as measured by utterance lengths in previous works) but also the semantic coherence of the upcoming utterance (e.g., whether the whole utterance reaches the termination of the proposition). From a perspective of message planning, our data suggest that the PU-initial pitch-related encoding at the onset of articulation has already anticipated not only the length of the upcoming discourse but also the projected propositional completeness of the prosodic unit (i.e., the RIGHT effect). That is, the PU-initial prosodic encoding indicates that what has been active in the speaker's mind can be a propositionally complete unit, prima facie evidence for proposition-based pitch-related preplanning in spontaneous speech production.

The anticipatory relationship between PU-initial f0 height and PU-final propositional completeness provides additional evidence in favor of a "prosody-first" language production model (Keating & Shattuck-Hufnagel, 2002). Keating and Shattuck-Hufnagel (2002) have presented compelling evidence showing that many speech phenomena in word form encoding (e.g., phonetic effects of the prosodic unit edges) often require a pre-existing prosodic structure. These observations lead to their hypothesis of a two-stage prosodic model: speakers build a skeletal prosody based on the high-level information (i.e., syntax/semantics) and then restructure the prosodic details in word form encoding when the high-level information becomes more substantive (e.g., more lexicalized). According to Keating and Shattuck-Hufnagel, the initial prosodic structure may reply on the intended syntax to be included in the utterance, which in turn is restructured on the basis of more substantive lexical choices. This two-stage prosodic processing for speech production highlights the precedence of prosody in language processing. Our current analysis supports this prosody-first model. In our analysis of the cross-boundary f0 shifting, we have found that at the onset of the PU, speakers have created a skeletal prosodic structure (i.e., the initial f0 height), which indicates whether they are about to reach a proposition completion point within the PU. While it remains unclear to what extent the details of the intended proposition (e.g., the word forms included in the proposition) have been preplanned before articulation, it is clear that the initial prosodic structure has been first created on the basis of the general semantic structure of the PU (i.e., its propositional completeness).

On the other hand, the carried-over relationship between LEFT and PU-final f0 height is also interesting in the sense that utterance-final f0 shifting is rarely discussed in relation to the semantic property at the utterance-initial position. Utterances analyzed in the previous works on read speech are often structurally complete sentences. As utterance-initial is always sentence-initial in the controlled speech, it is less clear from these experimental studies how the utterance-initial semantics (e.g., starting an utterance from the middle of a sentence) may be connected to the utterance-final prosodic encodings. Our analysis has demonstrated that left-aligned PUs in general show higher PU-final f0 heights, indicating that speakers are aware of the fresh-start of a proposition, leaving a carried-over prosodic cue even at the terminal of the articulated PU. Following the assumption that the information verbalized in

one PU should have been active in the speaker's mind at the onset of the articulation (Chafe, 1994), the carried-over effect of LEFT on PU-final prosodic encoding may be analyzed as preliminary evidence for pitch-related preplanning of a proposition: the PU-final prosodic encoding has been pre-arranged on the basis of the semantics asserted in the PU.

In connection to this relationship between LEFT and *PU-final* f0 heights, our analysis has also shown that speakers consistently show higher *PU-initial* f0 heights if they initiate a proposition with the PU. We posit that any initiation of another proposition (i.e., DU) may introduce a pitch-reset in the articulation, rescaling the speaker's pitch baseline to a fresh start, thus consistently contributing to an overall higher f0 values in [+LEFT] PUs. We have demonstrated a positive correlation between LEFT and the PU-initial f0 heights, and a positive correlation between LEFT and the PU-final f0 heights, both yielding a consistent pitch-reset pattern (i.e., f0 up-shifting). Our hypothesis can be further supported by the existence of the relationship between LEFT and the PU-medial f0 heights. As a post-hoc analysis, we compared the f0 heights in the mid syllable of the PUs in relation to the PU-DU LEFT alignment (if the PU has an even number of syllables, we took the mean of the energy-based f0 values of the central two syllables). The PU-medial f0 heights were extracted by the same procedure used in our earlier analyses. Fig. 10 shows the mean values of the PU-initial, PU-medial, and PU-final f0 heights in relation to LEFT. The f0 heights of [+LEFT] PUs (i.e., all triangle-shaped dots in each panel of Fig. 10) are on average higher than their counterparts (i.e., [−LEFT] PUs), not only in terms of the PU-initial and PU-final f0 heights, but also the PU-medial f0 heights ($\beta = 0.51$, $t = 6.60$, df = 6887, $p < 0.01$, $r = 0.08$). In other words, the initiation of another proposition in speech production is prosodically distinctive in terms of the higher f0 values across the whole domain of the prosodic unit, suggesting that speakers may have renormalized their pitch register at the onset of each proposition. This carried-over f0-raising effect of LEFT on the f0 values across the entire PU domain (i.e., PU-initial, PU-medial and PU-final f0 heights) may be taken as strong evidence for speakers' sensitivity to the semantic unit, the proposition, in the prosodic prearrangement of speech production.

