

## Prosody and information structure in a tone language: an investigation of Mandarin Chinese

Iris Chuoying Ouyang & Elsi Kaiser

To cite this article: Iris Chuoying Ouyang & Elsi Kaiser (2015) Prosody and information structure in a tone language: an investigation of Mandarin Chinese, *Language, Cognition and Neuroscience*, 30:1-2, 57-72, DOI: [10.1080/01690965.2013.805795](https://doi.org/10.1080/01690965.2013.805795)

To link to this article: <http://dx.doi.org/10.1080/01690965.2013.805795>



Published online: 14 Jun 2013.



[Submit your article to this journal](#)



Article views: 403



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

## Prosody and information structure in a tone language: an investigation of Mandarin Chinese

Iris Chuoying Ouyang\* and Elsi Kaiser

Department of Linguistics, University of Southern California, Los Angeles, CA 90089, USA

(Received 2 April 2012; final version received 8 May 2013)

Prosody conveys discourse-level information, but the extent to which prosodic cues distinguish different kinds of information-structural concepts remains unclear. The prosodic encoding of information structure is even more complicated in tone languages, where acoustic cues such as F0, intensity and duration also distinguish lexical items (e.g. Mandarin). Prior work on Mandarin led to divergent findings regarding whether and what prosodic cues mark the distinctions between information-structural types. We conducted a production study on Mandarin to investigate whether (1) the presence/absence of corrective focus and (2) the distinction between new/given information are encoded prosodically. Our results show that correctiveness was reflected in all three acoustic parameters: corrective words had longer durations, larger F0 ranges and larger intensity ranges than non-corrective words. The new-given distinction was reflected only in lengthening and F0 range expansion, and only in the absence of correction (correctiveness-by-givenness interaction). This suggests that new information is encoded differently from corrective focus in Mandarin: only corrective focus is associated with intensity range expansion. Our results provide further evidence for the multi-functionality of acoustic-prosodic dimensions. Even in a language with lexical tones, which differ in F0, intensity and duration, all these dimensions also encode information structure. Furthermore, not only can prosodic cues indicate discourse importance, they also distinguish different types of information structure in Mandarin. Our findings highlight the fine-grained ability of the language production system to utilise different aspects of acoustic dimensions with great efficiency.

**Keywords:** corrective focus; new-information focus; prosodic encoding; intensity range expansion

### Introduction

Across languages, it is widely accepted that prosodic structure can convey discourse-level information (e.g. Gussenhoven, 1983; Ladd, 1980; Watson, 2010). Acoustically, there are three dimensions that are commonly regarded as providing cues about information structure: duration, fundamental frequency (F0) and intensity. In English, signals of prosodic prominence – such as longer duration, changes in F0 movement and greater intensity – appear on elements that are semantically or pragmatically prominent (e.g. Breen, Fedorenko, Wagner, & Gibson, 2010; Cooper, Eady, & Mueller, 1985). For instance, consider the sentence ‘he bought mangos’ in the following contexts:

- (1) a. What did Peter buy at the market?  
b. He bought mangos (narrow new-information focus).
- (2) a. What did Peter buy at the market? Did he buy mangos?  
b. Yes, he bought mangos (given information).
- (3) a. What fruits did you and Peter buy at the market?

- b. I bought apples and oranges; he bought mangos (contrastive information).

In response to (1a), *mangos* in (1b) is new information that cannot be inferred from the preceding discourse context. In contrast, the same word *mangos* in (2b) in response to (2a) is given information because it has been mentioned in preceding discourse. When responding to question (3a), *mangos* in (3b) is contrastive information: it belongs to the same set (i.e. fruits sold in the market) as some other information provided in the previous utterances (i.e. apples and oranges) and hence creates a contrast. Prior work on English indicates that prosody can distinguish between different information-structural properties: Katz and Selkirk (2011) recently show that contrastive focus has stronger effects than new information on words’ duration, F0 movement and intensity.

While most researchers agree that contrastive focus and new-information focus are associated with increased duration and intensity, the relationship between F0 contours and different focus types is less well understood. It has traditionally been argued that

\*Corresponding author. Email: [chuoyino@usc.edu](mailto:chuoyino@usc.edu)

different types of discourse information occur with distinct pitch accents. Pierrehumbert and Hirschberg (1990) conclude that the stressed syllables of words in new-information focus receive an H\* pitch accent, whereas the stressed syllables of words in contrastive focus receive an L+H\* pitch accent. However, recent evidence suggests that the mapping between information-structural types and pitch accents is not a straightforward one-to-one relation. Watson, Tanenhaus, and Gunlogson's (2008b) visual-world eye-tracking study reveals that listeners look towards contrastive referents when they hear an L+H\* pitch accent, whereas hearing an H\* accent leads listeners to consider both new and contrastive referents. While these findings confirm that the patterns of F0 movement provide cues for new-information focus and contrastive focus in English, Watson et al.'s results also show that new-information focus and contrastive focus do not map straightforwardly onto different pitch accents.

#### Information structure and prosody in Mandarin Chinese

The question of how prosodic cues map on to different information-structural types becomes even more complex when we consider tone languages, where duration, F0 and intensity also distinguish between lexical items (e.g. African languages: Zerbian, Genzel, & Kugler, 2010; Cantonese: Bauer, Cheung, Cheung, & Ng, 2004; Vietnamese: Jannedy, 2007; cross-linguistic discussion: Hartmann, 2007). In Mandarin Chinese, for example, four pitch patterns – commonly referred to as 'tones' – function as phonemes: high (Tone 1), rising (Tone 2), low (Tone 3) and falling (Tone 4). They can alter lexical meaning, as shown in (4). In addition to the four-way distinction based on F0 movement, lexical tones in Mandarin also differ in amplitude and length. Tone 2, Tone 3 and Tone 4 are perceptible solely on the basis of their amplitude contours (Whalen & Xu, 1992), and Tone 3 is 1.5 times longer than the other tones when produced in isolation (Xu, 1997).

- (4) Tone 1 ma [High] 'mother'  
 Tone 2 ma [Rising] 'hemp'  
 Tone 3 ma [Low] 'horse'  
 Tone 4 ma [Falling] 'scold'

Given that intensity, duration and F0 are used to distinguish lexical items in Mandarin, we are faced with the question of whether these acoustic dimensions also function as signals to information structure, and if so, how they accomplish this dual role. As will become clear in the rest of this section, existing research on different information-structural categories in Mandarin has led to divergent results regarding the specifics of

which prosodic cues distinguish different information-structural categories.

Regarding F0, most researchers agree that in Mandarin, information structure is conveyed not by the *shapes* of F0 contours (as is the case in English), but rather by their *ranges* (e.g. Chen & Braun, 2006; Jin, 1996). For example, Chen and Gussenhoven (2008) found that the F0 shapes specified by different lexical tones are still distinct from one another within an information-structural type. This makes sense given that the shapes of F0 contours are the major cue for lexical tones in Mandarin, and thus need to be maintained to ensure successful spoken word recognition. However, if it is the ranges and not the shapes of the F0 contours that mark information structure in Mandarin, this means that the approaches based on data from English (e.g. Pierrehumbert & Hirschberg's view that L+H\* and H\* map onto different information-structural categories) cannot be applied directly to Mandarin, raising fundamental questions for our understanding of how human languages represent the interface between prosody and information structure.

Let us now review existing work on how prosody encodes information structure in Mandarin, starting with claims regarding *new-information focus*, as *mangos* in (1b), repeated here as (5b). Researchers have compared words that are new information in narrow focus to those in broad focus, exemplified in (6). Here, the entire response in (6b) can be construed as new information.

- (5) a. What did Peter buy at the market?  
 b. He bought mangos (narrow new-information focus).  
 (6) a. What happened?  
 b. He bought mangos (broad new-information focus).

Prior research has found that words that are new information in narrow focus have longer duration (Jin, 1996), larger F0 ranges (Jin, 1996; Xu, 1999) and higher mean F0 (Chen, Wang, & Xu, 2009), when compared to words that are new information in broad focus. The results for intensity in this domain are less clear: Chen et al. (2009) claim that new information in narrow focus has a higher mean intensity than new information in broad focus, but this is not found by Jin (1996). It is important to note that in general, these studies compared narrow new-information focus (5b) and broad new-information focus (6b), rather than new-information focus and non-focus (i.e. given information). Consequently, their results shed light on the differences between narrow new-information focus and broad new-information focus, but do not allow us to draw conclusions regarding the nature of the

prosodic differences between focused material and non-focused (given) material.

