*Article*

# English Focus Perception by Mandarin Listeners

**Ratree Wayland [1],\*, Chelsea Guerra [1], Si Chen [2] and Yiqing Zhu [1]**

[1] Department of Linguistics, University of Florida, Gainesville, FL 32611-5454, USA; chelseaguerra@ufl.edu (C.G.); zhuyiqing@ufl.edu (Y.Z.)

[2] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China; sarah.chen@polyu.edu.hk

\* Correspondence: ratree@ufl.edu

check for updates

**Abstract:** This study compared how well native Mandarin and native English speakers can perceive prosodically marked focus in English echo questions. Twenty-five yes–no echo questions were produced with a sentence focus, a verb focus, and an object focus. After hearing each sentence, they were asked to choose a correct response. Native English listeners were more accurate than native Mandarin on verb and object focus, but not on sentence focus. More importantly, both groups confused object focus with sentence focus and vice versa. However, confusion between object and verb focus, and between object and sentence focus was infrequent. These results suggest that, in some cases, (1) acoustic prominence on the head of a phrase or its internal argument can project to the entire phrase and make the entire phrase focused, and (2) parallel transmission of the two functions of intonation, and cross-linguistic variation in focus marking (prosodically versus syntactically) may contribute to their perceptual ambiguity.

**Keywords:** intonation; pitch accent; perception; Mandarin; English

## 1. Introduction

Processing accuracy and speed of an utterance in discourse is facilitated by the ability to distinguish new from background information that is already shared by the conversation partners and how they are structured and conveyed in a sentence. The means by which this is accomplished is language specific. Some languages rely on syntactic structures to mark the focus constituent, whereas others may use focus-marking particles or prosodic features, such as phrasing and accentuation as focus-marking devices (e.g., Büring 2009; Zimmermann and Onea 2011; Lee et al. 2015; Hagoort 2019). In English, new or relevant information is typically pitch-accented. For example, as a felicitous response to the question, "What did John buy?", BICYCLE in the response, "John bought a BICYCLE.", is pitch-accented. Accent placement, in turn, is closely linked to new versus already-shared or given information. Specifically, focus allows the speaker to present to the listener part(s) of an utterance that s/he deems semantically and pragmatically prominent (Bishop 2012). Focused constituents may vary in breadth (broad or narrow) and in type (e.g., contrastive, non-contrastive).

This study investigates the ability to perceive broad sentence focus, verb narrow focus and object narrow focus in American English yes–no echo questions by native speakers of Mandarin Chinese. Echo questions allow all three types of focus to be examined in morphosyntactically identical utterances. Unlike English, focus is preferably realized syntactically, rather than prosodically, in Mandarin (Xu 2004). This cross-linguistic difference in focus markings may lead to perceptual difficulty among Mandarin listeners learning English. The remainder of the paper is organized as follows. Section 2 presents a brief overview on intonation phonology, information structure, and focus markings in English and Mandarin. The study is described in Section 3. The results are reported in Section 4 and discussed in Section 5. Finally, Section 6 summarizes and concludes.

## 2. Intonation, Information Structure, Focus, and Prominence

Phonological theories of speech prosody proposed that the utterance is hierarchically organized, with a larger prosodic unit dominating a smaller one (e.g., Selkirk 1978, 1980, 1981, 1984, 1995; Nespor and Vogel 1983). At the lower level, weak and strong syllables are grouped together to form feet. In turn, feet are grouped into phonological words, phonological words are bundled into phonological phrases, and phonological phrases into the entire utterance. However, research differs on how these prosodic units are determined. In one approach, prosodic units are largely identified based on syntactic structure, whereas, in the other, they are determined by their intonation features (see (Jun 1993, 1998) for a discussion on the two approaches).

Different prosodic hierarchies have been proposed over the years (see (Shattuck-Hufnagel and Turk 1996) for an overview). According to (Grice 2006, p. 778), each well-formed utterance contains at least one intonation phrase (IP), the highest prosodic unit above Word (W), represented by "tonal marking of some kind at one or both of its edges and/or at least one prominent within it." In turn, an IP may contain one or more intermediate phrase (ip). The Pierrehumbert's autosegmental–metrical (AM) model of intonation and its corresponding ToBI transcription system posit that each ip has at least one pitch accent (T*), which is linked to a stressed syllable. An ip can have additional pitch accents, and the final pitch accent in an ip is called the nuclear pitch accent. There are no pitch accents after the nuclear accent; therefore, any words following the nuclear accent are (postnuclear) unaccented and any accents before the nuclear accent are called prenuclear accents (Ayers 1996). The end of an ip is marked by a phrasal accent (T-) and the end (and rarely the beginning) of an IP is marked by a boundary tone (T%) (see Figure 1) (Pierrehumbert 1980; Beckman and Pierrehumbert 1986; Beckman and Ayers 1994).
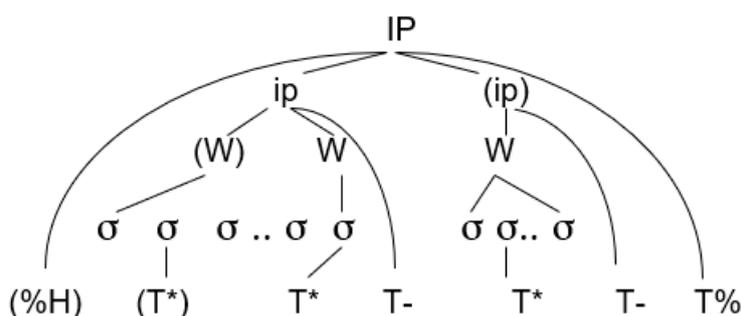


**Figure 1.** English intonation structure (Jun 2005, p. 31).

In English ToBI, five pitch accents have been proposed: L*, H*, L*+H, L+H*, H+!H* (Jun 2005). H tones denote high pitch targets corresponding to F0 peaks, L tones are low pitch targets associated with F0 troughs or valleys, and * indicates alignment with stressed syllables of the accented word. Pitch-accent assignment is determined post-lexically based on the meaning of the utterance in the discourse. In other words, pitch accents are assigned, not to every word with stress, but to words that are semantically and/or pragmatically prominent in discourse, and different types of pitch accents deliver different meanings (Jun 2005; Pierrehumbert and Hirschberg 1990). In a neutral production of an utterance, the last pitch accent, known as nuclear pitch accent, of an intermediate phrase is the most prominent pitch accent within the ip (Jun 2005).

Phrasal accents (L-, H-) occur at intermediate phrase boundaries, whereas boundary tones (L%, H%) are associated with the right edge of a full intonation phrase. H denotes the high pitch level in the local pitch range and L refers to low pitch level in the local pitch range (Beckman and Hirschberg 1994). Since an intonation phrase is usually comprised of one or more intermediate phrases and a boundary tone, a full intonation phrase boundary consists of two final tones: L-L%, L-H%, H-L% and H-H% (Beckman and Hirschberg 1994).

An utterance is often part of a longer interaction between the interlocutors; therefore, some of the information it contains is already shared between the speaker and the hearer while new information is

being added to the conversation (Hagoort 2019). Information is "given" if it has been earlier mentioned in the same or preceding sentence and all other information is "new" (Chafe 1974). Information structure refers to the way in which shared and new information are packaged in the sentence (Seuren 2009), but how this is accomplished varies from language to language. In some languages, syntactic structures are used to mark focus constituents, while focus-marking particles or prosodic features such as phrasing and accentuation may be employed in others (Hagoort 2019).

