# Exploring the Complexity of the L2 Intonation System: An Acoustic and Eye-Tracking Study

*Di Liu[1]\* and Marnie Reed[2]*

[1] Teaching and Learning Department, College of Education and Human Development, Temple University, Philadelphia, PA, United States, [2] Wheelock College of Education and Human Development, Boston University, Boston, MA, United States

Phonological research has demonstrated that English intonation, variably referred to as prosody, is a multidimensional and multilayered system situated at the interface of information structure, morphosyntactic structure, phonological phenomena, and pragmatic functions. The structural and functional complexity of the intonational system, however, is largely under-addressed in L2 pronunciation teaching, leading to a lack of spontaneous use of intonation despite successful imitation in classrooms. Focusing on contrastive and implicational sentence stress, this study explored the complexity of the English intonation system by investigating how L1 English and Mandarin-English L2 speakers use multiple acoustic features (i.e., pitch range, pitch level, duration, and intensity) in signaling contrastive and implicational information and how one acoustic feature (maximum pitch level) is affected by information structure (contrast), morphosyntactic structure (phrasal boundary), and a phonological phenomenon (declination) in L1 English and Mandarin-English L2 speakers' speech. Using eye-tracking technology, we also investigated (1) L1 English and Mandarin-English L2 speakers' real-time processing of lexical items that carry information structure (i.e., contrast) and typically receive stress in L1 speakers' speech; (2) the influence of visual enhancement (italics and bold) on L1 English and Mandarin-English L2 speakers' processing of contrastive information; and (3) L1 English and Mandarin-English L2 speakers' processing of pictures with contrastive information. Statistical analysis using linear mixed-effects models showed that L1 English speakers and Mandarin-English L2 speakers differed in their use of acoustic cues in signaling contrastive and implicational information. They also differed in the use of maximum pitch level in signaling sentence stress influenced by contrast, phrasal boundary, and declination. We did not find differences in L1 English and Mandarin-English L2 speakers' processing of contrastive and implicational information at the sentence level, but the two groups of participants differ in their processing of contrastive information in passages and pictures. These results suggest that processing limitations may be the reason why L2 speakers did not use English intonation spontaneously. The findings of this study also suggest that Complexity Theory (CT), which emphasizes the complex and dynamic nature of intonation, is a theoretical framework that has the potential of bridging the gap between L2 phonology and L2 pronunciation teaching.

Keywords: intonation, sentence stress, complexity theory, pronunciation teaching, applied linguistics, linear mixed effect model, L2 Phonology, eye-tracking

# INTRODUCTION

The past 30 years have witnessed evolutionary and far-reaching advances in the field of L2 pronunciation, including establishment of the intelligibility principle (Munro and Derwing, 1995; Murphy, 2014), development of a holistic approach acknowledging the importance of both segmental and suprasegmental features (Anderson-Hsieh et al., 1992; Derwing et al., 1998; Levis and Levis, 2018), and development of technologies such as speech visualization (Levis and Pickering, 2004), automated speech recognition (ASR) (Cucchiarini and Strik, 2018) and speech synthesis (Ding et al., 2019).

Despite significant advances, challenges remain. Levis (1999), for example, stated that "[p]resent intonational research is almost completely divorced from modern language teaching…" (p. 37). The lack of a guiding theory causes many issues. For example, teachers may only focus on the aspects that they are conscious of using, resulting in an overemphasis on the attitudinal and emotional aspects in the teaching of suprasegmental features (Levis, 1999). Further, while teachers may assume that successful imitation and reproduction of target intonation patterns in classrooms leads to spontaneous production, students "may walk out of the class without having accepted the system at all. Or they may think intonation is simply decorative" (Gilbert, 2014). Gilbert's observation echoes what Allen (1971) had noted half a century ago: "there is little carry-over into the students' own conversations outside the classroom and the listen and repeat approach has never yielded satisfactory long-term results" (p. 79).

One main issue that sets research and teaching apart is the lack of applicability of research theories and models. In the past few decades, the field of L2 phonology has seen a number of highly influential theories, approaches and models including Autosegmental-Metrical phonology (AM) (Pierrehumbert, 1980), the Systemic Functional Approach (Halliday, 2015), and Discourse Intonation (Brazil, 1980) as well as the PENTA model (Xu, 2004), the Kiel intonation model (KIM) (Kohler, 1995), and the Fujisaki model (Mixdorff, 2000). While these theories and models have gained wide popularity in research and software development, direct transfer of these theories and models into classroom teaching has faced tremendous difficulties. For example, in his attempt to apply Discourse Intonation in language teaching, Chapman (2007) found that both students and teachers encountered difficulties such as identifying rising and falling tone as well as locating tone-unit boundaries and prominence. The core of this issue is that L2 phonology and L2 pronunciation teaching, albeit sharing the same underlying subject of investigation, focus on different issues and have different goals. L2 phonology analyzes speech samples in an effort to develop a theory or a model that explains underlying schema of speech production. L2 pronunciation teaching, on the other hand, focuses on the development of learners' abilities in navigating the system of intonation in spontaneous speech.

Larsen-Freeman (2017) pointed out that the field of second language acquisition is dominated by an approach which "seeks to understand phenomena by taking them apart" (Larsen-Freeman, 2017, p. 22). However, it could be the case that

"it is from the components and their relationships that the system we are trying to understand emerges. If we isolate components artificially, we lose the essence of the phenomena we are attempting to describe" (Larsen-Freeman, 2017, p. 29). To promote L2 speakers' spontaneous use of intonation, a systematic view of intonation is needed. Complexity Theory (CT), which views the relationship among phonological features and phenomena as an interrelated and dynamic system, thus is a more appropriate theoretical framework for L2 pronunciation teaching.

# LITERATURE REVIEW

## Intonation

Scholars have long acknowledged that English intonation relates to multiple acoustic features and perceptual phenomena. Palmer (1922), for example, stated that "all phenomena connected with this musical pitch or tone [such as word-prominence, word-group prominence, intensity, command, doubt, concession, reassurance, etc.] are designated by the term Intonation" (p. 7). Halliday (2015) argued that intonation is a system with three phonological and systemic variables: tonality, tonicity, and tone. Pierrehumbert (1980) views the intonation system as having three components—a grammar of phrasal tunes, a metrical representation of the text, and rules that line up tune with the text. Over the past century, numerous scholars defined intonation from different perspectives and with different focuses. Some scholars focused on the pragmatic meaning and phrasal structure (Gussenhoven, 2004; Levis and Wichmann, 2015), others emphasized pitch patterns or tones (Kingdon, 1958; O'Connnor and Arnold, 1973), still others highlighted the emotional and attitudinal aspects (Bolinger, 1989) (see **Table 1**).

Despite differences in approaches and focuses, scholars generally agree that the system of intonation includes multiple suprasegmental features and is closely related to information structure, morphosyntactic structure, and pragmatic functions (Gilbert, 2014; Levis and Wichmann, 2015). Gilbert (2014), for example, stated that "[i]n English, prosodic cues serve as navigation guides to help the listener follow the intentions of the speaker. These signals communicate emphasis and make clear the relationship between ideas (new and old information) so that listeners can readily identify these relationships and understand the speaker's meaning" (p. 123). Wennerstrom (1998) proposed that there is an intonation system in English that functions at the discourse level to signal relationships in information structure and to mark interdependencies among constituents; she proposes a model in which intonation functions as a grammar of cohesion. These studies pointed out the importance of intonation and the necessity of teaching English intonation to L2 English speakers.

## Sentence Stress

One essential component of English intonation is sentence stress (Kingdon, 1958; Pierrehumbert, 1980; Pickering, 2018), which is also commonly referred to as *prominence* (Celce-Murcia et al., 2010), *pitch accent* (Pierrehumbert, 1980; Pierrehumbert and Hirschberg, 1990), *focus* (Grant and Yu, 2017), *nucleus* (Palmer, 1922; Cruttenden, 1990; Cruz-Ferreira, 1998; Wells, 2006), and

**TABLE 1 |** Scholars' definitions of intonation and the suprasegmental features included.

| Scholars | Definition | Suprasegmental features | Related variables |
|---|---|---|---|
| Ladd (2008) | "The use of *suprasegmental* phonetic features to convey 'postlexical' or *sentence-level* pragmatic meanings in a *linguistically structured* way" (p. 4). | Suprasegmental phonetic features | Pragmatic meanings, linguistic structure |
| Pickering (2018) | "The term *intonation* is narrowly defined in English as the use of pitch structure over the length of a given utterance" (p. 2). "Intonation is the grammatical system that includes our use of pitch, pause, and prominence (or sentence stress)…" (p. 3). | Pitch, pause, prominence | |
| O'Connnor and Arnold (1973) | "When we talk about English intonation we mean the pitch patterns of spoken English, the speech tunes or melodies, the musical features of English" (p. 1). | Pitch, tunes/melody, musical features | |
| Kingdon (1958) | "The active elements of intonation are the Tones, which always occur in association with stresses" (p. 3). | Tones, stress | |
| Levis and Wichmann (2015) | "The use of pitch variations in the voice to communicate phrasing and discourse meaning in varied linguistic environments" (p. 139). | Pitch | Phrasing and discourse meaning |
| Gussenhoven (2004) | "Intonation is treated as the use of phonological tone for non-lexical purposes, or—to put it positively—for the expression of phrasal structure and discourse meaning" (p. 12). | Phonological tone | Phrasal structure and discourse meaning |
| Bolinger (1989) | "Intonation manages to do what it does by continuing to be what it is, primarily a symptom of how we feel about what we say, or how we feel when we say" (p. 1). | | Emotions |
| Palmer (1922) | "All phenomena connected with this musical pitch or tone [such as word-prominence, word-group prominence, intensity, command, doubt, concession, reassurance, etc.] are designated by the term Intonation" (p. 7). | Pitch/tone, prominence, intensity | Attitudes |

*primary stress* (Hahn, 2004; Celce-Murcia et al., 2010). Sentence stress denotes the relative emphasis or prominence a word receives primarily by the manipulation of fundamental frequency (F0), duration, and intensity of the stressed syllable as well as the modification of vowel quality.