Our analysis so far builds upon the assumption that the semantic contents of the PU provide a clue to what has been active in speakers' mind before articulation, thus shedding light on the planning unit in speech production. Our results have suggested that the prosodic encodings of the PU significantly correlate with the initiation and termination of the propositions, indicating (1) speakers' sensitivity to this basic semantic unit, and (2) their capacity of preplanning a proposition before articulation. It is less clear whether speakers are capable of preplanning semantic units beyond a single proposition (e.g., multi-proposition discourse structure). The next section discusses this important issue by connecting our second analysis of complex PUs to our third question.

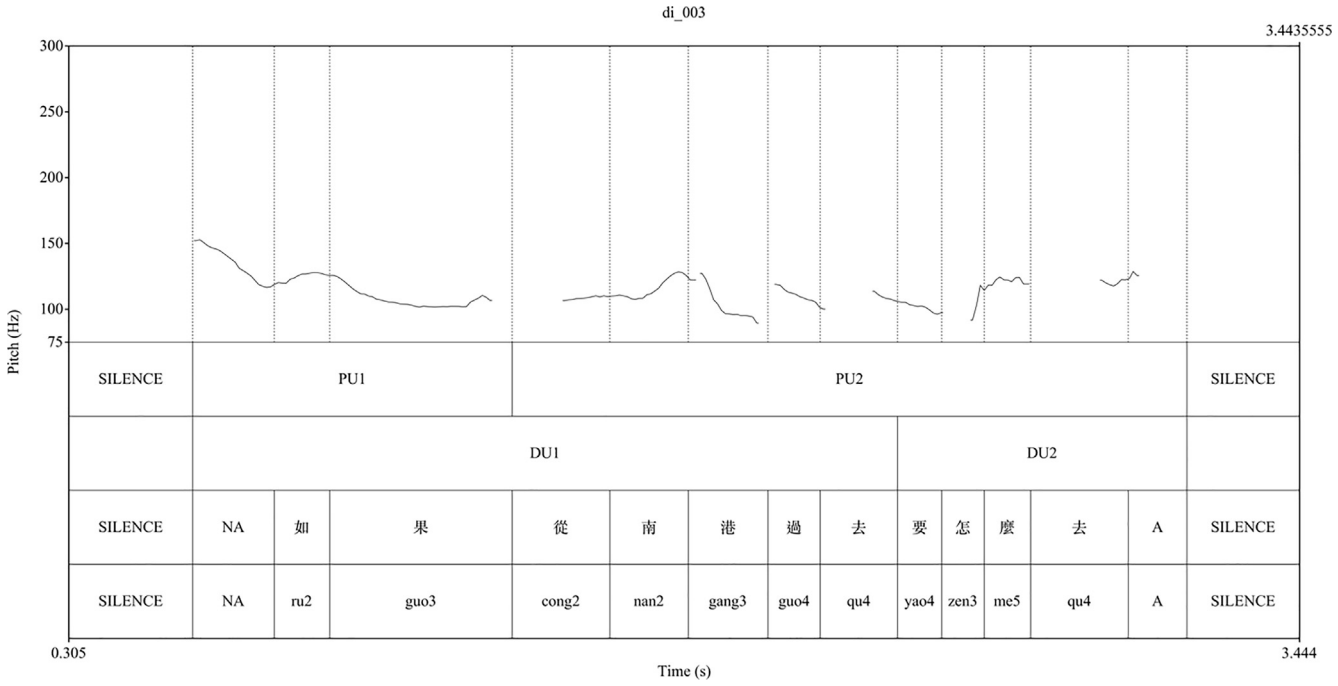
5.3. Advanced planning beyond one proposition