Accent placement is connected to focus, which, in turn, is closely related to "new" versus "given" or "shared with the listener" information (Ayers 1996, p. 22). In English, new or relevant information is typically pitch-accented. For example, as an answer to the question, "What did Mary cook for dinner?", the focus constituent, "STEAK," in "Mary cooked STEAK for dinner." is pitch-accented to convey the new information and may be perceived as being prominent by the listeners.

Perceived prominence is "subjective impression of 'prosodic strength,'" which can be driven by a bottom–up process based on acoustic cues or a non-signal-based, top–down process based on abstract patterns found in production (Bishop 2012, p. 5). Increased segmental duration, greater intensity, salient F0 features such as F0 peaks, F0 valleys, F0 movement or contours are acoustic cues that have been shown to be associated with perceived prominence. For the top–down process, it has been found that lexical frequency and the size of the focus constituents influenced perceived degrees of prominence (e.g., Bishop 2012; Cole et al. 2010). For example, Bishop (2012) found that the same object was perceived as being more prominent when it is narrowly focused than when it was part of a broad verb phrase and a sentence focus.

## 2.1. English Focus Marking

Focus is an important feature conveying a sentence's information structure (Bishop 2012). It is a means of emphasizing part of an utterance to specify new or informative information (Bolinger 1972; Gussenhoven 1983a; Selkirk 1995), "the information that the speaker presents *as semantically or pragmatically* prominent" (Bishop 2012, p. 238, emphasis is original). Focus may vary in two dimensions: size and type or contrastiveness (e.g., Gussenhoven 2008; Krifka 2008). The size of a focus constituent is the syntactic constituent that conveys the information determined by the speaker to be informative which, in turn, depends on the discourse context in which the sentence is produced (Bishop 2012). For example, the focus constituent of the answer (1D) depends on the contexts (1A–C) in which it is uttered:

1.   A: What happened?
      B: What did John do?
      C: What did John buy?
      D: John bought a bicycle.

In response to 1A, the focus constituent of 1D is the entire sentence or *sentence focus*. In the context of 1B, only the verb phrase (bought a bicycle) forms the focus constituent and is referred to as *VP focus*. Finally, as a response to 1C, the object noun-phrase (bicycle) forms the focus constituent and is referred to as *object focus*. Larger focus constituents such as sentence focus and VP focus are also referred to as broad focus compared to a narrow focus on a single word such as the object focus (Ladd 1980; Selkirk 1984). In other words, narrow focus refers to cases where only the agent, the action, the patient, etc., of an event is focused, whereas broad focus refers to cases where the entire event is focused.

Regarding focus type, two categories can be differentiated: non-contrastive and contrastive focus. Non-contrastive focus refers to the information required by WH- questions as in (1) where explicit alternative is not mentioned. Contrastive focus, on the other hand, refers to cases where an explicit alternative is mentioned (also referred to as 'corrective focus') as in 'bicycle' in 2B:

2.   A: Did John buy a car?
      B: (No), he bought a bicycle.

Contrastive focus often involves cases of narrow focus (e.g., Baumann et al. 2007; Hanssen et al. 2008). However, a focus of any size can be contrastive or non-contrastive. That is, focus size and focus type are orthogonal properties of informational structure (Bishop 2012). For instance, (2B) could have contrastive VP focus if it is uttered as a response to such questions as, "Did John fix the car?"

However, the distinction between contrastive and non-contrastive focus is not recognized by all researchers. Rooth (1992), for instance, argued that there is no principled difference between the two types of focus. According to Rooth (1992), two semantic representations are evoked for each expression: its actual meaning and a set of alternatives. When a constituent in an expression is focused, the alternative set consists of the expression itself and all expressions with an alternative substituted for the focus-marked constituent. If there is no focus within the expression, then the alternative set consists only of the expression itself (Breen et al. 2010).

In English, prosodically focused words are produced with an expanded pitch range, higher intensity, longer duration and/or a pause before or after it (e.g., Eady and Cooper 1986; Jun 2005; Xu and Xu 2005; Xu et al. 2012). In an acoustic study of focus production on the first and the last noun phrase in English statements and questions, Eady and Cooper (1986) observed an increase in duration for focused words in both statements and questions. For sentences with broad or sentence-final focus, the F0 peak on the last key word was significantly higher in questions than in statements. On the other hand, for sentences with focus on the first key word, the peak F0 on the focused word itself was comparable between statements and questions. However, while the F0 toplines of statements lower, they remain high for the remaining of the utterance for questions. In addition, the F0 contour of focused words is rising in questions but falling in statements. Xu and Xu (2005) examined focus realization in American English declarative statements and reported similar results; specifically, the pitch range is expanded under focus, reduced after focus, but remains unchanged before focus. A reduction in pitch range in post-focus position is known as post-focus compression (PFC) (Xu et al. 2012).

Foci of different sizes are acoustically differentiated. For example, to distinguish narrow focus on an object from broad VP focus, native English speakers may place more emphasis on the object and produce it with a higher F0 peak (Ladd 2012). Other studies e.g., (Eady et al. 1986; Sityaev and House 2003; Xu and Xu 2005; Breen et al. 2010), have also shown that focus constituents of different size are systematically distinguished by acoustic features. Eady et al. (1986) found that native English speakers produced narrow-focused words with a higher F0, and a longer duration compared to the same words uttered in neutral- or broad-focus sentences. However, broad focus on the verb phrase is pronounced with an increase in duration without an accompanying higher F0 peak. F0 peaks of narrow-focused words were also found to be comparable in single- and dual-focus sentences. However, unlike single-focus sentences, the word following an initial focused item in dual-focus sentences did not exhibit the low F0 value characteristic of post-focused words. Sityaev and House (2003) found a small but significant phonetic difference between broad focus, narrow focus and contrastive focus, with a greater difference between contrastive and non-contrastive (broad and narrow), particularly in sentence final position. Breen et al. (2010) found that focus location (subject, verb, and object) and focus type are reliably signaled by native speakers of English with greater acoustic intensity, longer duration, and higher mean and maximum F0, particularly when they were aware of prosodic ambiguity between focus type.

However, despite acoustic prominence, ambiguity between broad and narrow focus could still occur through the focus projection process (Selkirk 1984, 1995). According to Selkirk, an accented word is Focus (F)-marked where F-marking is based on a semantic feature to be used in interpretation. F-marking of higher constituents is, however, projected according to a set of rules: F-marking of the head of a phrase licenses F-marking of the entire phrase; and F-marking of an internal argument of a head licenses the F-marking of the head. Thus, according to these principles, ambiguity between a reading where only the object is narrowly focused and the one where the entire verb phrase is focused could arise in a clause containing a transitive verb whose direct object is accented, since F-marking on

the object can license F-marking of the verb phrase. Supports for the focus projection principles have been previously reported (e.g., Birch and Clifton 1995; Gussenhoven 1983b; Welby 2003).