Sentence stress plays a central role in English intonation because of its close connection with discourse meaning and information structure. It is frequently used to signal given vs. new (Pierrehumbert and Hirschberg, 1990) and contrast (Liu, 2020). It is also commonly used to make corrections (Ip and Cutler, 2016) or help listeners to anticipate what the speaker is going to say (Levis and Levis, 2018).

Sentence stress affects intelligibility, which is defined as "the extent to which a listener actually understands an utterance" (Derwing and Munro, 2005, p. 385). For example, investigating L1 English speakers' processing, comprehension, and evaluation of speech with correctly placed, misplaced, and missing sentence stress, Hahn (2004) found that the same speaker is more intelligible when sentence stress was used correctly. One of the multiple functions of intonation that impacts intelligibility is sentence stress used to show contrast. Levis and Levis (2018) argued that contrastive stress is a "high-value pronunciation feature" that should be given more attention in L2 pronunciation teaching.

## L2 Sentence Stress

Prior studies investigating Mandarin-English L2 speakers' intonation showed that even advanced level speakers face difficulties in using English sentence stress effectively (Chun,

1982; Wennerstrom, 1998; Pickering, 2001, 2004). This issue relates to both sentence stress realization and placement. Pickering (2001), for example, found that the pitch structure of Chinese international teaching assistants' speech is relatively flat and monotone compared to native speaker teaching assistants. Chun (1982) found that "Chinese speakers sometimes failed to place sentence stress on the appropriate word or syllable" (p. 386), pointing out the issue of stress placement.

It was assumed by some scholars that Mandarin-English L2 speakers' ineffective use of English intonation is due to the lexical tone system in Mandarin (Clennell, 1997). The claim is that Mandarin uses pitch variation to signal lexical tone, prohibiting it from using the same acoustic feature to indicate sentence stress. However, recent studies investigating the Mandarin sentence stress system suggested that this may be a false assumption (Xu, 1999; Chen and Gussenhoven, 2008; Kabagema-Bilan et al., 2011). Ouyang and Kaiser (2015), for instance, found that all three suprasegmental features that English uses to indicate sentence stress (fundamental frequency (F0), duration, and intensity) are all used to encode sentence stress in Mandarin. They further concluded that Mandarin uses sentence stress to indicate discourse-level information and make contrasts. Analyzing five types of focus in Mandarin and English, Ip and Cutler (2016) found that Mandarin speakers showed greater increase in pitch range and pitch level for new-information focus.

Despite similarities between English and Mandarin sentence stress, there is evidence of lack of transfer of suprasegmental features from L1 Mandarin to L2 English. For example, comparing L1 English speakers' use of sentence stress to Mandarin-English L2 speakers' sentence stress in both English

and Mandarin, Liu (2020) found that Mandarin-English L2 speakers did not use pitch to indicate English sentence stress even though they resemble L1 English speakers in the signaling of sentence stress when speaking in L1 Mandarin.

It is worth noticing that the lack of transfer of similar phonological features from L1 to L2 is not a language specific issue. For example, investigating the use of prosodic cues in German L2 English speakers' English and German speech, O'Brien et al. (2014) found that German-English L2 speakers do not transfer all prosodic uses from L1 to L2. The findings suggested that even speakers of Germanic languages that share similar morphosyntactic structure and uses of phonological cues with English do not transfer intonational cues directly from L1 to L2.

One factor that may account for the lack of transfer of phonological cues from L1 to L2 is the complex and dynamic nature of language. As Ortega-Llebaria and Colantoni (2014) stated, "…acquiring intonation in a L2 not only is an issue of learning to perceive and produce the target melody but, crucially, involves a new mapping between form and meaning that is affected by L1 transfer" (p. 351). The fact that intonation involves multiple acoustic cues and needs to be used with consideration of the information structure, morphosyntactic structure, and phonological phenomena may pose a significant challenge to L2 speakers. Complexity Theory (CT), which views the system of intonation as complex and dynamic, thus will be informative in L2 phonology research and L2 pronunciation teaching.

## Complexity Theory

Language is a complex dynamic system that "emerges bottom-up from interactions of multiple agents in speech communities" (Larsen-Freeman, 2017, p. 49). Language is also a social endeavor constantly influenced and shaped by the interaction and accommodation among different individuals, speech communities, and communities of practice (CoP) (Gumperz, 1971; Labov, 1972; Giles et al., 1987; Lippi-Green, 1989; Holmes and Meyerhoff, 1999). In the system of intonation, two types of complexity have been identified: structural complexity and functional complexity.

Structural complexity, also referred to as inherent or absolute complexity (Housen et al., 2019) captures the intrinsic complexity of a system. Encompassing multiple interrelated and interacting variables, the system of English intonation involves structural complexity. For example, when a speaker stresses a constituent, multiple segmental and suprasegmental features (i.e., vowel quality, pitch, duration, and intensity) are manipulated. The changes in these features affect not only the use of prosodic features at the syllable and word level, but also the intonational contour of the entire intonational phrase.

Another level of complexity is derived from the dynamic relationship between intonation and other variables such as information structure, morphosyntactic structure, and phonological phenomena. We list six examples of this functional complexity. Intonation dynamically encodes information structure (e.g., new vs. old, contrast, etc.) (Pierrehumbert and Hirschberg, 1990; Hahn, 2004; Levis and Wichmann, 2015), indicates grammatical structure (Pickering, 2018),

signals discourse level meanings and implications (Levis and Wichmann, 2015; Pickering, 2018), regulates speaker-listener interactions (Wennerstrom, 2001; Hellermann, 2003), directs listeners' attention (Chun, 1988; Gilbert, 2014), and expresses emotions and attitudes (Horley et al., 2010; Pell and Kotz, 2011). Functional complexity poses challenges to L2 speakers because not only can intonation be optionally used for these functions, it is also governed and dynamically shaped by all these aspects. In this sense, using L2 intonation is not simply manipulating a collection of acoustic cues, but navigating the entire linguistic repertoire while representing and balancing the influence of numerous linguistic and non-linguistic variables in real-time.

## The Present Study

The present study explores the structural and functional complexity of the intonation system by comparing L1 English and Mandarin-English L2 speakers' use of acoustic cues in speech production and visual processing of contrastive and implicational information that typically receives stress in L1 English speakers' speech. The present study both serves as an example for research investigating intonation from a CT perspective and offers insights into Mandarin-English L2 speakers' use of English intonation and the dynamic mapping between L1 and L2.

In their discussion of a developing cognitive system, DeBot et al. (2007) asserted, "the system is in constant complex interaction with its environment and internal sources. Its multiple interacting components produce one or many self-organized equilibrium points, whose form and stability depend on the system's constraints" (p. 14). As applied to L2 phonological phenomena, sentence stress, situated at the nexus of information structure, morphosyntactic structure, and other linguistic and non-linguistic variables, is appropriate for investigation. Treating every speech sample as an equilibrium point, acoustic analysis reveals information about how phonological features are used. From a Complexity Theory (CT) perspective, the present study investigated L1 English and Mandarin-English L2 speakers' use of various acoustic features in signaling English sentence stress and the influence of information structure, morpho-syntactical structure, and phonological phenomena.

Eye-tracking technology has been used widely in the field of second language research as a means to investigate L1 and L2 speakers' parsing of temporarily ambiguous sentences (Papadopoulou and Clahsen, 2003; Dussias and Sagarra, 2007); processing of lexical and morphological cues (Kambe et al., 2001; Lew-Williams and Fernald, 2010), and processing of spoken language (Tanenhaus, 2007; Ito and Speer, 2008). Researchers found "systematic relations between fixation duration and the characteristic of the fixated words" (Dussias, 2010, p. 150), providing information about the incremental processing of sentence comprehension. Using eye-tracking technology, this study explored real-time processing of lexical items (1) that are contrastive or non-contrastive, (2) at different positions within written sentences, (3) that do or do not carry implications, and (4) that appear with or without visual enhancement. The guiding research questions are:

- How do L1 and Mandarin-English L2 speakers use intonational features (i.e., pitch range, pitch level, duration, and intensity) to signal contrast and implication in speech production?
- How do L1 and Mandarin-English L2 speakers use a phonological feature (maximum pitch level) influenced by information structure (contrast), morphosyntactic structure (sentence boundaries), and phonological phenomena (declination)?
- How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?
- How do L1 and Mandarin-English L2 speakers orally produce and visually process contrastive information written with and without visual enhancement (italics and bold)?

## METHODS

### Participants

Ten subjects participated in the study. Five were L1 English speakers and five were Mandarin-English L2 speakers enrolled in degree programs in a university in the US. Participants' biographical information is summarized in **Tables 2**, **3**.

