Computational Linguistics and Chinese Language Processing
Vol. 27, No. 1, June 2022, pp. 1-26

The Uniqueness in Speech: Prosodic Highlights-prompted Information Content Projection in Continuous Speech Speech

Helen Kai-yun Chen[*] and Chiu-yu Tseng[+]

Abstract

Recently, it has been identified that perceived prosodic highlights in continuous speech can function alternatively as the projector of key/focal information allocation. This view provides a novel interpretation to the long-held claim that prominence is used predominantly to mark key information and alludes to the significance of information content planning prompted by perceived prominence. Exploring further information content planning and allocation prompted by prosodic highlights, this study focused on the information content planning unit—"projector" (PJR) and its respective "projection" (PJN) (henceforth PJR-PJN units)—across four diverse Mandarin speech genres. Using the corpus linguistic approach and quantitative analyses, the current study conducted acoustic correlates analyses of F0 realization and pause duration, also the calculation of emphasis-attributed weighting scores based on emphasis levels consistently annotated in the speech data. While the main goal of the study was to profile consistent acoustic realizations across the PJR-PJN units, further confirmation of the patterned deployment of information content in continuous speech was verified. Ultimately, the current results foregrounded the underlying mechanism for information prosody and features unique to speech.

Keywords: Continuous Speech and Discourse, Spoken Corpora and Annotations, Information Content Planning and Allocation, Prosodic Highlights-prompted Projection, Emphasis-attributed Weighting Scores, Information-attributed Weighting Scores.

[*] Language Center, National Central University
E-mail: helenkychen@gmail.com
[+] Institute of Linguistics, Academia Sinica, Taipei, Taiwan
E-mail: cytling@sinica.edu.tw

1. Introduction

The current study focused on information content planning and allocation, which is initiated by and associated with perceived prosodic highlights in continuous speech. In speech and discourse, one of the keys to communication is in how interlocutors plan "ahead of time" the allocation of focal information in speech production and perception: for speakers, this mostly concerns how they distinctively and effectively allocate key/focal information to facilitate comprehension. On the other hand, listeners are oriented to salient cues in prosodic manifestations, including the ups and downs of the melody, the pace of the speech output, and other perceptually distinctive cues to help pinpoint the most crucial information in the speech flow. It is our belief that to plan and identify information content in the speech context, perceivable prosodic saliency, particularly prosodic highlights, plays a crucial role. For this reason, we chose to concentrate on prominence[1] in speech and how it is incorporated to project information content allocation in this study.

In order to examine perceived prosodic highlights and their roles in information content projection functions within the speech context, we adopted an unconventional approach to discourse prosody. Specifically, instead of incorporating traditional methods to treat prosodic manifestations with certain predefined phonetic or syntactic units and examine only their face value, we took a holistic, top-down approach and paid attention to the role of upper-level discourse associations. To account for prosodic variations in continuous speech, we adopted the recently proposed hierarchical prosodic phrase grouping (HPG) framework as described in Tseng *et al*. (2005) and Tseng (2010). The main justification for resorting to such a framework was that it could better accommodate and account for the unique features in speech data: as a continuous flow of perceivable signals, the composition of discourse prosody can go beyond the mere concatenations of lower-level linguistic units. By incorporating the HPG framework, our goal was to capture features belonging to speech inclusive of prosody for information content planning at the upper level of discourse realization, as opposed to linguistic prosody that is constrained by lower-level units that are grammar based.

This article will report the follow-up acoustic analyses from a recent study that focused on perceived prosodic highlights-projected information content allocation in the speech context (cf. Chen & Tseng, 2021). Through examining diverse speech data that was annotated for the same discourse-level prosody in hierarchical relationships (i.e. using the HPG framework) and tagged for consistently perceived levels of prominence, it was demonstrated that perceived prosodic highlights involve the indexing function for key/focal information allocation, and thus project

---

[1] In this study we use "prosodic highlights" and "prominence" interchangeably in referring to the same concept of distinctively perceived segments in continuous speech signals. Note that, in this case, prosodic highlights and prominence used here are not the same as word-level stress and prominence.

information content planning at higher discourse-level prosody (Chen & Tseng, 2021). Based on the same set of speech data and annotation systems, this extended study will report further results of the acoustic analyses of prominence-prompted information content projection units (cf. Chen & Tseng, 2021). Ultimately, the goal was to foreground prominence-correlated information content planning through discourse-level prosody realizations and demonstrate how patterns and features were eventually be derived from "speaking" (i.e., the "parole" in de Saussure, 1966). In the end, we were able to derive prosody specifically for information content planning from speech and discourse, which went beyond seemingly random linguistic prosody in its surface values and realizations.

1.1 Discourse Prosody and Information Content Planning

Prosody, a unique feature of speech, has posed a major challenge in relevant studies concentrating on discourse perception and production. Given that discourse production can go beyond more than just a sequence of sentences (cf. Swerts & Geluykens, 1994) and that continuous speech happens in a highly spontaneous context and is unplanned, how to capture speech prosody in its highly variant realizations is crucial. In most cases, the prosodic realizations of speech are considered and processed by sound units that are segmented by their meta forms or, at most, from units that are syntactically predefined. This follows earlier studies and the long tradition of examining continuous speech signals through syntactic prosody (cf. Lehiste, 1970; see also Cutler *et al*., 1997 for a review on prosody of spoken languages).

With regard to the allocation of (focal) information by the prosodic realizations in discourse, oftentimes the discussions in relevant literature have focused on focal/new information is directly marked by prominence, for example, pitch accent (such as that conventionally annotated as H*; Silverman *et al*., 1992; see also Halliday, 1967; Pierrehumbert & Hirschberg, 1990; Watson *et al*., 2008). However, to associate a high pitch accent with focal information, the pitch accents are aligned with word-level stresses in most cases. In other words, the corresponding unit for prominence realizations are held at the lexical level. On the other hand, Swerts and Geluykens (1994) examined the role of prosody in the structuring of information (i.e., the topic flow and topic changes) and investigated prosodic variables including intonation and pause. In their experiment, the relative pitch height and pause length were associated with information flow markers (Swerts & Geluykens, 1994). Although the speech used in their study was spontaneous, the elicitation of the speech data was controlled for the design and purpose of the experiment.

The recently proposed HPG framework for discourse prosody i n continuous speech by Tseng et al. (2005) was suggested as an alternative approach (see the additional explanations in Tseng & Su, 2008; Tseng, 2010, 2013). The main strength of the HPG framework is that it is not text-bounded, nor is the relationship between discourse-prosodic units (DPUs) predetermined grammatically. Instead, the target is how continuous speech signals can be processed from a global viewpoint. According to Tseng (2013), the merit of adopting the HPG framework is to purposely distance it from the possible connotations associated with lower levels of linguistic information, while foregrounding the contribution to higher discourse-level prosody, which also includes discourse-paragraph associations and information content planning. The HPG framework has been adopted in several recent studies focusing on the prosodic features of higher discourse levels in various continuous speech genres (cf. Tseng & Su, 2012, Chen *et al*., 2016).