*2.2. Mandarin Focus Marking*

F0 is used to distinguish meaning at the syllable/word level in Mandarin. For example, the syllable /ma/ means 'mother,' 'hemp,' 'horse,' and 'admonish or scold' when produced with a high-level (Tone 1), a rising (Tone 2), a falling–rising (Tone 3), and a falling (Tone 4) F0, respectively. In contrast to English, focus is preferably realized syntactically, rather than prosodically, in Mandarin (Xu 2004). Xu (2004) differentiated two types of focus in Mandarin, informational and contrastive focus. Informationally focused words are placed in the default, sentence-final position (see 3B), whereas words that are in contrastive focus are surrounded by focus markers *shi . . . de* (Xu 2004; Paul and Whitman 2008) (see 4A).

3. 　A. Ni gangcai he-le shenme?
   　　You just-now drank what
   　　'What did you drink just now?'
   　　B. Wo he-le *kafei*
   　　I drank coffee
   　　'I drank coffee.' (Xu 2004, pp. 280–81)

4. 　A. Ta *shi* tan gangquin *de*
   　　3SG *shi* play piano *de*
   　　'He is a pianist.' (Paris 1998, p. 149)

Unlike English, phonological prominence (i.e., prosodically prominent word) is not required in Mandarin when a sentence has broad focus, or when narrow focus is marked syntactically (Baker 2010; Paul and Whitman 2008; Xu 2004).

Acoustic analysis of prosodic focus in Mandarin showed that prosodically marked, narrow contrastive focus is expressed by an increased duration and expanded F0 range for the focused word, a lower maximum F0, and reduced F0 range for all following post-focused words with little to no change to words in pre-focused positions (e.g., Wang and Xu 2006; Liu and Xu 2005; Xu et al. 2004; Yuan 2004; Xu 1999). According to (Xu 1999, p. 95), "the full realization of a focus requires that the F0 of all words after the focus be suppressed." When this cannot be implemented, as when focus is on the last word, for example, utterances with final narrow focus may not be reliably differentiated from utterances with broad focus (Xu 1999; Cooper et al. 1985; Jin 1996) due to a lack of a reliable F0 as perceptual cues (Bunnell et al. 1997). Post-focus compression is a highly effective perceptual cue for focus in Mandarin (Xu et al. 2004, 2012; Liu and Xu 2005; Chen et al. 2009).

Prosodic realization of focus varies from tone to tone in Mandarin (e.g., Shih 1988; Xu 1999). For tones with a high pitch target (e.g., Tone 1), its pitch is raised under focus, but it remains unclear whether the pitch is lowered or remains the same for a focused low tone (e.g., Tone 2, Tone 3) (Shih 1988; Xu 1999; Chen and Gussenhoven 2008). Lee et al. (2016) found Tone 3 to be the least accurately identified (77.1%) in comparison to Tone 1 (90.8%), Tone 2 (90.2%), and Tone 4 (92.5%). The relatively low identification accuracy rate for Tone 3 is attributed to its smaller pitch range expansion, smaller increases in intensity, and increases in the duration, intensity, and pitch range of surrounding syllables within the same phrase due to a local dissimilarity effect (Lee et al. 2016).

However, focus perception accuracy is more affected by its position in the utterance than by tone, with sentence-medial focus being the most accurately identified (92.9%) followed by sentence-initial (87.2) and sentence-final (75.5%) focus (Yuan 2004).

Few studies have examined the ability to process English phrase-level prosody by non-native listeners. This study fills this research gap by investigating Mandarin Chinese listeners' ability to determine focus locations in English interrogative intonation. Differences in focus marking preference between Mandarin and English could impair English focus identification among Mandarin listeners.

*2.3. Non-Native Perception of Phrasal Prosody*

The empirical evidence available suggests that native listeners process sentence accent and discourse implication conveyed by native prosody with ease (Cutler et al. 1997). However, few studies have investigated perception of non-native post-lexical prosody. Results of some of these previous studies suggested that listeners have difficulty perceiving non-native prosody (e.g., Cruz-Ferreira 1987; Akker and Cutler 2003; Baker 2010). Cruz-Ferreira (1987) examined comprehension of non-native intonation in Portuguese and in English, by native English and native Portuguese speakers, respectively, and found that the non-natives showed hesitation in assigning meaning to the L2 intonation patterns and performed significantly worse than the natives for intonation patterns that do not exist in their native language. Akker and Cutler (2003) found Dutch learners of English to be less efficient in processing accented and focused words in English than native speakers. More recently, Baker (2010) found that Korean and Mandarin learners of English had difficulty matching intended pitch accent to focus location. Particularly, they had difficulty identifying subject narrow focus and sentence broad focus. For the subject narrow focus context, they erroneously accepted a response with the nuclear-accented object to a question that asks about the subject. For example, they accepted the answer, "Kim bought a FAN.", with focus on the object as a felicitous response to a question narrowly focused on the subject, "Who bought a fan?" The fact that an object is nuclear-accented in a wide variety of contexts in English including the sentence broad focus, the verb phrase broad focus, and the object narrow focus may have led non-native speakers to extend the pattern to a wrong context (Baker 2010). Both Korean and Mandarin speakers also failed to reject an answer with the nuclear pitch accent on the subject as a response to a sentence with broad focus. For example, they accepted the answer "KIM bought a fan." as a felicitous response to a question, "What happened?", suggesting that non-native speakers may accept nuclear pitch accents (i.e., KIM) that are within the focused constituent (the entire sentence) even when they are not placed in the standard location (i.e., "fan") (Baker 2010). However, they were relatively more successful at rejecting an answer with incorrect nuclear pitch accent placement on the subject in the verb broad focus context. For example, they accurately rejected the answer "KIM bought a fan." as a felicitous answer to the question "What did Kim do?" In this case, the nuclear accent (i.e., KIM) is not in a common location or within the focused constituent (the verb phrase), so they may have seemed more clearly incorrect." (Baker 2010, p. 4).

**3. This Study**

Similar to Baker (2010), this study explores the ability to identify location of focus in an utterance. Three focus types, namely broad sentence focus, narrow verb focus and narrow object focus are examined. It complements Baker (2010)'s study by investigating non-WH echo question instead of WH-questions. Echo questions are questions that are structurally the same as the previous utterance (Artstein 2002; Repp and Rosin 2015). By using echo interrogative questions as stimuli, all three focus types can be examined in morphosyntactically identical utterances. In addition, "the entire echo question is given, so none of its parts needs to be marked with focus; therefore, focus can serve the purpose of indicating disputed (rather than new) material" (Artstein 2002, p. 98). Echo questions can be asked for various reasons including auditory failure or emotional arousal. For example, the speaker might request a clarification about the reference of the expression in the previous utterance (e.g., A: "She likes the baby."; B: "She likes WHO?"), the speaker might not believe what s/he heard and want to double-check, or the speaker might want to express his or her emotion (e.g., amazed, indignant) in response to what s/he heard. Intonational markings for different types of echo questions have been found to vary. Except for reference echo questions, it has been suggested that echo questions in German and English exhibit an obligatory rise at the end (H-H%) and the in situ wh-phrase is narrowly focused, e.g., (Reis 2012), and carry either the (L+)H* or the L* nuclear accent. (L+)H* followed by H-H% typically signals an auditory failure, whereas emotional arousal due to disbelief or surprise is typically signaled by a L* accent followed by H-H% (Repp and Rosin 2015). In this study, the focus is

on echo questions without wh-phrases (e.g., A: "She liked the sweater."; B: "She liked the SWEATER?") signaled by L* H-H% intonation contour (See Figure 1).

