Journal of Phonetics 39 (2011) 595–611

Contents lists available at ScienceDirect

Journal of Phonetics

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

Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese

Bei Wang [a,*], Yi Xu [b]

[a] Institute of Chinese Minority Languages, Minzu University of China, 27 Zhongguangcun Nadajie, Beijing 100081, PR China
[b] Department of Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences, University College London, London, UK

ARTICLE INFO

Article history:
Received 27 March 2009
Received in revised form
16 March 2011
Accepted 25 March 2011

ABSTRACT

This paper reports an experimental investigation of the prosodic encoding of topic and focus in Mandarin by examining disyllabic subject nouns elicited in four discourse contexts. These subject nouns also varied in terms of their tonal composition as well as the length of their embedding sentences. The central questions addressed are how prosodic effects of topic and focus differ from each other and how they interact with sentence length, downstep and newness to determine sentence-initial F_0 variation. Sixty short discourses were recorded with variable focus, topic level, newness, downstep, and sentence length conditions by six native speakers. The results of extensive acoustic analyses show that (1) the difference between topic and focus lies in that focus both raises on-focus F_0 and lowers post-focus F_0 while topic raises the F_0 register at the beginning of the sentence while allowing F_0 to drop gradually afterwards, (2) topic has higher pitch register in isolated and discourse-initial sentences than in non-initial sentences, (3) longer sentences have higher sentence-initial F_0 than shorter sentences, but the differences are small in magnitude and are independent of topic and focus, (4) the effect of downstep is independent of topic and focus, but is large in magnitude and accounts for a significant amount of the F_0 declination in a sentence, (5) newness has no F_0 manifestation independent of other factors, but a newly mentioned word is slightly longer than a previously mentioned word, and (6) the effects of topic, focus, downstep and sentence length are largely cumulative. We argue that these findings are harmonious with an articulatory-functional view of speech prosody represented by the Parallel Encoding and Target Approximation model (PENTA).

1. Introduction

It has long been known that F_0 is typically higher at the beginning of a sentence than at the end of the sentence. Early phonetic accounts have attempted to explain the phenomenon in terms of declination (e.g., Cohen & 'tHart, 1967). Later research has suggested, however, that the global F_0 decline is unlikely due to a single mechanism. First, it has been shown that the initial F_0 of a sentence can help listeners identify paragraph-initial sentences from later sentences (Lehiste, 1975; Silverman, 1987). Other research has linked initial F_0 to prosodic manifestation of discourse structure, involving topic and turn taking (Grenié & Grenié, 2003; Lehiste, 1975; Oliveria & Freitas, 2008; Silverman, 1987; Tseng, 2008; Tseng, Pin, & Lee, 2005; Umeda, 1982; Yang, 1995). Second, sentence initial F_0 can be raised due to focus (Bruce, 1982a; Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Shih, 1997; Xu, 1999). Third, new information has been found to be related to F_0 height (Bruce, 1977; Féry & Kügler, 2008; Halliday, 1967a, 1967b; Hirschberg & Pierrehumbert, 1986; Terken, 1984). Fourth, there have also been reports that initial F_0 height is correlated with sentence length: a longer sentence starts with higher F_0 and more gradual declination (Bosch, 2007; Bruce, 1982a, 1982b; Grenié & Grenié, 2003; Shih, 1997; Swerts, Strangert, & Heldner, 1996; Van Heuven, 2004). Fifth, further complication comes from downstep and local F_0 downshift across adjacent tones (Laniran & Clements, 2003; Xu, 1999). So far, there are no empirical studies on how the factors mentioned above, namely, focus, topic, sentence length, downstep and newness, interact with each other. In the following literature review, we will present the main findings on how these factors affect F_0 in sentence initial position.

1.1. Focus

Focus highlights part of a sentence against the rest of the sentence as motivated by a particular discourse situation (Bolinger, 1958; Eady & Cooper, 1986; Ladd, 1996; Xu, 1999, 2005). A focused word typically has higher F_0, longer duration and greater amplitude compared to its unfocused counterpart (for English: Cooper et al., 1985; Eady & Cooper, 1986; Eady, Cooper,

* Corresponding author. Tel.: +86 013641029534.
E-mail addresses: bjwangbei@gmail.com (B. Wang), yi.xu@ucl.ac.uk (Y. Xu).

0095-4470/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2011.03.006

Klouda, Mueller, & Lotts, 1986; Xu & Xu, 2005; for Mandarin: Liu & Xu, 2005; Xu, 1999; for German: Féry & Kügler, 2008). More importantly, the same studies have also reported sharp F_0 lowering and F_0 range compression in the post-focus words in the sentence.

However, in terms of the amount of F_0 raising by focus in sentence initial position, much uncertainty remains in literature. Cooper et al. (1985) and Eady and Cooper (1986), for example, have found that in English the F_0 peak in the first word of a sentence is relatively constant, regardless of whether the word is focused. In contrast, focus in sentence initial position has been found to raise F_0 of the focused word in other studies, such as Eady et al. (1986) and Xu and Xu (2005) for English, Xu (1999) for Mandarin, and Féry and Kügler (2008) for German. Eady et al. (1986) have explained the inconsistency across their own studies in terms of sentence length, as shorter sentences show a greater focus effect. Indeed, in Eady et al. (1986), Xu (1999) and Xu and Xu (2005), where F_0 raising by initial focus was found, the average sentence length was 5–8 syllables, whereas in Cooper et al. (1985) and Eady and Cooper (1986), where no focus-related F_0 raising has been found, the average sentence length was 11.4 syllables. However, there has not been any study specifically designed to examine how sentence length and focus interact with each other to determine sentence-initial F_0.

1.2. Topic

Each sentence has a topic, which is basically what the sentence is talking about (Lambrecht, 1994). However, when several sentences are strung into a larger unit, they form a discourse with a shared common topic (Smith, 2004). A large body of literature suggests that the structure of such a discourse is reflected in prosody. A well-known effect is that the beginning phrase starts with higher F_0 and a larger pitch range than the following phrases (Ayers, 1994; Brown, Currie, & Kenworthy, 1980; Grosz & Hirschberg, 1992; Sluijter & Terken, 1993; Terken, 1984; Tseng, 2008; Tseng et al., 2005; Umeda, 1982). As the discourse goes on, there is a tendency for F_0 to either decrease gradually or remain the same over the course of the paragraph (Bruce, 1982b; Cooper & Sorensen, 1981; Hirschberg & Pierrehumbert, 1986; Ladd, 1988; Nakajima & Allen, 1993; Sluijter & Terken, 1993; Thorsen, 1985). When the topic of a conversation is changed, the sentence-onset F_0 and pitch range are increased again (Brown et al., 1980; Nakajima & Allen, 1993; Oliveira & Freitas, 2008).

Different studies have used very different methods to analyze the organization of a discourse. Some have simply examined the effect of sentence position. Lehiste (1975) has shown that F_0 is higher in paragraph-initial position than in medial and final positions and in isolated sentences. Thorsen (1985) has further found that terminal declarative sentences have steeper sentence intonation contours than corresponding non-terminal clauses, and that the onset of the lower line in an isolated sentence is intermediate between initial and final sentences in a discourse. Umeda (1982), instead, has found a very small percentage of occurrences of obvious F_0 declinations in 43 short isolated sentences.

Another approach in analyzing discourse structure is to treat it as hierarchically organized (Grosz & Sidner, 1986; Swerts and Osterndorf, 1997; Tseng, 2008; Tseng et al., 2005). Depending on the unit of application (sentence or discourse), many make a distinction between sentence topic and discourse topic (Guijarro, 2001; Van Oosten, 1985). While sentence topic represents the proposition as to what information is given at a local level, discourse topic represents what a whole text or discourse is about and is defined as a cognitive schema which sequentially organizes and unifies all the local topics under the same topical frame (Downing et al., 1998; Guijarro, 2001; Van Oosten, 1985).

The third approach focuses on the transitions between turns or sentences, which are categorized into four classes by Nakajima and Allen (1993), that is, topic shift, topic continuation, elaboration and speech-act continuation. In topic shift, there is either no link between the current utterance and the previous utterances, or only a very weak link divided into three subclasses depending on the strength of the link: new topic, topic development and topic interruption. In continuation, the linkage between the current topic and the previous one is comparatively strong. They have found that topic shift has higher onset F_0 than topic continuation. Smith (2004) has applied this classification in a pre-designed discourse reading task and showed that topic shifts tend to involve lengthening of sentence-final words and pauses.

Although the three approaches just mentioned look quite different, their findings do not really contradict each other, but rather show much overlap. In particular, the raised F_0 on discourse-initial words is related to the properties of being initial, being a discourse-topic and introducing a new topic. Similarly, the less raised initial F_0 on the successive continuing topic is related to the properties of being in the middle of the discourse, being a sub-topic, and being a topic continuation. However, there have been no empirical studies showing how the raising of F_0 in a discourse-initial position differs from the raising of F_0 due to focus.

In a sentence, when the topic carries a contrastive meaning, it is not always easy to distinguish it from focus. Crocco (2009) and Tomioka (2009) have stated that thematic topic is typically given, presupposed, or anchored in speech situation. Contrastive topic, on the other hand, implies an alternative, hence a contrast. Tomioka (2009) further stated that a thematic topic refers to a contextually familiar or recoverable entity whereas a contrastive topic can be familiar or novel. Prosodically, it has been reported that thematic topic is usually unaccented, whereas contrastive topic has a focal part, e.g., B accent (or H*LH%) in English (Jackendoff, 1972, p. 258; Pierrehumbert, 1980, p. 25), L*H(H%) in German (Büring, 1997; Féry and Kügler, 2008; Gundel, 1994), and high tone on contrastive topic marker in Japanese (Tomioka, 2009). Büring (1997, 2003) and Wagner (2008) have argued that contrastive topic invokes alternatives and therefore has focus. If, as argued by these authors, focus is indeed involved in a contrastive topic, there would be the possibility that contrast and topic are confounded in the notion of contrastive topic. Therefore there is a need to systematically control for both in further research in order to clearly separate their distinctive prosody effects.

1.3. Sentence length

That sentence length may be a contributing factor to sentence initial F_0 was suggested by two additional findings about focus, which may also help to distinguish between topic and focus. One is that in a long sentence, an initial focus does not raise initial F_0 (Eady & Cooper, 1986). Another is that there is an upper limit on how much F_0 can be raised by focus (Chen & Gussenhoven, 2008). Thus it is possible that the amount of F_0 raising by topic in a discourse-initial sentence is positively related to sentence length, and beyond a certain length it may exceed the F_0 raising by a sentence-initial focus.

Some studies have found a positive correlation between initial F_0 height and sentence length (for Swedish: Bruce, 1982a; Swerts et al., 1996; for Mandarin: Shih, 1997; for Chumburung: Snider, 1998; for five Romance languages: Grenié & Grenié, 2003; Prieto et al., 2006), while others have found no such correlation (for English: Liberman & Pierrehumbert, 1984; for Danish: Thorsen, 1980, 1983; for Hausa: Lindau, 1986; for Spanish: Prieto, Shih, & Nibert, 1996; for Mambila: Connell, 2002, 2003, 2004). A positive

correlation between sentence-initial F_0 peak and sentence length would mean that speakers are able to plan the overall intonation by beginning with a higher F_0 when the sentence is longer than when it is shorter.

The strategy could be a simple and rough one, such as a *soft* pre-planning as proposed by Liberman and Pierrehumbert (1984). They argued that before a longer sentence, the speaker will take a deeper breath than before a shorter sentence. So it is the volume of air trapped inside the lungs that primarily determines the onset F_0 rather than some complex computational act the speaker performs on the F_0 contour (which would be *hard* pre-planning in Liberman and Pierrehumbert's terms) (also see Van Heuven, 2004). Two studies actually show evidence of such *soft* pre-planning. Whalen and Kinsella-Shaw (1997) have found a positive correlation between the duration of the sentence to be said and the duration of the inspiration in advance when the speaker is required to read aloud sentences using only one breath. But the correlation is not high and not consistent across speakers. Watson, Ciccia, and Weismer (2003) examined the relationship of speech breathing with other elements of speech production. They asked speakers to initiate speech from low, typical and high lung volume levels and found that with increasing lung volume initiation levels, average sound pressure level, average F_0 and declination rate all increased.

It is also possible that speakers can make precise plans according to sentence initial F_0, so that the degree of F_0 lowering is smaller in a long sentence than in a short sentence. Some experimental evidence is provided by Van Heuven (2004). He has studied downstep in Dutch enumerations of two to six items long and found that pre-planning is precise in the sense that the first downstep is exactly proportional to the number of items in the enumeration. However, he has also argued that speakers limit the burden of tonal planning to a minimum. For instance, they do not anticipate longer utterances by starting off at a higher F_0, nor do they execute a larger accent-lending rise on the first item in the enumeration.