1.2 Prosodic Highlight Prompted Information Content Projection

In an exploration of perceived prosodic highlights as an index of information content planning and projection, Chen *et al.* (2016) and Chen and Tseng (2021) reported the analyses of perceived prominence that was consistently annotated across continuous speech. Based on the data from four diverse speech genres that were preprocessed and annotated using HPG, the studies established two information-content indices prompted by prosodic highlights. It was demonstrated that far more tokens of prosodic words with perceivable prominence tags were incorporated into speech to forecast, ahead of time, speech planning and to "project" the allocation of focal information. For instance, in the following examples, the emphasis marked *zuizao de yipian* 'the earliest entry of' projects the following noun phrase (NP) *wenzhang* 'the article' in (1), whereas in (2), the perceived prominence-indexed *tixing nin* 'to remind you' projects the following clause, which contains key information such as *ziwaixian* 'UV rate' and *guoliangji* 'extreme level'[2]:

(1)

L: 那也是/最早的一篇/文章.

Na	yeshi	/zuizao	de	yipian/	wenzhang.
that	also.COP	earliest	DE	a.CL	article

'That is also the earliest entry of the article.' (Chen & Tseng, 2021: 197)

[2] The concept of prosodic highlights-prompted information projection is shown in the following example from Goodwin (1996: 372), in that the enhanced intonation from a specific part of the discourse is interpreted as projecting focal information:

(i) Nancy: Jeff made en asparagus pie
It was s: : so: goo:d.
Tasha: I love it.

According to Goodwin (1996), the prominently pronounced adverb "so" can be interpreted as a projector of the next bit of interaction, as it serves as a kind of prompt for the following adjective "good." In this example the adjective "good" serves as the main predicate, providing focal and possibly new information.

(2)

S: 特別/提醒您/目前白天紫外線都是過量級.

tebie	/tixing nin/	muqian	/baitian/	ziwaixian
especially	remind 2SG	at.the.moment	day.time	UV rate
doushi	/guoliangji/. (WB)			
all.COP	extreme level			

'Please be reminded especially that at the moment the UV rate during the daytime has reached the extreme level.'

As suggested in Chen and Tseng (2021), in addition to directly marking new/focal information, it has been found that prosodic highlights in continuous speech can be incorporated to index "specific parts of discourse" (e.g., Falk, 2014:8), and thus function to orient listeners' attention to focal information allocation in speech production. The advantage of incorporating such information projection prompted by prominence, according to Chen and Tseng (2021), is to help eliminate potential prediction errors in speech perception (i.e., Clark, 2013; Auer, 2015; Dilley, 2016) and hence facilitate successful communication.

With the assumption that the allocation of prosodic highlights directly reflects the deployment of information content in speech, Chen and Tseng (2021) conducted relevant analyses concentrating on the information content planning unit—"projector" (PJR) and its respective "projection" (PJN) (henceforth **PJR-PJN** units)—prompted by perceived prominence. It was demonstrated that while planning for prosodic highlights-prompted projection, speakers in general were oriented toward a "heavy-to-light" information-attributed weighting scores distributed across the PJR-PJN units (Chen & Tseng, 2021). The results showed that the prosodic highlights-correlated information content planning was realized in a fixed pattern. The main contribution of the study was clarifying that prosody-attributed information content planning in continuous speech takes place at a specific discourse-prosodic level based on the HPG framework (Chen & Tseng, 2021).

1.3 The Current Study: A Preview

The current study was a sequel to the findings on prosodic highlights-initiated PJR-PJN units reported in Chen and Tseng (2021). Following from the assumption of the direct correlation between perceived prosodic highlight distribution and information content allocation, this article will report further analyses based on the PJR-PJN units consistently annotated in data from continuous speech. With the same set of speech data annotated by the corresponding discourse-prosodic levels based on the HPG framework, and according to the same perceived prominence-level annotations, this extended study extracted acoustic measurements from the

PJR-PJN units. In addition, the results provided further validation of the calculation of information-attributed weighting scores across the PJR-PJN unit (cf. Chen & Tseng 2021). Based on the results, the findings suggested that there was an extended substantiation of prosody-attributed information content planning, especially at the higher level of DPUs in continuous speech.

In the present analyses, we chose to concentrate particularly on acoustic correlates across the PJR-PJN units, including a) F0 realization[3] and b) pause duration, among the possible acoustic correlates.[4] As for the validation of the emphasis-attributed weighting scores distribution across the PJR-PJN units, further statistical analyses were carried out. The main difference between Chen and Tseng's (2021) previous study and the results reported in this paper is mainly in that we included the PJR-PJN units in the projection trajectories of various sizes. Although the sizes of information content planning and projection differed from case to case, our analyses still demonstrated identifiable patterns of acoustic realizations and distributions of information-attributed weighting score. As will be shown, the results pinpointed information content planning in correlation with advance prosody prompting, not only in a patterned F0 contour but also in longer pause and heavier information loading that were required at the initiation of the PJR-PJN units. We believe that the results are significant in demonstrating the role of advance prosodic prompting in information projection. The current results will shed light on information content planning in online speech production, and the establishment of information prosody that is unique in speech.

2. Speech Data and Annotations

2.1 Speech Data

Continuous speech data in Taiwanese Mandarin from four diverse genres were incorporated for the purpose of the present analyses. Of the four speech genres, two were spontaneous speech and the other two were read speech . One of the two spontaneous speech genre was a university classroom lecture (henceforth SpnL), taught and delivered by a male professor (i.e., Tseng *et*

[3] Some of the preliminary observations regarding F0 realization throughout the PJR-PJN units have been reported earlier in Chen *et al*. (2016).

[4] Although we chose to focus on the acoustic cues of intonation (F0) and pause, this does not mean that other prosodic cues are irrelevant. The justification for concentrating on only these two acoustic correlates was mainly that each PJR-PJN unit identified was an independent case and had different length (ranging from one prosodic phrase to three prosodic phrases; see the results in Section 4.1.). Since each PJR-PJN was an independent unit, other cues (such as final lengthening at the end of PJR-PJN units) were not the focus of the current analysis. Moreover, with regard to amplitude, given that the sizes of the PJR-PJN units differed case by case, we assumed that it would be difficult to generate consistent amplitude results from the tokens identified.

al., 2008), whereas the other speech genre was a spontaneous informal interaction (SpnC) taken from a corpus of face-to-face interaction in Taiwanese Mandarin (Chen *et al*., 2012). The read speech, on the other hand, included data from the tasks of prose reading (CNA) and weather broadcast simulations (WB), both of which were culled from the Sinica COSPRO corpus (Tseng *et al*., 2003; Tseng *et al*., 2005). Note that we incorporated speech data from different genres for the purpose of comparing features that belong to read speech and spontaneous speech/discourse. Table 1 summarizes the total duration of the data from each speech genre, with additional information on the equivalent number of syllables:

Table 1. Summary of total time and number of syllables in the data from four speech genres

Corpora/ Genres	Total Time (min.)	Total Number of Syllables
SpnL	145	33,306
SpnC	54	10,756
CNA	50	22,988
WB	28	14,083

Although the total duration of each genre differed and was not balanced across speech genres, we ensured that there were ample acoustic features present in the target annotated tokens, especially for the purpose of the current acoustic analyses.[5]

2.2 Data Preprocessing and Annotations

First, the selected speech data underwent automatic preprocessing of force alignments using the HTK Toolkit. The output was followed-up by manual spot-checking and then adjusted by the trained transcribers. The next step of data preprocessing involved the annotations of prosody-related information in independent layers. These tasks were carried out by experienced annotators[6] who tagged the data for the following information: (i) level of DPUs; (ii) level of perception-based prosodic highlights; and (iii) information content planning PJR-PJN units (cf. Chen & Tseng, 2021).