Baker (2010) employed a question–answer prosodic matching task. The problem with such a metalinguistic judgement task is that the participants' interpretation of either the question or the answer is not known (Breen et al. 2010). Our study aimed to measure the participant's interpretation of the question. Specifically, our participants heard the question and were then asked to choose an appropriate response from three alternative forced choices written in standard English orthography without any indication of focus. In addition to obviating the need to train the participants on the meaning of prosody, the participant's interpretation of the question can be measured based on their answer choice. Two main questions guided our research: (1) Are the three focus sizes (broad-sentence focus, narrow-verb and narrow-object focus) equally differentiated?; (2) Is the ability to differentiate the three focus sizes affected by how focus is marked in the native language? Based on Baker (2010)'s findings, we hypothesized that verb narrow focus (VF) will be more accurately perceived than both the object narrow focus (OF) and the sentence broad focus (SF) among both groups of listeners due to the prevalence of nuclear-accented objects in English and the absence of PFC. In addition, consistent with Selkirk's focus projection principle, ambiguity between object focus and sentence focus, as well as between object focus and verb focus, is expected. However, with native experience, native listeners are expected to outperform non-native listeners.

### 3.1. Methods

#### 3.1.1. Participants

Sixteen native Mandarin Chinese speakers (7 males) participated in this experiment. The mean age for this group of participants was 24.8 (SD = 3) years. None of them had been in the US for more than 2 years, and most participants had only been in the US for 5 months. All reported having at least an intermediate level knowledge of English, most with 10–15 years of experience learning English, while 1 participant reported having 23 years of experience. All are graduate students at the University of Florida where a minimum TOEFL score of 550 (paper test); 80 (internet-based), International English Language System (IELTS) score of 6, Michigan English Language Assessment Battery (MELAB) score of 77, and a verbal Graduate Record Examination (GRE) score of 140 are required for admission. None reported prior knowledge of any other West Germanic languages.

Twelve (3 males, 9 females) native English speakers, also students at the University of Florida, participated in this experiment. The mean age for this group was 21.8 (SD = 2) years. None of the participants in either of the two groups reported having any known problems with their hearing, reading, or speech. None reported prior experience with Mandarin. All subjects gave their informed consent for inclusion before they participated in the experiment. The study was conducted in accordance with the University of Florida Institutional Review Board.

#### 3.1.2. Materials

The stimuli were 25 sets of English yes–no, echo question–answer pairs as shown in (5). Only the questions were recorded and aurally presented to the participants. Q1 is produced with a broad sentence focus, Q2 a narrow verb focus, and Q3 was uttered with a narrow focus on the object. The questions were produced and recorded (44,100 Hz sampling rate, 16-bit quantization rate) in Praat (Boersma and Weenink 2018) by a male native English speaker with phonetics training. Spectrograms, pitch tracks, and ToBI transcriptions of the three questions are shown in Figure 2. A total of 75 stimuli (25 questions x 3 focus types) were produced.

5.　　Q1: [She loved the new sweater?]$_{Focus}$
　　　A1: No, she didn't.
　　　 Q2: She [LOVED] $_{Focus}$ the new sweater?
　　　A2: No, she hated the new sweater?

Q3: She loved the new [SWEATER?]~Focus~
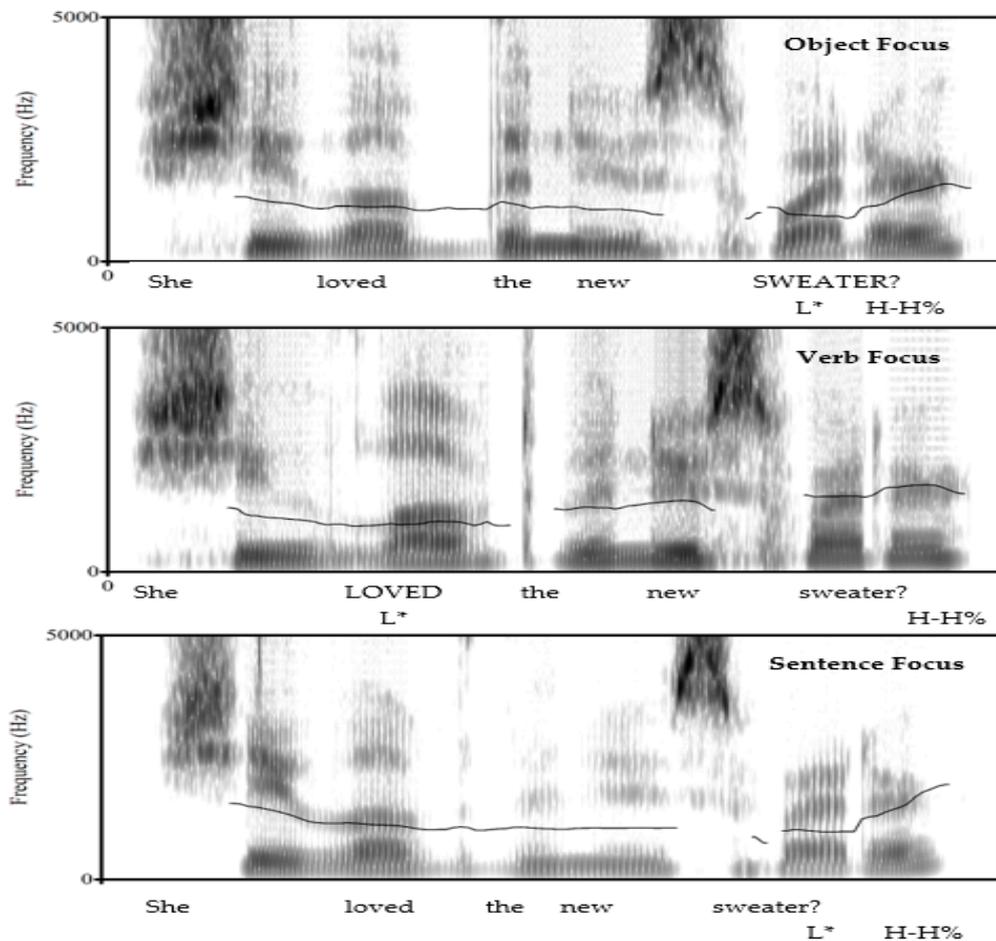A3: No, she loved the new TV.



**Figure 2.** Spectrograms and pitch tracks of the question "She loved the new sweater?" produced with a narrow focus on the object (**top**), the verb (**middle**), and a broad sentence focus (**bottom**).

F0 values of the target words in each utterance were automatically extracted using ProsodyPro, a custom-written script for the Praat program by Xu (2013). To avoid consonant onset F0 perturbation, the F0 extraction was taken from the first full vocal pulse to the last vocal pulse of the vowel. In addition, F0 values were extracted separately for the first and the second syllables of the last disyllabic nouns.