If speakers are capable of preplanning semantic units beyond one single proposition, we would expect a similar anticipatory effect of RIGHT on the initial f0 values for complex PUs. However, our analysis has demonstrated that speakers do not show significant variation in initial f0 values when the complex PU terminates a proposition, as compared to when it does not. The results suggest that speakers show few signs of preplanning semantic units beyond a single proposition. Whether they are [+RIGHT] or [−RIGHT], complex PUs show similar PU-initial f0 heights, suggesting that similar amount of semantic contents may have been active when speakers articulate these complex PUs. One of the shared semantic properties of all complex PUs is that they all span at least one proposition completion point. We can conclude that the lack of relationship between RIGHT and the initial f0 heights of complex PUs indicates speaker's limited capability of preplanning information beyond one single proposition[4].

Our negative answer to the third question above may look counter-intuitive to the assumption that the semantic contents of the PU showcase the amount of information that has been active in the speaker's mind before articulation (i.e., how much has been preplanned). If speakers combine more than one proposition within a single PU, one may wonder what contributes to the production of these more semantically-loaded PUs. As the prosodic encodings of these complex PUs provide little evidence of pitch-related preplanning, an alternative explanation is that the semantic contents asserted in these complex PUs may not be as informationally complex as we have expected. A closer examination of the semantic contents asserted in these complex PUs indicates that the first proposition asserted in the complex PUs tends to be a given event, state, or a reactive act in conversation, which has often been topical in the discourse context (i.e., information that has often been activated in the interactional context). These starting propositions often form the basis for the development of the subsequent proposition embedded within the same complex PU. For example, before the utterance in (8), the conversational participants were discussing different means of transportation to commute from Nang-Gang (i.e., a district in Taipei City) to their workplace. The second PU of (8), PU2, was a complex PU, spanning two propositions (i.e., DU1 and DU2 in (8)). It is clear that the second proposition, *yao4 zen3me5 guo4 qu4* 'how to get there', was the foregrounded information the speaker aimed to communicate, while the first proposition, *na4 ru2guo3 cong2 nan2gang3 guo4qu4* 'if (you) depart from Nang-Gang', served as a given information, i.e., a conditional scenario as a topic to which the question was relevant.

[4] Please note that the distinct prosodic structure observed at the juncture of supra-clausal discourse segments (e.g., topic structures in Nakajima and Allen (1993)) only suggests that speakers are "sensitive" to the discourse juncture, but it does not necessarily give us clear evidence of preplanning the whole discourse structure (i.e., the whole global prosody may not necessarily have been preplanned before the onset of the articulation.) This difference is similar to the two separate effects identified in our study: the LEFT effect on PU-initial f0 suggests speakers' sensitivity to DU boundaries; LEFT effect on PU-final f0 gives clearer evidence of proposition-based preplanning. To study the preplanning of global prosody at the higher discourse level, we may need more acoustic evidence of the latter type, which is yet to be found. It is difficult to identify a production unit whose contents can be argued to have been active in the speaker's mind before articulation, and at the same time whose prosodic domain can span (several) discourse topics.

(8) DI_003



Translation: "If (you) depart from Nang-Gang, / how did you get there?"

We argue that integrating more than one proposition within a single PU may be attributed to the interactional nature of conversational speech. The act of placing more than one proposition within a single PU is referred to as "rush-through" by Schegloff (1982, p. 76). Schegloff analyzed the rush-throughs in conversation as an interactional strategy of floor-maintaining so as to more effectively lead to the speaker's intended proposition or opinion. This rush-through nature of these complex PUs is further supported by its significantly shorter syllable durations (106 ms, SD = 30), compared to simple PUs (200 ms, SD = 90, t = -30.50, df = 4625.7, $p < 0.01$). The light-loaded semantic contents of the first proposition in the complex PUs (e.g., more given/topical) and the rush-through articulation of these complex PUs (e.g., faster speech-rates) indicate that speakers' prosodic arrangement in spontaneous speech may also show their sensitivity to the inter-subjective collaborative work of conversation.