[5] As the current speech data were taken from six different speakers (three male and three female speakers) in total, in the following acoustic analyses we normalized the measurements in order to avoid the problem of speaker idiosyncrasy.

[6] The "experienced taggers" (and trained transcribers) in this study were annotators who had undergone preliminary training for at least three to six months. After the training, these annotators continued working with the same data for at least one year. When working on each annotation task, they had to reach a minimum level of consistency rate from the initial training of a certain task before continuing on (see also the sections on annotations to follow).

2.2.1 Annotation Scheme for Discourse-Prosodic Units (DPU)

We first annotated all the speech data for prosody-based breaks and boundaries following the framework of the HPG framework, according to which, five DPU levels with hierarchical relationship were distinguished, and these were marked B1 through B5, corresponding respectively to syllable (SYL), prosodic word (PW), prosodic phrase (PPh), breath group (BG) and multiple phrase speech paragraph (PG). Beyond the lexicon-based and grammar-correlated PW and PPh levels in the HPG framework, there were two more higher-level units and one was at the BG level, which was defined as a physio-linguistic unit constrained by a change of breath while speaking continuously (cf. Lieberman, 1967; Tseng, 2010). As for the highest-level PG, it was mostly discourse based and was predominantly defined by major topic changes. By default, the boundary breaks, prosodic units, and their relationships within the hierarchy were accounted for as follows: SYL/B1 < PW/B2 < PPh/B3 < BG/B4 < PG/B5 (cf. Tseng, 2010).

In the current study, the annotation of the DPUs was carried out by marking boundary breaks in hierarchical relationships, instead of predetermined by any type of lexical or syntactic relationship. To ensure that the annotations reached a certain level of consistency, the participating annotators[7] had to reach at least an 80% consistency rate during the initial training to continue the task. During and after the annotation process, both intra- and inter-annotator consistency were constantly checked to ensure the agreement was reached and accuracy maintained at a level of least 95% agreement among the annotators.

2.2.2 Annotation Scheme for Perceived Prosodic Highlights

In a separate layer, all the speech data were additionally annotated for perception-based emphasis and non-emphasis tokens (ETs/non-ETs). Following the definition described in Tseng *et al*. (2011) and Tseng (2013), this annotation scheme for perceived prominence was marked by strength levels, from reduction to the most emphasized, and divided into four relative degrees, defined respectively as follows:

- E0 -- reduced pitch, lower volume, and/or contracted segments
- E1 -- normal pitch, normal volume and clearly produced segments
- E2 -- raised pitch, louder volume and irrespective of the speaker's tone of voice
- E3 -- higher raised pitch, louder volume, and with a change in the speaker's tone of voice

The rationale behind adopting this scheme for annotating prominences was based on the belief that only a limited number of contrastive degrees can be consistently perceived while processing

---

[7] At least ten annotators participated in the task of annotating the DPUs in the current speech data.

continuous speech.[8] In the annotation of perception-based prominence, the trained annotators simply tagged the speech data in a string that consisted of ETs (i.e., E2 and E3) and non-ETs (i.e., E0 and E1).[9] Among the four speech genres, only spontaneous speech (i.e., SpnL and SpnC) was tagged for the additional level of reduction (E0), as we assumed that speakers rarely carried out reduction in reading tasks.

For the annotations of perceived prominence, at least eight annotators[10] were involved in the task. In order to carry out the reliability check, we first assigned one to two "reliable" annotators who were more sensitive to prominence-level distinctions. Their tagging results were considered the "gold standard." As for the rest of the annotators, they had to reach at least an 80% agreement level compared with the reliable annotators' tagging results, to continue with the task. For the final annotation, the accuracy level had to reach at least 95% of agreement among the annotators.

2.2.3 Identification of Information Content Planning Units

The information content planning PJR-PJN units were annotated via a separate task in yet another independent layer. First, we started with the identification of the prosodic highlights-prompted PJR. The identification of the PJR index was based on the ETs (i.e., E2 and E3) that had already been annotated in the current data. Each E2 and E3 were broken up by a PW unit. Following the principles of categorizing prominence-prompted information content planning proposed by Chen and Tseng (2021), the PJR units were instances in which the speakers incorporated emphasis in a particular PW unit to head-up the deployment of key information in speech planning. In the following examples, the speech strings in between the slashes are the PW units with an E2 prominence level tagged under the current annotation scheme. In (3), the PW unit *bingbu zhidao* 'not (really) know' is categorized as a PJR unit. Moreover, in (4), which

---

[8] Since the annotation of prominence levels was mainly perception based, the annotators were not given specific instructions to correspond a prominence level to any absolute acoustic value (i.e., they were never given the instruction that an E2 tag would equal a fixed range of F0 measurements in number). We wanted the annotation of prominence to closely and faithfully reflect the perception of the speech signals. Moreover, given that the level of contrast degree was limited, in general the annotators working with this annotation scheme did not have much difficulty in deciding, for example, a two-way distinction between E1 and E2.

[9] Since in Mandarin the language does not actually carry pitch accent at the word level, our annotation scheme was distinguished from the model of prosody-related prominence proposed by Kohler (1997) and the framework discussed in Baumann *et al.* (2016), in that the current tagging scheme for prominence level was not lexically based nor syntactically predefined.

[10] Some of the annotators who worked on the DPU annotation also worked on the task for prominence-level annotation. However, those annotators did not work on the two tasks simultaneously. In other words, they trained for the two tasks and worked on each separately.

is repeated from (2), the PW unit *tixing nin* ‘to remind you’ is identified as a PJR unit.

(3)

L: 中文是，中文的文字是一堆字. 那麼你/並不知道/哪裏是一個詞.

Zhongwen	shi	zhongwen	de	wenzi	shi	yidui	zi.
Chinese	COP	Chinese	DE	text	COP	a.CL	character
Name	ni	/bingbu	zhidao/	nali	shi	yige	ci. (SpnL)
then	2SG	not	know	where	COP	a.CL	lexical.word

‘(As for) Chinese, the texts in Chinese are presented as a bunch of characters. Thus, you don’t really know which part equals a word.’

(4)

S: 特別/提醒您/目前白天紫外線都是過量級.

tebie	/tixing nin/	muqian	/baitian/	ziwaixian
especially	remind 2SG	at.the.moment	day.time	UV rate
doushi	/guoliangji/. (WB)			
all.COP	extreme level			

‘Please be reminded especially that at the moment the UV rate during the daytime has reached the extreme level.’

Following the identification of PJR units, we turned to the delineation of the respective PJN units, which were identified as anticipated syntactic/semantic/prosodic completion whose trajectory covered at least a piece of focal information (cf. Chen & Tseng, 2021). As suggested by the discussion of prosodic-highlights prompted projection in Chen and Tseng (2021), the projection trajectory of each PJR unit was realized by a different size, from the local to the global. The current study adopted the similar term “projector-projection” (PJR-PJN) coined in Chen and Tseng (2021: 197) to refer to the prosodic highlights-indexed PJR unit, which was followed immediately by the respective PJN unit. Two additional examples below illustrate the PJR-PJN units by the proposed definition:

(5)

L: 那也是/最早的一篇/文章.