Figure 3 shows time-normalized pitch tracks for all three focus types averaged across 25 utterances each. A multivariate ANOVA analysis indicated that: (1) the mean F0 of the first words (e.g., *she*) is significantly higher for OF questions than for VF questions (151 Hz vs. 135 Hz, $p = 0.03$); (2) the mean F0 of the verb (e.g., *loved*) is significantly lower for VF questions than for OF questions (117 Hz versus 135 Hz, $p = 0.025$); and, (3) the mean F0 of word #3, the article (e.g., *the*), word #4, the adjective (e.g., *new*), and the first syllable of the object (e.g., *sweat*), or the first word of the two-word compound (e.g., *bus*), are significantly higher for VF questions than for OF and SF questions ($ps < 0.000$). However, mean F0 of the last syllable (e.g., *er)*, or the last word (e.g., *stop*), of the object is comparable across the three focus types ($p > 0.10$).
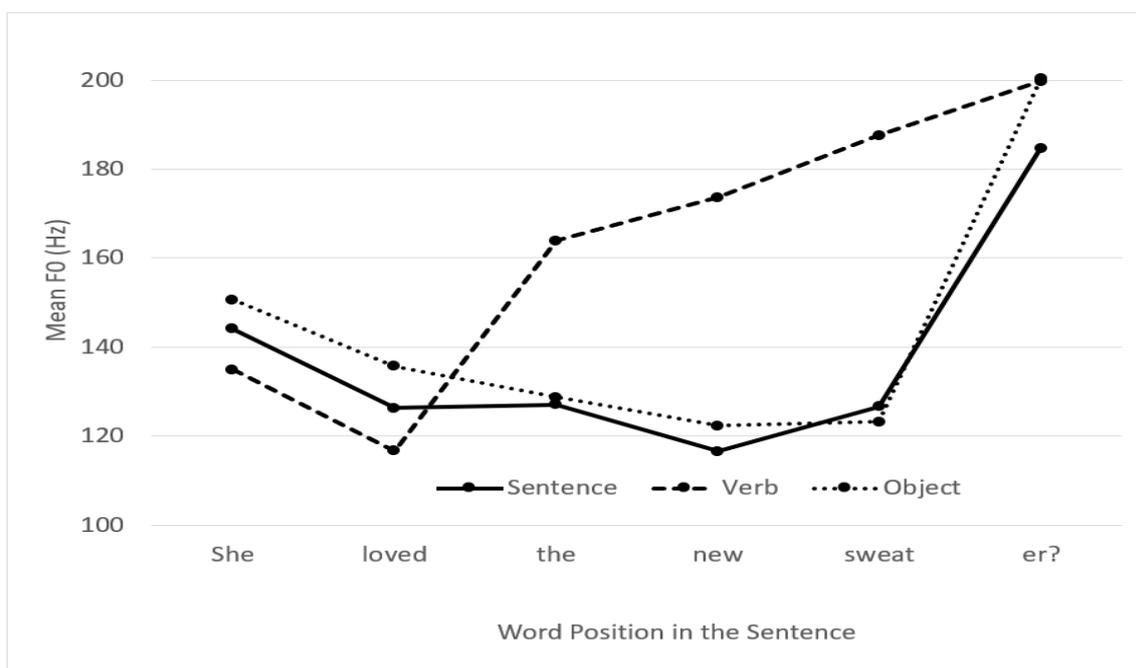
**Figure 3.** Time-normalized pitch tracks of broad, verb-narrow and object-narrow focus questions, averaged across 25 utterances each.

3.1.3. Procedures

Participants were tested individually in a quiet room. All three questions in each set were randomly presented aurally through headphones (Sennheiser HD 280 Pro) one at a time using Eprime (2.0). After hearing each question, they were instructed to choose one of the three possible answers (See A1, A2, and A3 in (5)) printed on an answer sheet in standard English orthography with no focus location indicated. A total of 75 questions (25 sets) were presented.

To familiarize the participants with the task, 6 practice trials preceded the experimental trials. They can repeat the trial by pressing the number 3 on the number pad. They are asked to press 5 to move on to the next question. Responses were coded as correct (1) or incorrect (0) for statistical analyses. Data from one English participant were excluded as an outlier for poor performance (2 SD below the group mean).

**4. Results**

Log-transformed proportions of response accuracy for each focus type by the two groups of listeners are shown in Figure 4. Overall, native English speakers outperformed native Mandarin speakers on all three focus types. For both groups, VF was the most accurately perceived, whereas OF focus was the least.
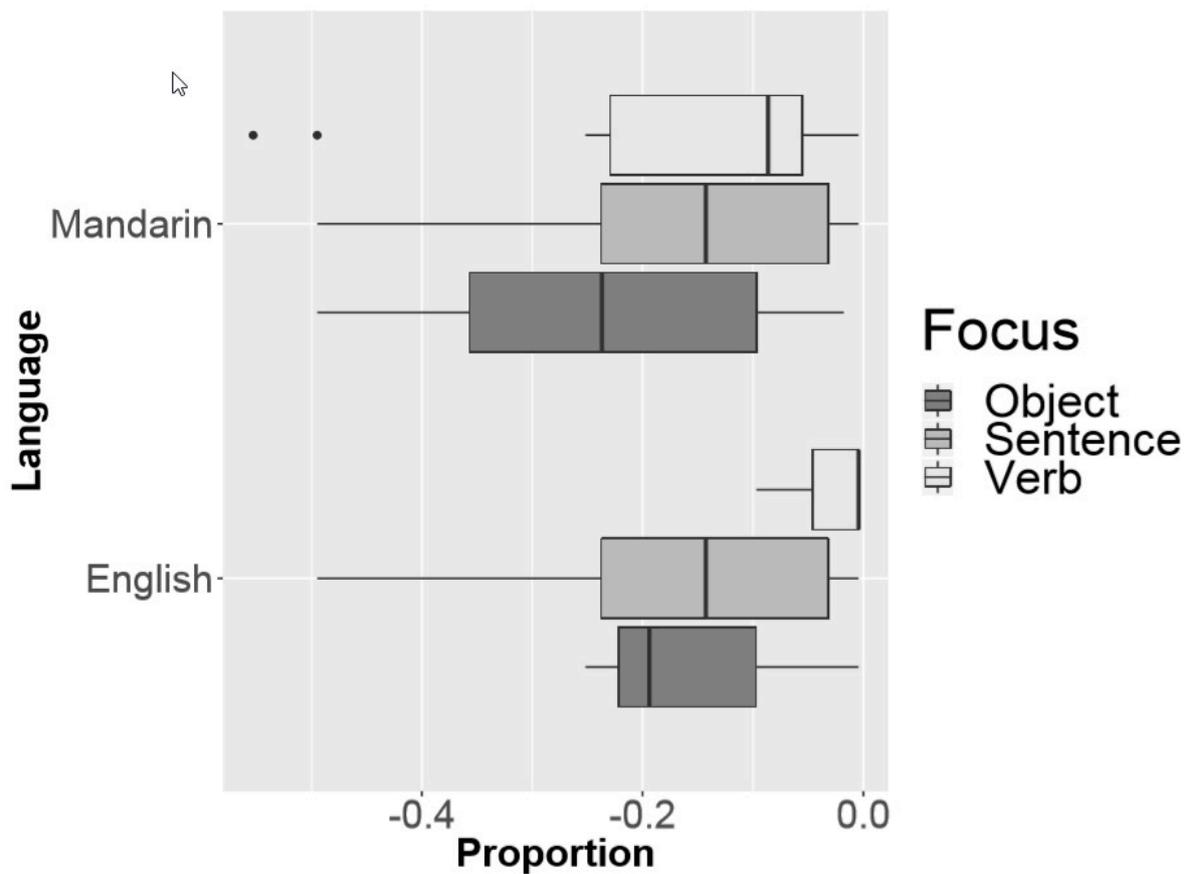
**Figure 4.** Proportions of correct response for three types of focus for the two language groups.