The limited prosodic effect of PU-DU co-extensiveness on complex PUs found in our analysis is the connection between the PU-final f0 height and RIGHT (Fig. 9). This prosodic pattern differs from that found in previous studies on read speech or scripted speech, where the utterance-final f0 values were reported to be invariant to speakers (Liberman & Pierrehumbert, 1984) or utterance durations (Yuan & Liberman, 2014). Our finding indicates that speakers show significantly lower PU-final f0 heights when they produce a complex PU which terminates a proposition. When combining more than one proposition within a PU, speakers may not necessarily prosodically mark the initiation of the starting proposition, but they still consistently mark the termination of the finishing proposition. Therefore, our data suggest that the PU-final f0 heights may not be invariant to speakers but still connected to the completeness of the intended semantic contents. Furthermore, this may be attributed to the interactive nature of spontaneous speech (Asu et al., 2016; Swerts et al., 1996). Due to the immediate pressure from message planning and turn-taking, speakers in conversational discourses are often subject to disfluencies and interruptions, resulting in smaller chunks of production units. In our data, simple PUs do outnumber complex PUs by about six to one (Simple: 6921; Complex: 1118). Comparing the f0 declination in two speaking modes (i.e.,

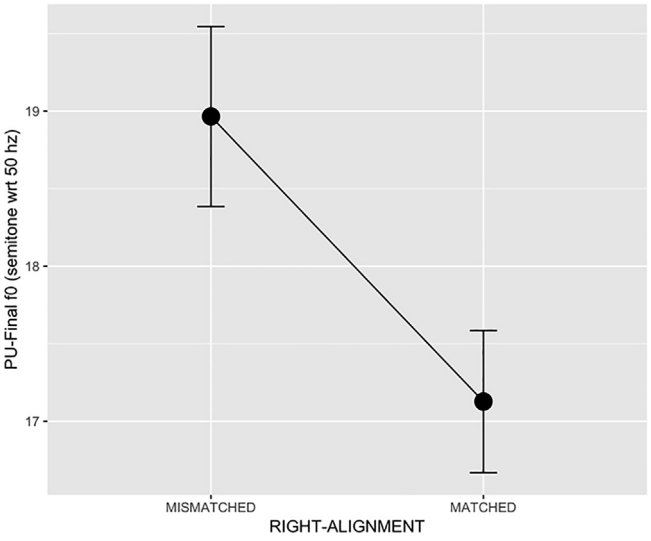


Fig. 9. Mean values of final f0 heights of complex PUs in relation to RIGHT with 95% confidence intervals. The bars represent the mean values of each sub-group, with confidence intervals delineated by the whiskers.