Na	yeshi	/zuizao	de	yipian/	wenzhang. (SpnL)
that	also.COP	earliest	DE	a.CL	article

‘That is also the earliest entry of the article.’ (Chen & Tseng, 2021: 197)

(6)

L: /為什麼直-/直接比對/字也有/困難?因為我們的/詞的/結構是非常 flexible 的.

/Weisheme	zhi-/	zhijie bidui		/zi	ye	you/	kunnan?	
why	di-	direct match		word	also	have	difficulty	
Yinwei	women	de	/ci			de/	jiegou	shi
because	1PL	DE	lexical word			DE	structure	COP
feichang	flexible	de.		(SpnL)				
quite	flexible	DE						

'Why is there difficulty in matching words directly? (It is) because the composition of the word structure is quite flexible.' (Chen & Tseng, 2021: 197)

In (5), which is repeated from (1), the prosodic highlights-prompted PW unit *zuizaode yipian* 'the earliest entry' as a PJR unit has a respective projection trajectory ending with the NP *wenzhang* 'article', as explained earlier. Turning to (6), the prosodic highlights-indexed PW unit *weishenme* 'why' is also categorized also as a PJR unit, and the prosodic highlights-prompted PJR unit entails a projection, with its trajectory extending to the end of the following clause which is initiated by the connective *yinwei* 'because'. According to the definition by Chen and Tseng (2021), the PJN unit's trajectory in (6) covers at least one piece of focal information (including examples such as *zi* 'character' and *ci* 'lexicon' and the foreign word 'flexible'). Hence both (5) and (6) demonstrate that the PJR-PJN units are of different sizes, from the immediate local projection (as shown in [5]) to the more global one (as shown in [6]). Figure 1 presents an illustration of the annotation for (6) taken from Chen and Tseng (2021), inclusive of the DPU levels and prosodic highlight annotations:

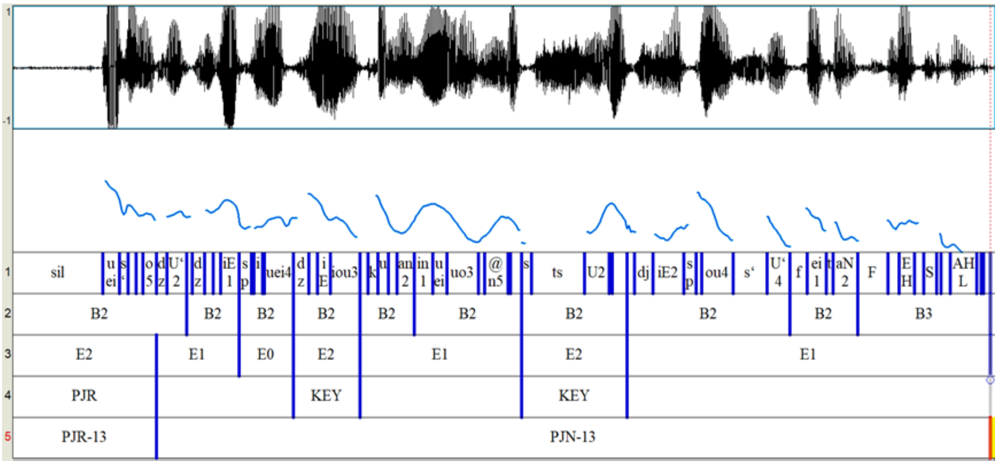

Figure 1. Illustration of the annotation schemes for the DPU levels (in the second layer beneath the spectrogram), prominence levels (in the third layer beneath the spectrogram), and PJR-PJN units using PRAAT (Boersma & Weenink, 2015)

Finally, in terms of annotation consistency rate checking, the identification of prosodic highlights-prompted PJR units was carried out by at least five annotators.[11] For the categorization of the PJR tags, the results had to reach 80% agreement among the annotators, and then the PJN trajectory of each PJR instance was demarcated. The annotators checked and discussed each case separately until a final consensus on the trajectory range of each PJN unit was reached.

3. Methodology

3.1 Acoustic Features Extraction

The methodology incorporated in the current analyses involved mainly the extraction of acoustic features, including F0 and pause duration, among other acoustic correlates. First, F0 values (in semitone) across the PJR-PJN units were automatically extracted using the software program PRAAT (© Boersma & Weenink, 2015). In order to facilitate further comparison and eliminate factors from speakers' discrepancy and idiosyncrasy, all the extracted F0 values were subjected to Z-score normalization. Then the next step was to calculate the average F0 values derived from the sampling points, including (i) the PJR at the initiation of the prosodic highlights-prompted projection; (ii) the ending PW at the completion of the PJN; and (iii) the PW units at the pre-/post-PPh boundaries, depending on the trajectory size of the projection (by PPh unit). Figure 2 illustrates the sampling points of a PJR-PJN unit:

Figure 2. Illustration of F0 sampling points of a PJR-PJN unit with a projection of three-PPh units

After deriving the average F0 values, we further attempted the removal of the intonation effect from the higher-level DPUs. This was carried out by remodeling the F0 slope based on PPh units, via turning the value of the F0 slope into 0.

For pause duration, we extracted the duration of silent pauses (in millisecond) located in the following positions: (i) the initiation of the PJR, which was defined as from the off-set of

[11] For this annotation task, the annotators included the first author of the paper. As for the other annotators, they had also worked on the DPU and prominence-level annotation tasks. Hence all annotators were quite familiar with the annotation scheme.

the PW unit immediately preceding the PJR to the onset of the PJR; and (ii) the initiation of the PJN, which was defined as from the off-set of the corresponding PJR to the onset of the PJN. After the pause durations were derived, we further obtained the mean values of the pause durations in both positions.

3.2 Emphasis-attributed Weighting Scores Calculation

To calculate the emphasis-attributed weighting scores, we followed a similar rationale for modeling prominence-correlated distribution of information-attributed weighting scores proposed by Tseng (2010) and Chen and Tseng (2021) and assumed that there was a direct association between the levels of perceived emphasis annotations and information-attributed weighting scores. Adhering to this assumption, the weighting scores were arbitrarily assigned by using the following formula:

(7)

$$\text{Score}(t_n) = \begin{cases} 0, \text{if label} = \text{E0} \\ 0, \text{if label} = \text{E1} \\ 1, \text{if label} = \text{E2} \\ 2, \text{if label} = \text{E3} \end{cases}$$

In the formula above, the *t* represents each ET annotated across the current speech data. One additional note is that, as explained in Section 2.2.2, in annotating the perceived prominence degrees of the current spontaneous speech data, the SpnL and SpnC were both tagged with one extra level of reduction (E0). In order to calculate the information-attributed weighting scores on the basis of the same set of prominence levels, initially we merged the E0 tags with the E1 tags in the SpnL and SpnC and assigned a score of 0 to both.[12]

After the scoring assignment, we calculated the average information-attributed weighting scores across the PJR-PJN units by PW units and averaged the weighting scores derived from each PW within the PJR-PJN units, which ranged from one to three PPh units according to the projection trajectory size. Finally, we conducted correlation analysis to examine the relationship between the average weighting scores and the PJR-PJN units with different trajectory sizes.