Tables 1 and 2 below show the confusion matrices for the perceptual accuracy of the three focus types for English and Mandarin speakers, respectively. English speakers were more accurate at detecting VF than either OF or SF. SF is also frequently confused with OF and vice versa. Mandarin speakers are less accurate than English speakers overall, but in contrast to English speakers, their ability to locate VF and SF is equally accurate, whereas their ability to identify OF was the worst. Like English speakers, they frequently confused OF with SF and vice versa.

**Table 1.** Confusion matrix for the perceptual accuracy of the three focus types for English speakers.

| Focus Type | Verb | Object | Sentence |
|---|---|---|---|
| **Verb** | **260 (94.54%)** | 8 (2.91%) | 7 (2.55%) |
| **Object** | 22 (8%) | **197 (71.64%)** | 56 (14%) |
| **Sentence** | 15 (3.75%) | 60 (21.82%) | **200 (72.73%)** |

**Table 2.** Confusion matrix for the perceptual accuracy of the three focus types for Mandarin speakers.

| Focus Type | Verb | Object | Sentence |
|---|---|---|---|
| **Verb** | **292 (73%)** | 39 (9.25%) | 71 (17.75%) |
| **Object** | 38 (9.5%) | **241 (60.25%)** | 121 (30.25%) |
| **Sentence** | 30 (7.5%) | 85 (21.25%) | **285 (71.25%)** |

A generalized linear mixed-effects (logistic regression) model was fitted to the response accuracy data (1 = correct, 0 = incorrect) in R (R Development Core Team 2008) using the package lme4 (Bates et al. 2015). The maximal model that converged included random intercepts for subjects and items:

full.model <- glmer(Response ~ factor(L1 background, levels = c(Mandarin, English))*factor(Focus type, levels = c(sentence, object, verb)) + (1|Subject) + (1|Item), data = datafile, family = binomial).

The results obtained (Table 3) show that, for both groups combined, object focus perception is significantly less accurate than sentence focus (the default focus type in the model). As shown in Tables 1 and 2, these results are largely driven by the performance of Mandarin speakers. Unlike English speakers, they are less accurate on object focus than on sentence focus.

**Table 3.** Logistic mixed-effects model output.

|  | Estimate | Std. Error | z Value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 1.16 | 0.30 | 3.86 | <0.001 ** |
| (L1 background) English | 0.06 | 0.45 | 0.12 | 0.90 |
| (Focus type) Object | **−0.60** | **0.17** | **−3.63** | **<0.001 **** |
| (Focus type) Verb | 0.10 | 0.17 | 0.61 | 0.54 |
| (L1 background) English: (focus type) Object | **0.54** | **0.27** | **2.02** | **0.043 *** |
| (L1 background) English: (focus type) Verb | **2.01** | **0.36** | **5.64** | **<0.001 *** |

** indicates statistical significance at $p < 0.001$ level, and * indicates statistical significance at $p < 0.05$ level.

The analysis also confirmed that English speakers are significantly more accurate than Mandarin speakers in identifying both object focus and verb focus.

## 5. Discussion

Native speakers of Mandarin and native speakers of American English were evaluated for their ability to recognize sentence broad focus, verb narrow focus and object narrow focus in English yes–no echo questions. Both verb narrow focus and object narrow focus were produced with a low (L*) nuclear pitch accent and the sentence broad focus bears a default L* nuclear pitch accent on the object. Native speakers of English were more accurate than native Mandarin speakers overall, particularly on the verb and the object narrow focus, but not on the sentence broad focus.

However, both groups struggled to differentiate between the sentence-final object narrow focus and the sentence broad focus. This finding agrees with that of Liu and Xu (2005), who found that Mandarin speakers frequently confused neutral focus in questions with final focus in statements. It is also consistent with the observation that interpretation of the focused constituent in sentence-final position is often ambiguous, not only in English, but also in other languages including German, Dutch, Spanish, European Portuguese and Italian. (Nooteboom and Kruyt 1987; Birch and Clifton 1995; Welby 2003; Zubizarreta 2014; Frota 1997; Avesani and Vayra 2003). For example, Welby (2003) found that native English listeners judged a sentence like "I read the DISPATCH." with acoustic prominence on 'dispatch' as a felicitous response to a question narrowly focusing the object, "What newspaper do you read?", as well as to a question broadly focusing the entire event, "How do you keep up with the news?" Similar results were reported by Birch and Clifton (1995). These findings are predicted by the focus projection principles proposed by Selkirk (1984, 1995). A similar claim was also made by Gussenhoven (1983b, 1999). Specifically, Selkirk (1995) argued that, through the focus projection process, an acoustic prominence on the head of a phrase or its internal argument can project to the entire phrase, thus making the entire phrase focused. The focus projection process appears to also account for the confusion between narrow object focus and broad sentence focus among both Mandarin and English speakers in our study.

A lack of perceptible difference between object focus and sentence focus may have also been responsible for the result. Gussenhoven (1983b) found that perceptible difference between narrow and broad focus exists in some productions, but that the listeners cannot reliably use it to tell in which context (narrow or broad focus) the sentence was produced. In contrast to Gussenhoven's finding, Rump and Collier (1996)'s listeners can accurately discriminate narrow and broad focus using pitch

cues. Unfortunately, our English and Mandarin listeners could not rely on pitch cues to discriminate between object narrow focus and sentence broad focus. As shown in Figure 2, no significant difference in pitch values were found for any portions of the question uttered in the narrow object focus and the broad sentence focus contexts. In addition, Bunnell et al. (1997) found F0 to be less effective in focus perception at the end of the utterance (Bunnell et al. 1997). Furthermore, Xu (1999) argued that the full realization of a focus requires that the F0 of all words after the focus be suppressed. A lack of post-focus compression in utterances with focus on the last word can, therefore, lead to perceptual confusion between utterances with final narrow focus and utterances with broad focus (Cooper et al. 1985; Jin 1996).

Besides a lack of difference in acoustic prominence, the prevalence of nuclear pitch accent on the object in English may have also contributed to object and sentence focus confusion. As Baker (2010) pointed out, a nuclear pitch accent object occurs in subject broad focus, verb broad focus and object narrow focus contexts. It appears that natives and non-natives alike easily mistake one for the other, and alterations in acoustic prominence produced by the speaker were insufficient and/or were treated as irrelevant for the metrical structuring of the two focus types.

The confusion between object focus and sentence focus is also consistent with the functional view of intonation (Xu 2005; Liu and Xu 2005). According to this view, components of intonation or speech melody are individually and independently defined and organized by their communicative functions (e.g., lexical stress, focus, sentence type). More importantly, these functions are encoded in parallel, each with its own distinctive encoding schemes, specifying its surface acoustic/articulatory values. However, due to limitations in acoustic/articulatory dimensions and space (e.g., F0, intensity, duration), interactions among the encoding schemes of these functions frequently occur, leading to a delicate balance between functions that share the same articulatory/acoustic parameters, as is the case with sentence-final focus and question (Xu and Xu 2005). In turn, ambiguity or confusion may result from overlapped articulatory/acoustic characteristics between two or more communicative functions. According to this framework, ambiguity between object focus and sentence focus found in our study may have resulted from interference in transmission of the acoustic/articulatory features associated with two communicative functions, namely questions and focus. Specifically, the F0 rise at the end of the utterance required to signal a question obscures or raises the low pitch target on the object (see Figure 2), and reduces its perceived prominence.

