

# The Effect of Fundamental Frequency on Mandarin Intelligibility by L2 Learners in Quiet and Noise Environments: A Pilot Study



Kaidi Chen and Chunsheng Yang

**Abstract** Fundamental frequency (F0), listening environment, and semantic context are three important factors for both tonal and non-tonal language intelligibility by native speakers. However, it remains unclear how these factors affect second language (L2) learners of Mandarin Chinese and whether there are differences between native and L2 Mandarin speakers. Through speech re-synthesis and sentence counterbalancing, this study investigated the possible effects of F0 (i.e., natural F0 versus flattened F0) on the intelligibility of Mandarin speech by L2 Mandarin learners from different proficiency levels in quiet and white noise conditions when controlling for sentence context. A mixed-effect statistical model confirmed the main effects of F0 contour, listening environment, and proficiency level. That is to say, the lack of natural F0 contour, the presence of noise, and the lower proficiency level would predict the reduction in intelligibility when adjusting for the other two variables. However, no significant interactions were found. Specifically, the hypothesis that flattened sentences are as intelligible as natural sentences for more advanced learners was not supported due to the change of experimental subjects from native speakers to L2 speakers. It was proposed that compared to native speakers, L2 speakers' underdeveloped utilization of secondary cues and semantic contexts, due to a developing proficiency level, may lead to non-significant interactions. The finding of the effect of F0 on intelligibility also illustrates the importance of tone accuracy and diversifying L2 learners' linguistic input in Chinese pronunciation teaching and learning.

**Keywords** Intelligibility · Fundamental frequency · Listening environment · Proficiency

---

K. Chen (✉) · C. Yang  
University of Connecticut, Storrs, CT, USA  
e-mail: [kaidi.chen@uconn.edu](mailto:kaidi.chen@uconn.edu)

© Springer Nature Singapore Pte Ltd. 2021  
C. Yang (ed.), *The Acquisition of Chinese as a Second Language Pronunciation*,  
Prosody, Phonology and Phonetics, [https://doi.org/10.1007/978-981-15-3809-4\\_10](https://doi.org/10.1007/978-981-15-3809-4_10)

213

## 1 Introduction

### 1.1 *Intelligibility*

Intelligibility is one of the most important constructs in second language (L2) pronunciation research. However, there is no universally agreed definition and measure of intelligibility (Munro and Derwing 1999; Pickering 2006; Chen 2011, among others), likely due to its confusion with comprehensibility. Smith and Nelson (1985) defined intelligibility as listeners' ability to recognize individual words or utterances. They further pointed out that miscommunication occurs when people only recognize words and utterances but fail to understand the meaning (termed as comprehensibility), or the pragmatic meaning behind them (termed as interpretability). In Smith and Nelson's definition, intelligibility and comprehensibility are closely related to each other but refer to speech understanding at different levels.

In another line of literature, intelligibility was broadly defined as the extent to which a speaker's message is actually understood by a listener (Munro and Derwing 1999; Derwing and Munro 2005). Levis (2018) interpreted it as "the extent to which a speaker is understandable" in a "narrow sense," and "whether the particular words used by a speaker are successfully decoded (the lexical level intelligibility)" in a "broad sense." It is measured by orthographic transcription tasks, i.e., percentage of words correctly transcribed (Munro and Derwing 1999; Derwing and Munro 2005; Yang 2016, among others). Different from Smith and Nelson (1985), comprehensibility was defined by Derwing and Munro as listeners' perception of the degree of difficulty in understanding an utterance. Comprehensibility is usually measured by scalar judgment tasks, from "extremely easy to understand" to "extremely difficult to understand" (Derwing and Munro 2005). However, it is worth pointing out that even if we recognize every single word and utterance, it does not mean that we can understand it when listeners do not have enough background knowledge. Even if we understand the utterance in the context, it does not mean that we need to recognize every single word, and in many cases, we do not have to do so. Thus, definitions and measures of intelligibility in both narrow sense and broad sense were considered in this study.