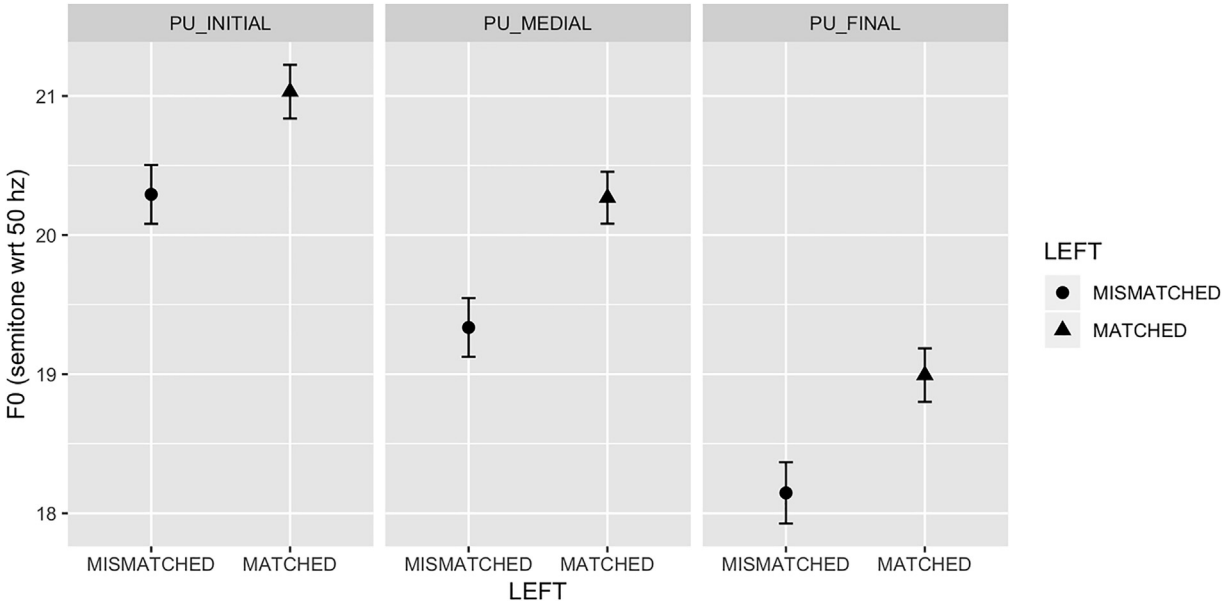


Fig. 10. Mean values of f0 heights at the PU-initial, PU-medial and PU-final positions in relation to LEFT (The bars represent the mean values of each sub-group, with 95% confidence intervals delineated by the whiskers). The values in the initial and final positions are the same as those reported in Section 4.

read and spontaneous speech) in Swedish, Swerts et al. (1996) consistently observe more drastic f0 changes in read speech, e.g., steeper f0 slopes, and stronger pitch resetting. According to their results, spontaneous speech tends to have less steep f0 slopes than read speech. Due to the prevalence of short utterances in spontaneous speech, it is likely that speakers more often end at a higher pitch register within their pitch range in these short utterances (e.g., simple PUs). It is thus reasonable to posit that speakers may be more likely to reach the low bound of their individual pitch range in uninterrupted longer utterances in spontaneous speech. Integrating multiple propositions into one prosodic contour, complex PUs are evidently longer than simple PUs (cf. Table 1 and 2), thus maximizing the likelihood of reaching speakers' low-bound pitch register at the propositional completion point. A comparison of the PU-final f0 heights of simple and complex PUs confirmed our prediction: the final f0 height of complex PUs is significantly lower than that of simple PUs by 0.72 semitone ($W = 3618300$, $p < 0.01$, $r = -0.04$)[5]. A post-hoc analysis also shows that the f0 range[6] of complex PUs is significantly larger than that of simple PUs by 2.72 semitones ($W = 4126100$, $p < 0.01$, $r = -0.24$). Therefore, the significant RIGHT effect on the PU-final f0 height of complex PUs reiterates the importance of semantic coherence of the PUs in the prosodic arrangement of speech production.

6. Conclusions

Our analyses have been built upon an implicit assumption that a PU is a functional speech unit in both speech production and message planning (Chafe, 1994). By examining the relationship between the prosodic encodings and the semantic contents of this speech segment in spontaneous speech, we are able to infer the possible advanced planning operative in speaker's incremental message arrangement. In this study, prosodic encodings included the PU-initial and PU-final f0 heights; semantic contents were defined on the basis of the co-extensiveness of the PU and a semantic unit, i.e., the proposition. Our analyses have identified several interesting relationships between f0 shifting and PU-DU alignment.

First, our results have shown that speakers consistently show sensitivity to the boundaries of the propositions (i.e., the initiation and termination of a proposition) in the cross-boundary prosodic encodings of the PUs. Second, our analyses have also identified two understudied long-distance relationships: (a) there is a PU-initial anticipatory f0 lowering connected to the PU-terminal semantic completeness (i.e., whether the PU terminates at a proposition completion point); (b) there is a PU-final carried-over f0 raising connected to the PU-initial semantic completeness (i.e., whether the PU initiates at a proposition onset point). These prosodic patterns suggest that speakers not only demonstrate sensitivity to the conceptual planning unit of a proposition, but also show signs of proposition-based preplanning in prosodic encodings. Finally, our analysis suggests that preplanning beyond a single proposition is not acoustically supported. As we have controlled a range of factors that may potentially influence the pitch variation (e.g., utterance length, tonal scaling, pitch declination, and topic structure), the prosodic effects resulting from the semantic configuration of the PUs should be independent, meaningful, and more likely to be a cross-speaker phenomenon, thus favoring the hypothesis of *hard* preplanning in speech production (Liberman & Pierrehumbert, 1984; Prieto et al., 2006).