However, the fact that native English speakers were more accurate than Mandarin speakers on object narrow focus relative to sentence broad focus suggested that native experience reduced the negative impact of a lack of acoustic distinctiveness due to parallel transmission of these two focus contexts. More specifically, compared to native English speakers, Mandarin speakers are less familiar with associating low F0 value with focus. In addition, unlike English, no prosodically marked word is required in Mandarin, particularly when a sentence has broad focus, or when narrow focus is marked syntactically (Baker 2010). In addition, when prosodically marked, focus raises rather than lowers the pitch of high tones. For low tones, it remains inconclusive whether focus decreases their low pitch targets.

Finally, a lack of ambiguity between verb focus and object focus, as well as between verb focus and sentence focus, are inconsistent with the focus projection principle. According to Selkirk (1984) and Gussenhoven (1983b), ambiguity between an object-only focused reading and a reading where the entire verb phrase is focused is predicted in a clause containing a transitive verb with an accented direct object. Similarly, ambiguous interpretation between a verb-only focus reading and a reading where the entire event is focused is also expected. While support for the focus projection principles have been reported in previous studies e.g., (Birch and Clifton 1995; Gussenhoven 1983b; Welby 2003), our findings are inconsistent with the hypotheses. In our case, the focus projection process, both to and from a constituent containing a narrowly focused verb carrying an L* nuclear pitch accent, the rarest type of pitch accent (Jun 2006) appeared to have been blocked. That is, English and Mandarin listeners

exhibited a preference for verb focus when it is prominently marked with a low pitch target. Whether this is also true for other types of pitch accent awaits further research.

## 6. Summary and Conclusions

This study compared how well native Mandarin and native English speakers can perceive prosodically marked focus. They were asked to choose a correct response for twenty-five English echo yes–no questions produced with a sentence-medial verb narrow focus, a sentence-final object focus, and a sentence broad focus associated with a low pitch accent (L*). Native English listeners were more accurate than native Mandarin listeners on verb and object narrow focus, but not on sentence broad focus. More importantly, both groups confused object narrow focus with sentence broad focus and vice versa. However, confusion between object focus and verb focus, and between verb focus and sentence focus, was relatively infrequent. These results suggested that, in some cases, acoustic prominence on the head of a phrase or its internal argument can project to the entire phrase and make the entire phrase focused, that parallel transmission of the two functions of intonation, and that cross-linguistic variation in focus marking (prosodically versus syntactically) contribute to their perceptual inaccuracy, the latter among Mandarin speakers, and native exposure reduces but does not remove its effect among native English speakers.

**Author Contributions:** All four authors contributed substantially to the work: R.W.: conceptualization, methodology, visualization, acoustic analyses, writing, review and editing; S.C.: statistical analyses, visualization, review and editing; C.G.: conceptualization, methodology, data curation, and original draft preparation; Y.Z.: statistical analyses, visualization, review and editing.

**Conflicts of Interest:** The authors declare no conflict of interests.

## References

Akker, Evelien, and Anne Cutler. 2003. Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition* 6: 81–96. [CrossRef]

Artstein, Ron. 2002. A focus semantics for echo questions. In *Workshop on Information Structure in Context*. Stuttgart: IMS, University of Stuttgart, vol. 98, p. 107.

Avesani, Cinzia, and Mario Vayra. 2003. Broad, narrow and contrastive focus in Florentine Italian. Paper presented at 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9; Rome: Consiglio Nazionale delle Ricerche, vol. 2, pp. 1803–6.

Ayers, Gayle. 1996. Nuclear accent types and prominence: Some psycholinguistics experiments. Ph.D. thesis, The Ohio State University, Columbus, OH, USA.

Baker, Rachel. E. 2010. Non-native perception of native English prominence. Paper presented at Speech Prosody 2010—Fifth International Conference, Chicago, IL, USA, May 10–14.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2015. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8. Available online: http://CRAN.R-project.org/package=lme4 (accessed on 1 September 2019).

Baumann, Stefan, Johannes Becker, Martine Grice, and Doris Mücke. 2007. Tonal and articulatory marking of focus in German. Paper presented at 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, August 6–10; Saarbrücken: Pirrot GmbH Dudweiler, pp. 1029–32.

Beckman, Mary, and Gayle Ayers. 1994. *Guidelines for ToBI Labelling*. Version 2.0. Manuscript and Accompanying Speech Materials. Columbus: Ohio State University.

Beckman, Mary E., and Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3: 255–309. [CrossRef]

Beckman, Mary E., and Julia Hirschberg. 1994. *The ToBI Annotation Conventions*. Manuscript and Accompanying Speech Materials. Columbus: Ohio State University.

Bishop, Jason. 2012. Information structural expectations in the perception of prosodic prominence. In *Prosody and Meaning*. Edited by G. Elordieta and P. Prieto. Berlin: Mouton de Gruyter, pp. 239–70.

Birch, Stacy, and Charles Clifton. 1995. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 38: 365–91. [CrossRef] [PubMed]

Boersma, Paul, and David Weenink. 2018. Praat: Doing Phonetics by Computer. Version 6.0.37. Available online: http://www.praat.org/ (accessed on 1 March 2018).

Bolinger, Dwight. 1972. Accent is predictable (if you're a mind-reader). *Language* 48: 633–44. [CrossRef]

Breen, Mara, Evelina Fedorenko, Michael Wagner, and Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25: 1044–98. [CrossRef]

Bunnell, H. Timothy, Steven R. Hoskins, and Debra Yarrington. 1997. Interactions among *F*0, Duration, and Amplitude in the Perception of Focus. *The Journal of the Acoustical Society of America* 102: 3203. [CrossRef]

Büring, Daniel. 2009. Towards a typology of focus realization. In *Information Structure*. Oxford: Oxford University Press, pp. 177–205.

Chafe, Wallace. L. 1974. Language and consciousness. *Language* 50: 111–33. [CrossRef]

Chen, Yiya, and Carlos Gussenhoven. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36: 724–46. [CrossRef]

Chen, Szu-Wei, Bei Wang, and Yi Xu. 2009. Closely related languages, different ways of realizing focus. Paper Presented at Tenth Annual Conference of the International Speech Communication Association, Brighton, UK, September 6–10.

Cole, Jennifer, Yansook Mo, and Mark Hasegawa-Johnson. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1: 425–52. [CrossRef]

Cooper, William E., Stephen J. Eady, and Pamela R. Mueller. 1985. Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America* 77: 2142–56. [CrossRef]

Cruz-Ferreira, Madalena. 1987. Non-native interpretive strategies for intonational meaning: An experimental study. *Sound Patterns in Second Language Acquisition* 36: 103–20.