Apart from the initial F_0 value being related to sentence length, a number of investigations have shown that the slope of declination is less in a longer intonation than in a shorter one (Maeda, 1976; Pierrehumbert, 1979; Sorensen & Cooper, 1979). Furthermore, it is found that the final local F_0 maximum and minimum are constant across texts of different lengths (Bruce, 1982b; Féry & Kügler, 2008).

1.4. Downstep

Downstep refers to the phenomenon that in a sequence of High–Low–High (HLH) tones, the second H is much lower in F_0 than the first, which is initially reported for a number of African tone languages (Hyman, 1973; Meeussen, 1970; Stewart, 1965, 1983). The phenomenon has later been found in other languages, e.g., Shih (1988, 1997) for Mandarin, Pierrehumbert (1980) for English, Pierrehumbert and Beckman (1988) for Japanese, Prieto et al. (1996) for Spanish, and Féry and Kügler (2008) for German. Downstep is conceptually different from declination. Declination is a continuous F_0 decline across the whole utterance (Ladd, 1984; Pierrehumbert, 1980; Shih, 1997; Thorsen, 1983). Downstep, on the other hand, is localized at specific junctures and is said to be conditioned by the tonal, lexical, morphological, and/or syntactic structure of the utterance in which it applies (Laniran & Clements, 2003; Liberman & Pierrehumbert, 1984; Prieto et al., 1996; Shih, 1997). There is also evidence that downstep actually consists of two effects: the lowering of subsequent F_0 by a L tone, known as carryover-lowering (Gandour, Potisuk, & Dechongkit, 1994; Xu, 1999), and the raising of the preceding F_0 by a L tone, known as anticipatory raising, or H-rising (Gandour et al., 1994; Laniran & Clements, 2003; Xu, 1997, 1999). Both anticipatory-raising and carryover-lowering would contribute to a sharp initial decline in F_0 when there is a L tone near the beginning of an utterance.

The issue of whether speakers of tone languages do pre-planning in regard to downstep has been a matter of controversy. Laniran and Clements (2003) have summarized two strategies: (1) the F_0 of H tones in a downstepping sequence is sensitive to the number of following downsteps (Stewart, 1965 for Akan); (2) the F_0 of the first H tones is "normally phonetically the same" regardless of the number of following downsteps (Schachter, 1965, for Akan). There is some evidence for the first strategy (Laniran & Clements, 2003 for Yoruba, Rialland & Somé, 2000 for Dagara, and Van Heuven, 2004 for Dutch).

Mandarin offers an opportunity to separate the effect of downstep from those of topic and focus. In a sentence with all H tones, the effect of downstep is absent, whereas in a sentence with a sequence of HL or LH tones, the downstep effect is present. The comparison between them should provide evidence on how downstep interacts with the other factors that affect sentence-initial F_0.

1.5. New vs. Given

To further complicate matters, the topic and focus distinction is often confounded by a possible division of given and new. Entities newly introduced into the discourse are considered as new, and those already in the discourse context as given (Chafe, 1974, 1976; Clark & Haviland, 1977; Halliday, 1967a; Prince, 1981). In Halliday (1967a), given is defined as "recoverable either anaphorically or situationally" (p. 211), while new is not only "not being recoverable from the preceding discourse" (p. 204) but also "contrary to some predicted or stated alternative" (p. 206) and "replacing the wh-element in a presupposed question" (p 226). Furthermore, Chafe (1976) proposes that "given information is the knowledge that a speaker assumes to be in the consciousness of the addressee at the time of the utterance. So called new information is what the speaker assumes he is introducing into the addressee's consciousness by what he says" (p. 30).

Many researchers have proposed that there is a close correspondence between information status and (de)-accentuation in Germanic languages in that given information is deaccented while new information is mostly accented (Brown, 1983; Chafe, 1976; Féry & Kügler, 2008; Fowler and Housum, 1987; Halliday, 1967b; Hirschberg, 1993; Ladd, 1996; Nooteboom & Kruyt, 1987; Nooteboom and Terken, 1982; Prince, 1981). However, as pointed out by Ladd (1996), the alleged given/new contrast is confounded with accentuation, and when accentuation is removed, there is little acoustic difference between given and new. Terken (1984) suggested that apart from newness, there are additional factors which also affect accent distribution. Xu and Xu (2005) argued further that accentuation itself is a concept that confuses between focus and lexical stress, so that accented units are usually stressed syllables that are focused. Thus acoustic differences solely due to given/new contrast cannot be known unless focus is explicitly controlled.

1.6. An articulatory-functional perspective

The multiple factors that affect sentence-initial F_0 reviewed above seem to present a rather perplexing picture. One way to find coherence in such complexity is to take an articulatory-functional view of speech, as represented by the Parallel Encoding and Target Approximation model (PENTA) (Xu, 2005). According to PENTA, speech conveys a multitude of communicative meanings by implementing encoding schemes of individual communicative functions *in parallel*, through a dynamic articulatory

process of Target Approximation (TA). A basic assumption of PENTA is that factors that shape the surface prosody of speech are either communicative or articulatory. While the communicative factors convey meanings, the articulatory factors do not. Of the five factors reviewed above, topic, focus and newness are apparently communicative, whereas downstep and sentence length are at least partially due to articulatory mechanisms (although the tonal sequence and sentence length that trigger these effects are communicative themselves). More importantly, PENTA assumes that the encoding scheme of each and every communicative function has to be unique, so as not to overlap with those of other functions. Such uniqueness is often achieved by having multiple components in the encoding scheme of a particular function. The encoding of focus in Mandarin and English, for example, is done not only by increasing the pitch range and intensity of the focused component, and by compressing the pitch range and intensity of the post-focal components (Chen, Wang, & Xu, 2009; Cooper et al., 1985; Xu, 1999; Xu & Xu, 2005). It is thus possible that such a multi-component encoding scheme could effectively distinguish focus from topic, which presumably raises sentence-initial F_0 without abruptly lowering subsequent F_0. The uniqueness of the communicative functions means that they are relatively independent of each other. However, such independence does not necessarily mean a lack of any interactions between the functions. In Mandarin, for example, focus raises the F_0 of the H tone, but lowers the F_0 of the L tone (Xu, 1999). Thus whether two communicative functions interact in their control of the phonetic details has to be empirically discovered.

Also, the PENTA model does not assume that it can directly predict the exact forms of individual encoding schemes, because how and even whether a communicative function is prosodically encoded in a language is the product of not only various basic communicative mechanisms, but also the evolutionary history of the language. The post-focus compression of pitch range reported for languages like Mandarin and English, for example, has recently been found to be absent in Taiwanese (Chen et al., 2009), Cantonese (Wu & Xu, 2010) and a number of other languages (Zerbian, Genzel, & Kügler, 2010). The exact details of each encoding scheme in a particular language, therefore, can be discovered only through systematic empirical investigations, as has been done previously (Chen et al., 2009; Lee and Xu, 2010; Liu & Xu, 2005, 2007; Wu & Xu, 2010; Xu, 1999; Xu & Wang, 2009; Xu & Xu, 2005). On the other hand, parallel encoding of multiple functions would predict that their combined effects would be cumulative. For example, given that tone already uses F_0 contours as its main acoustic correlates, F_0 encoding of extra-lexical functions such as focus and topic would have to involve pitch changes beyond those of the lexical tones.

1.7. Aims of the study

The literature review above shows that sentence-initial F_0 is determined not by any single factor, but by multiple mechanisms. At the same time, it is apparent that each study has focused on only one or two factors, and there often seem to be overlaps of the factors examined in different studies. As a result, it remains unclear how exactly sentence-initial F_0 is shaped by various factors. The PENTA model reviewed above may provide a theoretical framework which allows clear separation of multiple factors (Xu, 2005). In particular, PENTA assumes that encoding schemes of individual communicative functions have to be unique so as to be implementable in parallel. However, PENTA makes no predictions about the specific forms of encoding schemes, because, as explained in Section 1.6, they are likely to be products of the evolutionary history of a language, and thus can be discovered only through systematic empirical investigations. PENTA also assumes that individual encoding schemes interact both with each other and with various articulatory mechanisms. As a result, a complete picture of an encoding scheme cannot be seen unless articulatory mechanisms involved in its implementation as well as other frequently co-occurring functions are taken into full consideration. Nevertheless, PENTA would predict that the effects of these independent factors would be cumulative, i.e., each accounting for part of the total effect, and the observed surface form would be dividable into specific effects from all individual functions that are involved. The general goal of this study is therefore to answer the following questions about the multiple contributors to sentence-initial F_0:

(1) What are the different prosodic manifestations of topic and focus in terms of F_0 in sentence initial position?
(2) Is there an independent effect of sentence length on sentence-initial F_0? If yes, how does it interact with topic and focus in determining sentence-initial F_0?
(3) How does downstep interact with topic and focus in determining sentence-initial F_0?
(4) Dose a sentence start with higher F_0 when the initial word is new than when it is given?
(5) For discourse topic, how do topics at different discourse levels differ in terms of sentence-initial F_0?
(6) Are the effects of all these factors on sentence-initial F_0 cumulative?

Based on prior empirical evidence as discussed in Sections 1.1–1.5, the following predictions can be made. For question (1) it is predicted that focus not only expands the pitch range of the focused word but also compresses and lowers the pitch range of the post-focus words, while topic only raises the sentence-initial pitch range without compressing subsequent pitch range. The lack of post-topic compression has never been explicitly tested. For (2) it is predicted that sentence-initial F_0 is higher in longer sentences than in shorter sentences. For (3) it is predicted that downstep contributes to a large portion of what has previously been reported as topic- or focus-related sentence-initial F_0. For (4) it is predicted that a target word has a higher F_0 if it has never been mentioned before than if it has already been mentioned, but the difference is likely to be small. For (5), it is predicted that a topic at a higher discourse level leads to higher sentence-initial F_0 than a topic at a lower discourse level. Since topic at discourse level may coincide in position with a sentence topic, we will analysis topic at sentence level as well. Especially, we will examine if topic with a contrastive meaning has specific F_0 correlates. However, a full investigation of contastive topic is beyond the scope of the present study. For (6), it is predicted that, consistent with prediction (3), the effects of the significant factors found in this study will be cumulative, i.e., the observed total effect can be dividable into separate effects by each of the individual factors.

2. Method

The overall strategy of the study is to have speakers read aloud a set of pre-designed fully meaningful sentences in different discourse contexts and perform detailed acoustic analysis of the recorded utterances. Five factors were manipulated in these sentences: (1) focus, (2) topic, (3) downstep, (4) sentence length and (5) newness. Four different contexts were constructed to contrast between focus and topic, topics at different discourse levels, and given/new status. For each context, sentences with five

tone combinations and three lengths were constructed that would manifest the effects of downstep and sentence length.

Contrastiveness of topic was not one of the main factors to be tested, and this is for a number of reasons. First, as discussed in Section 1.2, the contrastiveness of topic overlaps conceptually with focus, and as a result, it also overlaps with newness (see Section 1.5). Second, despite the extensive theoretical speculations, there has been no clear experimental evidence for contrastive topic independent of other factors, and this makes it difficult for us to design proper experimental conditions for manipulating topic contrastiveness. Third, identifying the acoustic difference between focus and topic — one of the key objectives of the present study, will go a long way toward establishing a methodological basis on which future research can develop even more fine-grained strategies that can adequately address the issue of contrastive topic.

2.1. Reading materials

We used short and simple discourses composed of one or two sentences. Because there are no generally agreed categories of topic as reviewed in Section 1.2, we adopted a simple strategy of finding a set of conditions that can generate sufficiently different amount topic effects. Based on the results of several pilot tests, we found it effective to manipulate the location of sentence in a discourse, which could bring out three levels of topic: discourse-initial, non-initial, and sentence in isolation. Also from previous literature, we know that topic in a discourse-initial position is very likely to have higher F_0 than topic in a discourse non-initial position. But it was not our goal, nor was it possible for us, to establish "correctly" defined topic categories in this study. Instead, our goal was to investigate how sentence initial F_0 related to topic is different from that due to other factors, especially focus.

We composed sentences with a simple SVO structure in which the first disyllabic word is the target word. In each case, a personal name was used as the target word and the sentence was talking about what this person did. A context sentence determined whether the target word was *focus or topic*, *given or new, and in an isolated, discourse-initial or non-initial sentence.* For ease of F_0 analysis, the two syllables of the target words mostly had sonorant onsets. Table 1 lists the features of the target word in each condition. A sentence like "Wangying can-guan che-jian. (Pinyin spelling of the Chinese pronunciation. English translation: Wangying visits the workshop)" was used as an example, in which Wangying is the target word. Condition A was the base-line condition, in which the target word bears three default features: topic, non-initial and new. In the other three conditions, one of these features was altered, while the other two remained the same. More detailed descriptions of the four conditions are as follows.