The only existing experimental work in Mandarin that we have seen directly comparing new and given information (termed ‘rheme’ and ‘theme’ in their study) is Chen and Braun (2006). They found that new information has longer duration and larger F0 ranges than given information. However, their data were elicited through a reading task where the participants read question–answer pairs; to the best of our knowledge, there is no prior work investigating this issue with a more natural design.

Turning now to *contrastive focus*, we first need to point out that most prior work has focused on *corrective focus*, as exemplified in (7):

- (7) a. What fruit did Peter buy in the market? Did he buy apples?  
b. No, he bought *mangos*.

Here, *mangos* in (7b) is corrective information that is intended to replace the wrong information in a previous utterance (i.e. apples). Correction is considered to be a subtype of contrastive focus, since it provides an alternative to some information in the context (e.g. Dik, 1997). We will refer to this subtype of contrastive focus as ‘corrective focus’ henceforth. A number of studies on corrective focus realisation in Mandarin compared correctively-focused words to non-focused (given) words embedded in an utterance that contains some kind of focus, e.g. new-information focus as in (8), such as Chen (2006), or corrective focus as in (9), such as Chen and Gussenhoven (2008). Thus, these studies compared corrective ‘mangos’ in a context like (7b) to unfocused/presupposed ‘mangos’ in contexts like (8b) or (9b).

- (8) a. What did Peter do to the mangos?  
b. He bought mangos (new-information focus on ‘bought’).  
(9) a. Did John buy mangos?  
b. Peter bought mangos (corrective focus on ‘Peter’).

As a whole, these studies found that correctively-focused words have longer durations (Chen, 2006; Chen & Gussenhoven, 2008) and larger F0 ranges (Chen & Gussenhoven, 2008) than non-focused words. This result fits well with data from other languages as well as the general intuition that corrective words tend to be more prominent (in various ways) than words that refer to already mentioned or presupposed information. However, there seems to be no prior work investigating the intensity of correctively-focused words in Mandarin.

Some existing Mandarin studies have also compared new-information focus and corrective focus on

the other hand, but with conflicting results. Greif (2010)<sup>1</sup> found that correctively-focused words had longer durations than words in new-information focus, but did not differ reliably in terms of their F0 ranges. In contrast, when Chen and Braun (2006)<sup>2</sup> compared corrective focus to new-information focus, they found differences in F0 ranges but no differences in duration. Neither Greif (2010) nor Chen and Braun (2006) looked for differences in intensity between correctively-focused words and new-information words.

In sum, while existing findings for Mandarin generally agree that (1) narrow new-information focus involves increases in F0 displacement and duration when compared to broad new-information focus and that (2) corrective focus similarly involves increases in F0 displacement and duration when compared to *unfocused* words, there is as of yet no consensus about how the two focus types (new-information focus and corrective focus) differ from each other, nor about how constituents in new-information focus differs from non-focused constituents.

In light of the outcomes of Greif (2010) and Chen and Braun (2006), one might conclude that perhaps new-information focus and corrective focus do *not* differ reliably in their acoustic encoding. However, because the details of the target sentences, the designs and the information-structural manipulations in these two studies differ from each other (because their general aims were different), we should be careful in comparing them directly. In essence, then, existing work has not led to a conclusion about whether and how corrective focus and new-information focus differ from each other.

It is quite striking that the prosodic properties of two major types of information structure that have received the most attention in research on English (e.g. Calhoun, 2006; Katz & Selkirk, 2011; Pierrehumbert & Hirschberg, 1990; Watson et al., 2008b) – new information and contrastive focus – are not yet well understood in Mandarin. An understanding of this question in a cross-linguistic context is important for theories of information structure. This question relates to fundamental issues about whether new-information focus and corrective focus are semantically distinct categories or variants of the same category, and whether their prosodic realisations are categorically different or variants on a continuum (see Watson et al., 2008b for related discussion).

### Aims of this study, predictions

As we saw in the preceding section, it remains unclear to what extent prosodic cues differentiate one type of discourse information from another in Mandarin. To

shed light on this issue, and more generally on the question of how information structure is realised prosodically in Mandarin, a tone language where all three prosodic dimensions – duration, F0 and intensity – already serve lexical purposes, we conducted a psycholinguistic production study that investigates three main questions.

*First*, do words in new-information focus vs. corrective focus differ from each other in terms of their prosodic realisation in Mandarin, and if so, which parameters (e.g. duration, mean F0, F0 range or intensity) encode these differences? *Second*, we also investigated whether and how the distinction between ‘new’ and ‘given’ influences the acoustic realisation of both corrective and non-corrective words. As we saw in the preceding section, prior work does not yield a clear picture.

Broadly speaking, we expected new information to be prosodically more prominent than given information in Mandarin, based on existing work in other languages (Brown, 1983; Fowler & Housum, 1987) as well as Chen and Braun (2006)’s findings regarding increased duration and larger F0 ranges.

However, the question of whether and how corrective/contrastive focus differs from new-information focus is more open. As we saw above, existing work on Mandarin has not reached a consensus (Greif, 2010 vs. Chen & Braun, 2006). A comprehension experiment on English by Watson et al. (2008b) found that F0 shapes in English do not map neatly onto contrastive vs. new-information focus, and a production study by Katz and Selkirk (2011) further suggests that contrastive focus and new-information focus differ in prosodic prominence. To contribute to our understanding of this phenomenon in a cross-linguistic setting, we wanted to see how and whether the two focus types differ in Mandarin, a language where discourse-driven prosodic cuing is constrained by the existence of lexical tone.

The *third* key aim of our study is a more inclusive analysis of the acoustic parameters relevant to information structure, to ensure that potentially crucial distinctions are not inadvertently overlooked. This aim has two sub-parts:

*First*, we analysed not only duration and F0 but also *intensity*. Prior studies on Mandarin mostly focused only on duration and F0. The two studies that did look at intensity (Chen et al., 2009; Jin, 1996) analysed mean intensity but did not look at potential effects of information structure on intensity ranges (i.e. difference between maximum and minimum intensity) – and even in the domain of mean intensity, their results do not agree with each other. Given that intensity contours, as well as F0 contours, are associated with lexical tones in Mandarin, we expected that intensity ranges could reflect discourse-level information

just like F0 ranges do. Thus, we investigated how and whether all three prosodic dimensions encode information structure.

The *second* sub-part of this third aim has to do with *how F0 range and intensity expansion is accomplished*. Theoretically, there are three possible ways of expanding the range of an excursion: (1) raising the maximum and lowering the minimum, (2) only raising the maximum and (3) only lowering the minimum, as illustrated schematically in Figure 1. If one finds, say, F0 range expansion for both new-information focus and corrective focus, then in order to assess whether the phenomena are really the same or not, we need to investigate the underlying source of the range expansion, i.e. which of the three ‘strategies’ is being used. Failing to do this could result in incorrectly grouping together two phenomena that are underlyingly different. Thus, we conducted detailed analyses not only of F0 and intensity ranges but also maxima, minima and means.

#### Production study: method

To investigate the prosodic encoding of information structure in a tone language, we conducted a production study on Mandarin. Participants produced instructions based on pictures and arrows shown on a computer screen. They were told to imagine that another person in another room would be listening to their instructions and moving the objects on the screen accordingly (though this person might sometimes make mistakes), and that the movements made by the listener would be visible on the participants’ computer screen. Using pictures allowed us to avoid presenting participants with written sentences which can result in unnatural ‘reading’ intonation. In the following subsections, we will first discuss the experimental design and the stimuli, and then go over the procedure.

In this study, we focus on two major distinctions between information-structural types: (1) Do new and corrective elements differ from each other in terms of their prosodic realisation in Mandarin? (2) Are new and given elements marked differently in prosody? Although these kinds of issues have been investigated

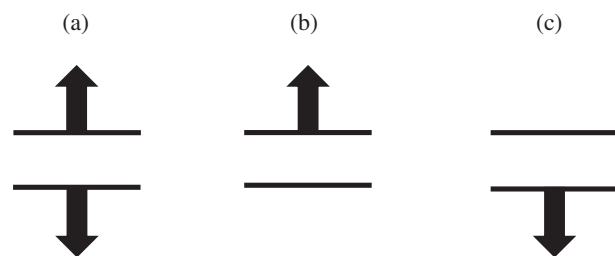


Figure 1. Possible ways of expanding ranges.

in prior work, the existing results do not yield a clear picture.