Cutler, Anne, Delphine Oahan, and Wilma van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40: 141–201. [CrossRef]

Eady, Stephen J., William E. Cooper, Gayle V. Klouda, Pamela R. Mueller, and Dan W. Lotts. 1986. Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29: 233–51. [CrossRef]

Eady, Stephen J., and William E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America* 80: 402–15. [CrossRef]

Frota, Sonia. 1997. On the prosody and intonation of focus in European Portuguese. In *Issues in the Phonology and Morphology of the Major Iberian Languages*. Washington: Georgetown University Press, pp. 359–92.

Grice, Martine. 2006. Intonation. In *Encyclopedia of Language and Linguistics*, 2nd ed. Edited by Brown Keith. Oxford: Elsevier, vol. 5, pp. 778–88.

Gussenhoven, Carlos. 1983a. Focus, mode and the nucleus. *Journal of Linguistics* 19: 377–417. [CrossRef]

Gussenhoven, Carlos. 1983b. Testing the reality of focus domains. *Language and Speech* 26: 61–80. [CrossRef]

Gussenhoven, Carlos. 1999. On the limits of focus projection in English. In *Focus: Linguistic, Cognitive, and Computational Perspectives*. Edited by Peter Bosch and Rob van der Sandt. Cambridge: Cambridge University Press, pp. 43–55.

Gussenhoven, Carlos. 2008. Notions and subnotions in information structure. *Acta Linguistica Hungarica* 55: 381–95. [CrossRef]

Hagoort, Peter. 2019. The neurobiology of language beyond single-word processing. *Science* 366: 55–58. [CrossRef]

Hanssen, Judith, Jörg Peters, and Carlos Gussenhoven. 2008. Prosodic effects of focus in Dutch declaratives. Paper presented at Speech Prosody, Boston, MA, USA, May 31–June 3; Campiñas: Editora RG/CNPq, pp. 609–12.

Jin, Shunde. 1996. An Acoustic Study of Sentence Stress in Mandarin Chinese. Ph.D. dissertation, The Ohio State University, Columbus, OH, USA.

Jun, Sun-Ah. 1993. The Phonetics and Phonology of Korean Prosody. Ph.D. dissertation, The Ohio State University, Columbus, OH, USA.

Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15: 189–226. [CrossRef]

Jun, Sun-Ah. 2005. Prosody in sentence processing: Korean vs. English. *UCLA Working Papers in Phonetics* 104: 26–45.

Jun, Sun-Ah, ed. 2006. Prosodic typology. In *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, vol. 1, pp. 430–48.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55: 243–76. [CrossRef]

Ladd, Robert. 1980. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. Robert. 2012. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lee, Yong-Cheol, Ting Wang, and Mark Liberman. 2016. Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in Psychology* 7: 1058. [CrossRef] [PubMed]

Lee, Yong-Choel, Bei Wang, Sisi Chen, Martine Adda-Decker, Angélique Amelot, Satoshi Nambu, and Mark Liberman. 2015. A crosslinguistic study of prosodic focus. Paper presented at 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, Australia, April 19–24; Piscataway: IEEE, pp. 4754–58.

Liu, Fang, and Yi Xu. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62: 70–87. [CrossRef] [PubMed]

Nespor, Marina, and Irene Vogel. 1983. Prosodic structure above the word. In *Prosody: Models and Measurements*. Springer: Berlin/Heidelberg, pp. 123–40.

Nooteboom, Sieb G., and Johanna G. Kruyt. 1987. Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *The Journal of the Acoustical Society of America* 82: 1512–24. [CrossRef] [PubMed]

Paris, Marie- Claude. 1998. Focus operators and types of predication in Mandarin. *Cahiers de Linguistique-Asie Orientale* 27: 139–59. [CrossRef]

Paul, Waltraud, and John Whitman. 2008. Shi . . . de focus clefts in Mandarin Chinese. *The Linguistic Review* 25: 413–51. [CrossRef]

Pierrehumbert, Janet B. 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, Massachusetts Institute of Technology, Boston, MA, USA, September 9.

Pierrehumbert, Janet, and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*. Edited by Philip Cohen, Jerry Morgan and Martha Pollack. Cambridge: MIT Press, pp. 271–311.

Reis, Marga. 2012. On the analysis of echo questions. *Tampa Papers in Linguistics* 3: 1–24.

Repp, Sophie, and Lena Rosin. 2015. The intonation of echo wh-questions. Paper Presented at 16th Annual Conference of the International Speech Communication Association, ISCA speech, Dresden, Germany, September 6–10; Red Hook: Curran Associates, pp. 938–42.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1: 75–116. [CrossRef]

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rump, Hans H., and Rene Collier. 1996. Focus conditions and the prominence of pitch-accented syllables. *Language and Speech* 39: 1–17. [CrossRef]

Selkirk, Elizabeth. 1978. On prosodic structure and its relation to syntactic structure. In *Nordic Prosody II*. Trondheim: TAPIR.

Selkirk, Elizabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. *Juncture* 7: 107–29.

Selkirk, Elizabeth O. 1981. On the nature of phonological representation. *Advances in Psychology* 7: 379–88.

Selkirk, Elizabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.

Selkirk, Elizabeth. 1995. Sentence prosody: Intonation, stress, and phrasing. In *The Handbook of Phonological Theory*. Edited by J. A. Goldsmith. Oxford: Blackwell Sci, pp. 550–69.

Seuren, Pieter A. 2009. *Language from Within: Vol. 1. Language in Cognition*. Oxford: Oxford University Press.

Shih, Chilin. 1988. Tone and intonation in Mandarin. In *Working Papers of the Cornell Phonetics Laboratory*. Ithaca: Cornell University, vol. 3, pp. 83–109.

Shattuck-Hufnagel, Stephanie, and Alice. E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25: 193–47. [CrossRef] [PubMed]

Sityaev, Dmitry, and Jill House. 2003. Phonetic and phonological correlates of broad, narrow and contrastive focus in English. Paper Presented at 15th International Congress of Phonetic Sciences ICPhS, Barcelona, Spain, August 3–9.

Yuan, Jiahong. 2004. Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling. Ph.D. thesis, Cornell University, Ithaca, NY, USA.

Wang, Bei, and Yi Xu. 2006. Prosodic encoding of topic and focus in Mandarin. Paper Presented at Speech Prosody: 3rd International Conference, Dresden, Germany, May 2–5.

Welby, Pauline. 2003. Effects of pitch accent position, type, and status on focus projection. *Language and Speech* 46: 53–81. [CrossRef] [PubMed]

Xu, Yi, Szu-wei Chen, and Bei Wang. 2012. Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review* 29: 131–47. [CrossRef]

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27: 55–105. [CrossRef]

Xu, Yi, Ching X. Xu, and Xuejing Sun. 2004. On the temporal domain of focus. Paper Presented at International Conference on Speech Prosody, Nara, Japan, March 23–26.

Xu, Yi. 2004. Transmitting tone and intonation simultaneously-the parallel encoding and target approximation (PENTA) model. Paper Presented at International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China, March 28–31.

Xu, Yi, and Ching X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159–97. [CrossRef]

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220–51. [CrossRef]

Xu, Yi. 2013. *ProsodyPro—A Tool for Large-Scale Systematic Prosody Analysis*. Aix-en-Provence: Laboratoire Parole et Language.

Zubizarreta, Maria L. 2014. Nuclear stress and information structure. In *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.

Zimmermann, Malte, and Edgar Onea. 2011. Focus marking and focus interpretation. *Lingua* 121: 1651–70. [CrossRef]