This study has several limitations. One important issue that has not been addressed is the mapping of our PU to the prosodic units in other hierarchical prosodic frameworks. It is worth acknowledging that our prosodic units may not precisely replicate the segmentation of the *intermediate phrases* in the ToBI

[5] With the imbalanced sample sizes of simple and complex PUs, we conducted a non-parametric Wilcoxon rank-sum test.

[6] For each syllable in the PU, we extracted the f0 values at the energy max of all the syllables. The f0 range of each PU was computed based on the difference of the maximum and the minimum of these energy-controlled f0 values.

transcription framework. A more operationalized mapping scheme may be necessary to ensure the replicability of our present findings. Also, the strong relationships between the prosodic encodings and the semantic configurations of the PUs may suggest that some of our PUs may correspond to a larger prosodic juncture in the autosegmental framework (e.g., *intonational phrases*). Our model can be extended to analyze the higher-level discourse prosody by further examining the additional effects of the higher-level discourse junctures (e.g., themes) on the prosodic encodings of the PUs. The prosodic hierarchy of our spontaneous speech model can be acoustically built based on the layering effects of the configurations of the PUs at different semantic-pragmatic levels (e.g., PU co-extensiveness with these higher-level discourse structures).

Secondly, the extraction of the f0 values in our study may not be optimal. While energy-based method is helpful in identifying a reliable f0 point, it does not necessarily imply that the extracted f0 is representative of the prosodic contour of the PU. Many sophisticated stylization methods have been applied to the f0 extraction in linguistic analyses to ensure the validity of the f0 values under examination (de Ruiter, 2011; Grabe et al., 2007; Tseng et al., 2005; Wang & Xu, 2011). A critical comparison of f0 values computed by different methods may be needed for future research. Another issue concerns the pitch-shifting effects of the PU-DU alignment on the internal f0 values within the PU. The prosodic pattern identified in this study so far applies to the cross-boundary f0 encodings. Although limited support has also been provided for the PU-medial f0 values, it is not clear how the semantic configuration of the PU may contribute to the overall variability of the PU-internal f0 contours (e.g., the f0 slopes). Finally, this study has not considered the impact of prominence (e.g., contrastive focus) on prosody (Wang & Xu, 2011; Wang et al., 2018). There are a very limited number of utterances with strong pragmatic focuses in conversational discourse. With more data, it would be interesting to see how PU-DU alignment and focus may interact with each other to determine the cross-boundary f0 encodings.

The clause schema has been assumed in almost all syntactic theorizing (Thompson & Couper-Kuhlen, 2005). It is a basic linguistic structure to encode a basic semantic unit, a proposition. While the internal structuring and configuration of the clause has received central attention in many syntactic theories, few studies have aimed to seek acoustic support for the functional role of the clause in spontaneous speech production. As usage-based studies have provided systematic patterning in the co-construction of clauses by conversational participants in support of the importance of the clause schema (Thompson & Couper-Kuhlen, 2005), this study further contributes the evidence from prosodic encodings in spontaneous speech production. By highlighting the important relationships between the PU-initial and final pitch variation and the semantic configuration of these PUs, we hope to shed more light on the psycholinguistic importance of a proposition-based advanced planning in language production.

Acknowledgements

This work was supported by funds from the Taiwan Ministry of Science and Technology (MOST103-2410-H-018-006, granted to the first author, and MOST100-2911-I-001-504, granted to the second author).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2019.100912.

References

Asu, E. L., Lippus, P., Salveste, N., & Sahkai, H. (2016). F0 declination in spontaneous Estonian: Implications for pitch-related preplanning in speech production. In *Proceedings of Speech Prosody 2016* (pp. 1139–1142).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2006). The original ToBI system and the evolution of the ToBI framework. In S.-. A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). Oxford, England: Oxford University Press.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: University of Chicago Press.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics, 36*(4), 724–746.

Clark, H. H., & Tree, J. E. F. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition, 84*, 73–111. https://doi.org/10.1016/s0010-0277(02)00017-3.