### 1.2 *Factors Affecting Intelligibility*

#### 1.2.1 **Fundamental Frequency**

Fundamental frequency, referred to as F0, is the lowest frequency of a complex periodic sound. F0 determines pitch contour generally and expresses intonation (broadly speaking) linguistically. Although F0 does not influence segmental parts (consonants and vowels) of the speech, its prosodic feature has many linguistic functions: distinguishing lexical meaning (only for tonal languages), discriminating declarative

and interrogative sentences, marking emphasis, and paralinguistic functions such as expressing emotions (such as F0 increase in anger or fear, and F0 decrease in grief, sorrow, or depression). According to Lehiste (1970, cited from Binns and Culling 2007), important content words will be accented in normal speech and the corresponding F0 tends to be above the average F0 of the sentence. In this sense, the content words will be acoustically clearer than surrounding words, together with the contributions from the factor that content words are often articulated louder and more slowly. However, when F0 is flattened, none of the words in the sentence are accented and all of them are at the same F0 (Binns and Culling 2007). Without F0 cues, it would be difficult to find where the content or important words are. In an inverted F0 contour, the accented content words will go to opposite directions: a fall will be a rise and vice versa; F0 above the average will be below the average F0 and vice versa. As a result, no F0 cues will highlight important words in monotonous sentence; the F0 cues in inverted sentences will be misleading and highlight words that are not important to the meaning of the sentence (Binns and Culling 2007).

Previous studies (Maassen and Povel 1984; Laures and Weismer 1999) have investigated the role of intact fundamental frequency (F0) contours on the intelligibility of non-tonal languages and indicated that lack of intact F0 will decrease intelligibility. Laures and Weismer (1999) tested a typical group who did not self-report hearing loss or professional training in speech science and experimental psychology. Their results showed that the intelligibility of English sentences in terms of both word transcription and interval scaling were significantly lower when F0 contour was flattened, as compared with naturally varying contours. Maassen and Povel (1984) explored the role of fundamental frequency on the intelligibility in atypical population, namely, deaf children who are frequently reported to have monotonous voice. The overall results showed that when the original F0 contour of Dutch sentences from the deaf utterances was replaced by artificial contours, the percentage of the identified words increased significantly (although the change is small). It led to the conclusion that intonation correction yields significant improvement of intelligibility.

## 1.2.2 Listening Environment

Many studies (Laures and Bunton 2003; Binns and Culling 2007; Watson and Schlauch 2008; Miller, Schlauch and Watson 2010) included listening environment when examining the role of F0 on intelligibility. They consistently demonstrated that dynamic F0 contours are significant to speech intelligibility when taking background noise into account. Results from Laures and Bunton (2003) showcased that the absence of fundamental frequency variation has a significant impact on overall speech intelligibility. A flattened fundamental frequency contour negatively influences intelligibility when taking account of the competing listening background (white noise and multi-talker babble noise). Watson and Schlauch (2008) had similar findings that sentences with flattened F0 yielded poorer intelligibility than the unmodified ones in white noise. Their study also tested the effects of resynthesized F0 that reflected the average low F0, the median F0 and the average high F0. Sentences

flattened at the average high F0 yielded poorer intelligibility than that at the median F0, and the average low F0 yielded better intelligibility than that at the median F0. Binns and Culling (2007) compared the effects of intact F0 contour on intelligibility with flattened F0 and inverted F0 in adverse listening conditions. They found that against speech-shaped noise, flattened F0 has no significant impact on speech reception thresholds (SRTs) while inverted F0 does increase SRTs significantly, compared to intact F0 contour; however, when against single-talker interferer, both flattened and inverted conditions have greater effects and significantly increase SRTs. Therefore, it was concluded that intact F0 actually improves the intelligibility in noise, as compared to monotone or inverted F0. Building upon research on flattened and inverted F0, Miller, Schlauch and Watson (2010) further investigated how the F0 manipulations affect intelligibility in background noise. They had unmodified F0, flattened F0 at the median, natural but exaggerated F0, inverted F0, and sinusoidally frequency modulated F0. The results showed that the last two F0s (which create misleading cues) have more detrimental effect on speech intelligibility than flattened F0 and intact F0 in background noise.