### Procedure

The study was conducted in a Language Acquisition and Visual Attention lab at a large research university in the northeast. Participants were seated in front of a PC connected to an *Eyelink 1000 Plus* eye-tracker. All participants went through a calibration process. Then, the participants were presented a scenario with pictures contextualizing the first experiment.

In the first experiment, participants saw 18 sentences on the computer screen in random order. Participants saw one sentence at a time and each sentence was presented in a single line. Participants were asked to first read the sentence silently. Then, when they were ready, they read the sentence aloud. In the second experiment, the participants saw three sets of sentences in random order. Each set of sentences contains two sentences of the same wording, but different meanings or implications presented in parentheses. Participants were asked to first read silently and then read aloud the sentences to express the meanings or implication included in the parentheses. In the third experiment, participants read silently and then read aloud two short passages with contrastive information. In the fourth experiment, participants saw an eight-frame picture cartoon that describes a single story. They were asked to look at the pictures silently and prepare to tell the story. Then, they were asked to tell the story in their own words. The experiment materials and data analysis are further discussed within each experiment in the following sections.

While participants silently read the sentences and passages (in experiments I—III) and viewed the pictures (in experiment IV), the eye-tracker documented the fixation count, fixation percentage, dwell time, dwell percentage, run count, and regression of the Areas of Interest (AOI), which were the contrastive/implicational information or equivalent places in the distractors. The eye-tracker was recalibrated before each experiment. When the participants read aloud the information, they were audio-recorded using both an audio recording application software (*Audacity*) and a handheld recorder (*Zoom H4N*).

## Data Elicitation and Analysis

Participants' fixation count (total number of fixations within the interest area), fixation percentage (percentage of total fixations in a trial falling within the current interest area), dwell time (total time (in milliseconds) spent on the current interest area), dwell percentage (percentage of trial dwell time spent on the current interest area), run count (number of times the interest area was entered and left), and regression on the focused areas were documented using the eye-tracker (definitions elicited from *EyeLink Data Viewer User's Manual*, Version 1.11.900, p. 39–40). Participants' speech data were analyzed using *Praat* version 6.0.37 (Boersma and Weenink, 2018). Maximum pitch level, minimum pitch level, and pitch range in Hertz and in semitones relative to 1 Hz, as well as the duration of words in seconds and the intensity of words in decibels (dBs) were elicited using a *Praat* script. Participants' pitch in Hertz and pitch in semitones were normalized based on each participant's average pitch level. We used *R* (R Core Team, 2017) and *lme4* (Bates et al., 2015) to analyze the speech and processing data. Linear mixed-effects models were constructed for the statistical analyses; *P*-values were obtained by likelihood ratio tests.

## EXPERIMENT 1

In Experiment 1, we investigated L1 and Mandarin-English L2 speakers' production and processing of sentence stress using a contextualized sentence read aloud task. The scenario we used was a Christmas Tree decoration task adapted from Ito and Speer (2008). The participants were told that a friend of theirs, Martin, is decorating a Christmas tree using items from an ornament board. Martin is trying to select two of the ornaments at a time and the participants were tasked with telling him which ones to hang. In the contextualizing process, the participants were asked to tell Martin what to hang based on circled items on pictures of the ornament board. Then, after participants understood the scenario, they were asked to complete the task using 18 written sentence prompts.

The sentences belong to six sentence sets. Each set has three sentences: one sentence with contrastive information presented not at phrasal boundaries, one sentence with contrastive information presented at phrasal boundaries, and a distractor that does not include contrastive information (see **Appendix A** for a complete list of sentences).

- First, hang the blue drum, then hang the yellow drum ("blue" and "yellow" are contrastive and not at phrasal boundaries).
- First, hang the blue drum, then hang the blue ball ("drum" and "ball" are contrastive and at sentence phrasal boundaries).
- First, hang the blue drum, then hang the pink ball (distractor: no contrastive information in the sentence).

**TABLE 2 |** L1 English speakers' biographical information.

| Gender | Age | L1 | Other language(s) | Major | Degree |
|---|---|---|---|---|---|
| Female | 21 | English | Spanish (advanced); French (intermediate) | Neuroscience | Bachelor |
| Male | 24 | English | Amharic (bilingual) | Biomedical engineering | Bachelor |
| Female | 21 | English | ASL (intermediate) | Psychology | Bachelor |
| Female | 20 | English | Mandarin (intermediate) | International relations, linguistics | Bachelor |
| Female | 21 | English | Spanish (beginner) | Health science | Bachelor |

**TABLE 3 |** Mandarin-English L2 speakers' biographical information[*].

| Gender | Age | L1 | Other language(s)/years of learning | Major | Degree | Standardized Test Scores | Speaking | Listening | Percentage of English use daily |
|---|---|---|---|---|---|---|---|---|---|
| Male | 28 | Mandarin | English (10 years) | Computer Science | Master | TOEFL 100 | TOEFL 18–25 | TOEFL 22–30 | 20–40% |
| Male | 24 | Mandarin | English (10 years) | Advertising | Bachelor | TOEFL 109 | TOEFL 18–25 | TOEFL 22–30 | 20–40% |
| Male | 26 | Mandarin | English (15 years) | System Science | Bachelor | IELTS 6.5 | IELTS 5.5 | IELTS 5.5 | 0–20% |
| Female | 24 | Mandarin | English (17 years) | Advertising | Master | TOEFL 111 | 18–25 | 22–30 | 20–40% |
| Female | 20 | Mandarin | English (15 years), Japanese (intermediate) | Biology | Bachelor | TOEFL 100 | 18–25 | 22–30 | 60–80% |

*A total of 6 L1 English speakers and 11 Mandarin-English L2 speakers were recruited to participate in the study. 1 L1 English speaker and 6 Mandarin-English L2 speakers were excluded from the dataset because of technological/calibration failure primarily due to the reflection of participants' eyeglasses.*

**TABLE 4 |** Sample sentences with contrastive and non-contrastive information at different locations.

| | | | Position 1 | Position 2 | | | | Position 3 | Position 4 |
|---|---|---|---|---|---|---|---|---|---|
| 1. First, | hang | the | *Blue* Contrastive + not at boundary | *Drum,* Non-contrastive + boundary | then | hang | the | *Yellow* Contrastive + not at boundary | *Drum* Non-contrastive + boundary |
| 2. First, | hang | the | *Blue* Non-contrastive + not at boundary | *Drum,* Contrastive + boundary | then | hang | the | *Blue* Non-contrastive + not at boundary | *Ball* Contrastive + boundary |
| 3. First, | hand | the | *Blue* Non-contrastive + not at boundary | *Drum,* Non-contrastive + boundary | then | hang | the | *Pink* Non-contrastive + not at boundary | *Ball* Non-contrastive + boundary |

All sentences were scrambled and presented to the participants in random order. For each sentence, the participants were asked to read the sentence silently to themselves first, and then produce the sentence as if they were providing Martin directions on which items to hang on the Christmas tree.

Our first question was:

(1) How do L1 and Mandarin-English L2 speakers use intonational features (i.e., pitch range, pitch level, duration,

and intensity) to signal contrast and implication in speech production?

To answer this question, we analyzed speech data using linear mixed-effects models that predicted normalized pitch range, normalized maximum pitch level, duration, and intensity. We used different L1s, contrastive or non-contrastive, and the interaction between L1 and contrast as the fixed effects. Individual participants, sentences, and words were included in

the model as the random effects. The results showed that in L1 English speakers' speech, there were statistically significant differences between the contrastive and non-contrastive information regarding the use of pitch range (estimate = 1.136, SE = 0.107, df = 445.824, t = 10.609, ***$p$ < 0.0001), maximum pitch level (estimate = 0.77, SE = 0.111, df = 455.753, t = 6.932, ***$p$ < 0.0001), and duration (estimate = 0.112, SE = 0.012, df = 460.980, t = 9.284, ***$p$ < 0.0001). However, there was no significant difference in the intensity of contrastive and non-contrastive information (estimate = 0.243, SE = 0.382, df = 457.689, t = 0.635, $p$ = 0.526).

In Mandarin-English L2 English speakers' speech, however, there were no significant differences between the contrastive and non-contrastive information in terms of pitch range (estimate = −0.137, SE = 0.107, df = 443.794, t = −1.272, $p$ = 0.204), maximum pitch level (estimate = −0.0247, SE = 0.111, df = 453.505, t = −0.222, $p$ = 0.824), or duration (estimate = −0.008, SE = 0.012, df = 461.264, t = −0.679, $p$ = 0.498). However, there was a significant difference between the contrastive and non-contrastive information signaled by intensity in Mandarin-English L2 speakers' speech (estimate = 1.092, SE = 0.383, df = 458.084, t = 2.853, *$p$ = 0.005) (see **Figure 1**).

The results showed differences in the acoustic cues that L1 and Mandarin-English L2 speakers used to signal contrastive information. L1 speakers used pitch and duration to indicate contrastive stress. Mandarin-English L2 speakers, on the contrary, did not signal contrast using pitch or duration. Their

use of greater intensity may have been intended to signal contrastive information.

Our second research question was:

(2) How do L1 and Mandarin-English L2 speakers use a phonological feature (maximum pitch level) influenced by information structure (contrast), morphosyntactic structure (sentence boundaries), and phonological phenomena (declination)?