Collier, R. (1975). Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America, 58*, 249–255. https://doi.org/10.1121/1.380654.

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics, 33*, 839–882. https://doi.org/10.1515/ling.1995.33.5.839.

de Ruiter, L. E. (2011). Polynomial modeling of child and adult intonation in German spontaneous speech. *Language and Speech, 54*, 199–223. https://doi.org/10.1177/0023830910397495.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes, 22*, 1151–1177. https://doi.org/10.1080/01690960701461293.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57–84. https://doi.org/10.1006/jmla.2001.2797.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, England: Sage.

Fon, J., Johnson, K., & Chen, S. (2011). Durational patterning at syntactic and discourse boundaries in Mandarin spontaneous speech. *Language and Speech, 54*(1), 5–32.

Ford, C. E. (1993). *Grammar in interaction: Adverbial clauses in American English conversations*. Cambridge, England: Cambridge University Press.

Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics, 41*, 29–47. https://doi.org/10.1016/j.wocn.2012.08.007.

Givón, T. (1984). *Syntax: A functional and typological introduction* (Vol. 1) Amsterdam, Netherlands: John Benjamins.

Givón, T. (1993). *English grammar: A function-based introduction* (Vol. 2) Amsterdam, Netherlands: John Benjamins.

Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech, 50*, 281–310. https://doi.org/10.1177/00238309070500030101.

Huang, S., & Chui, K. (1997). Is Chinese a pragmatic order language? *Chinese Languages and Linguistics, 4*, 51–79.

Iwasaki, S., & Tao, H. (1993). A comparative study of the structure of the intonation unit in English, Japanese, and Mandarin Chinese. Paper presented at the The 67th Annual Meeting of the Linguistic Society of America, Los Angeles, CA.

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

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge, England: Cambridge University Press.

Ladd, D. R., & Johnson, C. (1987). Metrical factors in the scaling of sentence-initial accent peaks. *Phonetica, 44*, 238–245. https://doi.org/10.1159/000261801.

Lee, E.-K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition, 129*, 544–562. https://doi.org/10.1016/j.cognition.2013.08.007.

Lehmann, C. (1988). Towards a typology of clause linkage. In J. Haiman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 181–225). Amsterdam, Netherlands: John Benjamins.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: MIT Press.

Liu, Y.-F., & Tseng, S.-C. (2009). Linguistic patterns detected through a prosodic segmentation in spontaneous Taiwan Mandarin speech. In S.-. C. Tseng (Ed.), *Linguistic patterns in spontaneous speech* (pp. 147–166). Taipei, Taiwan: Institute of Linguistics, Academia Sinica.

Liu, Y.-F., Tseng, S.-C., Jang, J. S. R., & Chen, A. C.-H. (2010). Coping imbalanced prosodic unit boundary detection with linguistically-motivated prosodic features. *Proceedings of INTERSPEECH, 2010*, 1417–1420.

Maeda, S. (1976). *A characterization of American English intonation* (Doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology.

Matsumoto, K. (2000). Intonation units, clauses and preferred argument structure in conversational Japanese. *Language Sciences, 22*, 63–86. https://doi.org/10.1016/S0388-0001(99)00004-2.

Mo, Y., Cole, J., & Lee, E.-K. (2008). Naïve listeners' prominence and boundary perception. In *Proceedings of Proceedings of Speech Prosody, Campinas, Brazil* (pp. 735–738).

Nakajima, S. Y., & Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica, 50*, 197–210. https://doi.org/10.1159/000261940.

Ono, T., & Thompson, S. A. (1995). What can conversation tell us about syntax? In P. W. Davis (Ed.), *Alternative linguistics: Descriptive and theoretical modes* (pp. 213–271). Amsterdam, Netherlands: John Benjamins.

Ono, T., & Thompson, S. A. (1996). Interaction and syntax in the structure of conversational discourse: Collaboration, overlap, and syntactic dissociation. In E. H. Hovy & D. R. Scott (Eds.), *Computational and conversational discourse* (pp. 67–96). Berlin, Germany: Springer-Verlag.