**Design and stimuli**

Participants saw coloured pictures on the computer screen. Objects were presented in circles, each with its name shown below. There were six pictures on each screen. Arrows were used to indicate the commands participants should produce. For example, in Figure 2a, the arrow points from the cigarette to the lounge chair, so participants should say: ‘Move the cigarette next to the lounge chair’. After they produced the instruction, participants saw a moving event on the computer screen that responded to the instruction either correctly or incorrectly. For example, in Figure 2b, the cigarette is moved next to the lounge chair, which is a correct response.

To examine discourse-level intonation across lexical tones, we manipulated the information structure of the target words and controlled their tonal combinations. Specifically, a repeated-measures within-subjects design with two independent variables were used: (1) *correctiveness* (with two levels: presence or absence of correction) and (2) *givenness* (with two levels: new or given information). Target words were bisyllabic, with one of the three tonal combinations: High–High (HH), High–Low (HL) or Low–High (LH). A third of the target words were HH, a third were HL and a third were LH. All sentences were produced in the frame illustrated in (10).<sup>3</sup> For instance, the sentence ‘*ba xiangyan (cigarette) fangdaolfangzai tangyi (lounge chair) pangbian*’ would be produced for the display in Figure 2.

- (10) ba OBJECT fang-dao/-zai LOCATION pangbian  
 BA OBJECT put-PREP LOCATION side  
 ‘Move the OBJECT next to the LOCATION’

A target word always appeared in the OBJECT role in a sentence. Table 1 shows the summary of the four conditions. Target sentences were the last two sentences in a trial, i.e. the sentences in bold in Table 1. Next, let us consider the four conditions in more detail.

Broadly speaking, there were two types of target trials: new-information trials and given-information trials. The new-information trials were composed of three spoken instructions (sentence (a), (b) and (c) in Table 1). First, a participant saw an image with an arrow and produced the corresponding sentence, e.g. *Move the cigarette next to the lounge chair* (sentence (a)). The object moved correctly in the display, and after that, another arrow appeared. To convey the information represented by the second arrow, the participant produced another sentence, e.g. *Move the juice next to the crow* (sentence (b)). After the second sentence was uttered, an incorrect object moved to the location. For example, instead of the juice, the pacifier moved next to the crow. To correct the moving event, the participant repeated the instruction, e.g. *Move the JUICE next to the crow* (sentence (c)). This time, the correct object moved on the screen. The corrective sentence was the last sentence in a trial.

Note that on these new-information trials, the target word (*juice* in this case) had not been mentioned until sentence (b) was uttered for the first time, i.e. neither of the two nouns mentioned in sentence (a) (*cigarette* and *lounge chair* in this case) was the target word in sentence (b). Thus, ‘*juice*’ was new information when it was mentioned in sentence (b), which we refer to as the *non-corrective new information condition*. Later, when the participant repeated the instruction in order to correct the incorrect moving event, thus producing sentence (c), we refer to this as the *corrective new information condition* since the target word (*juice*) here was uttered in a corrective context. As a result, the distinction between new and given in this study is defined from the *hearer’s perspective*. From the

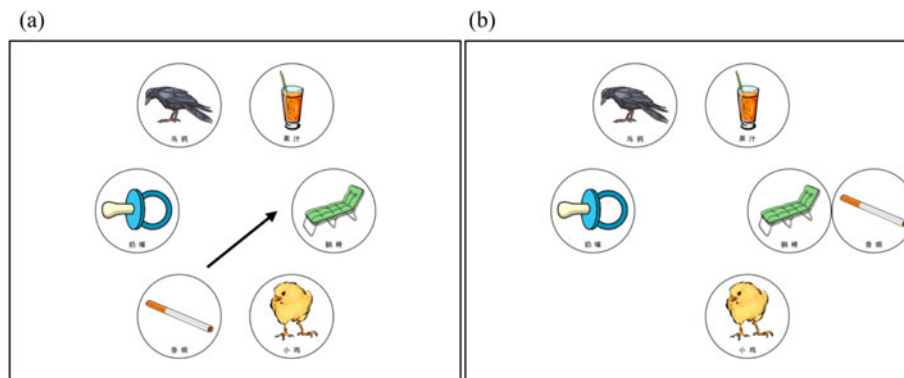


Figure 2. Sample display.

Table 1. Structure of target trials.

Trial type	New information	Given information
1st sentence 1st visual event	(a) Move A next to B (Correct moving)	(d) Move A next to TARGET (Correct moving)
2nd sentence 2nd visual event	[Non-corrective new] (b) Move TARGET next to C (Wrong object is moved next to C)	[Non-corrective given] (e) Move TARGET next to C (Wrong object is moved next to C)
3rd sentence 3rd visual event	[Corrective new] (c) Move TARGET next to C (Correct moving)	[Corrective given] (f) Move TARGET next to C (Correct moving)

perspective of the speaker, by the time they got to sentence (c), ‘juice’ was already given information because the speaker had already uttered it in sentence (b). However, from the perspective of the hearer, ‘juice’ was presumably still new information at the point where sentence (c) was uttered, since the hearer apparently did not hear sentence (b) correctly and moved an incorrect object instead of the ‘juice’. Thus, the fact that a wrong object was moved implies that the hearer did not pay attention to sentence (b) and misheard the object to be moved – i.e. it was still new information to the hearer in sentence (c) (we discuss this more below).

Having considered the new-information trials, let us now turn to the given-information trials. The given-information trials had the same structure as the new-information trials, except for the information-structural properties of the target words. Specifically, in the given trials, the target word had already been mentioned in the LOCATION role of the first sentence (sentence (d) in Table 1). In other words, it had already been involved in an earlier moving event, as shown in Table 1. Thus, in the given-information trials, the second spoken instruction (sentence (e) in Table 1) is in the *non-corrective given information condition*, and the third sentence (sentence (f) in Table 1) is in the *corrective given information condition*.

In our design, we used the distinction between the speaker’s perspective and the hearer’s perspective to differentiate the corrective given and corrective new conditions. Importantly, there is a considerable body of work showing that speakers’ prosodic realisations are indeed sensitive to the hearer’s perspective and attentional state. For example, in instruction-giving tasks where participants gave instructions to confederates about where to place objects (e.g. ‘the teapot goes on red’), longer pronunciations were used for the determiner when the addressees were not anticipating them (Arnold, Kahn, & Pancani, 2012) and for the object to be moved when the addressees were multitasking (Rosa, Finch, Bergeson, & Arnold, 2013). In addition, Ito and Speer (2006) showed that if the addressee makes a

mistake, the correction by the speaker is produced with a contrastive accent. Other studies using interactive designs also found that listener-oriented processes affect prosodic prominence in the speaker’s output in various ways (e.g. Lam & Watson, 2010; Watson, Arnold, & Tanenhaus, 2008a). In addition, corpus work on word order patterns and non-canonical constructions (in English and other languages) shows that people are very sensitive to whether or not something has been mentioned in preceding discourse/successfully introduced into the discourse model (e.g. Birner & Ward, 1998, see also Prince, 1992). Based on these findings, it seems reasonable to assume that speakers understand that in our corrective new condition, the critical target noun is new to the hearer, whereas in the corrective given condition, the critical target noun has already been introduced to the discourse (and the listener’s mental model of the discourse) by virtue of being successfully moved after the first sentence (sentence (d)).

*Definition of new/given.* It is important to note that in our design, the distinction between ‘given’ and ‘new’ information is drawn in terms of discourse-status, i.e. whether the noun has already been mentioned in the preceding discourse or not (e.g. Birner & Ward, 1998; Kaiser & Trueswell, 2004; Prince, 1992). Thus, any entity that has already been mentioned in the prior discourse, regardless of grammatical position, is by definition discourse-old/given information. In particular, the target word in the given conditions first occurs in the LOCATION role and then in the OBJECT role. Prior work has identified connections between prosodic cues and discourse-status (e.g. Féry, Kaiser, Hörnig, Weskott, & Kliegl, 2009 on F0 contours in German) as well as word order patterns and discourse-status (e.g. Birner & Ward, 1998 on English). This definition of old vs. new differs from Terken and Hirschberg (1994) and Schwarzschild (1999), who note that a change in syntactic role (and/or surface position) can render something ‘new’ for purposes of accent assignment. According to their view, the target nouns in our given conditions should in fact count as ‘new’ for purposes of

prosodic prominence because their syntactic role has changed. However, an eye-tracking study by Dahan, Tanenhaus, and Chambers (2002) shows that when listeners hear temporarily ambiguous discourse-old nouns realised in a new syntactic position with prosodic prominence, they do not interpret these prosodic cues as potentially referring to discourse-new referents. In other words, Dahan et al.'s results suggest that discourse-old entities mentioned again in a different syntactic position *do not* pattern in the same way that discourse-new nouns do, seemingly in contrast to the claims of Terken and Hirschberg (1994).