- In *condition A*, the structure is of the type "Da-jia dou you shi-qing yao zuo. Wangying can-guan che-jian. (Everybody has something to do. Wang Ying visits the workshop)". The first sentence introduces the background, and the second sentence talks about what Wangying did. Wangying is hence the topic. According to Büring (1997), Kiss (1998), Wagner (2008) and Zubizarreta (1998), this topic also involves contrast in that the speakers need to select one out of a possible set of persons (not mentioned explicitly in the preceding discourse) as the topic of the sentence. However, because visiting the workshop is also new information, Wangying is not a single initial focus of the sentence, hence it clearly differs from the focus condition (condition B). In addition, because the target word Wangying is not explicitly mentioned, it is new. Also here Wangying is in a non-initial sentence.
- In *condition B,* the background context elicits a narrow focus on the target word. The structure is of the type "Bu shi Lixiao. Wangying can-guan che-jian (It is not Lixiao. Wangying visits the workshop)". In such a structure, Wangying is in corrective focus according to Gussenhoven (2004). Also, because Wangying is not mentioned in the first sentence, it is new and is in a non-initial sentence.
- In *condition C,* the target sentence, e.g., "Wangying can-guan che-jian (Wangying visits the workshop)", is uttered in isolation. As a result, Wangying is new and is the topic of the isolated sentence.
- In *condition D,* the target word is directly mentioned in the background context. The structure is like "Wangying he Lixiao dou you shi-qing yao zuo. Wangying can-guan che-jian. (Both Wangying and Lixiao have things to do. Wangying visits the workshop)". In the first sentence, Wangying is new and is considered as the topic of the discourse-initial sentence, whereas Wangying in the second sentence is given and is the topic in a non-initial sentence. Similar to condition A, the topic in the second sentence involves a contrast as the speakers need to select one out of the two possible persons in the preceding discourse as the topic of the sentence. Also as in condition A, because 'to visit the workshop' is new information too, the second sentence does not contain a single initial focus.

The other two controlled factors, sentence length and downstep, were manipulated by varying the sentence length and lexical tones of the base sentences. Sentences with three different lengths were used: 6, 14 and 20 syllables. The short sentences were the base sentences. The medium and long sentences were constructed by adding modifiers before the object, keeping the

Table 1
Illustrations of stimulus conditions. Condition A is the base-line and "*Wangying*" is the target word with the features of being topic, new, and non-initial. In each of conditions B, C, D[a] and D[b], there is just one of these features different from those in condition A, which is marked in bold letters. The last column presents whether the topic is contrastive or thematic on the sentence level.

Conditions	Example	Focus/Topic	New/Given	Discourse topic	Sentence topic
A	Dajia dou you shiqing yao zuo. **Wangying** canguan chejian (Everybody has something to do. Wangying visits the workshop)	Topic	New	Non-initial	Contrastive
B	Bu shi Lixiao, **Wangying** canguan chejian (It is not Lixiao. Wangying visits the workshop)	**Focus**	New	Non-initial	-
C	**Wangying** canguan chejian (Wangying visits the workshop)	Topic	New	**Isolated**	**Thematic**
D	**Wangying[a]** he Lixiao dou you shiqing yao zuo **Wangying[b]** canguan chejian (Both Wangying[a] and Lixiao have something to do. Wangying[b] visits the workshop)	[a]Topic [b]Topic	[a]New **[b]Given**	**[a]Discourse-initial** [b]Non-initial	**[a] Thematic** [b] Contrastive

subject (the first word), the verb (the second word) and the object (the last word) the same as in the short sentences. All the experimental sentences consisted of only disyllabic content words.

The four lexical tones (not including the neutral tone) in Mandarin, H (also known as Tone 1), R (Tone 2), L (Tone 3), and F (Tone 4), have the pitch contours of high-level, mid-rising, low-dipping, and high-falling, respectively, when produced in isolation (cf. Chao, 1968). The basic stimuli consisted of five base sentences with different tone combinations (HH, HL, LH, FR, and RF). Compared with the HH sentences, the other sentences would all present a downstep effect with varying degrees according to previous research (e.g., Xu, 1999). A full list of all the experimental sentences is shown in Appendix A.

In total, there are 180 utterances for each speaker: 5 (tone combinations) × 3 (sentence lengths) × 4 (contexts) × 3 (repetitions)= 180 (utterances).

2.2. Participants

Participating in the experiment were four female and two male native speakers of Mandarin. Two of the speakers were from Beijing, and the others were from North China (Dongbei and Henan province) with no noticeable accent in their Mandarin, although they spoke their own dialects as well. They did not report any speech and hearing impairments. They were students at University of Potsdam, aged 22–30. They had stayed in Germany for about two years, and had continued to speak Mandarin over the last two years. They were paid a small amount of money for their participation.

2.3. Recording procedure

The participants were recorded individually in a quiet room at University of Potsdam. They were asked to read aloud both the context sentences and the target sentences at a normal speed and in a natural way. Importantly, they were instructed to read the target sentences in a single breath without pause. For sentences in condition B, they were reminded during the practice that they should emphasize the first word of the sentence although the context would already naturally elicit initial focus (see: Cooper et al., 1985; Xu, 1999). This is a measure only to reduce noise in the data, as we are not testing a new focus-conditioning factors previously not reported. For sentences in the other three topic conditions, apart from being reminded not to specifically emphasize the subject as they would do in condition B, they were not given any specific instruction on how to utter them.

The subject wore a head-mounted microphone and sat in front of a computer monitor on which the test sentences were displayed. During the recording, when the experimenter determined that a particular sentence was not said properly, the subject was asked to say the whole discourse again. All sentences were digitized directly into the computer and saved as wav files. The sampling rate was 48 kHz and the sampling format was one channel 16-bit linear. Each participant read the sentences four times in four sessions, with 5-min breaks in between. In each session, all the 60 sentences were randomized, and each speaker had a different random order. The total recording time was about 45 min. The last three repetitions were used for data analysis.

2.4. Acoustic measurement

The individual target sentences were extracted and saved as separate wav files. The acoustic analysis procedures were similar to those in Xu and Xu (2005). A Praat script (Xu, 2005–2010) was used to take F_0 and duration measurements from the target sentences. To extract continuous F_0 contours, the vocal cycles were first marked by Praat (Boersma & Weenink, 2005) and then hand checked by the first author for errors, such as double-marking and period skipping. While checking for vocal pulse markings, segmentation labels were also added to mark the syllable boundaries. The vocal periods were converted into F_0 values by the Praat script, which then smoothed the resulting F_0 contours using a trimming algorithm (Xu, 1999, 2005–2010). The vocal pulse marking, segment labels, and F_0 values for each utterance were saved in text files. Those files were then processed by a custom-written Perl program to calculate the highest and lowest F_0 and the duration of each syllable.

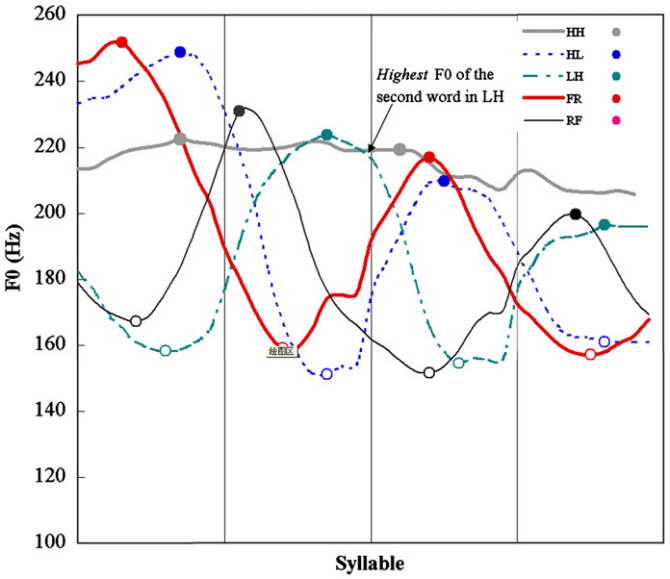

Fig. 1. Illustration of measurement points for maximum F_0 (solid dots) and minimum F_0 (empty dots) in the first two words of the five tone combinations.

The highest and lowest F_0 values of each word were measured. Due to tonal variations, these values were calculated differently for different tones. Fig. 1 gives an example of F_0 contours of two adjacent words for the five tone combinations. Note that one cannot simply treat the highest F_0 of a word as its maximum F_0 in the LH sentences, because the highest F_0 of the second word is at the boundary between the first and second words (see Fig. 1), which is carried over from the preceding word due to articulatory inertia (Gandour et al., 1994; Xu, 1997). To avoid this problem, we applied the same method as Chen and Gussenhoven (2008). The highest F_0 points of the H tone in HL and LH words are calculated as word maximum F_0, and the lowest F_0 points of the L tone are calculated as word minimum F_0. Similarly, for sentences with FR or RF words, the maximum F_0 was taken from the F tone and the lowest F_0 from the R tone. For the all-H-tone sentences, only maximum F_0 was obtained. The F_0 values of Hz were converted to semitone (st), using 50 Hz as the reference for all the speakers in the following:

$$st = 12 \times \log_2(F_0/50) \quad (1)$$

3. Analysis and results

3.1. General analysis strategy

Our first goal is to investigate how topic and focus interact with the other two factors, i.e., sentence-length and downstep, in

determining the F_0 height and the amount of F_0 lowering in sentence-initial position (Section 3.2). Three steps are taken to compare condition A (topic) and condition B (focus). First, an overall comparison is made to detect main factors and interactions using a three-way repeated measures ANOVA with context, sentence length and tone as the independent variables. Based on the overall statistical results, we will further examine whether the effect of sentence length on sentence initial F_0 applies in both topic and focus conditions (Section 3.2.1). We will then examine the effect of tonal variation on the encoding of topic and focus (Section 3.2.2).

The second research goal is to investigate whether given/new is an independent factor in determining sentence-initial F_0 height. Topic as given (condition A) and new (condition D[b]) are compared in Section 3.3.

The last goal is to investigate whether topics at different discourse-levels differ in sentence-initial F_0 (Section 3.4). Topics in discourse-initial (condition D[a]), non-initial (condition A) and isolated sentences (condition C) are compared. The issue of whether contrastive topic is encoded prosodically will also be discussed in this section.

3.2. Topic vs. Focus

First, we present an overall analysis of topic (condition A) and focus (condition B). The two conditions differ only between topic and focus while all the other potential factors, such as status of newness and location of sentence in discourse, are controlled (see Table 1).

The F_0 contours of focus (condition B) and topic (condition A) in non-initial sentences are displayed in Fig. 2. Short, medium and long sentences are displayed in columns, while sentences with HH, HL, LH, FR, and RF tone sequences are displayed in rows. The F_0 plots in Fig. 2 were generated by a Praat script (Xu, 2005–2010), with ten points taken from each syllable at equal proportional intervals. For each point, the F_0 values were averaged across three repetitions by six speakers applying formula (2) so that the contributions of different speakers, especially males and females were equally weighted (Xu, 2005). Also the time value of each point was averaged across both the repetitions and the speakers so that we could see the tonal contours of different sentences without losing sight of the actual duration of