### 1.2.3 Semantic Context

Semantic context is a factor often considered to help listeners recognize and understand an utterance. For example, Cole and Perfetti (1980) used the task that children and adults listen to mispronunciations in a children's story to test the role of context on words recognition. It is suggested that children detected mispronunciations more accurately when they occurred in highly predictable context, and all age groups detected the mispronunciations more quickly in predictable words. Craig, Kim, Rhyner and Chirillo (1993) examined the interaction of acoustic information with contextual information during speech perception. The results showed that predictability-high (PH) words were recognized earlier and with greater confidence than predictability-low (PL) words for all ages ranging from 5 to 83. Later, researchers also started to combine listening conditions with context. Fernald (2001, cited from Zhou, Li, Liang, Guan, Zhang, Shu and Zhang 2017) claimed that previous work showed that children as early as two years old are able to use semantic context to assist speech recognition in quiet. Pichora-Fuller and Daneman's (1995) experiment illustrated that old adults derived more benefit from supportive context (with sentence-final words that were either predictable context or unpredictable context) than young adults in babble background. Sheldon, Pichora-Fuller and Schneider (2008) explored how younger and older adults benefited from context when identifying target words in noise-vocoded sentences. The first type of context is either highly predictable or not predictable sentence-final target words, and the second type of context is either with priming or not. The results indicated that younger and older adults benefited from each type of context and with the most benefit gained when both types were combined. Similarly, Dubno, Ahlstrom and Horwitz (2000) found that both older and younger adults with normal hearing derived equivalent benefit from context given equivalent speech audibility in noise. Benichov, Cox, Tun and

Wingfield (2012) included more factors than previous researches. They confirmed the robust role of linguistic context to aid spoken word recognition when taking age, hearing acuity, verbal ability, and cognitive function into consideration.

### 1.2.4 Intelligibility of Mandarin Chinese

As we can see from the above literature review, fundamental frequency, listening background, and semantic context are three important factors jointly affecting intelligibility. However, most of previous studies examined the intelligibility of non-tonal languages, primarily English. In tonal languages, such as in Mandarin Chinese, tones are lexically specified and lexical tones distinguish lexical meanings from otherwise identical strings of phonemes (Wang 1973; Wang, Shu, Zhang, Liu and Zhang 2013; Xu, Zhang, Shu, Wang and Li 2013). Different from lexical tones in tonal languages, F0 or intonation in non-tonal languages is mainly used for pragmatic purposes, such as sentence modality, emphasis, and emotion (Cutler, Dahan and Donselaar 1997). In this sense, it is expected that F0 may play a more important role in the intelligibility of tonal languages than that of non-tonal languages.

Only a very limited number of studies (Liu and Samuel 2004; Patel, Xu and Wang 2010; Wang et al. 2013; Xu et al. 2013; Chen, Wong and Hu, 2014; Zhou et al., 2017) have investigated the intelligibility of Mandarin Chinese. Liu and Samuel (2004) found that in whispered speech, identification of tonal patterns remains “surprisingly” good when the F0 information is neutralized. Native Mandarin listeners can use secondary cues (i.e., duration and amplitude) when the primary cue (F0) is unavailable. The prediction from the finding of whispered speech to flat-F0 speech was questioned by Patel et al. (2010) in that flat-F0 has voicing while whispered speech does not and F0 gives prominent cue for tone perception (Whalen and Xu 1992, cited from Patel et al. 2010). Patel et al. (2010) conducted their own experiment on the role of intact and flattened F0 when controlling for listening environments. They found that for native Mandarin listeners, monotonic speech is just as intelligible as natural speech in a quiet background, but the flat-F0 speech became substantially less intelligible than natural speech when noise was added. Their finding was corroborated by behavioral experiments by Xu et al. (2013) in which listeners (native Mandarin speakers with minimal music experience) rated monotone sentences as equally intelligible as normal sentences; it was also supported by Chen et al. (2014) which found that normal hearing listeners (native Mandarin speakers) perfectly recognized Mandarin sentences produced with modified tone contours (flat tone or tone randomly selected from the four Mandarin lexical tones) in a quiet environment, but their performance declined in noise. Furthermore, the fMRI result by Xu et al. (2013) provided an explanation for the equative intelligibility of flat F0 and natural F0 (regardless of listening background). Monotone sentences elicited greater activation in the left planum temporale (PT), demonstrating the automatic use of additional neural resources to recover the phonological loop from altered tonal patterns. However, the preceding studies did not explain what cues are utilized for comprehension when the sentence is flattened. Wang et al. (2013) investigated