Pre-empting our findings somewhat, let us note already that our results indicate that Mandarin speakers *are* indeed sensitive to the given vs. new distinction defined in terms of discourse-information status and independently of grammatical role or surface position. Although the given information manipulation in our experiment may not represent the 'highest degree' of givenness (due to the change in grammatical role), the key observation (discussed more in the results section) is that the difference between our new and given information conditions is reflected in participants' prosody.

Each condition contained nine items, three in each of the three tone combinations. (See Appendix for the list of target words.) There were 18 target trials, each including a non-corrective target sentence and a corrective target sentence. The dependent variables that we measured were the duration, F0 range and intensity range of the target word region in a target sentence. In the target trials, either three or four of the objects were mentioned in a particular trial. The extra two to three pictures in a display were presented to ensure that participants could not predict which picture was going to be involved in the next instruction. The experiment also included 36 filler trials, which differed from the target trials along one or more of the following parameters: the number of sentences in a trial, whether and in which sentence a wrong moving event occurs, on which noun the correction needs to be made in a corrective sentence, and the lexical tones of the nouns.

Among all the trials, any two of the nouns co-occurred less than four times, and any three of the nouns co-occurred less than once in a trial. The tonal combinations used in target words (HH, HL and LH) appeared 56 times each, and the tonal combination which was only used in filler words (LL) appeared 48 times. All the words (i.e. names of the pictures) were concrete nouns that denoted movable objects and had a frequency no higher than 15.23 counts per million according to Cai and Brysbaert (2010). We controlled for various factors including the number of syllables, the combination of tones, semantic properties and word frequency. This placed severe constraints on the choice of words, and thus voiceless and sibilant

consonants in the words could not be entirely avoided. The positions and directions of arrows on the displays were counterbalanced across trials.

### Procedure and participants

Participants were told to give instructions to move objects based on the pictures and arrows on the computer screen, to check whether their instructions were carried out correctly, and to provide a correction if their instructions were not followed. They were asked to only use the sentence frame in (10) during the entire experiment, and to speak as naturally as possible. Participants were told to imagine that they were speaking to a person in another room, in front of another computer connected to the participants' computer, and that the listener, who might sometimes get distracted and make mistakes, would move the objects according to the participants' instructions. This was done in order to make the task as natural as possible; our assumption was that people would be most likely to mark information-structural cues in their prosody in a communicative situation.

Ten adult native speakers of Mandarin, five women and five men were participated. All were either born in Beijing or had lived in Beijing since age 13 or younger. All of them were students or visiting scholars at University of Southern California who left Beijing no longer than 2 years before. The participants received \$10 for their participation.

### Data analysis

Acoustic analyses were done using the Praat software with the ProsodyPro script (Xu, 2005–2011). Duration, F0 and intensity were extracted by the script. Repeated measure ANOVAs and paired *t*-tests were conducted on the duration, F0 ranges (maximum F0 minus minimum F0) and intensity ranges (maximum intensity minus minimum intensity) of target words.<sup>4</sup> All ANOVAs presented in this paper had correctiveness (correction or non-correction) and givenness (given or new) as independent variables. In the by-subject analyses, the tonal combination of each target word (HH, HL or LH) was also included as a control variable.<sup>5</sup>

To make sure that our data or conclusions are not distorted by the occurrence of creaky voice, we manually removed the markings for aperiodic pulses (resulting from irregular vibration of vocal folds) before the F0 values were computed. This is because pitch tracking for aperiodic waveforms is often inaccurate, which could potentially impact the analysis of F0, as low tones prevalently bring about creakiness in tone languages (e.g. Yoruba: Welmers, 1973, p. 109;



Cantonese: Vance, 1977; Mandarin: Belotel-Grenié & Grenié, 1994). To minimise this problem, all of our F0 analyses focus on the non-creaky portions. Since the minimum F0 in the HL and LH tonal combinations appears during the low tone component, our data might not accurately reflect the actual F0 ranges. In a HL or LH word that contained a creaky region, the actual minimum F0 could be *lower* than what was measured, and if so, the F0 range calculated by deducting minimum F0 from maximum F0 would be smaller, and the effects of information structure on F0 ranges would be consequently *underestimated*. Thus, creaky voice could potentially obscure effects of information structure on F0 ranges because it may cause the ranges to be underestimated. Nevertheless, as will become clear in the rest of the paper, our results clearly indicate that F0 ranges *do* play a significant role signalling information structure (see the following subsections). Thus, we do not regard the occurrence of creaky voice as a problem for our findings, as it did not obscure the patterns of F0 ranges in this study.

### Main results: duration, F0 ranges and intensity ranges

In this section, we first present the results for duration of target words. Then we turn to F0 and intensity measures, which we analysed in three different ways: (1) the range of F0 and intensity measures on the target words (difference between maximum and minimum on a given word), (2) the maximum and minimum measures for F0 and intensity and (3) the mean values for F0 and intensity on the target word.

### Duration

Overall, as can be seen in Figure 3, words in the corrective conditions (the two bars on the left) are longer than words in the non-corrective conditions (the two tall bars on the right). Within the non-corrective conditions, words that are given information have shorter durations than words that are new information, but this distinction between given and new does not

appear in the corrective conditions. The observations are confirmed by statistical analyses: ANOVAs show a main effect of correctiveness ( $F(1,9) = 20.020, p < 0.01$ ), with corrective conditions showing significantly longer duration than non-corrective conditions, and no main effect of givenness ( $F(1,9) = 2.189, p = 0.173$ ). There is a significant interaction between correctiveness and givenness ( $F(1,9) = 6.260, p < 0.05$ ). More specifically, planned comparisons reveal that the correctiveness effect on duration occurs in both the new information conditions (corrective new has longer duration than non-corrective new:  $t(9) = 4.177, p < 0.01$ ) and the given information conditions (corrective given has longer duration than non-corrective given:  $t(9) = 4.641, p < 0.01$ ), but the givenness effect on duration emerges only when the words are non-corrective (non-corrective new has longer duration than non-corrective given:  $t(9) = 3.333, p < 0.01$ ) and not when the words are corrective (corrective new does not differ from corrective given:  $t(9) = 0.331, p = 0.748$ ). In sum, while non-corrective words show an effect of givenness, no effect of givenness is detected on corrective words.

### F0 ranges

Having considered duration, let us now turn to the findings for F0 ranges. Overall, F0 ranges show a similar pattern as duration, as can be seen in Figure 4. Mirroring the results of duration, words in the corrective conditions have larger F0 ranges than words in the non-corrective conditions; given information has smaller F0 ranges than new information in the non-corrective conditions, but this given/new distinction is not present in the corrective conditions. These observations are again confirmed statistically: ANOVAs show a main effect of correctiveness ( $F(1,9) = 22.232, p < 0.01$ ) but no main effect of givenness ( $F(1,9) = 0.749, p = 0.409$ ). There is a significant interaction between correctiveness and givenness ( $F(1,9) = 5.892, p < 0.05$ ). Planned comparisons reveal that the correctiveness effect on F0 ranges occurs in both the new information conditions (corrective new has a larger range than

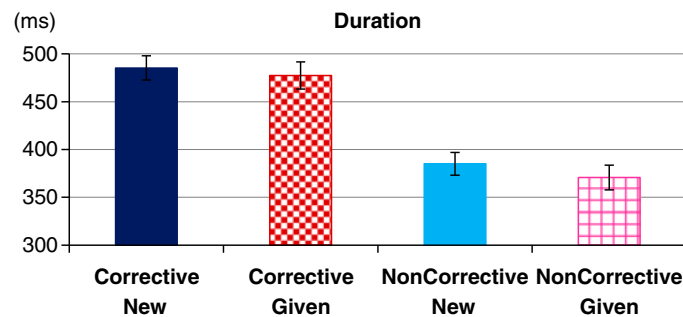


Figure 3. Average duration of the target words in each condition (error bars show  $\pm 1$  SE).