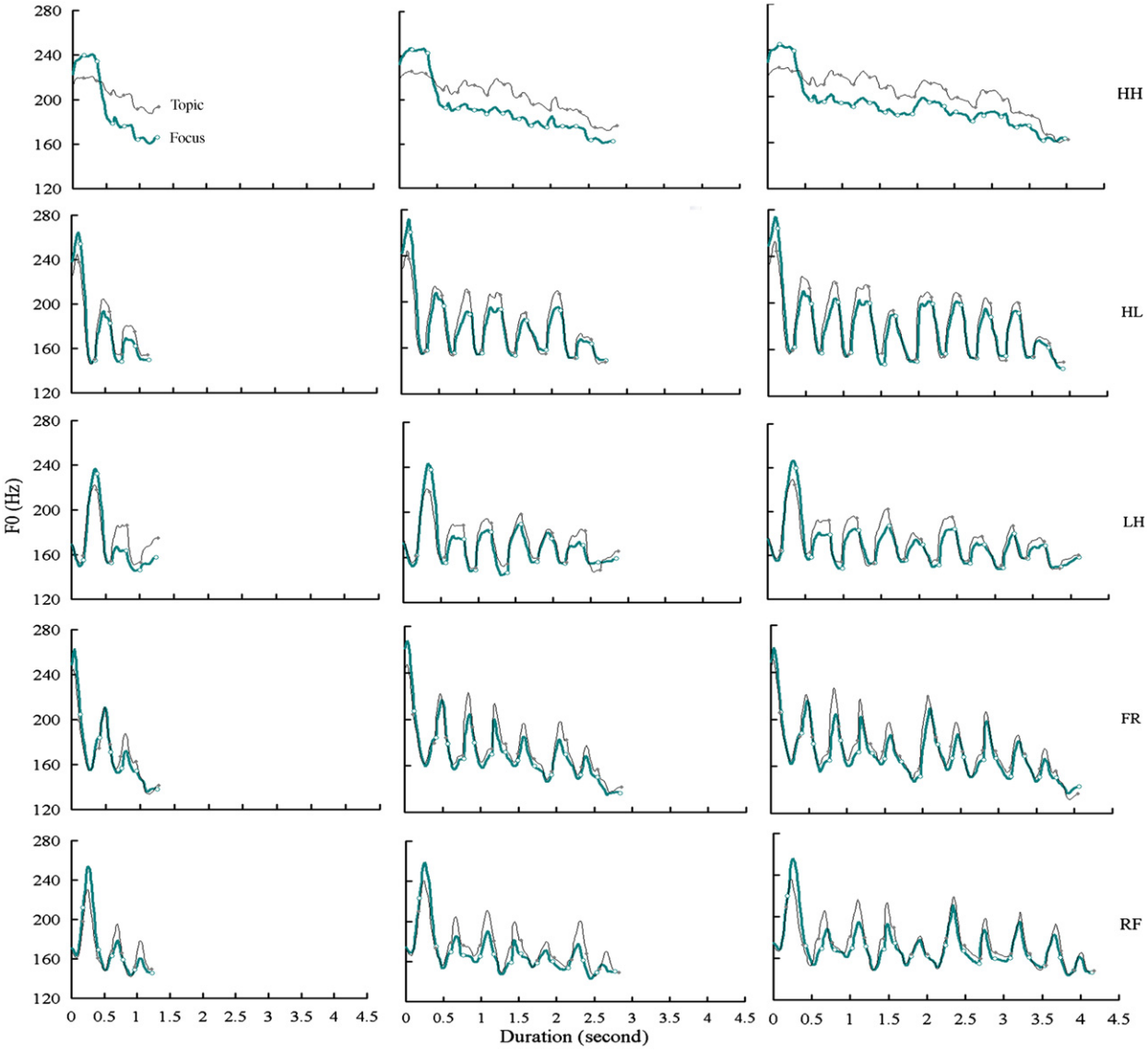

Fig. 2. Overlaid F_0 contours of non-initial sentences with the target words as topic (conditions A) or focus (condition B). Sentences with different lengths and tone combinations are displayed separately. The thin gray lines are from the topic condition (A), and the thick green lines are from the focus condition (B). The small circles indicate syllable boundaries (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

each syllable. In the graphs, the *x*-axis displays time in second, and the *y*-axis displays F_0 in Hz

$$F_0 = 10^{(\sum_1^6 \log F_{0_i})/6} \quad (2)$$

Table 2
The mean values of maximum and minimum F_0 of the *first* word in the target sentences broken down by tone, sentence length and focus conditions (topic vs. focus, condition A and B, respectively).

	Topic (condition A)				Focus (condition B)			
	Short	Medium	Long	Average	Short	Medium	Long	Average
$MaxF_0$								
HH	25.9	26.2	26.2	26.1	27.4	27.8	27.6	27.6
HL	27.6	27.6	28.3	27.8	29.0	29.5	29.7	29.4
LH	26.0	25.8	26.4	26.1	27.0	27.5	27.7	27.4
FR	27.7	27.9	28.1	27.9	29.0	29.4	28.9	29.1
RF	26.7	27.4	27.0	27.0	28.2	28.6	28.4	28.4
Average	26.8	27.0	27.2	27.0	28.1	28.6	28.4	28.4
$MinF_0$								
HL	18.4	19.3	19.7	19.1	18.4	19.3	19.6	19.1
LH	18.8	19.1	19.5	19.1	19.1	19.6	19.7	19.5
FR	19.4	19.9	20.1	19.8	19.5	20.3	20.5	20.1
RF	20.4	20.9	20.9	20.7	20.1	20.6	20.7	20.5
Average	19.3	19.8	20.0	19.7	19.3	20.0	20.1	19.8

Table 3
The mean values of maximum and minimum F0 of the *second* word in the target sentences broken down by tone, sentence length and focus conditions (topic vs. focus, condition A and B, respectively).

	Topic (condition A)				Focus (condition B)			
	Short	Medium	Long	Average	Short	Medium	Long	Average
$MaxF_0$								
HH	25.3	24.9	25.4	25.2	22.6	23.9	23.5	23.3
HL	24.7	25.4	26.1	25.4	23.7	24.8	25.2	24.6
LH	23.0	23.2	23.5	23.2	21.1	22.3	22.4	21.9
FR	25.1	26.0	26.0	25.7	25.1	25.6	25.5	25.4
RF	23.7	24.6	24.7	24.3	22.2	22.9	23.1	22.7
Average	24.4	24.8	25.1	24.8	22.9	23.9	23.9	23.6
$MinF_0$								
HL	19.2	19.4	19.9	19.5	18.6	19.5	19.7	19.3
LH	19.1	19.7	20.1	19.6	19.1	19.4	19.6	19.4
FR	19.6	20.2	20.2	20.0	19.1	19.8	19.6	19.5
RF	18.7	19.6	19.2	19.2	18.7	19.3	19.3	19.1
Average	19.2	19.7	19.8	19.6	18.9	19.5	19.6	19.3

The first observation of Fig. 2 is that focus (condition B) has higher F_0 than topic (condition A) in sentence initial position, which holds true for all sentence lengths and tone combinations. To test the observation statistically, the mean values of maximum and minimum F_0 of the *first* word in sentences of the two conditions are presented in Table 2, broken down by sentence length and tone. Similar results from the *second* word are presented in Table 3.

Three-way repeated measures ANOVAs on maximum and minimum F_0 were carried out separately for the first and the second word, followed by post-hoc tests (Student–Newman–Keuls) and simple-effect tests for significant interactions. The independent variables are context, sentence length and tone. All the statistical results are presented in Table 4.

For the three main factors, Table 4 shows significant effect of context in both the first and second words on maximum F_0 but not on minimum F_0. As can be seen in Table 1, the first word has lower maximum F_0 in the topic condition than in the focus condition (27.0 vs. 28.4 st). In the second word, maximum F_0 is higher in the topic condition than in the focus condition (24.8 vs. 23.6 st). The effect of sentence length on sentence initial F_0 is shown in both the first and the second word as well; however, the effect is shown in minimum F_0 of the first word and maximum F_0 of the second word. Finally, the effect of tone on both words holds only for maximum F_0. For minimum F_0, a significant effect of tone is found only in the first word. Interestingly, only one interaction is found, indicating that the three factors are mostly independent of each other. Below we will give a detailed analysis on how topic and focus interacts with the other two factors, namely, sentence length and downstep.

3.2.1. *Interaction of sentence length with focus and topic*

The second question raised in this study is whether there is an independent contribution of sentence length to sentence-initial F_0. Fig. 3 shows maximum and minimum F_0 of the first and second words in topic and focus conditions with sentences of three lengths (data from Tables 2 and 3).

As shown in Table 4, sentence length has an effect on minimum F_0 but not maximum F_0 of the first word. Results of the post-hoc tests show that medium and long sentences have higher minimum F_0 than short sentences in the first word. Moreover, no interaction between sentence length and context is found. As can be seen in Fig. 3, the raising of minimum F_0 due to sentence length applies in both topic and focus conditions.

Table 4
The *F*- and *p*-values of the three-factor repeated measures ANOVAs of the maximum and minimum F_0, carried out separately in the first, and the second word, together with significant comparisons by corresponding Post-hoc tests (Student–Newman–Keuls).

	First word		Second word	
	Main effect	Post-hoc test (st)	Main effect	Post-hoc test (st)
$MaxF_0$				
Context	$F(1, 5)=18.807$, $p=0.007$	–	$F(1, 5)=20.81$, $p=0.006$	–
Length	$F(2, 10)=5.438$, *n.s.*	–	$F(2, 10)=6.871$, $p=0.044$	Short (23.7) < Medium (24.4), Long (24.5)
Tone	$F(4, 20)=13.012$, $p<0.001$	HH (26.8) & LH (26.7) < FR(27.7), HL(28.6), FR(28.5)	$F(4, 20)=14.409$, $p<0.001$	LH(22.6) < HH (24.3), RF (24.2) < HL (25.0), FR (25.6)
Interaction	no significant interaction	–	Context & Tone: $F(4, 4)=4.842$, $p=0.001$	Topic: LH(23.3) < RF(24.4) < HH(25.4), HL(25.4), FR(25.7); Focus: LH(21.9), RF(22.7) < HH (23.3), HL(24.6) < FR(25.4)
$MinF_0$				
Context	$F(1, 5)=0.219$, *n.s.*	–	$F(1, 5)=2.467$, *n.s.*	–
Length	$F(2, 10)=12.88$, $p=0.012$	Short(19.3) < Medium(19.9),ng(20.1)	$F(2, 10)=4.31$, *n.s.*	–
Tone	$F(3, 15)=9{,}899$, $p=0.001$	HL(19.1),LH(19.3),FR(19.8) < RF(20.7)	$F(3, 15)=0.491$, *n.s.*	–
Interaction	No significant interaction	–	No significant interaction	–

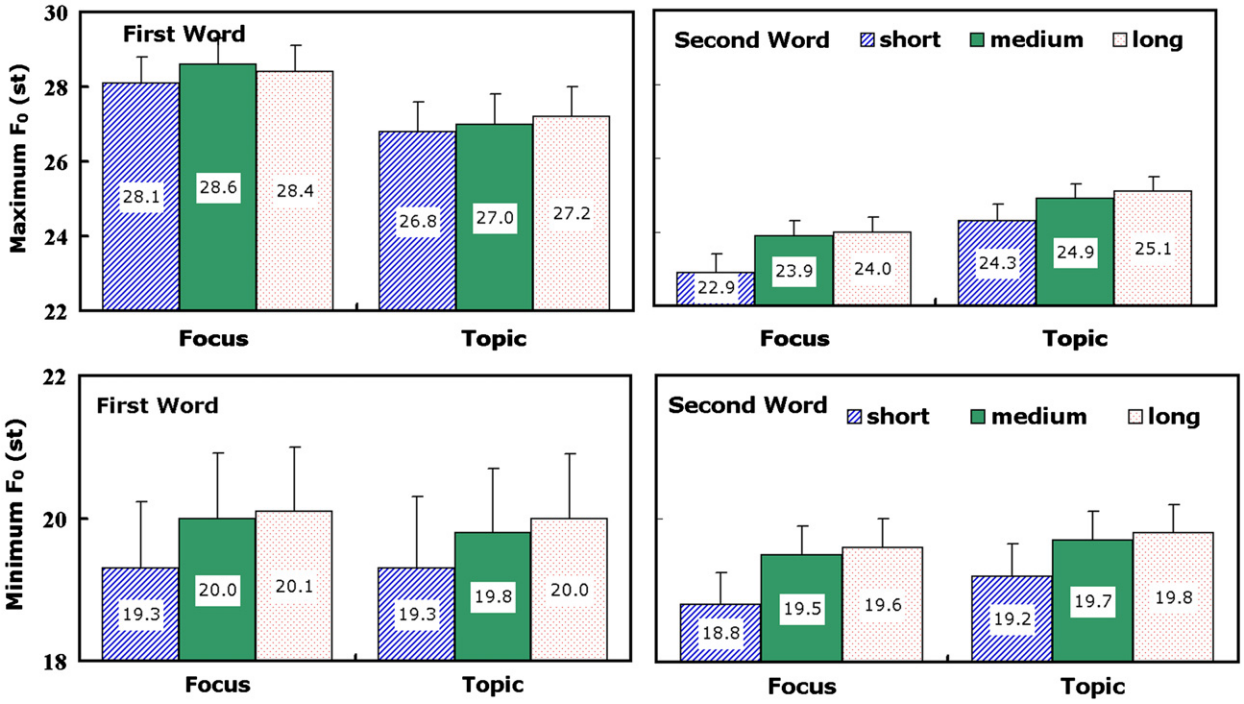

Fig. 3. Maximum and minimum F_0 of the first (left column) and second (right column) words in focus (condition B) and topic (condition A) sentences with different lengths.

In the second word, as shown in Table 4, sentence-length has an effect on maximum but not minimum F_0, and there is no interaction with context. Fig. 3 shows that maximum F_0 of the second word is higher in longer sentences than in shorter sentences, which holds in both topic and focus sentences. The results indicate that there is less post-focus lowering in longer sentences. Moreover, post-hoc tests reveal that such a difference exists only between short and medium/long sentences, but not between medium and long sentences.