To answer question (2), we analyzed participants' use of maximum pitch level of contrastive and non-contrastive information at different positions within sentences (see **Table 4**). Specifically, we explored how L1 and Mandarin-English L2 speakers' pitch level is affected by (1) information structure (contrastive vs. non-contrastive), (2) morphosyntactic structure (at sentence boundary vs. not at sentence boundary), and (3) phonological phenomena (declination). We used the position of the words, contrastive or non-contrastive and the interaction between position and contrast as the fixed effects to predict normalized maximum pitch level. Individual participants, sentences, and words were entered into the model as random effects.

In L1 English speakers' speech, there was a significant difference between the maximum pitch level of the contrastive and non-contrastive information (estimate = 0.632, SE = 0.255, df = 222.926, t = 2.476, *$p$ = 0.014). We also found differences among positions. Compared to the words in position 1, words in position 3 had significantly lower pitch level (estimate = −0.882,
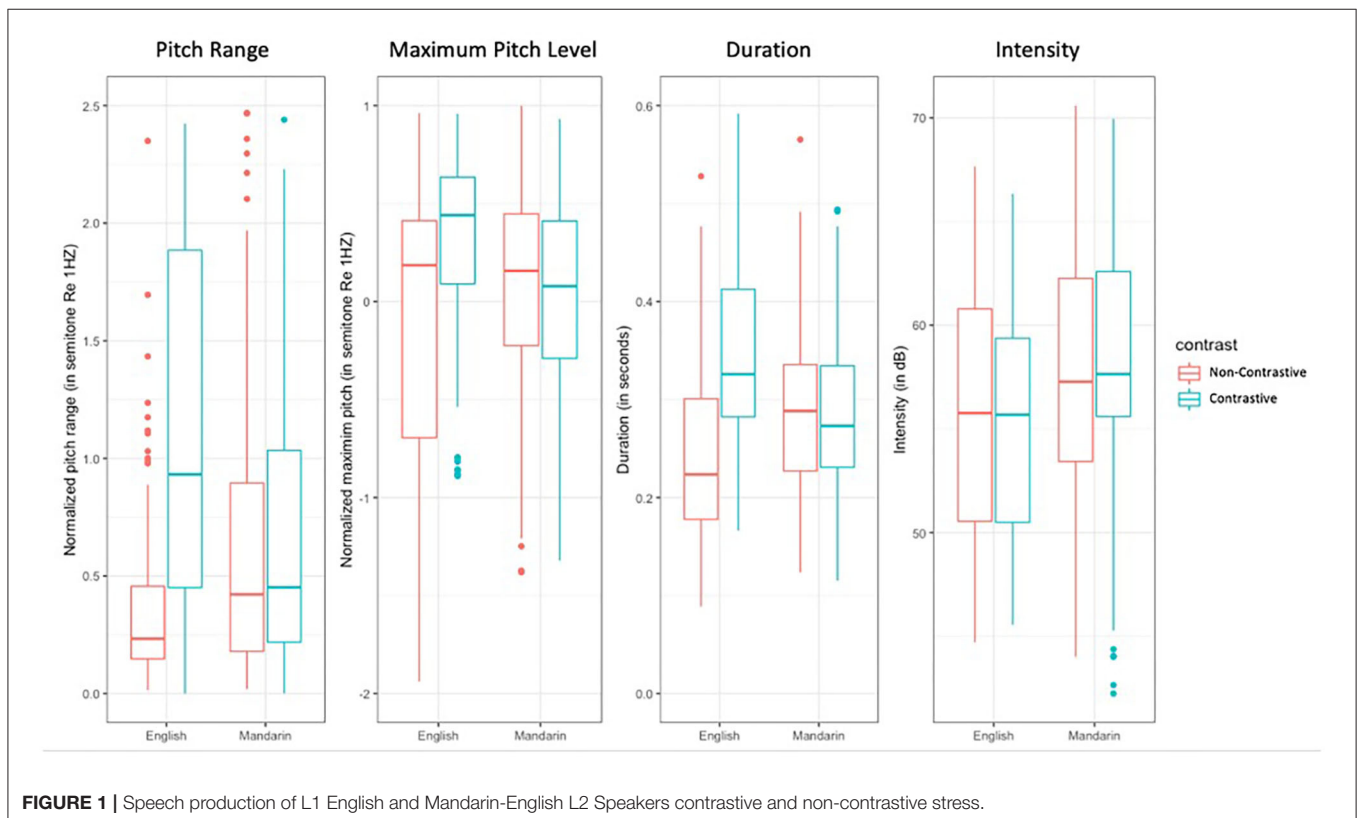


**FIGURE 1 |** Speech production of L1 English and Mandarin-English L2 Speakers contrastive and non-contrastive stress.

SE = 0.255, df = 222.926, t = 3.457, ***$p$ = 0.0006). There was also a significant difference between words in position 1 and words in position 4 (estimate = −1.148, SE = 0.262, df = 223.083, t = −4.379, ***$p$ < 0.0001). The position differences suggested the effect of declination. There were no significant differences between words in position 1 and words in position 2 (estimate = 0.256, SE = 0.255, df = −222.926, t = 1.005, $p$ = 0.316), suggesting an influence of a H- (high) boundary tone at phonological boundary (Pierrehumbert and Hirschberg, 1990) used to signal non-finality.

In Mandarin-English L2 English speakers' speech, however, there was no significant difference between contrastive and non-contrastive information (estimate = −0.067, SE = 0.141, df = 53.646, t = −0.478, $p$ = 0.635). There was a significant difference between words in position 1 and words in position 3 (estimate = −0.3646, SE = 0.128, df = 202.882, t = −2.842, **$p$ = 0.00493), suggesting potential influence of declination. However, there was no significant difference between words in position 1 and words in position 4 (estimate = −0.251, SE = 0.153, df = 41.98, t = −1.632, $p$ = 0.1101). There was also no significant difference between words in position 1 and words in position 2 (estimate = −0.0345, SE = 0.149, df = 38.088601, t = −0.232, $p$ = 0.81805) (see **Figure 2**).

These results showed that in L1 English speakers' speech, one individual intonational cue—maximum pitch level—is affected by multiple variables including information structure (i.e., contrast), morphosyntactic structure (i.e., phrasal boundary), and phonological phenomena (i.e., declination). Although Mandarin-English L2 speakers' maximum pitch level reflected potential influence of declination, L2 speakers did not show schema that reflect functional complexity at a level comparable to L1 speakers.

The third and fourth research questions were:

(3) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?
(4) How do L1 and Mandarin-English L2 speakers orally produce and visually process contrastive information written with and without visual enhancement (italics and bold)?

To answer questions (3) and (4), we analyzed the subset of sentences with visual enhancement (i.e., bold and italics) signaling the contrastive information (two sets with italics and two sets with bold). We hypothesized that the orthographical conventions used in English may have different implications or functions in the Chinese logographic writing system (Mair, 1996, p. 200). These differences may affect the processing of written information and the production of contrastive information.

Sample sentence set with italicized words

- First, hang the *yellow* tree, then hang the *white* tree ("yellow" and "white" are contrastive, signaled by italics, and not at sentence boundaries).
- First, hang the yellow *tree*, then hang the yellow *star* ("tree" and "star" are contrastive, signaled by italics, and at sentence boundaries).

- First, hang the yellow tree, then hang the white bell (no contrastive information, no visual enhancement).

Sample sentences set with words in bold

- First, hang the **green** egg, then hang the **brown** egg ("green," and "brown" are contrastive, signaled by bold, and not at sentence boundaries).
- First, hang the green **egg**, then hang the green **sock** ("egg," and "sock" are contrastive, signaled by bold, and at sentence boundaries).
- First, hang the green egg, then hang the orange tree (no contrastive information, no visual enhancement).

Participants' pitch range was analyzed using L1, visual enhancement (no visual enhancement, italics, and bold), and contrast (contrastive vs. non-contrastive) as the fixed effect and individual participants, sentences and words as the random effects. The results show that L1 English speakers used a significantly greater pitch range to signal contrastive information regardless of the use of visual enhancement (estimate = 1.04, SE = 0.2, df = 216.14, t = 5.09, ***$p$ < 0.0001). Mandarin-English L2 speakers, on the other hand, did not use pitch range to signal contrastive information even when the information was enhanced by bold or italics (estimate = −0.1, SE = 0.17, df = 223.56, t = −0.62, $p$ = 0.54) (see **Figure 3**).