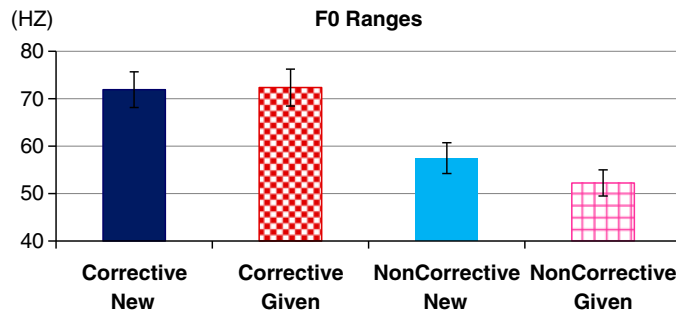


Figure 4. Average F0 ranges of the target words in each condition (error bars show  $\pm 1$  SE).

non-corrective new:  $t(9) = 3.536$ ,  $p < 0.01$ ) and the given information conditions (corrective given has a larger range than non-corrective given:  $t(9) = 5.059$ ,  $p < 0.01$ ). However, the givenness effect on F0 ranges emerges only when the words are non-corrective (non-corrective new has a bigger range than non-corrective given:  $t(9) = 3.348$ ,  $p < 0.01$ ) and not when they are corrective (non-corrective new does not differ from non-corrective given:  $t(9) = -0.669$ ,  $p = 0.521$ ).

### Intensity ranges

Finally, let us move on to the findings for our third parameter, intensity ranges, shown in Figure 5. Interestingly, the intensity range patterns differ from the patterns observed for F0 ranges and duration. On one hand, the distinction between corrective and non-corrective conditions remains: words in the corrective conditions have larger intensity ranges than words in the non-corrective conditions. However, new information does not differ from given information on intensity ranges, in either the corrective or non-corrective conditions. ANOVAs show a main effect of correctiveness ( $F(1,9) = 9.659$ ,  $p < 0.05$ ). There is no main effect of givenness ( $F(1,9) = 0.130$ ,  $p = 0.727$ ) and no interaction between correctiveness and givenness ( $F(1,9) = 0.563$ ,  $p = 0.472$ ).

### Further analyses of F0 and intensity ranges: F0 and intensity maxima and minima

In contrast to prior work where the distinction between correction and new information was only found in one prosodic dimension (duration in Greif, 2010; F0 ranges in Chen & Braun, 2006), the results presented in the preceding sections show that corrective focus and new-information focus differ in all three prosodic dimension. However, in order to better understand the underlying nature of the F0 and intensity ranges, further analyses are needed. As mentioned in the 'aims' section, F0 ranges or intensity ranges are not a single parameter in and of themselves. To understand how F0 and intensity ranges are altered acoustically, we need to examine their 'components' – the maxima and minima of F0 and intensity. Inspecting the maxima and minima allows us to see how range expansion is accomplished, e.g. (1) by raising the maximum, (2) lowering the minimum or (3) both? (Figure 1). We aim to answer two questions: First, are F0 range expansion and intensity range expansion achieved in the same way? Second, do new-information focus and contrastive focus involve different 'strategies' of range expansion for either F0 or intensity? In this section, we present the results of maximum and minimum F0, and maximum and minimum intensity.

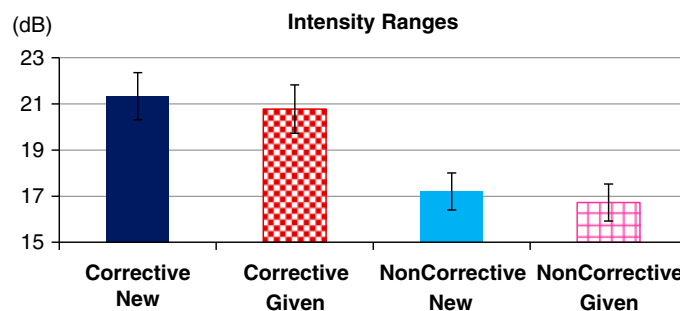


Figure 5. Average intensity ranges of the target words in each condition (error bars show  $\pm 1$  SE).

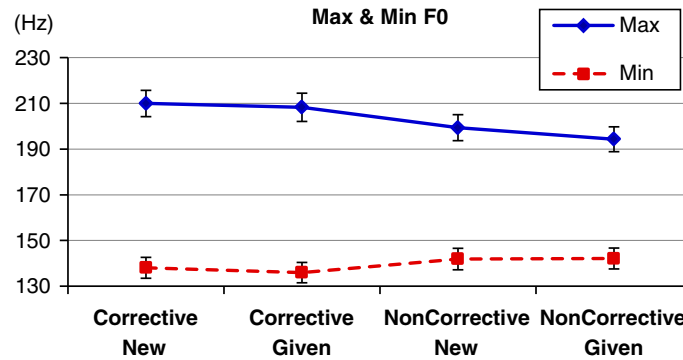


Figure 6. Average maximum and minimum F0 of the target words in each condition (error bars show  $\pm 1$  SE).

### Maximum and minimum F0

Breaking F0 ranges down into maximum and minimum F0, we see that different types of information structure enlarge F0 ranges through different means. As shown in Figure 6, words in the corrective conditions have overall higher maximum F0 and lower minimum F0 than words in the non-corrective conditions – in other words, the *increased F0 range that we observe for correction is accomplished by both raising the maximum and lowering the minimum*. Taking a closer look at the conditions, we see that the corrective given and corrective new conditions do not differ from each other in terms of their F0 maxima or minima ( $|t(9)|s < 0.643$ ,  $p's > 0.536$ ). However, when we look at the non-corrective conditions, we see that non-corrective given and non-corrective new differ in terms of their maximum F0 (higher for new information) but not in terms of their minimum F0. Thus, *the F0 range expansion that we reported above for words that are new information in non-corrective contexts is accomplished by raising the maximum F0 without changing the minimum F0*.

The statistical analyses confirm these observations to a large extent. ANOVAs show a main effect of correctiveness on maximum F0 ( $F(1,9) = 27.844$ ,  $p < 0.01$ ) and minimum F0 ( $F(1,9) = 7.415$ ,  $p < 0.05$ ), with maximum F0 being significantly higher and minimum

F0 significantly lower in the corrective words than the non-corrective words. Somewhat unexpectedly, there is no significant interaction between correctiveness and givenness in either maximum F0 ( $F(1,9) = 2.994$ ,  $p = 0.118$ ) or minimum F0 ( $F(1,9) = 0.802$ ,  $p = 0.394$ ). Nevertheless, given and new information in the non-corrective conditions differ in maximum F0 by 5 Hz, which is in magnitude the same as the statistically significant difference in minimum F0 between corrective and non-corrective words.

### Maximum and minimum intensity

Turning now to the question of intensity ranges, we find a very different pattern: in contrast to F0, maximum intensity stays the same among different types of information structure, as indicated in Figure 7. The presence and absence of correction is reflected *only in minimum intensity*: minimum intensity is lower in corrective words than non-corrective words. ANOVAs show a main effect of correctiveness on minimum intensity ( $F(1,9) = 9.013$ ,  $p < 0.05$ ) but not on maximum intensity ( $F(1,9) = 0.059$ ,  $p = 0.813$ ). Consistent with the patterns of intensity ranges, there is no interaction between correctiveness and givenness in

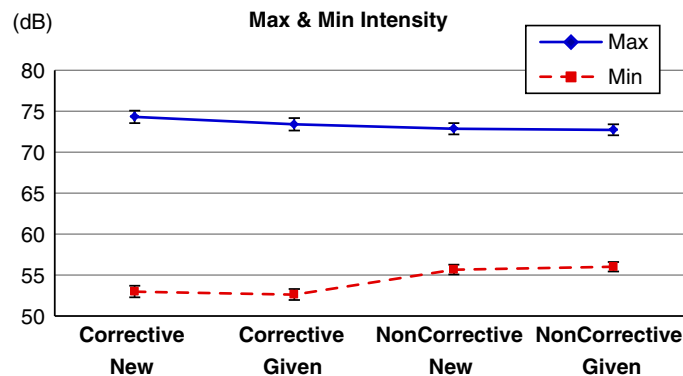


Figure 7. Average maximum and minimum intensity of the target words in each condition (error bars show  $\pm 1$  SE).

either minimum intensity ( $F(1,9) = 1.119, p = 0.318$ ) or maximum intensity ( $F(1,9) = 0.802, p = 0.394$ ).