Finally we can see in Table 4 that the effect of sentence length on sentence initial F_0 does not interact with tone. In Tables 2 and 3, we can see that the higher F_0 in longer sentences is true of almost all tone conditions.

3.2.2. *Interaction of downstep with focus and topic*

Here we will first look at the HL, LH and HH sentences, since the effect of downstep due to the L tone is well established in previous studies (see Section 1.4). As can be seen in Fig. 2, F_0 drops greatly after the first word in HL and LH sentences, but not much in HH sentences, which is known as the downstep effect. Table 5 presents the sentence-initial F_0 drop, which is calculated as the difference in maximum F_0 (st) between the H tone of the first two words in the HH, HL and LH sentences divided by focus (condition B) and topic (condition A) condition in short, medium and long sentences.

A repeated measures ANOVA on the amount of F_0 drop is carried out with three main factors, context (topic vs. focus), sentence length (short, medial and long) and tone (HH, HL and LH). The results show no effect of sentence length. Significant effects are found for context ($F(1, 5)=21.533$, $p=0.006$) and tone ($F(2, 5)=5.672$, $p=0.023$) The effect of context can be seen in Table 5, as F_0 drop is smaller in topic context than in focus context. The effect of tone is due to the fact that F_0 drops are greater in HL and LH tone sequences than in HH tone sequence. A post-hoc test (Student–Newman–Keuls) shows that the difference between HH and HL, and between HH and LH are both significant, but non-significant between HL and LH sentences. There is no interaction between context and tone, suggesting that F_0 lowering due to downstep does not interact with topic and focus.

Table 5
Columns 1–3: Difference in maximum F_0 (st) between the H tone of the first two words in the HH, HL and LH sentences divided by focus (condition B) and topic (condition A) condition in short, medium and long sentences. Columns 4–5: Difference in the amount of F_0 lowering (st) between HL and HH sentences and between LH and HH sentences.

	HH	HL	LH	HL − HH	LH − HH
Short					
Topic	1.0	2.8	2.9	1.8	1.9
Focus	4.8	5.3	5.9	0.5	1.1
Medium					
Topic	1.2	2.2	2.6	1.0	1.3
Focus	3.9	4.7	5.2	0.8	1.3
Long					
Topic	0.8	2.1	3.0	1.3	2.1
Focus	3.6	4.5	5.3	0.9	1.7

The F_0 drop in a HL and LH sentence comes from at least two sources: downstep due to the low tone and the lowering due to topic or focus, whereas the F_0 drop in a HH sentence comes only from topic or focus. If we subtract the F_0 drop in the HH sentences from that in the HL and LH sentences, the effect of downstep can be calculated, as shown in the 6th and 7th column of Table 5. A similar three-way repeated measures ANOVA on these differences in F_0 drop shows that none of the three main effects, namely, context, length and tone, is significant, nor are there any interactions. This indicates that the amount of downstep is roughly the same across sentences with different lengths and between topic and focus contexts.

As mentioned in the introduction, downstep may not be a simple effect, but rather a combination of anticipatory raising and carryover lowering. To examine this possibility, we did a graphic analysis by putting the HH sentences together with the sentences with HL tones in both topic and focus conditions, as shown in Fig. 4. (The sentences with LH, FR and RF tones overlaid with HH sentences in Fig. 4 will be discussed later.) A similar time-normalization procedure as used in generating Fig. 2 was applied, except that, instead of using averaged real time for the *x*-axis, the ten mean F_0 values in each syllable are plotted evenly along the *x*-axis. The short and long sentences are not shown in the figure because they would make the plot crowded, and because we know from the results above that the effect of sentence length on F_0 in sentence initial position is independent of that of tone (Section 3.1).

About the HH and HL sentences (the first row in Fig. 4), two observations can be made. First, anticipatory-raising due to the following L tone works separately from the F_0 raising by focus, since the first H tone in the HL sentence is higher than the corresponding H tone in the HH sentence in both topic and focus sentences, see Table 2 (27.6 vs. 29.4 st in HH and HL in the focus condition, 26.1 vs. 27.8 st in the topic sentences). This observation is supported by a repeated measures ANOVA with sentence length (short and medium), context (topic and focus) and tone (HH and HL) as independent variables, which shows main effects of tone ($F(1, 5)=134.867$, $p<0.001$) and context ($F(1, 5)=24.506$, $p=0.004$), but not sentence length, nor are there any interactions. In Table 2, we can see that the HL tone sentences start with higher maximum F_0 across the three sentence lengths.

Second, The amount of anticipatory-raising and that of carryover-lowering are roughly the same, since for the HL sentences, from the second H tone on, the height of the H tones seems to be determined by both the raising and lowering effects, and as a result it has almost the same height as the H tones in the HH sentences in both topic and focus conditions, see Table 3 (24.6 vs. 23.3 st in HL and HH under the focus condition, and 25.4 vs. 25.2 st under the topic condition). The observation is supported by a similar repeated measures ANOVA on maximum F_0 of the second word, which shows an effect of sentence length ($F(2, 5)=8.696$, $p=0.006$) and context ($F(1, 5)=13.019$, $p=0.015$), but no effect of tone. Nor are there any interactions.

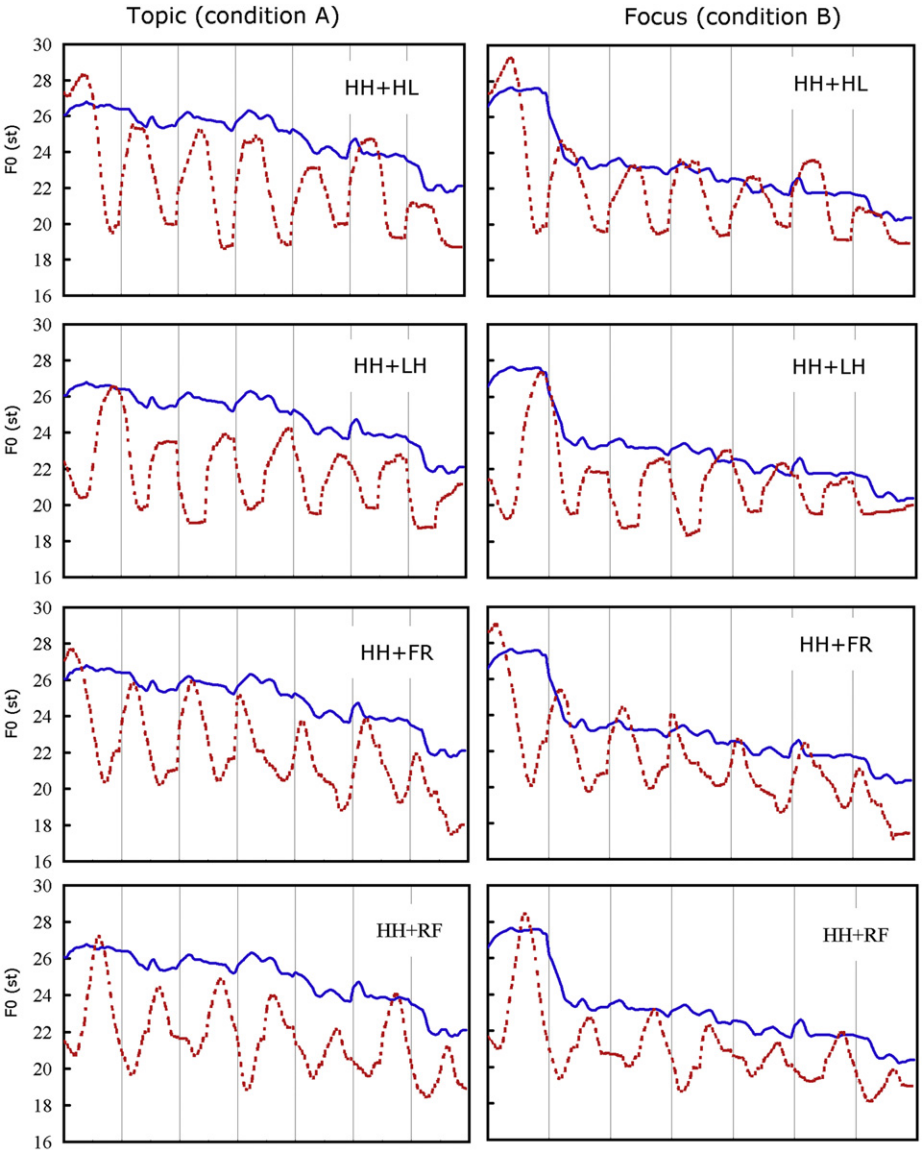

Fig. 4. F_0 contours of HH tone sentences overlaid with sentences of the other four tone conditions (from top to bottom: HL, LH, FR, and RF) in topic (condition A, left column) and focus (condition B, right column) conditions.

Moreover, the fact that from the second word on the H tones do not differ in F_0 between the HL and HH sentences also implies that the post-focus lowering is independent of downstep, because such an effect holds in both focus and topic conditions. Similar patterns can be seen in the LH and HH sentences as shown in the second row of Fig. 4.

For the FR and RF sentences, as displayed in Fig. 4, maximum F_0 is higher in FR sentences than in HH sentences in both topic (27.9 vs. 26.1 st) and focus conditions (29.1 vs. 27.6 st). Sentences with RF tones also start with higher maximum F_0 than HH sentences in the two conditions (27.0 vs. 26.1 st in the topic condition, and 28.4 vs. 27.6 st in the focus condition). These two observations are statistically supported by the corresponding post-hoc tests in Table 4. The second word, however, shows significant interaction between context and tone, which is due to different realization of FR and RF tones in the two conditions. As can be seen in Table 4, maximum F_0 of the second word is significantly higher in RF tone sentences than in LH sentences in the topic condition, but not in the focus condition. Besides, the second word of FR tone sentences has higher maximum F_0 than HH and HL sentences in the focus condition, but not in the topic condition.

In general, the data show that the realization of F_0 due to tonal articulation is mostly independent of topic and focus, especially in the first word.

3.3. New vs. Given

The third factor investigated in this study is whether givenness plays a role in determining sentence-initial F_0. Condition A and D[b] are compared. In condition A, the target word is not mentioned before, so it is new. In condition D[b], the target word is mentioned in the previous sentence, so it is given. In both conditions, the sentences are non-initial and the target words are the topic. Moreover, the first word has a contrastive meaning in both cases, since it is selected out of a group of other people as in condition A or two persons as in condition D[b]. However, the two conditions are not different in maximum F_0 (27.0 st in both conditions) or minimum F_0 (20.0 st in both conditions).

Table 6
Duration (ms) of the first word in new (condition A) and given (condition D[b]) topic sentences.

	Short	Medial-length	Long	Average
New	388	391	386	388
Given	377	369	365	370

Word duration of new (condition A) and given (condition D[b]) topic of all sentence lengths is shown in Table 6. A two-way repeated measures ANOVA with sentence length (short, medium, and long) and context (new vs. given topic) as independent factors shows an effect of context ($F(1, 5)=13.079$, $p=0.015$) but not length, with no interaction between the two. Newness thus seems to increase duration (5% on average).

3.4. Topic at different discourse levels

The last question to be answered is whether there are differences between topics at different discourse levels. As discussed in the literature review, a topic in an isolated sentence has very different F_0 from that in a discourse-initial or non-initial sentence. Here we will compare sentence initial F_0 in condition A (non-initial), C (isolated) and D[a] (discourse-initial). We need to mention that topic in condition A can be described as a contrastive topic, whereas it can be described as a thematic topic in condition C and D[a], which will be put into consideration when we interpret the results in general discussion. Since the discourse-initial sentences here have 11 syllables, which is roughly the same length as medium sentences (14 syllables), we will only analyze medium sentences in this section. Fig. 5 displays maximum F_0 and minimum F_0 of the first word in the three topic conditions, broken down by tone.

For maximum F_0, a two-way repeated measures ANOVA shows significant effects of topic-level ($F(2, 10)=6.212$, $p=0.018$), tone ($F(4, 20)=14.053$, $p<0.001$) and the interaction between the two factors ($F(8, 40)=2.34$, $p=0.036$). A post-hoc test (Student–Newman–Keuls) shows that the first word has significantly higher maximum F_0 in discourse-initial sentences (28.0 st) and isolated sentences (27.7 st) than in non-initial sentences (27.1 st), with no significant difference between the former two conditions. The interaction between tone and topic level lies in the fact that for HH tone sentences, the differences of maximum F_0 between each of the three topic levels are all significant, whereas for the other sentences, there is significant difference between non-initial and isolated, and between non-initial and discourse-initial, but not between isolate and discourse-initial conditions (see Fig. 5).