[12] Initially we merged E0 and E1 tags for the calculation of the weighting scores purely for the purpose of comparing the current read speech and spontaneous speech data with the same set of prominence levels. We also attempted the further calculation of contrast degree by acoustic cues (including F0, duration and intensity) between all the E0 and E1 tags from the SpnL and SpnC. It was found that all the acoustic features were significantly distinctive in the SpnC data, while for SpnL only the duration feature was distinguished. Hence in the analysis reported later in the paper, we further manipulated the E0 tags from spontaneous speech by assigning a score of -1 to all the reduction tags (see section 4.4.2).

4. Acoustic Profiles and Emphasis-attributed Weighting Scores of the PJR-PJN Units

This section will present the analyses of the acoustic realizations and the results of emphasis-attributed weighting scores derived from the information content planning PJR-PJN units. For the acoustic profiles, we focused on the realization of intonation contours throughout the PJR-PJN units in F0 and pause duration in correlation with the initiation of the PJR and PJN. In addition, we examined the correlation between the emphasis-attributed weighting scores and the projection trajectory size of the PJR-PJN units, which shed light on the overall distribution and planning of information content that was prompted by the prosodic highlights.

4.1 Calculation of PJR-PJN Units by PPh Units

Before the analyses, we took an initial step to examine the general distribution of the PJR-PJN unit across the speech data from the four different genres. As suggested previously, the trajectory size of the projection varied for each PJR-PJN unit (Chen & Tseng, 2021). It was thus essential to first identify the projection range distribution of all the PJR-PJN units. As shown in the results from Chen *et al*. (2016) and Chen and Tseng (2021), it was found that over 90% of the PJR-PJN units were accounted for by up to three PPh units. With the identification of a PPh unit as the basic planning DPU for the PJR-PJN units, we further calculated the total number of PJR-PJN units by PPh units compared with the total number of PPh units across the four speech genres. The results shown in Table 2 provide a further illustration of the proportion of PJR-PJN and PPh units in each speech genre:

Table 2. Summary of the total number of syllables in the PJR-PJN units by PPh units and the total number of PPh units in the four speech genres

Corpora/ Genre	Total Number of Syllables in the PJR-PJN Units by PPh Units	Total Number of PPh Units
SpnL	1,257 (28%)	4,535
SpnC	347 (25%)	1,372
CNA	821(48%)	1,702
WB	324 (38%)	861

Hence in the following analyses, we adopted the PPh unit as the base planning unit to estimate the acoustic correlates and weighting score calculation of the PJR-PJN units. To extend further the findings from Chen and Tseng (2021), we included the PJR-PJN units with projection trajectories ranging from one to three PPh units in the current data.

4.2 Acoustic Correlate: F0 Realizations (with and without Intonation Effect)

Following the methodology described in Section 3.1, we calculated the mean F0 values by the PW units at each sampling points , including the initial and final PW of the PJR-PJN units, as well as the PW units by the PPh boundaries in each PJR-PJN unit, and the results are summarized in Figure 3. On the other hand, Figure 4 presents the results of the F0 measurements at the same sampling points after removing the intonation effect from the higher-level DPUs. Note that both figures present the results according to the trajectory size of the PJR-PJN units, from one up to three PPh units.

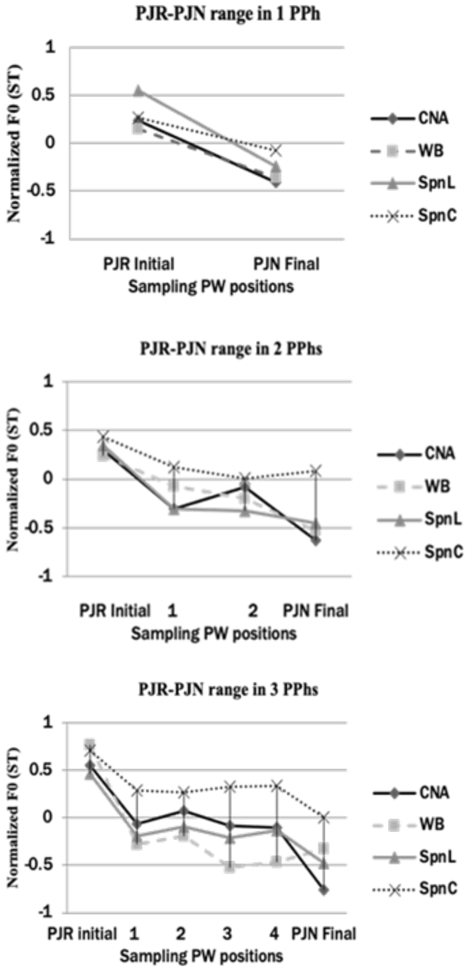

Figure 3. F0 of each PJR-PJN unit (sampling points by position: 1= PW prior to first PPh boundary; 2 = PW after first PPh boundary; 3 = PW prior to second PPh boundary; 4 = PW after second PPh boundary)

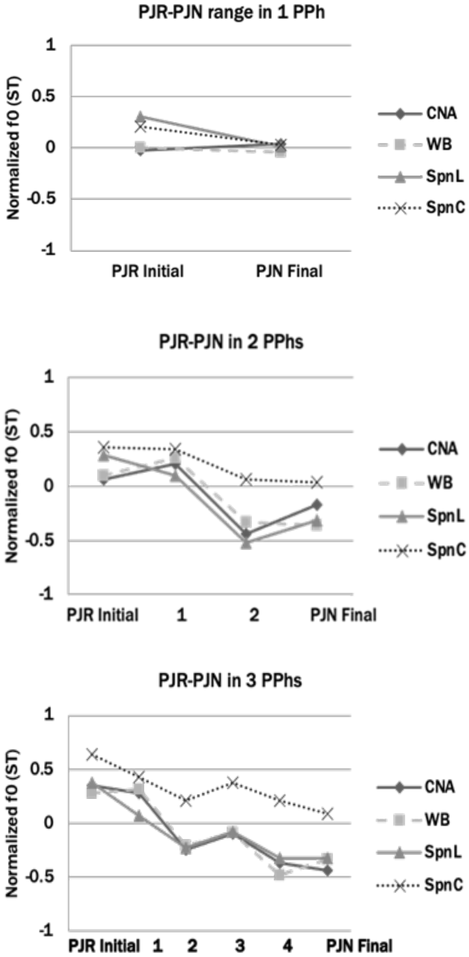

Figure 4. F0 of each PJR-PJN unit without intonation effect (sampling points by position: 1= PW prior to first PPh boundary; 2 = PW after first PPh boundary; 3 = PW prior to second PPh boundary; 4 = PW after second PPh boundary)

4.2.1 Results

First, a general tendency of a "high-to-low" pitch contour was observed across the PJR-PJN units as shown in the three panels of Figure 3. This falling contour was noticeable, regardless of the projection size. Although there were occasional exceptions when a slight rising contour was observed in the PJR-PJN units, (i.e., in the CNA data), when an information content unit extended to two PPh units, the rising contour never reached a point higher than the F0 derived from the initial point of the corresponding PJR unit. Slight final-rising contours were also observed in the SpnC data with two PPh units and in the WB data when the unit expanded to three PPh units. However, the final rising contours in both cases never reached a point higher than the F0 values extracted from the corresponding PJR initiation point. Above all, we found that the F0 values derived from the beginning of PJR units and the ending of PJN units were distinguished, regardless of the trajectory sizes. Further statistical tests indicated that significant differences were present (h=1, $p<0.05$ across all three panels in Figure 3) and thus substantiated the observation of the general falling intonation contour across the PJR-PJN units.