We then investigated whether there were processing differences indicated by the fixation percentage and dwell percentage of the AOI by using L1, visual enhancement, and the interaction between these two factors as the fixed effects, and individual participants, sentences, and words as the random effects in our models. The results suggest that there were no statistically significant differences between L1 and Mandarin-English L2 speakers in fixation percentage (estimate = 2.31, SE = 2.26, df = 20.19, t = 1.024, $p$ = 0.32) or dwell percentage (estimate = 2.82, SE = 2.32, df = 25.75, t = 1.217, $p$ = 0.235). Also, whether the contrastive information was visually enhanced by italics or bold did not lead to significant processing differences.
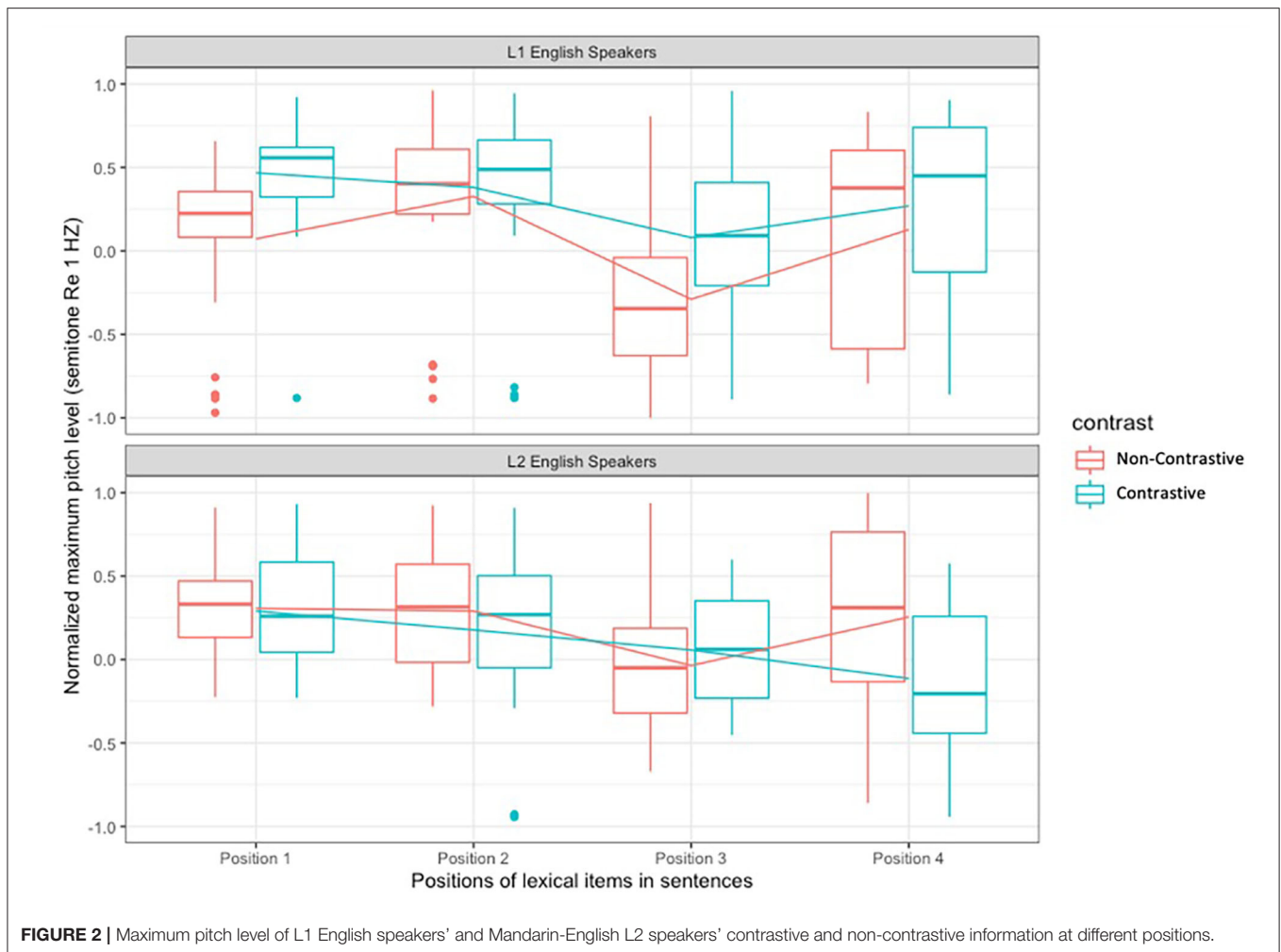
## EXPERIMENT 2

The second experiment is a read aloud task. Participants were given three sets of sentences (adapted from *POSE*-test). For each set of sentences, two different implications were given in parentheses and could be signaled by altering the stressed constituents in the same sentence (for a complete list of sentences used in Experiment 2, refer to **Appendix B**).

- My brother is a doctor (not my sister).
- My brother is a doctor (not a teacher).

The participants were asked to read the sentences silently first, and then read the sentence out loud to express the implications in the parentheses. When they were reading the sentences silently, their fixation and dwell time and percentage were measured using an eye-tracker. When they were reading aloud, participants were asked not to read the information in the parentheses, and the

**FIGURE 2 |** Maximum pitch level of L1 English speakers' and Mandarin-English L2 speakers' contrastive and non-contrastive information at different positions.

sentences were recorded. The research questions we asked in experiment 2 were:

(1) How do L1 and Mandarin-English L2 speakers use intonational features (i.e., pitch range, pitch level, duration, and intensity) to signal contrast and implication in speech production?
(3) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?

Linear mixed-effects models were constructed with L1s, different implications, and the interaction between L1 and implication as the fixed effects, and individual participants, sentences, and words as the random effects. The result shows that L1 speakers use statistically significantly greater pitch range (estimate = 1.2515, SE = 0.2687, df = 101.0587, t = 4.658, ****$p$ < 0.0001), higher maximum pitch level (estimate = 0.61312, SE = 0.20432, df = 100.838, t = 3.001, **$p$ = 0.003), and longer duration (estimate = 0.09075, SE = 0.02539, df = 102.385, t = 3.574, ***$p$ = 0.0005) to express the

information related to the implication. There was no significant difference between the intensity that L1 speakers used to encode implicational or non-implicational information. For Mandarin-English L2 speakers, there were no significant differences between implicational vs. non-implicational conditions for all prosodic features we analyzed.

In terms of the processing of implicational information, we established linear fixed effects models with language as the fixed effect to predict both the fixation percentage and the dwell percentage. The results showed that there were no significant differences between L1 and Mandarin-English L2 speakers' processing of the information related to the implication as indicated by their fixation percentage (estimate = 1.880, SE = 3.998, df = 7.895, t = 0.470, $p$ = 0.6508) and dwell percentage (estimate = 3.344, SE = 4.920, df = 7.918, $p$ = 0.680, $p$ = 0.516).

## EXPERIMENT 3

We used two passage read-aloud tasks to further investigate L1 and Mandarin-English L2 speakers' production and processing of contrastive information with and without visual enhancement
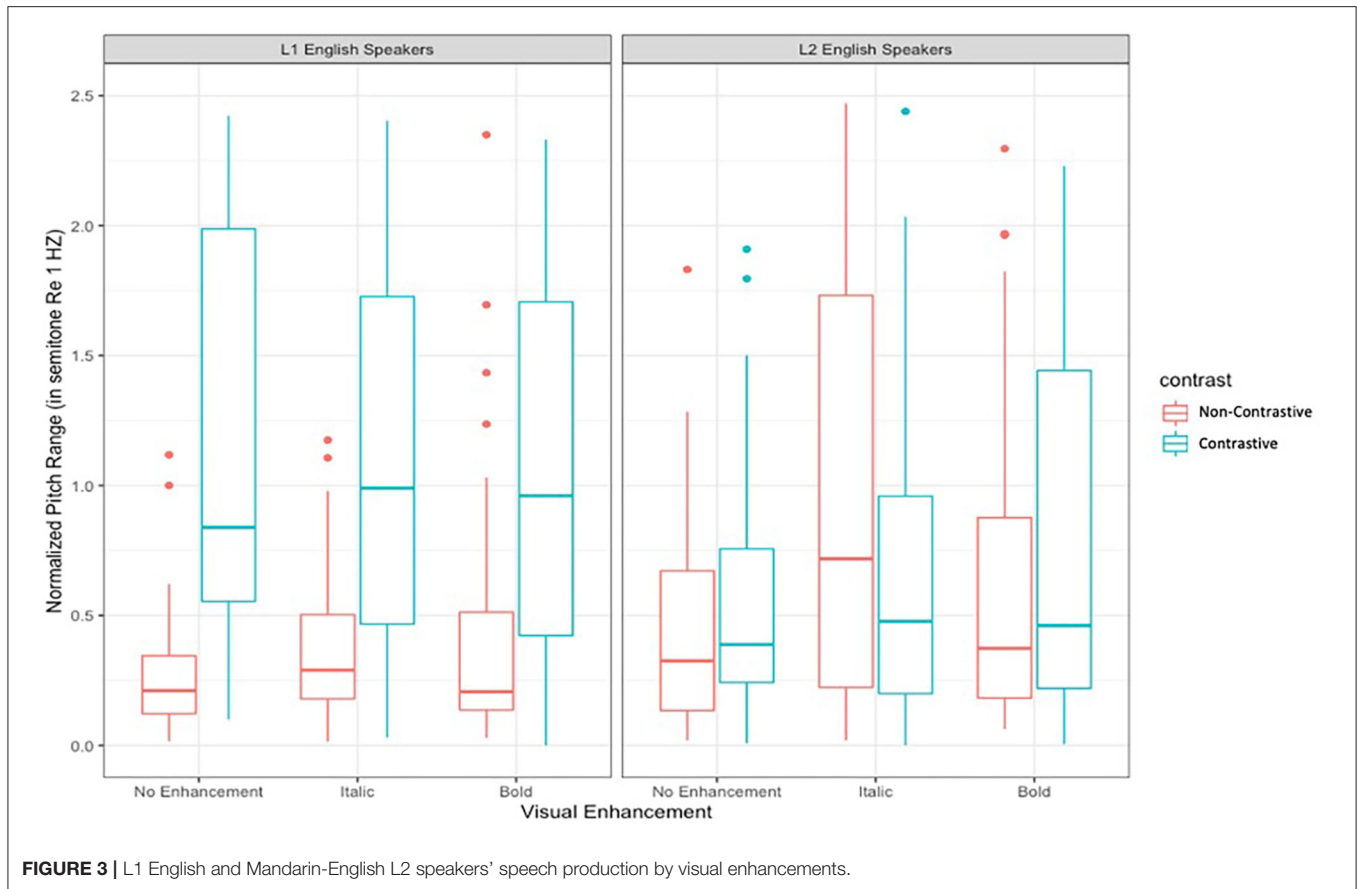
**FIGURE 3 |** L1 English and Mandarin-English L2 speakers' speech production by visual enhancements.

with less predictable text structures. The first paragraph was a passage adapted from a *New York Times* article. The contrastive information in this passage was signaled using italics. The lexical items analyzed were: "must," "may," "allows," and "requires."