A two-way repeated measures ANOVA on minimum F_0 shows an effect of discourse level ($F(2, 10)=10.206$, $p=0.004$), tone ($F(3, 15)=13.802$, $p<0.001$) and an interaction between them ($F(6, 30)=2.867$, $p=0.025$). A simple effect test shows that for LH, FR, RF tones, discourse-initial and isolated sentences start with higher minimum F_0 than discourse non-initial sentences. However, no difference is found in HL sentences.

In summary, topic at a higher discourse level (discourse-initial and isolated) has higher F_0 register than topic in a non-initial sentence.

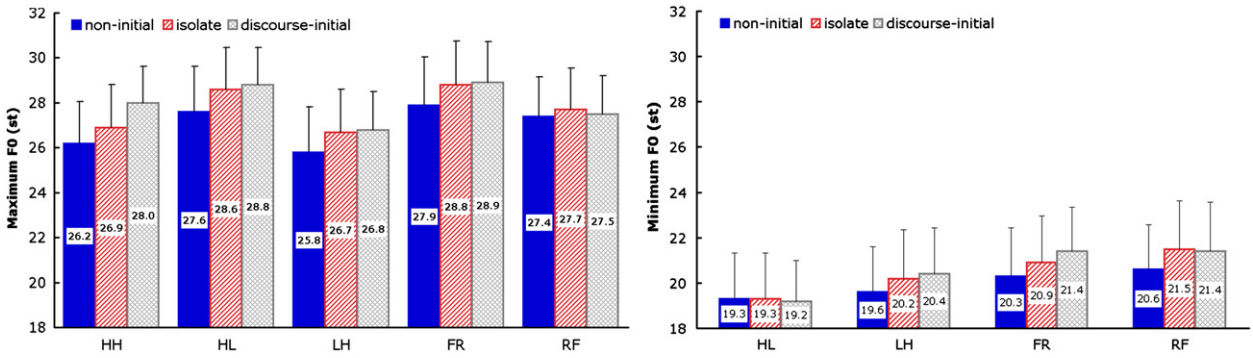

Fig. 5. Maximum F_0 and minimum F_0 of the first word in the three topic conditions, broken down by tone.

4. General discussion

Five questions were raised at the outset of the study, each in regard to a source that would potentially contribute to the sentence-initial declination often observed in previous research, which are repeated below for ease of reading.

(1) What are the different prosodic manifestations of topic and focus in terms of F_0 in sentence initial position?
(2) Is there an independent effect of sentence length on sentence-initial F_0? If yes, how does it interact with topic and focus in determining sentence-initial F_0?
(3) How does downstep interact with topic and focus in determining sentence-initial F_0?
(4) Dose a sentence start with higher F_0 when the initial word is new than when it is given?
(5) For discourse topic, how do topics at different discourse levels differ in terms of sentence-initial F_0?
(6) Are the effects of all these factors on sentence-initial F_0 cumulative?

In the following discussion we will examine the data obtained about each of the questions and see how well they meet our predictions in Section 1.7. Apart from the five questions, the effect of contrastiveness will also be discussed. In the end, we will attempt to integrate all the findings to see whether the overall picture will can be interpreted in terms the articulatory-functional view of speech prosody as represented by the PENTA model (Xu, 2005).

4.1. Topic versus focus

Two main conclusions can be drawn in regard to the F_0 manifestations of topic and focus: (1) Focus lowers maximum F_0 of post-focus words (4.3 st on average in the H-tone sentences), whereas topic leads to only gentle F_0 lowering after the sentence initial word (0.9 st on average in the H-tone sentences). (2) Sentence-initial focus raises maximum F_0 and leaves minimum F_0 intact, whereas topic at a higher discourse level shows both higher maximum F_0 and minimum F_0. In the following we will discuss these two points in greater detail.

4.1.1. Post-focus F_0 lowering and post-topic F_0 drop

Consistent with previous findings (e.g., Cooper et al., 1985; Féry & Kügler, 2008; Xu, 1999; Xu & Xu, 2005), robust post-focus F_0 lowering is observed in sentences with initial focus. Moreover, post-focus lowering seems to be executed with such a strong force that F_0 is lowered to the post-focus level well within the third syllable, as can be seen in the HH sentences in Fig. 2. But such a force sometimes is compromised by on-focus pitch range expansion force, which is presumably also strong. This is seen in FR sentences (see Fig. 2), where the on-focus force increases the rising slope of the R tone in syllable 2, generating an opposing force to that of post-focus lowering, which weakens the lowering of maximum F_0 in syllable 3 (cf. Chen & Xu, 2006 for detailed analysis of final velocity of a tone on the following tone). The post-focus force eventually manifests itself by strengthening the falling slope of the F tone in syllable 3, thus helping to lower the minimum F_0 in syllable 4. Such effect of opposing forces can be also seen when comparing HL and LH sentences, where the movement toward the L tone in syllable 3 helped the lowering of maximum F_0 in syllable 4 in the LH sentences.

In contrast to initial focus, the F_0 drop between the first and second words due to topic is small (0.9 st in all-H-tone sentences). The initial F_0 drop in HL and LH sentences is much bigger (2.7 st on average). However, as discussed in Section 3.2.2, most of this drop is due to downstep. Moreover, between the second and third words, there is almost no further F_0 drop as shown in Figs. 2 and 4. Thus F_0 drop from the first word to the second is much bigger than in the following declination, which is similar to previous reports for Dutch (Van Heuven, 2004) and Mandarin (Shih, 1997).

4.1.2. F_0 raising by focus and by discourse-initial topic

In the present data, topic in discourse-initial sentences does not have a higher maximum F_0 than initial focus (28.0 vs. 28.3 st). However, this does not necessarily mean that a discourse-initial topic cannot have higher F_0 than initial focus. Larger F_0 raising has been reported in sentences that start a new paragraph and a conversational turn (Lehiste, 1975; Nakajima & Allen, 1993; Umeda, 1982). It is possible that the experimental paradigm employed in the present study could not elicit the full effect of discourse-initial topic, as a two-sentence discourse may not be long enough. Another possibility is that raising F_0 dramatically to introduce a new topic is a skill in formal speech that has to be acquired, as suggested by Umeda (1982). The participants in the present study, being undergraduate and graduate students with little experience in lecturing or other form of public speech, probably have not yet mastered this skill. It would thus be interesting to carry out more research into how topic realization may vary with text type (e.g., news vs. story), speaker attribute (e.g., professional vs. non-professional) and speaking style (e.g., conversation vs. monologue speech), etc.

Nevertheless, the present data, for the first time, have presented direct evidence on the critical difference between focus and topic in terms of F_0 control: focus expands the pitch range of the focused word and compresses the pitch range of post-focus words, whereas discourse-initial topic raises the initial pitch register, and lets the pitch of subsequent words drop gradually. Thus our initial prediction on the first question was largely confirmed.

The results on the effect of sentence length are largely consistent with our initial prediction, although the magnitude of the effect is quite small. In the present data, the magnitude of F_0 raising due to focus in long sentences is roughly to the same degree as that in short sentences. However, Eady et al. (1986) reported that shorter sentences show a greater effect of focus than longer sentences, as found in their former studies (Cooper et al., 1985; Eady & Cooper, 1986). While this could be due to a language-specific difference, research on focus in both English and Mandarin has shown many similarities between the two languages (see Xu, 1999 for Mandarin, and Xu & Xu, 2005 for English). More research is therefore needed on the interaction between sentence length and sentence-initial focus in English.

These results also shed some light on the issue of pre-planning, as they are in general consistent with the limited pre-planning hypothesis (Liberman & Pierrehumbert, 1984; Van Heuven, 2004). First, initial maximum F_0 and minimum F_0 are slightly lower in short sentences than in medium and long sentences, which is in agreement with Bruce (1982b) and Shih (1997). However, there is no significant difference between medium and long sentences. The fact that both maximum F_0 (0.4 st on average) and minimum F_0 (0.6 st on average) are higher in shorter than in longer sentences indicates that the whole pitch range is raised while starting a longer sentence. Second, as shown in Fig. 3, both maximum and minimum F_0 of the second word are also lower in short sentences than in medial/long sentences. This suggests that the effect of sentence length is independent of focus and

topic, because the higher F_0 in both the first and second words in longer than in shorter sentences goes against the direction of pitch change between the first and second words due to either topic or focus as shown in Table 2. Such independence is further supported by the lack of interaction of sentence length with any other factors.

4.2. Downstep

As predicted, downstep introduces large F_0 drops in the sentence-initial position, and in fact, much of the initial F_0 drop in a sentence without focus comes from downstep rather than from topic. The evidence is multifold. First, in the HH sentences where downstep does not apply, the F_0 drop across the H tones in the first two words is only about 1 st on average. In contrast, the amount of F_0 drop in HL and LH sentences are much larger (2.4 st and 2.8 st on average, respectively). Second, the amount of downstep in sentence-initial position is unrelated to sentence length, indicating independence of the two effects. Finally, the amount of sentence-initial downstep is unaffected either by focus or by topic level, again indicating independence of downstep. The most important implication of these findings is that the magnitude of the effect of topic can be easily overestimated if downstep is not taken into full consideration. This is relevant not only for tone languages, but also for languages such as English where lexical stress, just like lexical tone, introduces local F_0 undulations (Fry, 1958; Xu & Xu, 2005). Such local F_0 undulation may also lead to downstep, as has already been recognized (e.g., Liberman & Pierrehumbert, 1984; Van Heuven, 2004). There is therefore a need for future studies of topic and focus in non-tone languages to pay special attention to potential contributions of downstep.

4.3. F_0 correlates of newness: are there any?

Contrary to our initial prediction, we did not find any clear F_0 difference related to newness. This seems to agree with Horne (1990) but contradicts many other studies, e.g., Prince (1981), Nooteboom and Terken (1982), Brown (1983), Nooteboom and Kruyt (1987) and Féry and Kügler (2008). The disagreement is likely due to the way newness is experimentally controlled. In this study, we made *direct comparisons* between given and new by keeping everything else equal. In a comparison between conditions A (new) and D[b] (given) (as shown in Table 1), the only variable is whether the information has been mentioned before. In the studies mentioned above, some other factors beside the givenness of the target words are not explicitly controlled. For instance, in Féry and Kügler (2008), the location and type of focus are also varied in the comparison. And in the spontaneous speech data of Brown (1983), *a* direct comparison between given and new is not possible.

Another way to compare given and new in this study is to compare conditions D[a] and D[b], which we may call a *sequential comparison*. The first word is new in D[a], but given in D[b]. The maximum F_0 of the first word in these two conditions is indeed different (28.0 vs. 27.1 st). However, we cannot conclude from this that given information has lower F_0 than new information, because other factors, such as the position of the embedded sentence in the discourse, sentence length and local tonal environment, are all different. Although sentence length can be still controlled in *sequential comparison*, location in discourse is always confounded with newness. Thus whatever the difference between D[a] and D[b], it has at least two sources: location in discourse and newness. As seen in the data discussed so far, location in discourse has a strong impact on F_0, and so could account for much of the previously reported differences between new and given.

4.4. Topic at different discourse levels

Consistent with our initial prediction, the results show that topic at a higher level (e.g., discourse-initial) raises pitch register more than topic at a lower level (e.g., non-initial). This is consistent with the findings of Lehiste (1975) and Thorsen (1985). However, unlike in those studies, where isolated sentences are found to start with lower F_0 than discourse-initial sentences, no significant difference is found between the two levels of topic in the present data. It is difficult to identify the source of this discrepancy, as there are various small differences across these studies, which may have influenced the detailed F_0 measurements. Further research may be needed to resolve the issue.

It could be argued that our isolated sentences have a broad focus rather than being focusless (Ladd, 1996). But the notion of broad focus seems to easily overlap with the notion of topic. A sentence like "Wangying canguan chejian" (Wangying visits the workshop) spoken in isolation is apparently introducing a new subject matter, hence, a new topic. The claim that the sentence is also broadly focused is hard to distinguish from the claim that the entire subject matter is being highlighted. More importantly, from a functional point of view (Hirst, 2005; Kohler, 2005; Xu, 2005), focus should be in contrast with no focus. A sentence-wide broad focus would therefore contrast with a total lack of focus in the entire sentence. From the findings of the present study, it is hard to conceive of a sentence with an F_0 pattern that can focally contrast that of an isolated sentence unless it is on a lower topic level.

Note that the finding of significant differences across the topic levels examined in the present study does not necessarily mean that these levels are categorical. They only suggest that the discourse levels we constructed were sufficiently different to exhibit significant effects. Had we been able to construct conditions with finer-grained differences, the effects main turn out to be even more gradient.