In sum, we find that the intensity range expansion for F0 and for intensity is accomplished in different ways, with intensity range expansion accomplished simply by lowering the minimum intensity and F0 range expansion showing a more complex pattern: the F0 range expansion observed for corrective focus is accomplished by both raising the maximum and lowering the minimum, whereas the F0 range expansion observed for non-corrective new-information focus is accomplished by raising the maximum F0 without changing the minimum F0.

### Mean F0 and mean intensity

Having inspected the ranges of F0 and intensity and their maxima and minima, we now look at the means of F0 and intensity. In the preceding section, we saw that only in some contexts did range expansion involve both raising maxima and lowering minima (i.e. F0 range expansion for corrective focus). In other contexts where the ranges were expanded by only raising the maxima (i.e. F0 range expansion for new-information focus) or only lowering minima (i.e. intensity range expansion for new-information focus), means could potentially also reflect information-structure. Somewhat surprisingly, mean F0 and mean intensity do not robustly differ between information-structural types. Despite a main effect of correctiveness on mean F0 ( $F(1,9) = 12.244, p < 0.01$ ), the difference in mean F0 is significant only between the given information conditions (i.e. corrective given is significantly higher than non-corrective given:  $t(9) = 4.307, p < 0.01$ ) but marginal between the new information conditions (i.e. corrective new is marginally higher than non-corrective new:  $t(9) = 1.858, p = 0.096$ ). There is no main effect of givenness on mean F0 ( $F(1,9) = 0.616, p = 0.453$ ). Also, neither correctiveness ( $F(1,9) = 0.110, p = 0.748$ ) nor givenness ( $F(1,9) = 0.486, p = 0.503$ ) has a main effect on mean intensity.

In sum, we find that – unlike ranges, maxima and minima – means of F0 and intensity do not provide reliable cues about information status in Mandarin. This finding highlights the importance of a comprehensive analysis of prosodic features; in our case, the major cues for information-structural distinctions would have been overlooked if one had only analysed the means of F0 and intensity.

### Results regarding lexical tone combinations

As mentioned in the design section, we used bisyllabic target words in one of three lexical tonal combinations:

HH, HL or LH. These tonal combinations were equally distributed among the target words and the different conditions, to ensure that our conclusions would not be restricted to a single tonal combination type. Because we controlled for factors such as word frequency and semantic properties, the identity of the segments in target words was not controlled across tonal combinations. This means that prosodic differences between tonal combinations may come from segmental variance rather than tonal properties. Thus, any comparison between tonal combinations must be viewed cautiously. The analyses in this section are included for the sake of completeness, but the reader should keep in mind that these are *post-hoc* analyses and the study was not designed with these analyses in mind (due to the variability in the segmental properties of the target words).

When we look at the prosodic features within each tonal combination, we see that the majority of the patterns discussed in the results section emerges within each tonal combination as well. We first consider the corrective vs. non-corrective manipulation, and then the given vs. new manipulation.

### Corrective vs. non-corrective manipulation

Paired *t*-tests were conducted comparing corrective new vs. non-corrective new for all three tonal combinations, and corrective given vs. non-corrective given for all three tonal combinations. Overall, the effects of correctiveness in both new and given information that we observed in the preceding sections appear in all tonal combinations, although some of them do not reach significance. The absence of significance is not surprising, given (1) the fact that identity of the segments in target words was not controlled across tonal combinations and (2) the reduction in power that comes from looking at a third of the entire dataset (since there are three tonal combinations), and even less when we split it into given vs. new and corrective vs. non-corrective. Nonetheless, the distribution of significance among the conditions shows compatible patterns with prior work that examines the dependence of focus-driven prosody on lexical tones (Chen & Gussenhoven, 2008). In fact, while all tonal combinations are significantly lengthened in corrective focus, only the HL combination robustly differs in the F0 and intensity dimensions between corrective and non-corrective conditions. The properties of tones seem to impose greater restrictions on LH and HH, than on HL, in terms of the extent to which F0 and intensity can be altered to encode information structure (see Chen & Gussenhoven, 2008 for relevant further discussion).

*Duration:* corrective words are longer ( $t(9)s > 3.707$ ,  $p's < 0.01$ ) in all tonal combinations than non-corrective words. *F0 ranges:* corrective words have larger F0 ranges ( $t(9)s > 2.659$ ,  $p's < 0.05$ ) than non-corrective words (except corrective new vs. non-corrective new in LH, which is marginal,  $p = 0.083$ , and corrective new vs. non-corrective new in HH which is ns). *Intensity ranges:* corrective words have larger intensity ranges than non-corrective words ( $t(9)s > 2.599$ ,  $p's < 0.05$ ; except for corrective given vs. non-corrective given in LH which is marginal,  $p = 0.057$ , and corrective new vs. non-corrective new in HH which is ns). *F0 maxima:* corrective words have higher F0 maxima than non-corrective words ( $t(9)s > 2.986$ ,  $p's < 0.05$ , except for corrective new vs. corrective given in LH which is ns). *F0 minima:* corrective words have numerically lower F0 minima than non-corrective words (ns). *Intensity maxima:* corrective words have numerically higher intensity maxima (corrective new vs. non-corrective new and corrective given vs. non-corrective given for HH reach significance,  $p's < 0.05$ ). *Intensity minima:* corrective words have lower intensity minima ( $t(9)s < -2.522$ ,  $p's < 0.05$ ; except for corrective new vs. non-corrective new LH which is marginal,  $p = 0.052$ , and corrective new vs. non-corrective new HH which is ns).

#### **Given vs. new manipulation**

Paired *t*-tests were conducted comparing corrective new vs. corrective given for all three tonal combinations, and non-corrective new vs. non-corrective given for all three tonal combinations. Numerically, the prosodic properties of the three tonal combinations largely mirror the data pattern presented in the preceding section, although the analyses do not reach significance<sup>6</sup>, which – as discussed above for the corrective manipulation – is not surprising. The descriptive statistics are largely consistent with our previous observations: within a tonal combination, most of the numerical differences between non-corrective new and non-corrective given conditions are towards the same tendencies as tested in the main analyses.

#### **Discussion**

The study presented in this paper investigates the prosodic cues for two kinds of distinctions between discourse-information structures in Beijing Mandarin: the presence or absence of corrective focus, and the new vs. given distinction. Although existing findings on Mandarin prosody generally agree that new-information focus and corrective focus both involve increases in F0 displacement and duration (relative to given words or words in broad focus), there is as of yet no consensus about whether and how the two focus types – new-

information focus and corrective focus – differ from each other. The role of intensity is also not well understood. A better understanding of these issues is important because they are involved in fundamental questions regarding the relationship of information structure and prosody, such as how the prosodic system represents different information-structural categories and lexical contrasts at the same time. In this section, we discuss how our results relate to our three key aims sketched out at the start of the paper, as well as their broader implications.

Our *first aim* was to investigate whether words in new-information focus vs. corrective focus differ from each other in terms of their prosodic realisation in Mandarin, and if so, which parameters (e.g. duration, mean F0, F0 range or intensity) encode these differences. Our *second aim* was to explore whether and how the distinction between ‘new’ and ‘given’ influences the acoustic realisation of both corrective and non-corrective words.

As regards the first aim, our results suggest that the cues for corrective focus *do* differ from those signalling the new vs. given distinction. Corrective focus and new-information focus are distinguished from each other both by the degrees of prominence they induce along the same acoustic dimensions and by the different acoustic dimensions they occupy. On one hand, corrective focus and new-information focus both affect duration and F0, but corrective focus has a stronger impact on these prosodic features than new-information focus. On the other hand, intensity cues only appear for corrective focus, not for new-information focus. Despite the fact that we did not focus on the same kinds of information-structural distinctions, our findings are consistent with Chen and Braun (2006) on the conceptual level. This indicates the full complexity of prosodic structure that allows many information-structural categories to be encoded distinctively.