To further validate the falling contours observed, we attempted the removal of the intonation effect from the higher-level DPUs. As presented in Figure 4, after removing the intonation effect, the falling pitch contour was still sustained. Even though there were also slight rising contours both within the projection trajectories and at the end of the projection trajectories in some of the data, the rising contours did not reach a point higher than the F0 values in the corresponding initial PJR units. The only noticeable exception was in the read speech genres (i.e., WB and CAN), in which the PJR-PJN units equaled one PPh unit. *T*-test results also confirmed that the F0 values of the beginning of the PJR units and the ending of the PJN units were distinguished, (all h=1, $p \leqq 0.05$), except for instances in which the projection trajectory was local and within one PPh unit in the read speech genres.

4.2.2 Discussion

The results above demonstrated that, when planning for prosodic highlights-prompted PJR-PJN units as the information content planning units, in general the speakers initiated the intended information content planning units from a higher F0 and continue with a gradual falling contour across the projection trajectories. Although there were cases in which slight rising contours were observed, the rising pitch never reached a point higher than the F0 values derived from the beginning of the PJR units. Furthermore, a general tendency was observed in that, the larger the projection was (i.e., when the trajectory expanded over two PPh units), the greater the difference between the mean F0 values from the beginning of the PJR units and the ending of the PJN units was. This in turn reflected that in the fore-planning of larger information projections, the speakers had to prepare to start the PJR unit at a higher F0 to allow for the further manipulation and allocation of the prosodic variations within the planned projection trajectory.

After removing the intonation effect from the higher-level DPUs, the falling pitch contour across the PJR-PJN units was still maintained. Interestingly, when the projection size was only within one PPh unit and of local planning, the falling contour was not as obvious: the F0 values of the projection trajectories of the initiations and endings were barely distinguishable. In the end, it was only when we considered the global projection of information content that the falling contour was of distinctive significance. Chen *et al*. (2016) reported their results from further calculations of the down-stepping degrees across the PJR-PJN units of different trajectory sizes, and a positive correlation between the down-stepping degrees and the projection trajectory sizes was identified. As shown in Table 3 repeated from Chen *et al*. (2016), the longer the projection trajectory size was, the larger the degree of differences derived from the beginning of the PJR units and the end of the PJN units was:

Table 3. Down-stepping degree across the PJR-PJN units, calculated by PPh units (Chen et al., 2016)

Down-stepping Degrees across PJR-PJN Units			
Genre	**Within a PPh**	**Across 1 PPh**	**Across 2 PPhs**
CNA	0.067	0.234	0.789
WB	0.049	0.452	0.614
SpnL	0.294	0.600	0.700
SpnC	0.173	0.316	0.553

According to Chen *et al.* (2016), the result from the down-stepping degree calculations further reinforced that the overall intonation planning across the PJR-PJN units was not due to the influence of the higher-level intonation effect. In other words, in the actual planning of the information content within a larger projection trajectory that was prominence-prompted, the speakers resorted to a noticeable falling contour and a larger down-stepping degree. This was for the purpose of accommodating more variations in the prosodic highlight allocations within the projection trajectories to reflect focal information allocations.

4.3 Acoustic Correlate: Pause Duration

The second acoustic feature we turned to was the pause duration. In particular, we focused on the duration of silent pauses located prior to the initiation of a PJR unit and in between PJR and PJN units for planning the projection trajectories. Similar to the findings on pause durations in topic flow in spoken discourse by Swerts and Geluykens (1994), it was hypothesized that the longer the PJR-PJN unit was, the more time required for the speaker to initiate the prosodic-prompted PJR unit and plan the projection trajectory; hence, the longer the silent pause duration prior to the initiation of both the PJR and PJN units. Here we focused on the estimation of the correlation between the average pause duration and size of the PJR-PJN units (by PPh unit).

4.3.1 Results

As demonstrated in Figure 5, a general tendency was observed in that, when the projection trajectory size increased, the pre-PJR pause duration was also longer. This was most obvious when comparing PJR-PJN units in the one to two PPh units range. Turning to the PJR-PJN units in the three PPh units range, there were exceptions from the speech genres of WB and SpnL, when the pre-PJR pause was slightly shorter than the average pause duration preceding the PJR-PJN units in the two PPh units range. To verify, we further performed *t*-tests between the average pause duration derived from the PJR-PJN units in the one to three PPhs range, and the results indicated that significant differences were found in the data from both read speech genres (i.e., CNA and WB, both h=1, $p<0.05$). As for the pre-PJN pause duration, the results shown in Figure 6 revealed a similar tendency in that the larger the projection size (i.e., up to three PPh units), the longer the pre-PJN pause was. Further statistical results also showed significant differences in the average pre-PJN pause duration for the PJR-PJN units in the one to three PPh units range, and the results were valid for all speech genres (h=1, $p<0.05$), except for WB.

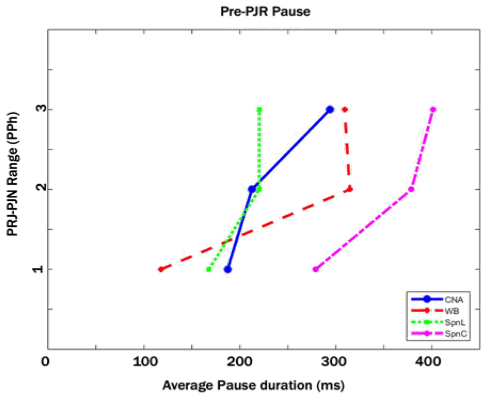

Figure 5. Average pre-PJR pause duration in correlation with projection size

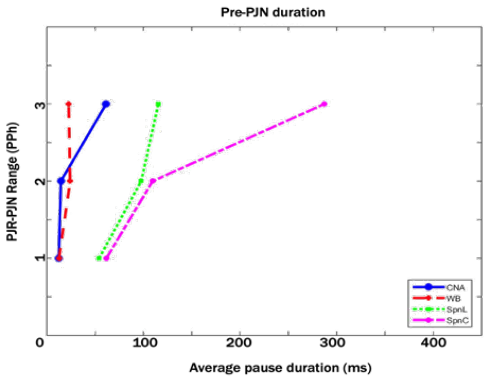

Figure 6. Average pre-PJN pause duration in correlation with projection size

4.3.2 Discussion

Based on the findings, it was suggested that when planning for a PJR-PJN unit as an information content unit, the speakers were mostly oriented to a longer pause in order to initiate the prosodic highlights-prompted PJR unit. This further alluded to a longer preparation time required to plan for a longer projection trajectory. Although there were cases when a PJR-PJN unit with three PPh units was preceded by a slightly shorter pause, the general tendency mostly held given the statistical results of the pause durations for the PJR-PJN units in the one to three PPh units range. Another possible explanation was related to specific speech genre features. For the pre-PJR pause duration, the statistical results pointed to the main differences between read and spontaneous speech. We surmised that this reflected a discrepancy in the design of the speaking tasks in the four different speech genres: in the production of read speech, the speakers were given enough time to prepare before the actual recording; hence they had a chance to preplan the acoustic realizations for information projection due to familiarity with the reading materials. On the other hand, in the spontaneous speech genres, the planning of prosodic deployment and information content was interrupted intermittently because the spontaneous action that was interaction-based.