4.5. Is contrastiveness of topic encoded in F_0?

Although it was not the focus of the present study, contrastiveness is something unavoidable when constructing sentences with different topic conditions. Several reviewers have rightly suggested that some analysis should be done in regard to contrastive topic. As can be seen in Table 1, the topic in condition A and D[b] could described as contrastive, while that in condition C and D[a] could be described as non-contrastive. The results show that sentence initial F_0 in condition C and D[a] (non-contrastive) is higher than that in condition A and D[b] (contrastive). Such a difference, however, can be mostly explained in terms of discourse level topic, since topic in condition C and D[a] are in the initial position of isolated sentences and discourse-initial sentences, whereas topic in condition A and D[b] are discourse non-initial. Moreover, once discourse position is controlled, a contrastive topic (condition A and D[b]) has lower F_0 compared to focus (condition B). Thus if we do assume it has an effect, contrastiveness seems to lower rather than raise F_0, which would contradict previous proposals discussed in Section 1.2.

Overall, therefore, the present data show that contrastiveness of topic, though a conceptually plausible communicative function, has no direct prosodic correlate in Mandarin. Further studies that focus more systematically on contrastive topic may be needed to verify our finding.

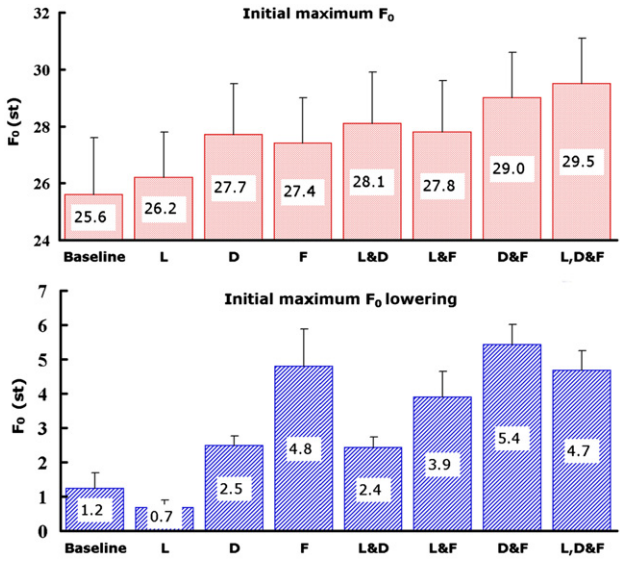

Fig. 6. Effects of the three main factors on maximum F_0 in the first word (top) and size of F_0 lowering between the first and second words (bottom). The three factors are sentence length (L), downstep (D) and focus (F). "&" means combined effects of two or three factors, e.g., L&D means combined effect of length and downstep.

4.6. *Parallel encoding of topic, focus, sentence-length effect and downstep*

Five sources of sentence-initial F_0 variations have been systematically studied in the present study. From the above discussion, we can see that focus, downstep and sentence length have significant and consistent effects on sentence-initial maximum F_0, while the effect of topic at higher discourse level is not very strong in this study. Moreover, newness does not show any effect on F_0. Nor does contrastiveness, although it is not systematically controlled in the study. Because the five main factors have been clearly separated in the study design, we can now test our prediction on the cumulative nature of these effects by calculating how much each of them contributes to the sentence-initial F_0 values.

In Fig. 6, the cumulative contributions of the three significant factors are graphically summarized. The base-line condition is the short HH sentences in condition A, without focus and downstep. Sentence length (L) is from the medial-length HH sentences in condition A (again, no focus or downstep). The downstep condition (D) is from the short HL sentences. And the effect of focus (F) is from the short HH sentences in condition B. Furthermore, combined effects of two factors can also be calculated, i.e., medium HL sentences in condition A show the effects of both sentence length and downstep (L&D); medium HH sentences in condition B show the effects of both length and focus (L&F), and the short HL sentences in condition B show the effect of both focus and downstep (F&D). The effect of all the three factors together is taken from the medium HL sentences in condition B (L, D&F). The initial maximum F_0 value and the maximum F_0 lowering (calculated as the difference between the first and the second words) are shown in the upper and lower parts of Fig. 6, respectively. We can see that F_0 raising due to sentence length (L) is very small, about 0.6 st. The amount of raising due to downstep (D) is about 2.1 st, which is almost as large as that due to focus (F) (about 1.8 st). When the effects of downstep and initial focus are combined (D&F), the amount of F_0 raising is 3.4 st, which is almost the same as adding the two single effects together: $2.1+1.8=3.9$ st.

The lower graph in Fig. 6 shows the amount of F_0 lowering from the first to the second word. The contribution of sentence length is negative (compared to the baseline), indicating that a longer sentence has a smaller lowering effect ($0.7-1.2=-0.5$ st). The amount of lowering due to downstep is about 1.3 st ($2.5-1.2=1.3$ st), and that due to focus is about 3.6 st ($4.8-1.2=3.6$ st). The combined effect of focus and downstep (F&D) is about 4.2 st ($5.4-1.2=4.2$ st), which is again similar to the sum of the two individual effects ($1.3+3.6=4.9$ st).

The cumulative contribution of multiple factors to sentence initial F_0 as predicted based on PENTA is therefore largely confirmed. According to PENTA, speech conveys a multitude of communicative meanings by encoding communicative functions *in parallel* through an articulatory process of target approximation. A basic assumption of PENTA is that the encoding scheme of each communicative function has to be unique, so as not to overlap with other functions. This is clearly seen in the present data. Topic and focus are both distinct from lexical tone because they specify global pitch ranges rather than local F_0 contours; topic is distinct from focus because it *raises* the entire sentence-initial pitch range and allows subsequent F_0 to drop gradually, whereas focus *expands* the on-focus pitch range and *compresses* the post-focus pitch range. The finding that newness does not exhibit clear F_0 manifestations independent of focus and topic suggests that at least it is not encoded with F_0. The small yet significant durational differences related to newness (Table 4: new is longer than given by 18 ms on average) may be due to an effort to speak the new item more clearly by assigning it more time to allow better target approximation, as suggested by a recent modeling study (Prom-on, Xu, & Thipakorn, 2009). In addition, contrastiveness does not show clear independent encoding with F_0.

PENTA also distinguishes factors directly related to communicative meanings from those due to articulatory constraints. Of the factors examined in the present study, topic and focus are clearly communicative, whereas sentence-length-related F_0 changes and downstep are likely to be consequences of articulatory mechanisms interacting with various communicative functions. Sentence length itself is determined by many factors. But once the length is known to the speaker, there seems to be an attempt to get prepared, probably by taking a deeper breath (Watson et al., 2003; Whalen & Kinsella-Shaw, 1997). But the fact that the effect is so weak (mostly less than 1.0 st) indicates that it can account for only a very small portion of sentence-initial F_0. In contrast, downstep, although also being articulatory, as it is a byproduct of producing certain lexical tone sequences, accounts for almost the same amount of the initial raising and subsequent lowering of F_0 as does focus. These findings thus provide further support for the basic assumption of PENTA that the encoding of communicative functions cannot be adequately understood unless articulatory mechanisms are taken into full consideration.

The present findings may also help highlight an aspect of PENTA that is often not directly obvious. That is, the assumption of uniqueness of encoding schemes by PENTA implies that whether and how a conceptually plausible communicative function is prosodically encoded in a language can be established only empirically, i.e., through experiments in which relevant factors are systematically controlled. The findings of the present study demonstrate that once focus and downstep are controlled, the effect of topic becomes much clearer than has been shown before, and that once focus, downstep and topic are controlled, there is no clear evidence for independent effects of newness and contrastive topic. Thus our conceptualization about what may make a plausible or implausible communicative function should never be taken as given unless it has stood the test of stringent empirical investigations. Also in support of such a theoretical position is the

finding mentioned in Section 1.5 that the kind of prosodic focus characterized with post-focus compression found in English, German, Swedish and Mandarin is empirically shown to be absent in many other languages, including Taiwanese, Cantonese and Taiwan Mandarin which are all rather closely related to Mandarin (Chen et al., 2009; Wu & Xu, 2010).

5. Conclusions

The results of the present study demonstrate how topic and focus are encoded as two different communicative functions and how topic is encoded at different discourse levels. Furthermore, the effects of topic and focus are examined together with the other three commonly co-occurring factors, i.e., sentence length, downstep and newness. Such a systematic experimental control allowed us to calculate how much each factor contributed to sentence-initial F_0 variations, which has led to the following main conclusions.

(1) Discourse-initial topic raises the initial pitch register and allows the subsequent F_0 to drop gradually, whereas initial focus expands the on-focus pitch range and compresses the post-focus pitch range.
(2) Topic in discourse-initial and isolated sentences has higher pitch register than in non-initial sentences.
(3) Longer sentences have higher sentence-initial F_0 than shorter sentences, but the difference is small in magnitude and is independent of topic and focus.
(4) The effect of downstep is independent of topic and focus, but it is large in magnitude and thus accounts for a significant amount of the F_0 declination in a sentence.
(5) Newness has no F_0 manifestation independent of other factors, but a newly mentioned word is slightly longer in duration than the same word that has been previously mentioned.
(6) The effects of topic, focus, downstep and sentence length are largely cumulative.

Overall, the present data have improved our understanding of sentence-initial F_0 contours in Mandarin, which complements previous findings on sentence-final F_0 contours in Mandarin (e.g., Liu & Xu, 2005). More importantly, the clear separation of the contributions of independent factors and their interaction with one another provide support for the articulatory-functional view of speech prosody as represented by the PENTA model.

Acknowledgments

We are grateful to Caroline Féry, Frank Kügler, Shinan Lv for their constructive discussions and comments. We are particularly thankful to the five anonymous reviewers for their valuable comments and suggestions. The research was partly supported by National Natural Science Foundation of China (Grant no. 60905062) to the first author and NIH Grant (no. DC006243-01A1) to the second author. The experiments and the first draft were completed with support of DFG project (Germany), SFB632: Information Structure (D2). Part of results of this study was presented at Speech Prosody 2006, Dresden, Germany.

Appendix A

See Table A1.

Table A1
Reading materials.

1. Target sentences with three different lengths

Sentences composed of HH tone words

(a) Wangying canguan chejian
Wangying visits the workshop
(b) Wangying canguan Sanxi Qingxiang yijie jiagong chejian
Wangying visits Sanxi Qingxiang first-street product workshop
(c) Wangying canguan Sanxi Qingxiang yijie Feiying gonsi jixin jiagong chejian
Wangying visits Sanxi Qingxiang first-street Feiying company hardware product workshop.

Sentences composed of HL tone words

(a) Yimeng xiuli zhuoyi
Yimeng fixes tables and chairs
(b) Yimeng xiuli Xiya Sanjing jichang jingmei zuoyi
Yimeng fixes Xiya Sanjing chicken farm high quality tables and chairs
(c) Yimeng xiuli Xiya Sanjing jichang Zhangzhong xinmai chukou jingmei zuoyi
Yimeng fixes Xiya Sanjing chicken farm President Zhang's newly purchased export high quality tables and chairs

Sentences composed of LH tone words

(a) Mawei ningkai dingdeng
Mawei turns on the lamp
(b) Mawei ningkai Beijing Jinggang xiaoqu lijian dingdeng
Mawei turns on Beijing Jinggang district's inner room's lamp
(c) Mawei ningkai Beijing Jinggang xiaoqu Jiangsan laojie xiaowu lijian dingdeng
Mawei turns on Beijing Jinggang district Jiangsan old street's small flat's inner room's lamp

Sentences composed of FR tone words

(a) Lumiao jielai liantiao
Lumiao borrows the chain
(b) Lumiao jielai zhuming Datong Liangqiao muchuan liantiao
Lumiao borrows the famous Datong Liangqiao wooden boat chain
(c) Lumiao jielai zhuming Datong Liangqiao Qingtang Liyang bianxie muchuan liantiao
Lumiao borrows tne famous Datong Liangqiao Qiangtang Liyang portable wooden boat chain

Sentences composed of RF tone words

(a) Liangna yanzhi yanliao
Liangna makes paint
(b) Liangna yanzhi Chongqing nanbu chengshi Hequ yanliao
Liangna makes Chongqing southern urban Hequ paint
(c) Liangna yanzhi Chongqing nanbu chengshi Fangxian Qizhen shise Hequ yanliao
Liangna makes Chongqing southern urban Fangxian Qizhen ten-color Hequ paint

2. Background sentences

Condition A:	da jia dou you shiqing yao zuo. (Everybody has something to do)
Condition B:	bushi Lixiao (It is not Lixiao.) (Here, the personal name is different for sentences with different tone combination)
Condition C:	The target sentence is an isolated sentence without any background sentence
Condition D:	** he Lixiao dou you shiqing yao zuo. (Both ** and Lixiao have something to do.) (Here ** stands for the personal name in the target sentences. Again, the name of the second person varies across sentences with different tones)

References

Ayers, G. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics*, *44*, 1–49.
Boersma, P., & Weenink, D. (2005). 〈http://www.praat.org/〉.
Bolinger, D. (1958). A theory of pitch accent in English. *Word*, *14*, 109–149.
Bosch, M. M. V. (2007). Effect of utterance length on F_0 scaling. In *Proceedings of the ICPhS XVI* (pp. 1165–1168). Saarbrücken, Germany.
Brown, G. (1983). Prosodic structure and Given/New distinction. In A. Culter, & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67–77). Berlin: Springer.
Brown, G., Currie, K., & Kenworthy, J. (1980). *Questions of intonation*. Baltimore: University Park Press.
Bruce, G. (1977). *Swedish word accent in sentence perspective*. Gleerup, Lund: Travaux de l'Institut de Linguistique de Lund 12.
Bruce, G. (1982a). Developing the Swedish intonaiton model. *Working Papers of Lund University Department of Linguistics*, *22*, 51–116.