Park, J. S.-Y. (2002). Cognitive and interactional motivations for the intonation unit. *Studies in Language, 26*, 637–680. https://doi.org/10.1075/sl.26.3.07par.

Prévot, L., Tseng, S.-C., Peshkov, K., & Chen, A. C.-H. (2015). Processing units in conversation: A comparative study of French and Mandarin data. *Language and Linguistics, 16*, 69–92. https://doi.org/10.1177/1606822X14556605.

Prieto, P., D'Imperio, M., Elordieta, G., Frota, S., & Vigário, M. (2006). Evidence for soft preplanning in total production: Initial scaling in Romance. *Proceedings of Speech Prosody, 2006*, 803–806.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Washington, DC: Georgetown University Press.

Scholz, F., & Chen, Y. (2014). Sentence planning and f0 scaling in Wenzhou Chinese. *Journal of Phonetics, 47*, 81–91. https://doi.org/10.1016/j.wocn.2014.09.004.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3*, 371–405.

Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation, analysis, modeling and technology* (pp. 243–268). Dordrecht, Netherlands: Kluwer Academic Publishers.

Smith, C. L. (2004). Topic transition and durational prosody in reading aloud: Production and modeling. *Speech Communication, 42*, 247–270.

Swerts, M. (1997). Prosodic features at discourse boundaries of different length. *Journal of the Acoustical Society of America, 101*, 514–521. https://doi.org/10.1121/1.418114.

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. In *Proceedings of the fourth international conference on spoken language processing* (pp. 1501–1504).

Syrdal, A. K., & McGory, J. (2000). Inter-transcriber reliability of ToBI prosodic labeling. In *Proceedings of 6th International Conference on Spoken Language Processing* (pp. 235–238).

Tao, H. (1996). *Units in Mandarin conversation: Prosody, discourse, and grammar*. Amsterdam, Netherlands: John Benjamins.

't Hart, J. (1979). Exploration in automatic stylization of F0 curves. *IPO Annual Progress Report, 14*, 61–65.

Thompson, S. A., & Couper-Kuhlen, E. (2005). The clause as a locus of grammar and interaction. *Discourse Studies, 7*, 481–506. https://doi.org/10.1177/1461445605054403.

Thompson, S. A., & Hopper, P. J. (2001). Transitivity, clause structure, and argument structure: Evidence from conversation. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 27–60). Amsterdam, Netherlands: John Benjamins.

Thorsen, N. G. (1980). A study of the perception of sentence intonation—Evidence from Danish. *Journal of the Acoustical Society of America, 67*, 1014–1030. https://doi.org/10.1121/1.384069.

Thorson, J. (2007). The scaling of utterance-initial pitch peaks in Puerto Rican Spanish: Evidence for tonal preplanning. *University of Rochester Working Papers in the Language Sciences, 3*, 91–97.

Tseng, C.-Y., Pin, S.-H., Lee, Y., Wang, H.-M., & Chen, Y.-C. (2005). Fluent speech prosdy: Framework and modeling. *Speech Communication, 46*, 284–309.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 423–440. https://doi.org/10.1037/a0018619.

Wang, Y.-F. (2002). The preferred information sequences of adverbial linking in Mandarin Chinese discourse. *Text, 22*, 141–172. https://doi.org/10.1515/text.2002.002.

Wang, B., & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics, 39*, 595–611. https://doi.org/10.1016/j.wocn.2011.03.006.

Wang, B., Xu, Y., & Ding, Q. (2018). Interactive prosodic marking of focus, boundary and newness in Mandarin. *Phonetica, 75*(1), 24–56. https://doi.org/10.1159/000453082.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication, 50*, 81–94. https://doi.org/10.1016/j.specom.2007.06.002.

Whalen, D. H., & Kinsella-Shaw, J. M. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica, 54*, 138–152.

Wichmann, A. (2000). *Intonation in text and discourse: Beginnings, middles, and ends*. London, England: Longman.

Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin broadcast news speech. *Speech Communication, 65*, 67–74. https://doi.org/10.1016/j.specom.2014.06.001.

Zellers, M., Post, B., & D'Imperio, M. (2009). Modeling the intonation of topic structure: Two approaches. *Proceedings of Tenth Annual Conference of the International Speech and Communication Association*, 2463–2466.