4.4 Correlation between Emphasis-attibuted Weighting Scores and Information Projection

In the third analysis, we examined the correlation between emphasis-attributed weighting scores and information content projection. Following the findings reported in Chen and Tseng (2021) concerning the calculation of emphasis-attributed density scores throughout the PJR-PJN units, we further validated the information content loading distributions by prosodic highlights-prompted PJR-PJN units. It was demonstrated previously that speakers devote maximal efforts to the planning of information content from the beginning of prosodic highlights-prompted PJR units, and such effort decreases gradually throughout the projection trajectory (Chen & Tseng, 2021). However, Chen and Tseng (2021) reported the results of the emphasis-attributed weighting scores only by PJR-PJN units in the one PPh unit range. To extend the claim further, we carried out the weighting scores calculation again and included all the PJR-PJN tokens with a similar rationale and methodology proposed in Chen and Tseng (2021). We then conducted additional analysis of the correlation between the weighting scores and the PJR-PJN units in the one to three PPh units range for a more solid verification.

4.4.1 Results

As summarized in Figure 7, further analyses confirmed that, when a PJR-PJN unit was extended by three PPh units, a lower average emphasis-attributed weighting score was arrived at by the ending of the PJN units. In other words, the general trend of a decreasing weighting score following an increase in projection size was confirmed. This finding was quite consistent across the data of the four speech genres. Most of all, further *t*-test results verified that the average weighting scores were distinguished between the PJR-PJN units with one PPh unit and three

PPh units. The statistical results were in general supported (all h=1, $p<0.05$), except for the spontaneous speech data from the SpnC genre.

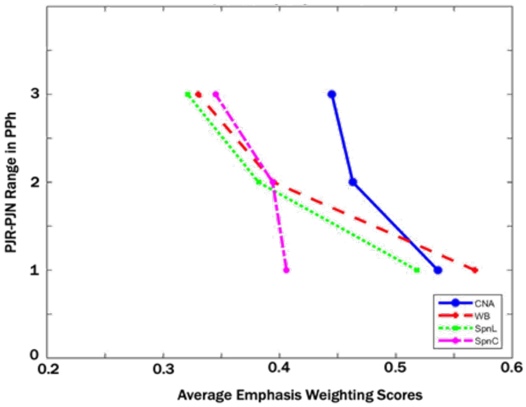

Figure 7. Correlation analysis between average emphasis-attributed weighting scores and the PJR-PJN units of the four speech genres (score assignment: E0=E1=0; E2=1; E3=2)

4.4.2 Discussion

Again, the above result confirmed that, when planning for prosodic-prompted projection for information content allocation, the speakers were oriented to a general pattern of heavy-to-light information loading across the PJR-PJN units, regardless of the projection trajectory size. When planning for a projection with a longer trajectory, the weighting scores decreased gradually toward the end of the projection, and hence information content loading diminished. Such findings in turn provided further confirmation of a PJR-PJN unit as the planning unit of prosodic highlights-correlated information content allocation and deployment in continuous speech. However, the statistical analysis did not find significant results for the SpnC data, which led us to wonder whether this may have had to do with the additional emphasis level of reduction (E0) annotated for the current spontaneous speech genres (i.e., SpnL and SpnC). We attempted a further test by re-assigning the weighting scores only for the spontaneous speech data. In particular, we assigned a score of -1 to the emphasis level of reduction (E0) annotated in the SpnL and SpnC genres, and then recalculated the average weighting scores. The results are summarized in Figure 8:

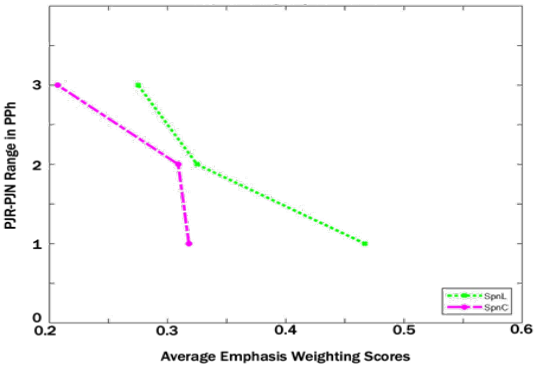

Figure 8. Correlation analysis between the PJR-PJN units and average emphasis-attributed weighting scores of the spontaneous speech genres (score assignment: E0=-1; E1=0; E2=1; E3=2)

Figure 8 presents a pattern similar to the above findings in that, the longer the PJR-PJN unit (i.e., up to three PPhs), the lower the average weighting scores derived from the ending of the PJN units. Further *t*-tests confirmed that the average weighting scores were distinguished between the PJR-PJN units with one PPh unit and three PPh units (both h=1, $p<0.05$). In other words, by taking into consideration the reduction annotation in the spontaneous speech genres, the heavy-to-light information allocation further stood out. With the attempt to faithfully model distinctive emphasis degrees in spontaneous speech signals, therefore, we were able to obtain even more solid evidence to support the current hypothesis regarding information content planning and allocation. This in turn verified the prosodic-prompted projections in association with information content deployment in continuous discourse and speech; above all, it was patterned on the prosodic highlights allotment in the speech context.

5. General Discussion and Summary

The current study focused on information content deployment that was prompted and projected by perceived prosodic highlights consistently annotated in continuous speech and discourse. In the first part of the analyses, we concentrated on the acoustic profiles of the information content planning of the PJR-PJN units, which was initiated and prompted by annotated tokens of prominence across four diverse speech genres. In terms of F0 realization, although the projection size differed in each PJR-PJN unit, we were able to derive a general falling contour starting with the PJR unit and throughout the whole projection trajectory. The further removal of higher-level intonation effects and the calculation of down-stepping degrees offered solid substantiations of the underlying intonation pattern. Above all, the current results demonstrated that only when we considered the information content planning unit of a global projection could we arrive at an identifiable falling contour with a clear down-stepping degree presented. Though the falling contour was within expectations and the results here are much in accordance with previous findings on prosody-based discourse units (i.e., Swerts & Geluykens, 1994), it was most crucial that we were able to further confirm that information content planning associated with prosody-prompted projections could possibly be established as a constant linguistic category with its own identifiable prosodic manifestation.

The second acoustic feature that we turned to was pause duration. As suggested, the duration of silent pauses located prior to the initiation of PJR and PJN units rendered some ideas about the relevant effort devoted to the planning of information content projection. It was demonstrated that, in order to plan for a longer projection, the speakers in general took more time prior to the initiation of the PJR and PJN units. Although not all pause-correlated results were presented with statistical significance, we assumed that the discrepancy was related to the task-specific features of the four difference speech genres.

Through the calculation of emphasis-attributed weighting scores, the third part of the analyses provided further validation of the "high-to-low" distribution of weighting score across the PJR-PJN units, which was similar to the finding from Chen and Tseng (2021). As previously indicated, the tendency of a higher weighting score for the initiation of information projection and a lower score for the end of information content projection reinforced the finding that the heaviest information loading was planned by prominence-prompted PJR units, with a gradually decreased planning effort demonstrated (Chen & Tseng, 2021). Here via the systematic modeling of prosodic highlights, including the reduction, our results faithfully reflected information content allocation and deployment for speech planning. Above all, the results showcased that only when taking into consideration the reduction feature in spontaneous speech could we arrive at a more significant distinction among the four speech genres with diverse features.