As regards the second aim, we found that the prosodic distinction between new and given information only emerged in non-corrective words in our study. There are several possible reasons for why correctively focused words do not show a distinction between given and new. Cognitively, correction might be more salient than ‘newness’ (Kaiser, 2011), which could in some sense ‘overwhelm’ the distinction between new and given information in a context where the words are also corrective. In principle, such a ‘ceiling effect’ could also be caused by physiological constraints. Since correction yields extremely strong prosodic prominence even when the information is given, it might be difficult or inefficient to further increase the prominence for new information. However, a prior study has found duration lengthening for corrective focus in two different emphatic degrees (Chen & Gussenhoven, 2008). This

suggests that physiology is unlikely to be hindering the prosodic realisation of different degrees of emphasis (for related work on speech perception, see Ladd & Morton, 1997).

Another explanation is that speakers in our experiment might define givenness from their own perspective (rather than from the hearer's perspective as intended), which would then effectively remove the distinction between new and given in the corrective focus conditions in our experiment. More specifically, recall that target sentences in the corrective conditions had been uttered by the speaker, although the listener apparently did not hear them properly the first time. If speakers fail to keep a log of the listener's knowledge state, then the prosodic differences between new and given information in non-corrective conditions might actually reflect whether the words had been uttered more than once, rather than whether the listener had heard the words (Lam & Watson, 2010; Watson et al., 2008a).

Lastly, as discussed in the method section, existing work has shown that given information is prosodically prominent when it appears in a different syntactic role (Schwarzchild, 1999; Terken & Hirschberg, 1994). In our study, a target word in the given-information conditions first occurs in the LOCATION role and then in the OBJECT role. Although we did find prosodic differences between given information and new information when there is no correction, the degree of givenness might not be large enough for given information to be substantially de-accented in corrective focus due to the change in syntactic roles. These are intriguing questions that deserve to be investigated further in future work.

Our *third aim* was to provide a more inclusive analysis of the acoustic parameters relevant to information structure, including analysis of intensity as well as a closer look at how F0 range and intensity expansion is accomplished (e.g. lowering minima, raising maxima or both). Regarding intensity, our results show that it can provide information-structural cues, though to a limited extent: while contrastive focus results in increased intensity range expansion, new-information focus does not do so.

Regarding range expansion, we found that intensity range expansion for F0 and for intensity were accomplished via multiple routes. As illustrated in Figure 1, the two parameters involved in the expansion of ranges – maximum and minimum – potentially form three ways of achieving range expansion: (1) raising the maximum and lowering the minimum, (2) only raising the maximum and (3) only lowering the minimum. Intuitively, one might expect that extending both the upper and lower bound of a range (option (a)) would require the least articulatory effort while accomplishing the largest range expansion, so this pattern might be

the most widely used. However, analyses of the production data from our experiment reveal that both maximum and minimum were employed, and all three possible ways of expanding ranges emerged in different portions of our data. Recall that F0 ranges were extended for both corrective focus and new-information focus, whereas intensity ranges were extended only for corrective focus. We found that *corrective focus* expands F0 ranges by raising the maximum and lowering the minimum (option (a)), *new-information focus* expands F0 ranges by only raising the maximum (option (b)), and *corrective focus* expands *intensity ranges* by only lowering the minimum (option (c)). Our findings rule out a simple hypothesis that the semantically or pragmatically prominent words are merely spoken slower, louder and with higher pitch. Indeed, the duration of the words does become longer, but F0 range expansion during a focused word results from extending not only the upper bound but also the lower bound of the range. Moreover, the expansion of intensity ranges is due to a decrease in intensity during some part of the word, rather than an increase.

In earlier work, Chen and Gussenhoven (2008) found that the expansion of F0 ranges for information structure in Mandarin is mainly accomplished by raising maximum F0, which is somewhat inconsistent with our findings. As pointed out by Chen and Gussenhoven (2008), this might have to do with the fact that F0 lowering in a low lexical tone leads to serious creakiness, which makes it difficult to assess the effect of emphasis on F0 minimum. In the section on 'Results regarding lexical tone combinations', our analyses of F0 minimum also show no significant difference between information-structural conditions within a tonal combination. However, a possibility that cannot be excluded is that the cues for range expansion do not necessarily exist at both sides of the range; we leave the question open for future work.

Let us now briefly consider the fact that in a tone language like Mandarin, discourse-level intonation and lexical tones potentially occupy the same acoustic dimensions in tone languages. Existing work on Mandarin has found that all three prosodic dimensions – duration, F0 and intensity – provide cues for discourse-level information, as well as make contrasts (i.e. lexical tones) between word meanings. Largely consistent with prior studies (Chen, 2006; Chen & Braun, 2006; Chen & Gussenhoven, 2008; Greif, 2010; Jin, 1996; Xu, 1999)<sup>7</sup>, we found lengthening and F0 range expansion in corrective focus and new-information focus. Furthermore, our results show that intensity ranges may also be expanded to emphasise words in an utterance: intensity excursions become larger when the speakers express a correction. In other words, there is no evidence for specialised functions where some

prosodic dimensions mark information structure and others mark lexical items, e.g. it is *not* the case that intensity ranges are used to mark only lexical contrasts, whereas F0 ranges are used to mark only information-structural distinctions. All three prosodic dimensions are multi-functional.

Our findings are also compatible with earlier observations regarding the manner in which prosodic cues in Mandarin encode discourse-information structure and lexical distinctions – in particular, the idea that these two kinds of information are encoded differently: the latter has to do with the *shapes* of F0 movement and intensity movement, and the former with the *ranges* of their movement. Earlier work has pointed out that, for different lexical tones, the shapes of F0 contours clearly differ, whereas with information-structural types, what vary are the ranges of F0 contours (Chen & Gussenhoven, 2008; Xu, 1997). Whalen and Xu (1992) suggest that F0 and intensity are positively correlated in lexical tones, which enables Mandarin speakers to perceive tones without the presence of contrastive F0 patterns. Indeed, if one inspects the contours of F0 and intensity in our data, both their shapes considerably differ between tonal combinations while staying similar across different conditions of information structure. Given our results showing that intensity ranges are used to differentiate information-structural types, there appear to be parallels between F0 and intensity in the specialisation of parameters. Lexical information is encoded by the shapes of F0 and intensity contours, whereas discourse information is marked by the ranges of F0 excursions and, as indicated by our findings, the ranges of intensity excursions. This highlights the fine-grained ability of the language production system to utilise different aspects of acoustic dimensions.

### Conclusions

On the basis of the production study reported in this paper, we can draw two main conclusions. *First*, our findings provide further evidence for the multi-functionality of acoustic-prosodic dimensions. Even in a language with lexical tones, which differ in F0, intensity and duration, all these dimensions nevertheless also encode information structure (see also Chen & Gussenhoven, 2008). Nevertheless, since the shapes of F0 and intensity contours mark lexical items (Whalen & Xu, 1992), these two prosodic dimensions are modulated in another way to mark discourse importance. Parallel to what has been found in F0, the ranges of which provide cues for focus (e.g. Jin, 1996), our results clearly show that the *ranges of intensity* encode discourse information as well. *Second*,

not only can prosodic cues indicate discourse importance, they also distinguish different kinds of information-structural distinctions in Mandarin (see also Chen & Braun, 2006). Lengthening and F0 range expansion occur in both corrective focus and new-information focus, whereas intensity range expansion only appears on correctively-focused words, regardless of their givenness. Taken together, our results show that even in a tone language – where the dimensions of F0, duration and intensity are used to mark lexical distinctions – prosodic information is nevertheless a rich source of cues about information structure, including different subtypes of focus. Our findings highlight the fine-grained ability of the language production system to utilise different aspects of acoustic dimensions with great efficiency.

### Acknowledgements

Earlier version of this work were presented at AMLaP (Architectures and Mechanisms for Language Processing) 2011, ETAP (Experimental and Theoretical Advances in Prosody) 2, WECOL (Western Conference on Linguistics) 2011 and the Annual Meeting of the Linguistic Society of America in 2012; we would like to thank the audience members for the valuable comments and suggestions. Thanks also go to the USC Language Processing Lab group for feedback during the development of this project. We would also like to thank the editors of ‘Language and Cognitive Processes’ and two anonymous reviewers for their helpful comments.