Bruce, G. (1982b). Textual aspects of prosody in Swedish. *Phonetica*, *39*, 274–287.
Büring, D. (1997). *The meaning of topic and focus: The 59th Street Bridge Accent*. London and New York: Routledge.
Chafe, W. L. (1974). Language and consciousness. *Language*, *50*, 111–133.
Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subject, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–53). New York: Academic Press.
Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.
Chen, Y. Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, *36*, 724–746.
Chen, S.-W., Wang, B., & Xu, Y. (2009). Closely related languages, different ways of realizing focus. In *Proceedings of the Interspeech 2009* (pp. 1007–1010). Brighton, UK.
Chen, Y., & Xu, Y. (2006). Production of weak elements in speech—Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica*, *63*, 47–75.
Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse processes: Advances in research and theory*, Vol. 1 (pp. 1–40). Ablex Publishers.
Cohen, A., & 'tHart, J. (1967). On the anatomy of intonation. *Lingua*, *19*, 177–192.
Connell, B. (2002). Downdrift, downstep and declination. In *Proceedings of the typolody of African prosodic systems workshop* (pp. 3–12). Bielefeld, Germany.
Connell, B. (2003). Pitch realization and the four tones of Mambila. In S. Kaji (Ed.), *Cross-linguistic studies of tonal phenomena* (pp. 181–197). Tokyo: Research Institute for the Language and Cultures of Asia and Africa.
Connell, B. (2004). Tone, utterance length and F_0 scaling. In *Proceedings of the TAL-2004* (pp. 41–44). Beijing, China.
Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, *77*, 2142–2156.
Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.
Crocco, C. (2009). Topic accent and prosodic structure. In L. Mereu (Ed.), *Informaiton structure and its interfaces* (pp. 15–50). Berlin, New York: Mouton de Gruyter.
Downing, A., Neff, J. A., & Carretero, M., et al. (1998). Structuring and signalling topic management. In S. Embleton (Ed.), *Lacus Forum XXIV* (pp. 267–278). The Linguistic Association of Canada and The United States.
Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, *80*, 402–415.
Eady, S., Cooper, W., Klouda, G., Mueller, P., & Lotts, D. (1986). Acoustical characterization of sentential focus: Narrow vs. broad and single vs dual focus environments. *Language and Speech*, *29*, 233–250.
Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*, 680–703.
Fowler, C., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*, 489–504.
Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech*, *1*, 126–152.
Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, *22*, 477–492.
Grenié, A. B., & Grenié, M. (2003). Declination line and tones variations in standard Chinese. *Report of phonetic research 2003* (pp. 142–145). China: Chinese Academy of Social Sciences.
Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In: *Proceedings of the second international conference-spoken language processing* (pp. 429–432). Banff.
Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.
Guijarro, A. J. M. (2001). Topicality chanins in two discours genres. *Estudios Ingleses de la Universidad Complutense*, *9*, 103–128.
Gundel, J. K. (1994). On different kinds of focus. In P. Bosch, & R. v.d. Sandt (Eds.), *Focus and natural language processing*, Vols. 1–3 (pp. 457–466). Heidelberg: IBM.
Gussenhoven, C. (2004). Types of focus in English. In D. Büring, M. Gordon, & C. Lee (Eds.), *Topic and focus: Intonation and meaning, theoretical and crosslinguistic perspectives*. Dordrecht: Kluwer Academic Publishers.
Halliday, M. A. K. (1967a). Notes on transitivity and theme in English, Part II. *Journal of Linguistics*, *3*, 199–244.
Halliday, M. A. K. (1967b). *Intonation and grammar in British English*. The Hague: Mouton.
Hirschberg, J. (1993). Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, *63*, 305–340.
Hirschberg, J., & Pierrehumbert, J. B. (1986). The intonational structuring of discourse. In *Proceedings of the 24th annual meeting of the ACL* (pp. 136–144). Association of Computational Linguistics.
Hirst, D. J. (2005). Form and function in the representation of speech prosody. *Speech Communication*, *46*, 334–347.
Horne, M. (1990). Accentual patterning in 'New' vs 'Given' subjects in English. *Working Papers of Department of Linguistics, Lund University*, *36*, 81–97.
Hyman, L. M. (1973). The role of consonant types in natural tonal assimilations. In L. M. Hyman (Ed.), *Consonant types and tone* (pp. 151–179). Los Angeles, CA: Department of Linguistics, University of Southern California.
Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
Kiss, K. (1998). Identificational focus versus information focus. *Language*, *74*(2), 245–273.
Kohler, K. (2005). Timing and communicative functions of pitch contours. *Phonetica*, *62*, 88–105.
Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology Yearbook*, *1*, 53–74.
Ladd, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, *84*, 530–544.
Ladd, D. R. (1996). *Intonational phonology—Cambridge Studies in Linguistics (Vol. 79)*. Cambridge: Cambridge University Press.
Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge University Press.
Laniran, Y. O., & Clements, G. N. (2003). Downstep and hihg raising: Interacting factors in Yoruba toe production. *Journal of Phonetics*, *31*, 203–250.
Lee, Y.-C., & Xu, Y. (2010) Phonetic realization of contrastive focus in Korean. In *Proceedings of the Speech Prosody 2010* (Vol. 100033, pp. 1–4). Chicago.
Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen, & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 195–206). Berlin: Springer-Verlag.
Liberman, M. Y., & Pierrehumbert, J. (1984). Intonational invariance under changes in pich range and length. In M. Aronoff, & R. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: MIT.
Lindau, M. (1986). Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America*, *80*, 757–764.
Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, *62*, 70–87.
Liu, F., & Xu, Y. (2007). Question intonation as affected by word stress and focus in English. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 1189–1192). Saarbrücken.
Maeda, S. (1976). *A characterization of American English intonation*. Doctoral Dissertation. Cambridge: MIT.
Meeussen, A. E. (1970). Tone typologies for West African Languages. *African Language Studies*, *11*, 266–271.
Nakajima, S., & Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica*, *50*, 197–210.
Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America*, *82*(5), 1512–1523.
Nooteboom, S. G., & Terken, J. (1982). What makes speakers omit pitch accents? An experiment. *Phonetica*, *39*, 317–336.
Oliveira, M. & Freitas, T. (2008). Intonation as a cue to turn management in telephone and face-to-face interactions. In P*roceedings of the Speech Prosody 2008* (pp. 485–488). Campiñas.
Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal Acoustics Society of America*, *66*(2), 363–379.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Ph.D. Dissertation. Cambridge, MA: MIT.
Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. Cambridge, MA: The MIT Press.
Prieto, P., D'Imperio, M., & Elordieta, G., et al. (2006). Evidence for soft preplanning in tonal production: Initial scaling in Romance. In R. Hoffmann, & H. Mixdorff (Eds.), *Proceedings of Speech Prosody 2006* (pp. 803–806). Dresden: TUD Press Verlag der Wissenschaften GmbH.
Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, *24*, 445–473.
Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole. (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.
Prom-on, S., Xu, Y., & Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America*, *125*(1), 405–424.
Rialland, A., & Somé, P. A. (2000). Dagara downstep: How speakers get started. In V. Carstens, & F. Parkinson (Eds.), *Advances in African linguistics. Trends in African Linguistics*, Vol. 4 (pp. 251–263). Trenton, NJ: Africa World Press.
Schachter, P. (1965). Some comments on J. M. Stewart's "The typology of the Twi tone system." *Bulletin of the Institute of African Studies* (Vol. 1). Legon, Ghana, unpublished (preprint).
Shih, C. L. (1988). Tone and intonation in Mandarin. *Working Papers, Cornell Phonetics Laboratory*, *3*, 83–109.
Shih, C. L. (1997). Declination in Mandarin, intonation: Theory, models and applications. In *Proceedings of an ESCA workshop* (pp. 293–296). Athens, Greece.
Silverman, K. E. A. (1987). *The structure and processing of fundamental frequency contours*. Ph.D. Thesis. University of Cambridge.
Sluijter, A., & Terken, J. (1993). Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica*, *50*, 180–188.
Smith, C. L. (2004). Topic transitions and durational prosody in reading aloud: Production and modeling. *Speech Communication*, *42*, 247–270.
Snider, K. (1998). Tone and utterance length in Chumburung: An instrumental study. In *Proceedings of the 28th colloquium on African languages and linguistics*. Leiden.
Sorensen, J., & Cooper, W. (1979). Syntactic coding of fundamental frequency in speech production. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.
Stewart, J. M. (1965). *The typology of the Twi tone system*. Legon, Ghana: Institute of African Studies, University of Ghana.
Stewart, J. M. (1983). Key lowering (downstep/downglide) in Dschang. *Journal of African Languages and Linguistics*, *3*, 113–138.

Swerts, M., & Osterndorf, M. (1997). Prosodic and lexical indications of discourse structure in human–machine interactions. *Speech Communcation*, *22*, 25–41.

Swerts, M., Strangert, E. & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. In *Proceedings of the ICSLP* (pp. 1501–1504). Beijing.

Terken, J. (1984). The distribution of pitch accents in instructions as a function of discourse. structure. *Language and Speech*, *27*, 269–289.

Thorsen, N. G. (1980). Intonation contourss and stress group patterns in declarative sentences of varying length in ASC Danish. *Annual Report of the Institute of Phonetics, University of Copenhangen*, *14*, 1–29.

Thorsen, N. G. (1983). Standard Danish sentence intonation—Phonetic data and their representation. *Folia Linguistica*, *17*, 187–220.

Thorsen, N. G. (1985). Intonation and text in standard Danish. *Journal of the Acoustical Society of America*, *77*(3), 1205–1216.

Tomioka, S. (2009). Contrastive topics operate on speech acts. In M. Zimmermann, & C. Féry (Eds.), *Information structure: Theoretical, typological, and experimental perspectives* (pp. 115–138). Oxford University Press.

Tseng, C. Y. (2008). Corpus phonetic investigations of discourse prosdy and higher level information. *Language and Linguistics*, *9*(3), 659–719.

Tseng, C. Y., Pin, S. H., & Lee, Y. H., et al. (2005). Fluent speech prosody: Framework and modeling. *Speech Communication*, *46*, 284–309.

Umeda, N. (1982). ''F0 declination'' is situation dependent. *Journal of Phonetics*, *10*, 279–290.

Van Heuven, J. V. (2004). Planning in speech melody: Production and perception of downstep in Dutch. In H. Quené, & J. V. van Heuven (Eds.), *On speech and language: Studies for Sieb G. Nooteboom* (pp. 83–93). The Netherlands: LOT Occasional series by Utrecht University.

Van Oosten, J. (1985). *The nature of subjects, topics and agents: A cognitive explanation*. Bloomington, Indiana: Indiana University Linguistic Club.

Wagner, M. (2008). A compositional analysis of contrastive topics. In M. Abdurrahman, A. Schardl & M. Walkow (Eds.), *NELS 38* (pp. 1–14).

Watson, P. J., Ciccia, A. H., & Weismer, G. (2003). The relation of lung volume initiation to selected acoustic properties of speech. *Journal of the Acoustical Society of America*, *113*(5), 2812–2819.

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

Wu, W. L., & Xu, Y. (2010). Prosodic focus in Hong Kong Cantonese without post-focus compression. *Paper presented at the Speech Prosody 2010*, Chicago, USA.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*, 55–105.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46*, 220–251.

Xu, Y. (2005-2010) *ProsodyPro.praat*. Available from: 〈http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/〉.

Xu, Y., & Wang, M. L. (2009). Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics*, *37*, 502–520.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*, 159–197.

Yang, L. C. (1995). Prosodic structures and discourse organization. In *Proceedings of the ICphS 95* (pp. 274–277). Stockholm.

Zerbian, S., Genzel, S., & Kügler, F. (2010) Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). In *Proceedings of the Speech Prosody 2010*, Chicago (Vol. 100976, pp. 1–4).

Zubizarreta, M. L. (1998). *Prosody, focus and word order. Linguistic inquiry monograph* (Vol. 33). Cambridge, MA: MIT Press.