In sum, in this study we examined prosodic highlights-prompted information content planning and projection by the recently identified PJR-PJN units in continuous speech. Solid accounts were provided for the specific acoustic features, including F0 and pause duration, as well as the information-attributed weighting scores in correlation with the projection size in the PJR-PJN units. As has been identified previously the PJR-PJN units for information content projection were planned at a higher discourse-prosodic level from the HPG framework (cf. Chen & Tseng 2021); ultimately the identification of the patterns enabled a better understanding of information content planning within the hierarchical framework of the prosody context. In future studies, we propose to explore the following: (i) other possible acoustic correlates that might be involved prominence-prompted information content projection; (ii) empirical validations of the correlation between perceived prosodic distinctiveness in limited degrees and information weighting (i.e., Kurumada *et al*., 2014); and (iii) the incorporation of the current analyses' results into the automatic modeling of discourse prosody based on a hierarchical relationship (i.e., Lin *et al*., 2019).

Acknowledgments

The authors gratefully acknowledge the assistance of Mr. Yen-Hsing Chen, Mr. Wei-Te Fang, and Dr. Chao-yu Su for the data analyses and relevant discussions in this study. Moreover, we would like to thank the research assistants from the Phonetics Lab at the Linguistic Institute, Academia Sinica (2015-2018) for the annotation tasks reported in this work. Most of all, we are grateful for the anonymous reviewers' and the editor's insightful comments on a previous version of the manuscript. All remaining errors, nevertheless, are ours.

About the Speech Corpora

To access the speech corpora incorporated in this study, the Sinica COSPRO corpus (for the read speech data) can be obtained through the database link from the Association for Computational Linguistics and Chinese Language Processing (http://www.aclclp.org.tw/use_mat.php#cospro). A sample of the spontaneous interaction (SpnC) data with the annotations described in this paper can be obtained via contacting the first author.

Abbreviations

2SG	second-person singular pronoun	COP	copula
1PL	first-person plural pronoun	DE	associative/complementizer de
CL	classifier		

References

Auer, P. (2015). The temporality of language in interaction: Projection and latency. In A. Deppermann & S. Günthner (Eds.), *Temporality in interaction*, 27-56. John Benjamins.

Baumann, S., Niebuhr, O., & Schroeter, B. (2016). Acoustic cues to perceived prominence levels: Evidence from German spontaneous speech. In *Proceedings of 8th Speech Prosody Conference*, 711-715. https://doi.org/10.21437/SpeechProsody.2016-146

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer. Retrieved November 20, 2015, from http://www.praat.org

Chen, H. K., Fang, W., & Tseng, C. (2016). Advance prosodic indexing-Acoustic realization of prompted information projection in continuous speeches and discourses. In *Proceedings of the 10th International Symposium on Chinese Spoken Language Processing (ISCSLP 2016)*. https://doi.org/10.1109/ISCSLP.2016.7918412

Chen, H. K., Prévot, L., Bertrand, R. Priego-Valverde, B., & Blache, P. (2012). Toward a Mandarin-French corpus of interactional data. In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogues*, 147-148.

Chen, H. K., & Tseng, C. (2021). From speech to language-An alternative corpus account of prosodic highlight in continuous speech. *Concentric: Studies in Linguistics*, *47*(2), 184-224. https://doi.org/10.1075/consl.00027.che

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181-204. https://doi.org/10.1017/S0140525X12000477

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141-201. https://doi.org/10.1177/002383099704000203

De Saussure, F. (1966). *Course in general linguistics*. McGraw-Hill Book Company.

Dilley, L. (2016). Rhythm, context effects, and prediction. In *Proceedings of Speech Prosody 2016 conference (SP 2016)*.

Falk, S. (2014). On the notion of salience in spoken discourse—Prominence cues shaping discourse structure and comprehension. *TIPA, Travaux Interdisciplinaires Sur La Parole et Le Langage [Interdisciplinary Works on Speech and Language]*, 30, 1-23. https://doi.org/10.4000/tipa.1303

Goodwin, C. (1996). Transparent vision. In E. Ochs, E. A. Schegloff & S. A. Thompson (Eds) *Interaction and grammar*. Cambridge University Press. https://doi.org/10.1017/cbo9780511620874.008

Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 1. *Journal of Linguistics*, *3*(1), 37-81. https://doi.org/10.1017/S0022226700012949

Kohler, K. J. (1997). Modelling prosody in spontaneous speech. In S. Yoshinori, N. Campbell & N. Higuchi, (Eds.), *Computing prosody: Computational models for processing spontaneous speech*, 187-210. Springer. https://doi.org/10.1007/978-1-4612-2258-3_13

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. (2014). Rapid adaptation in online pragmatic interpretation of contrastive prosody. In *Proceedings of the Annual Meeting of the Cognitive Science Society 2014*, 791-796

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Lieberman, P. (1967). *Intonation, perception, and language*. MIT Press.

Lin, C. H., You, C. L., Chiang, C. Y., Wang, Y. R., & Chen, S. H. (2019). Hierarchical prosody modeling for Mandarin spontaneous speech. *Journal of Acoustic Society America*, *145*(4), 2576–2596. https://doi.org/10.1121/1.5099263

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Morgan, J. L. Cohen, & M. E. Pollack (Ed.) *Intentions in communication*. MIT Press.

Silverman, K., Beckman, M., Pitrelli, J. Ostendorf, J. M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the 2nd International Conference on Spoken Language Processing(ICSLP 1992)*, 867-870.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*(1), 21-43. https://doi.org/10.1177/002383099403700102

Tseng, C. (2010). An F0 analysis of discourse construction and global information in realized narrative prosody. *Language and Linguistics*, *11*(2), 183-218.

Tseng, C. (2013). Output prosody-How information highlights are piggybacked by discourse structure. *Chinese Journal of Phonetics (中國語音學報)*, *4*, 109-124.

Tseng, C., Cheng, Y., Lee, W., & Huang, F. (2003). Collecting Mandarin speech databases for prosody investigation. In *Proceedings of the 2003 International Conference Oriental-COCCOSDA*.

Tseng, C., Lee, L., & Su, Z. (2008). Spontaneous Mandarin speech prosody - The NTU DSP lecture corpus. In *Proceedings of the 2008 International Conference Oriental-COCCOSDA*, 171-174.

Tseng, C., Pin, S., Lee, Y., Wang, H. & Chen, C. (2005). Fluent speech prosody: Framework and modeling. *Speech Communication*, *46*(3-4), 284-309. https://doi.org/10.1016/j.specom.2005.03.015

Tseng, C., Su, C. & Huang, C. (2011). Prosodic Highlights in Mandarin Continuous Speech—Cross-Genre Attributes and Implications. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association*, 1381-1384. https://doi.org/10.21437/Interspeech.2011-454

Tseng, C., & Su, Z. (2008). Discourse prosody and context–global F0 and tempo modulations. In *Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH 2008)*, 1200-1203. https://doi.org/10.21437/Interspeech.2008-361

Tseng, C., & Su, Z. (2012). Information allocation and prosodic expressiveness in continuous speech: A Mandarin cross-genre analysis. In *Proceedings of the 2012 International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 243-246. https://doi.org/10.1109/ISCSLP.2012.6423535

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L + H. *Cognitive Science*, *32*(7), 1232–1244. https://doi.org/10.1080/03640210802138755