### Notes

1. Greif (2010) compared narrow new-information focus with two types of corrective focus: semantic correction and pragmatic correction. Due to length reasons, we will not discuss Greif (2010)’s findings for what he calls ‘pragmatic correction’, since only semantic correction is similar to the corrective focus that we investigated.
2. Chen and Braun (2006)’s study is important, as it is the first prosodic investigation of information structure in Mandarin that investigated different information-structural categories in a way that closely tied them to theoretical work. Their information-structural notions and terminology are based on Steedman (2000): Theme, Rheme, Background and Focus. In our discussion, we refer to their ‘normal rheme focus’ and ‘corrective rheme focus’ as ‘narrow focus’ and ‘corrective focus’ for the sake of consistency. More specifically, Chen and Braun (2006) examine four categories of discourse information: theme background, theme focus, rheme background and rheme focus. Following Steedman (2000)’s framework, these four categories are based on two layers of information structure: a primary distinction between rheme and theme, and a secondary distinction between focus and background. While ‘rheme’ and ‘theme’

roughly correspond to the new and given information as defined in our study, the division between ‘focus’ and ‘background’ is based on prosody – focus is intonationally marked and background is not. Chen and Braun (2006) find that both rheme and focus are marked by lengthening and F0 range expansion, and that the distinction between rheme and theme is prosodically more prominent than the distinction between focus and background. Additionally, and most relevantly for us – they look into two subtypes of rheme focus (essentially, *new-information focus* and *corrective focus*), and find that correctively-focused rhemes have bigger F0 ranges than non-correctively-focused rhemes, but do not differ in duration – a finding which contrasts with Greif (2010). As a whole, Chen and Braun (2006)’s findings suggest that different information-structural distinctions can be encoded in the same prosodic dimensions with different degrees of prominence, as well as reflected in different prosodic dimensions.

3. For the verb ‘put’, the variant *fang* is also possible, in addition to *fang-dao* and *fang-zai*. These forms are interchangeable across speakers in this context. Participants were asked to use the one most natural to them; only one participant used the short form *fang*.
4. Three sentences are missing from the recordings due to technical problems, and two sentences were misspoken. They amount to 1.39% of the data.
5. In this paper, we focus on the by-subject analyses because the design of the study does not allow for by-item analyses of all targets, as the nine target words were ‘cycled’ through the 36 target items. However, if one analyses the nine target words in all four conditions, the statistical patterns closely resemble the by-subject analyses.
6. The analysis comparing given vs. new words in the non-corrective condition do not reach significance, except for the analysis of intensity range: for words in the HH tone group, non-corrective new information has a larger intensity range than non-corrective given information ( $t(9) = 2.519, p < 0.05$ ).
7. For corrective focus, Chen and Braun (2006) did not find lengthening and Greif (2010) did not find F0 range expansion, as discussed in the introduction.

## References

- Arnold, J. E., Kahn, J. M., & Pancani, G. (2012). Audience design affects acoustic reduction via production facilitation. *Psychological Bulletin & Review*, 19(3), 505–512. doi:10.3758/s13423-012-0233-y
- Bauer, R. S., Cheung, K.-H., Cheung, P.-M., & Ng, L. (2004). *Acoustic correlates of focus-stress in Hong Kong Cantonese*. Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society. Arizona State University, Program for Southeast Asian Studies.
- Belotel-Grenié, A., & Grenié, M. (1994). *Phonation types analysis in Standard Chinese*. Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP 1994), Yokohama, Japan.
- Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam/Philadelphia: John Benjamins.
- Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25, 1044–1098. doi:10.1080/01690965.2010.504378
- Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67–77). Berlin: Springer Verlag.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE*, 5(6), e10729. doi:10.1371/journal.pone.0010729
- Calhoun, S. (2006). *Information structure and the prosodic structure of English: A probabilistic relationship*. Edinburgh: University of Edinburgh dissertation.
- Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics*, 34(2), 176–201. doi:10.1016/j.wocn.2005.05.002
- Chen, Y., & Braun, B. (2006). Prosodic realization in information structure categories in standard Chinese. In R. Hoffmann & H. Mixdorff (Eds.), *Speech prosody 2006*. Dresden: TUD Press.
- Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36, 724–746. doi:10.1016/j.wocn.2008.06.003
- Chen, S.-W., Wang, B., & Xu, Y. (2009). *Closely related languages, different ways of realizing focus*. Proceedings of Interspeech 2009, Brighton, UK.
- Cooper, W., Eady, S., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, 77, 2142–2156. doi:10.1121/1.392372
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292–314. doi:10.1016/S0749-596X(02)00001-3
- Dik, S. (1997). *The theory of functional grammar*. Berlin: Mouton De Gruyter.
- Féry, C., Kaiser, E., Hörnig, R., Weskott, T., & Kliegl, R. (2009). Perception of intonational contours on given and new referents: A completion study and an eye-movement experiment. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 235–266). Berlin: Mouton de Gruyter.
- Fowler, C. A., & 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. doi:10.1016/0749-596X(87)90136-7
- Greif, M. (2010). Contrastive focus in Mandarin Chinese. *Proceedings of Speech Prosody 2010*, Chicago.
- Gussenhoven, C. (1983). Focus, mode, and nucleus. *Journal of Linguistics*, 19, 377–417. doi:10.1017/S0022226700007799
- Hartmann, K. (2007). Focus and tone. In C. Féry, G. Fanselow, M. Krifka (Eds.), *Interdisciplinary studies on information structure 6* (pp. 221–235). Potsdam: Universitätsverlag Potsdam.
- Ito, K., & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In S. Sudhoff et al. (Eds.), *Methods in empirical prosody research* (pp. 229–258). Berlin: Mouton de Gruyter.
- Jannedy, S. (2007). Prosodic focus in Vietnamese. In S. Ishihara, S. Jannedy, & A. Schwarz. (Eds.), *Interdisciplinary studies on information structure* (Vol. 8, pp. 209–230). Potsdam: Universitätsverlag Potsdam.
- Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese* (Doctoral dissertation). Columbus, OH: Ohio State University.
- Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalization and contrastive focus. *Language and Cognitive Processes*, 26, 1625–1666. doi:10.1080/01690965.2010.523082
- Kaiser, E., & Trueswell, J. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2), 113–147. doi:10.1016/j.cognition.2004.01.002
- Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*, 87, 771–816. doi:10.1353/lan.2011.0076
- Ladd, D. R. (1980). *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press.



- Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, 25, 313–342. doi:10.1006/jpho.1997.0046
- Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory and Cognition*, 38, 1137–1146. doi:10.3758/MC.38.8.1137
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thompson & W. Mann. (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295–325). Philadelphia: John Benjamins.
- Rosa, E. C., Finch, K., Bergeson, M., & Arnold, J. E. (2013). The effects of addressee attention on prosodic prominence. *Language and Cognitive Processes*. doi:10.1080/01690965.2013.772213
- Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, 7(2), 141–177. doi:10.1023/A:1008370902407
- Steedman, M. (2000). Information structure and the syntaxphonology interface. *Linguistic Inquiry*, 31, 649–689. doi:10.1162/002438900554505
- Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing ‘given’ information: Effects of persistence of grammatical function and surface position. *Language and Speech*, 37(2), 125–145. doi:10.1177/002383099403700202
- Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica*, 34, 93–107. doi:10.1159/000259872
- Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. *The Psychology of Learning and Motivation*, 52, 163–183. doi:10.1016/S0079-7421(10)52004-8
- Watson, D., Arnold, J. A., & Tanenhaus, M. K. (2008a). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, 106, 1548–1557. doi:10.1016/j.cognition.2007.06.009
- Watson, D. G., Tanenhaus, M. K., & Gunlogson C. A. (2008b). Interpreting pitch accents in online comprehension: H\* vs. L+H\*. *Cognitive Science*, 32, 1232–1244. doi:10.1080/03640210802138755
- Welmers, W. E. (1973). *African language structures*. Berkeley: University of California Press.
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25–47.
- 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. (2005–2011). ProsodyPro.praat. Retrieved from <http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/>.
- Zerbian, S., Genzel, S., & Kugler, F. (2010). Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). *Proceedings of Speech Prosody 2010*, Chicago.

### Appendix 1. Target words

Tone	Word	
HH (high + high)	xiang.yan	‘cigarette’
	wu.ya	‘crow’
	qing.wa	‘frog’
HL (high + low)	qiu.yin	‘earthworm’
	ying.wu	‘parrot’
	ban.ma	‘zebra’
LH (low + high)	gui.wu	‘ghost house’
	yu.yi	‘raincoat’
	hai.ou	‘seagull’