> There are roughly 6,000 languages in the world. Are they mostly the same or are they different from each other? Fifty years ago, a famous linguist pointed out a crucial fact about differences among languages. He said, "Languages differ essentially in what they *must* convey and not in what they *may* convey." What this means is this: if different languages influence our minds in different ways, this is not because of what our language *allows* us to think but rather because of what it habitually *requires* us to think *about*.

The second passage we presented is adapted from Hahn (2004). This paragraph has contrastive information not signaled by any orthographic symbols. The lexical items analyzed were: "personal (1)," "group (1)," "group (2)," "personal (2)."

> I will start by defining the topic for today, which is individualism and collectivism. Individualism concerns the placing of personal goals ahead of group goals. And collectivism concerns placing group goals ahead of personal goals. So let's suppose you have a conflict at work about break time. Let's say your co-workers want longer breaks, but you want shorter breaks. If you're a collectivist, you'll give in to the group. But if you're an individualist, you'll go against the group.

Participants were directed to read each of these passages silently first, during which an eye-tracker was used to measure their processing of the contrastive lexical items where the areas of interest (AOI) were set. Then the participants were asked to read the two passages aloud in a natural way. The results answer question (3) and (4) at the passage level:

(4) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?

(5) How do L1 and Mandarin-English L2 speakers orally produce and visually process contrastive information written with and without visual enhancement (italics and bold)?

Experiment III differs from Experiments I and II in that all analyzed lexical items are contrastive. Thus, comparisons were made between L1 and Mandarin-English L2 speakers' uses of intonational cues in signaling the contrastive information instead of how L1 and Mandarin-English L2 speakers used intonational cues to signal contrastive information as opposed to non-contrastive information.

We used L1, visual enhancement and the interaction between these two factors as the fixed effects, individual participants, sentences, and words as the random effects to predict normalized pitch range, normalized maximum pitch level, duration, and intensity in the analyses of speech production. We found

that L1 English speakers used a greater pitch range to signal the contrastive information compared to Mandarin-English L2 speakers when the contrastive information was visually enhanced (estimate = −1.423, SE = 0.305, df = 20.835, t = −4.668, ***p = 0.0001) or not (estimate = 0.9547, SE = 0.3048, df = 20.8352, t = −3.132, *p = 0.005). L1 English speakers also used longer duration in indicating the contrastive information compared to Mandarin-English L2 speakers when italics were used (estimate = −0.131, SE = 0.036, df = 14.724, t = −3.658, **p = 0.0024). However, there was no significant difference in the two groups' use of duration when the contrastive information was not visually enhanced (estimate = −0.059, SE = 0.036, df = 14.724, t = −1.66, p = 0.118). There was no significant difference between L1 and Mandarin-English L2 speakers in the maximum pitch level or intensity with and without visual enhancement.

In terms of the processing of contrastive information in passages, when contrastive information was not visually enhanced (in Passage 2), Mandarin-English L2 speakers had a significantly lower fixation percentage (estimate = −0.882, SE = 0.399, df = 71.738, t = −2.213, *p = 0.03) and dwell percentage (estimate = −0.857, SE = 0.405, df = 72.147, t = −2.114, *p = 0.0379) compared to L1 English speakers. When italics were used to signal contrastive information (in Passage 1), there was no significant difference in the two groups' fixation percentage (estimate = −0.744, SE = 0.399, df = 71.738, t = −1.867, p = 0.066) and a slightly significant difference in the two groups' dwell percentage (estimate = −0.818, SE = 0.405, df = 72.147, t = −2.019, *p = 0.0472).

The production data support findings in Experiments I & II by showing differences in the use of pitch range and maximum pitch level by L1 English and Mandarin-English L2 speakers in speech production of contrastive information. Processing data suggested that L1 English and Mandarin-English L2 speakers differ in the processing of contrastive information at the passage level. Specifically, Mandarin-English L2 speakers did not fix on the contrastive information at a percentage comparable to the L1 English speakers when processing passages. when italics were used to signal contrastive information in passages, there was no difference between the two groups. These findings suggested that Mandarin-English L2 speakers may face challenges processing information with rich contextual information and less predictable text structure. Further, the use of visual enhancement may facilitate visual processing of contrastive information at the passage level.

## EXPERIMENT 4

In experiment 4, we investigated L1 and Mandarin-English L2 speakers' processing and production of contrastive information in a picture narrative task. In this task, participants were given a set of eight pictures that describe a single story. Participants were asked to spend a couple of minutes looking at the pictures and then describe the story in English. This picture narrative task is adapted from Derwing et al. (2004) to elicit extemporaneous speech from speakers (see **Appendix C**). The research question that Experiment 4 answered was:

(3) How do L1 and Mandarin-English L2 speakers orally produce and visually process lexical items and pictorial information that are contrastive or implicational and that typically receive stress in L1 speakers' speech?

Two researchers listened to the recordings and selected the words being stressed with the assistance of speech visualization software *Praat*. The researchers found that L1 speakers used intonational cues to signal contrastive information. The following is a transcription of the story told by an L1 speaker with the stressed information in capitalized letters.

> "Two people are walking around in a big city. One man is leaving the building and one woman is ENTERING the building. But they bumped into each other at the entrance and they both dropped…their bags. Um… so they picked it up and walked away. And when the man gets home, he realizes that he has WOMAN's bag. And when the woman gets to work, she realizes that she has the man's bag." (L1 participant #1)

We found that when telling the story, some Mandarin-English L2 speakers did not specify as much contrastive information in their speech as the L1 speakers did. The transcript below illustrates this tendency.

> "In an apartment, um… um… a woman and a man run into each, each other. They took the package. They, they took the package. They make, make a mistake. So, when they return home, they open the package and found they make a mistake." (Mandarin-English L2 participant #3)

We also found that, in some cases, Mandarin-English L2 speakers included contrastive information in the story but did not use intonational cues to signal the contrastive information.

> "It's a big city, a woman and a man are walking on the street and carrying the same suitcases. And they crushed each other at the corner of the street, and the suitcases dropped on the ground. They stand up and pick up their suitcases and walked away. But when the man got home, he found a dress in the suitcase, so he actually got the woman's suitcase. And the woman is in… When the woman got home, she found a tie in her suitcase." (Mandarin-English L2 participant #4)

We used heatmaps to show the areas that the participants paid attention to, and noted differences. We found that L1 speakers paid attention to the details of the pictures in a more holistic way, especially when contrastive information was presented. For instance, in the top two frames in **Figure 4**, L1 speakers spent longer gazing at both the man and the woman as well as the items they found in the suitcases. The Mandarin-English L2 speakers, however, paid less attention to the details and not at the contrastive information. For example, in the bottom two frames in **Figure 4**, one Mandarin-English L2 speaker focused on the man and the item held by the woman and another Mandarin-English L2 speaker focused on the woman and the item held by the man (see **Figure 4**).

**FIGURE 4 |** L1 English and Mandarin-English L2 speakers' processing of pictorial information.
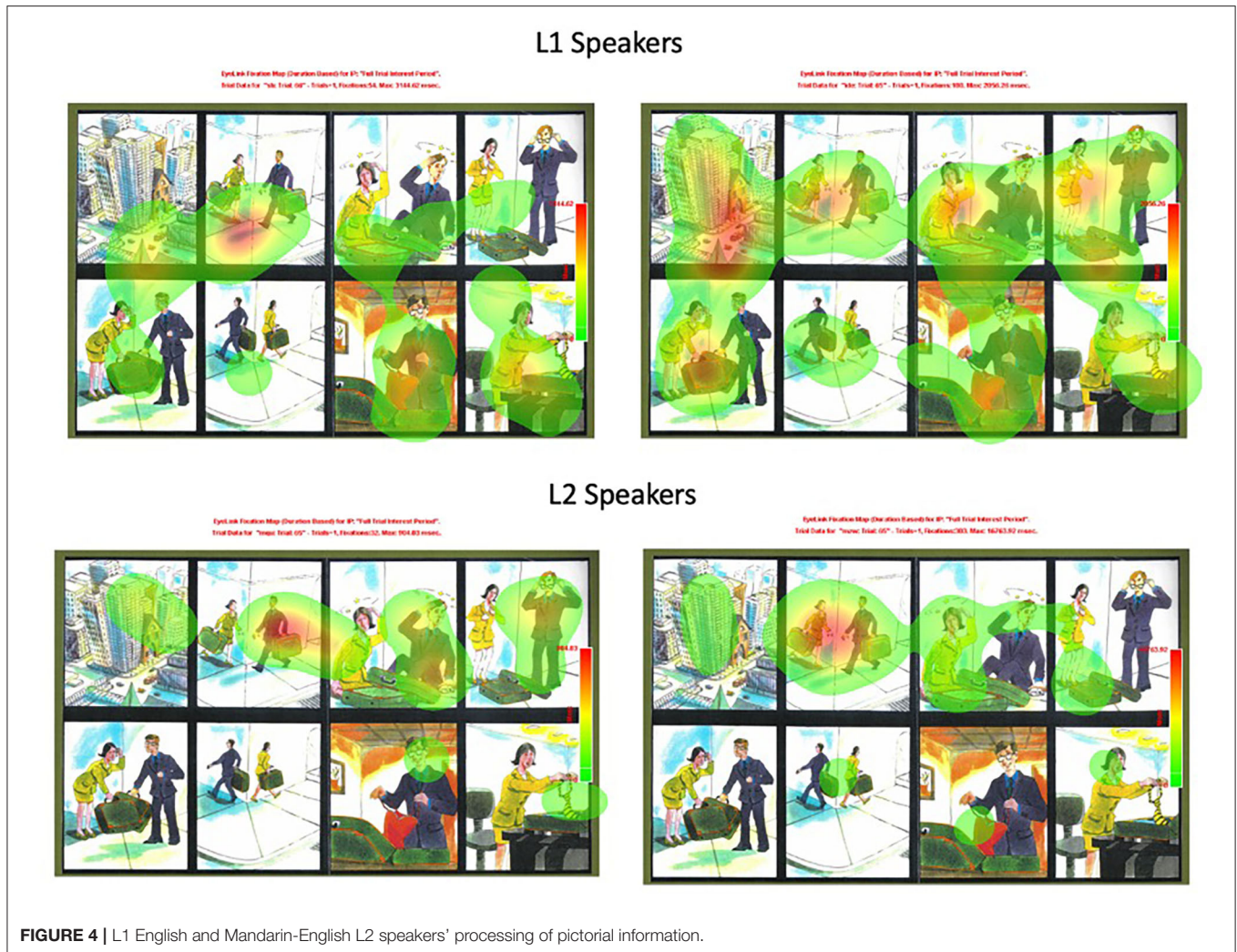
## DISCUSSION

Analyzing L1 English and Mandarin-English L2 speakers' speech production and processing of written and pictorial information from a Complexity Theory (CT) perspective, the present study found that sentence stress is a multidimensional and multifaceted feature dynamically connected with a series of linguistic and non-linguistic variables. To promote spontaneous use of intonational features like sentence stress, a systematic view that takes into consideration both structure and functional complexity is needed.

In Experiment 1 and Experiment 2, L1 English speakers used a collection of intonational cues to signal contrastive and implicational information in sentences including pitch range, maximum pitch level, and duration. Mandarin-English L2 speakers did not show differences in the use of any of these acoustic features to encode information structure. However, Mandarin-English L2 speakers used greater intensity when producing contrastive information in Experiment 1. The findings in Experiment 3 also showed that, compared to

Mandarin-English L2 speakers, L1 English speakers used greater pitch range and higher maximum pitch level when signaling contrastive information in passages. Altogether, these findings suggested that Mandarin-English L2 speakers have difficulties in the integrated use of multiple acoustic cues in signaling contrastive or implicational information. Mandarin-English L2 speakers may, as in Experiment 1, manipulate one acoustic cue (intensity) in an attempt to signal stress. However, intensity was an intonational feature that L1 speakers did not rely on when signaling sentence stress in the same context. These findings suggested the need to address the structural complexity of sentence stress.

To address the structural complexity of the system of intonation, the relationship between different acoustic features needs to be clarified. Intonation teaching will benefit from helping learners to understand intonation as a complex system encompassing multiple interacting features. In addition to chapters focusing on individual features, textbook chapters that adopt a systematic view of intonation and that summarize the relationship among different suprasegmental features will help

teachers and learners to develop a systematic view of intonation. Teacher training and preparation programs would do well to also focus more on the systematic use of intonation features and the complexity within the system of intonation.

The results of Experiment 1 showed that L1 English speakers' use of a single intonational cue (i.e., maximum pitch level) is affected by multiple interrelated variables and phenomena including information structure (contrast), morphosyntactic structure (phrasal boundary) and a phonological phenomenon (declination). Mandarin-English L2 speakers' speech, while still showing a general declination trend, did not reflect the influence of information structure or morphosyntactic structure. The results of Experiment 2 further demonstrated that L1 English speakers use intonational features to highlight the lexical item associated with the meanings and implications of the sentences whereas Mandarin-English L2 speakers did not signal implicational information with acoustic cues. These results suggested that learners either were unaware of the connection among intonation, meaning or implication, and information structure, or were incapable of navigating the system of intonation while taking into consideration all influential variables. Processing limitations may be a crucial factor. As O'Brien and Féry (2015) commented about L2 speakers' use of information structure, "[a]n appeal to processing limitations might predict that L2 learners, regardless of their L1s, may have difficulty coordinating all of the potential cues at their disposal when producing structures at the interface… it may be that L2 learners rely on a particular default strategy (e.g., making use of a single syntactic or phonological structure or the same article, regardless of discourse status) as the result of their being unable to integrate all of the types of information in real time" (p. 405). Thus, promoting L2 speakers' awareness about the functions of intonation and their ability in coordinating different segmental and suprasegmental cues for pragmatic proposes may help address the functional complexity of intonation and facilitate more target-like spontaneous use of English intonation.

In Experiments 1 and 2, we did not find significant differences in L1 English and Mandarin-English L2 speakers' processing of contrastive or implicational information as indicated by their fixation percentage and dwell percentage of the focused lexical items. These results support Ip and Cutler's (2016) statement that "Information structure is a linguistic universal" (p. 330). These findings also suggested that when the structure of sentences is relatively stable and predictable, L1 English and Mandarin-English L2 speakers did not differed in their processing of written information. However, in Experiment 3, we found that Mandarin-English L2 speakers' fixation percentage of the contrastive information was significantly lower than that of the L1 speakers when no visual enhancement was used to indicate the information in a passage. In Experiment 4, we found processing differences in the contrastive information in the pictures: L1 speakers focused more holistically on the contrastive information while Mandarin-English L2 speakers gazed at the information that is not contrastive. These results suggested that L1 English and Mandarin-English L2 speakers differ in visual processing when the information

structure and morphosyntactic structure were more complex and less predictable. These results further supported O'Brien and Féry's (2015) hypothesis that processing limitations may be the crucial factor in the processing and production of L2 intonation.

In Experiments 1 and 3, Mandarin-English L2 speakers did not use intonational cues in their speech to signal contrastive information even when the information is cued in text by visual enhancement such as italics and bold. The results suggested a lack of familiarity with the L2 orthographical conventions and the connection between visually enhanced information (italicized words) and speech production (sentence stress). Pronunciation textbooks use visual enhancements (e.g., italics) for pedagogical purposes (e.g., indicate the placement of sentence stress). However, the use of visual enhancement such as italics in authentic materials is often a conscious choice authors use to convey their intent (e.g., contrast, implication, etc.). When teaching pronunciation using pedagogical materials with visual enhancements, teachers need to explicitly point out the three-way connection among the constituents that are visually enhanced, the functions and rationale for the use of visual enhancement, and the role of intonation in conveying the meanings and functions in speech production.

## CONCLUSION

This study found that the structure and functional complexity of the system of intonation poses challenges to Mandarin-English L2 speakers. Complexity Theory (CT), which emphasizes the connection and interaction of interwoven variables within intonation, is an appropriate framework for L2 pronunciation research and teaching. A systematic view that highlights the complex and dynamic nature of intonation is recommended. Future studies researching the processing limitations of L2 speakers are needed. Further research investigating the dynamic mapping between L1 and L2 intonation is also recommended.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Boston University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DL and MR jointly identified the theoretical underpinning, designed, and carried out the experiments in the study. DL analyzed the data with the support from MR. DL and MR jointly

wrote the manuscript. All authors contributed to the article and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.627316/full#supplementary-material

## REFERENCES

Allen, V. F. (1971). Teaching intonation, from theory to practice. *TESOl Q.* 5, 73–81. doi: 10.2307/3586113

Anderson-Hsieh, J., Johnson, R., and Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Lang. Learn.* 42, 529–555. doi: 10.1111/j.1467-1770.1992.tb01043.x

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer. Version 6.0.37.* Available online at: http://www.praat.org (accessed March 14, 2018).

Bolinger, D. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse.* Stanford, CA: Stanford University Press.

Brazil, D. (1980). *Discourse Intonation and Language Teaching.* Harlow: Longman.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., and Griner, B. (2010). *Teaching Pronunciation: A Course Book and Reference Guide, 2nd Edn.* Cambridge, UK: Cambridge University Press.

Chapman, M. (2007). Theory and practice of teaching discourse intonation. *ELT J,* 61, 3–11. doi: 10.1093/elt/ccl039

Chen, Y., and Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *J. Phonet.* 36, 724–746.

Chun, D. M. (1982). *A contrastive study of the suprasegmental pitch in Modern German, American English, and Mandarin Chinese* (Doctoral dissertation). University of California, Berkeley, CA, United States.

Chun, D. M. (1988). Teaching intonation as part of communicative competence: suggestions for the classroom. *Die Unterrichtspraxis Teach. German* 21, 81–88. doi: 10.2307/3530748

Clennell, C. (1997). Raising the pedagogic status of discourse intonation teaching. *ELT J.* 51, 117–125. doi: 10.1093/elt/51.2.117

Cruttenden, A. (1990). The origins of nucleus. *J. Int. Phon. Assoc.* 20, 1–9. doi: 10.1017/S0025100300003984

Cruz-Ferreira, M. (1998). "Intonation in European Portuguese," in *Intonation Systems: A Survey of Twenty Languages,* eds D. Hirst and A. Di Cristo (Cambridge, UK: Cambridge University Press), 167–178.

Cucchiarini, C., and Strik, H. (2018). "Automatic speech recognition for second language pronunciation assessment and training," in *The Routledge Handbook of English Pronunciation,* eds O. Kang, R. I. Thomson, and M. J. Murphy (London: Routledge), 556–569.

DeBot, K., Lowie, W., and Verspoor, M. H. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism Lang. Cogn.* 10, 7–21. doi: 10.1017/S1366728906002732

Derwing, T. M., and Munro, M. J. (2005). Second language accent and pronunciation teaching: a research-based approach. *TESOL Q.* 39, 379–397.

Derwing, T. M., Munro, M. J., and Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Lang. Learn.* 48, 393–410. doi: 10.1111/0023-8333.00047

Derwing, T. M., Rossiter, M. J., Munro, M. J., and Thomson, R. I. (2004). Second language fluency: judgments on different tasks. *Lang. Learn.* 54, 655–679. doi: 10.1111/j.1467-9922.2004.00282.x

Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., et al. (2019). Golden speaker builder–an interactive tool for pronunciation training. *Speech Commun.* 115, 51–66. doi: 10.1016/j.specom.2019.10.005

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annu. Rev. Appl. Linguist.* 30:149. doi: 10.1017/S026719051000005X

Dussias, P. E., and Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism* 10:101. doi: 10.1017/S1366728906002847

Gilbert, J. (2014). "Myth 4: intonation is hard to teach," in *Pronunciation Myths: Applying Second Language Research to Classroom Teaching,* ed L. Grant (Ann Arbor, MI: University of Michigan Press), 107–136.

Giles, H., Mulac, A., Bradac, J., and Johnson, P. (1987). "Speech accommodation theory: the first decade and beyond," in *Communication Yearbook,* Vol. 10, ed M. McLaughlin (Beverly Hills, CA: Sage), 13–48.

Grant, L., and Yu, E. E. (2017). *Well Said Intro: Pronunciation for Clear Communication, 2nd Edition.* Boston, MA: Cengage Learning

Gumperz, J. J. (1971). *Language in Social Groups.* Stanford, CA: Stanford University Press.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation.* Cambridge, UK: Cambridge University Press.

Hahn, L. D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Q.* 38, 201–223. doi: 10.2307/3588378

Halliday, M. A. K. (2015). *Intonation and Grammar in British English,* Vol. 48. The Hague: Walter de Gruyter GmbH & Co KG.

Hellermann, J. (2003). The interactive work of prosody in the IRF exchange: teacher repetition in feedback moves. *Lang. Soc.* 32, 79–104. doi: 10.1017/S0047404503321049

Holmes, J., and Meyerhoff, M. (1999). The community of practice: theories and methodologies in language and gender research. *Lang. Soc.* 28, 173–183. doi: 10.1017/S004740459900202X

Horley, K., Reid, A., and Burnham, D. (2010). Emotional prosody perception and production in dementia of the Alzheimer's type. *J. Speech Lang. Hearing Res.* 53, 1132–1146. doi: 10.1044/1092-4388(2010/09-0030)

Housen, A., De Clercq, B., Kuiken, F., and Vedder, I. (2019). Multiple approaches to complexity in second language research. Second *Lang. Res.* 35, 3–21. doi: 10.1177/0267658318809765

Ip, M. H. K., and Cutler, A. (2016). "Cross-language data on five types of prosodic focus," in *Proceedings of Speech Prosody,* eds J. Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux (Boston, MA), 330–334. doi: 10.21437/SpeechProsody.2016-68

Ito, K., and Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J. Mem. Lang.* 58, 541–573. doi: 10.1016/j.jml.2007.06.013

Kabagema-Bilan, E., Lopez-Jimenez, B., and Truckenbrodt, H. (2011). Multiple focus in Mandarin Chinese. *Lingua* 121, 1890–1905. doi: 10.1016/j.lingua.2011.02.005

Kambe, G., Rayner, K., and Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: evidence from eye fixations. *Mem. Cognit.* 29, 363–372. doi: 10.3758/BF03194931

Kingdon, R. (1958). *The Groundwork of English Intonation.* New York, NY: Longmans.

Kohler, K. J. (1995). "The Kiel Intonation Model (KIM), its implementation in TTS synthesis and its application to the study of spontaneous speech," in *ATR Workshop on Computational Modelling of Prosody for Spontaneous Speech Processing* (Kyoto).

Labov, W. (1972). *Sociolinguistic Patterns.* Philadelphia, PA: University of Pennsylvania Press.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge, UK: Cambridge University Press.

Larsen-Freeman, D. (2017). "Complexity theory: the lessons continue," in *Complexity Theory and Language Development: In Celebration of Diane Larsen-Freeman*, eds L. Ortega and Z. Han (Philadelphia, PA: John Benjamins), 11–50.

Levis, J., and Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System* 32, 505–524. doi: 10.1016/j.system.2004.09.009

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Q.* 33, 37–63. doi: 10.2307/3588190

Levis, J. M., and Levis, G. M. (2018). Teaching high-value pronunciation features: contrastive stress for intermediate learners. *CATESOL J.* 30:139.

Levis, J. M., and Wichmann, A. (2015). "English intonation–form and meaning," in *The Handbook of English Pronunciation*, eds M. Reed and J. Levis (West Sussex UK: John Wiley & Sons, Inc), 139–155.

Lew-Williams, C., and Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *J. Mem. Lang.* 63, 447–464. doi: 10.1016/j.jml.2010.07.003

Lippi-Green, R. L. (1989). Social network integration and language change in progress in a rural alpine village. *Lang. Soc.* 18, 213–234. doi: 10.1017/S0047404500013476

Liu, D. (2020). Prosody transfer failure despite cross-language similarities: evidence in favor of a complex dynamic system approach on pronunciation teaching. *J. Second Lang. Pronunciation* 6, 1–24. doi: 10.1075/jslp.18047.liu

Mair, V. (1996). "Modern Chinese writing," in *The World's Writing Systems*, eds P. T. Daniels and W. Bright (New York, NY: Oxford University Press), 200–208.

Mixdorff, H. (2000). "A novel approach to the fully automatic extraction of Fujisaki model parameters," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, Vol. 3 (IEEE) (Istanbul). 1281–1284.

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System* 42, 258–269. doi: 10.1016/j.system.2013.12.007

O'Brien, M. G., and Féry, C. (2015). Dynamic localization in second language English and German. *Bilingualism Lang. Cogn.* 18, 400–418. doi: 10.1017/S1366728914000182

O'Brien, M. G., Jackson, C., and Gardner, C. E. (2014). Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. *Appl. Psycholinguist.* 35, 27–70. doi: 10.1017/S0142716412000252

O'Connnor, J. D., and Arnold, G. F. (1973). *Intonation of Colloquial English*. London: Longman.

Ortega-Llebaria, M., and Colantoni, L. (2014). L2 English intonation: relations between form-meaning associations, access to meaning, and L1 transfer. *Stud. Second Lang. Acquisition* 36, 331–353. doi: 10.1017/S0272263114000011

Ouyang, I. C., and Kaiser, E. (2015). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Lang. Cogn. Neurosci.* 30, 57–72. doi: 10.1080/01690965.2013.805795

Palmer, H. E. (1922). *English Intonation, With Systematic Exercises*. Cambridge: W. Heffer & Sons Limited.

Papadopoulou, D., and Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: a study of relative clause attachment in Greek. *Stud. Second Lang. Acquisition* 25, 501–528. doi: 10.1017/S027226310300 00214

Pell, M. D., and Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE* 6:e27256. doi: 10.1371/journal.pone.0027256

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Q.* 35, 233–255.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English Specific Purposes* 23, 19–43. doi: 10.1016/S0889-4906(03)00020-6

Pickering, L. (2018). *Discourse Intonation: A Discourse-Pragmatic Approach to Teaching the Pronunciation of English*. Ann Arbor, MI: University of Michigan Press. doi: 10.3998/mpub.6731

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (PhD thesis). Cambridge, MA: Indiana University Linguistics Club. Available online at: http://dspace.mit.edu/handle/1721.1/16065

Pierrehumbert, J., and Hirschberg, J. B. (1990). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, eds P. Cohen, J. Morgan, and M. Pollack (Cambridge, MA: MIT Press), 271–311.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Tanenhaus, M. K. (2007). "Spoken language comprehension: insights from eye movements," in *The Oxford Handbook of Psycholinguistics*, eds P. Levelt and A. Caramazza (New York, NY: Oxford University Press), 309–326.

Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge, UK: Cambridge University Press.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: a study of Chinese speakers of english. *Stud. Second Lang. Acquisition* 20, 1–25. doi: 10.1017/S0272263198001016

Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. New York, NY: Oxford University Press.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *J. Phon.* 27, 55–105. doi: 10.1006/jpho.1999.0086

Xu, Y. (2004). "Transmitting tone and intonation simultaneously-the parallel encoding and target approximation (PENTA) model," in *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages* (Beijing), 215–220.