the role of sentence context on intelligibility together with F0 contour and listening environment. It is revealed that for native Mandarin listeners, word list sentences with natural F0 contours were less intelligible than normal sentences counterparts in both quiet and noise conditions, indicating that sentence context improves speech intelligibility regardless of listening backgrounds; they also argued that sentence context partially explained the unchanged intelligibility of monotonous sentences in the quiet environment. Zhou et al. (2017) corroborated the influence of semantic context on intelligibility together with factors of F0 and listening environment by elementary and middle-school-aged children. Children of both age groups use semantic context to assist speech recognition; with flat F0 contours, younger children are worse in making use of context in recognizing speech than older children. Considering the interactions and joint impact of sentential semantic context, F0 contours and listening environments on Mandarin speech intelligibility by native Mandarin speakers, both children and adults, it would be interesting and worthwhile to examine how these factors affect L2 Mandarin speakers and whether there are any differences between native and L2 speakers.

This study attempts to investigate the effects of F0 (i.e., natural F0 versus flattened F0) on the intelligibility of Mandarin speech by L2 Mandarin learners in quiet and white noise conditions when controlling for sentence context. Intelligibility in present study is defined at two levels: it consists of both word and utterance recognition (Kirkparick et al., 2008), and to what extent a listener can understand the locutionary meaning of a message (Munro and Derwing, 1999). Previous studies have shown L2 Mandarin speakers' real-time perceptual development toward more native-like directions (in both reaction time and accuracy) on Mandarin tones AX-discrimination task (Wiener, 2017) and advanced L2 Mandarin learners' better perception of Mandarin intonation and better identification of intonation-superimposed tones as compared to the first- and second-year learners (Yang, 2016). Zhou et al. (2017) also showed the developmental changes of native Mandarin speakers' speech intelligibility (Zhou et al., 2017). To this end, we also want to examine how L2 learners of Mandarin at different proficiency levels differ in speech intelligibility.

We address the following questions in this study:

- (1) What are the effects of F0 (natural versus flat) and listening environment (quiet versus noise) on Chinese intelligibility when keeping semantic context constant?
- (2) How does proficiency level affect L2 Chinese listeners' intelligibility?
- (3) What are the interactions of F0, listening environment, and proficiency level in L2 Mandarin intelligibility?

Drawing upon the discussions above, we make the following predictions: noise, flat F0, and low proficiency will all reduce intelligibility when holding context constant. There are interactions among F0 variations, listening environment, and proficiency level. Specifically, in quiet background, flattened sentences are as intelligible as natural sentences for more advanced learners, but not for lower proficiency learners. When noise is added, the intelligibility of both flat and natural sentences will drop across all proficiency levels.

## 2 Methodology

### 2.1 Participants

Twenty L2 Mandarin learners, 4 at each of the 5 proficiency levels (level 1, 2, 2.5, 3, and 4), from an intensive summer program in the USA, were recruited for this study. At the beginning of the summer program, students were placed into these five levels according to their performance in the informal ACTFL standardized Oral Proficiency Interviews (OPI) conducted by the instructors of the summer school. They participated in the research at the end of the summer program.

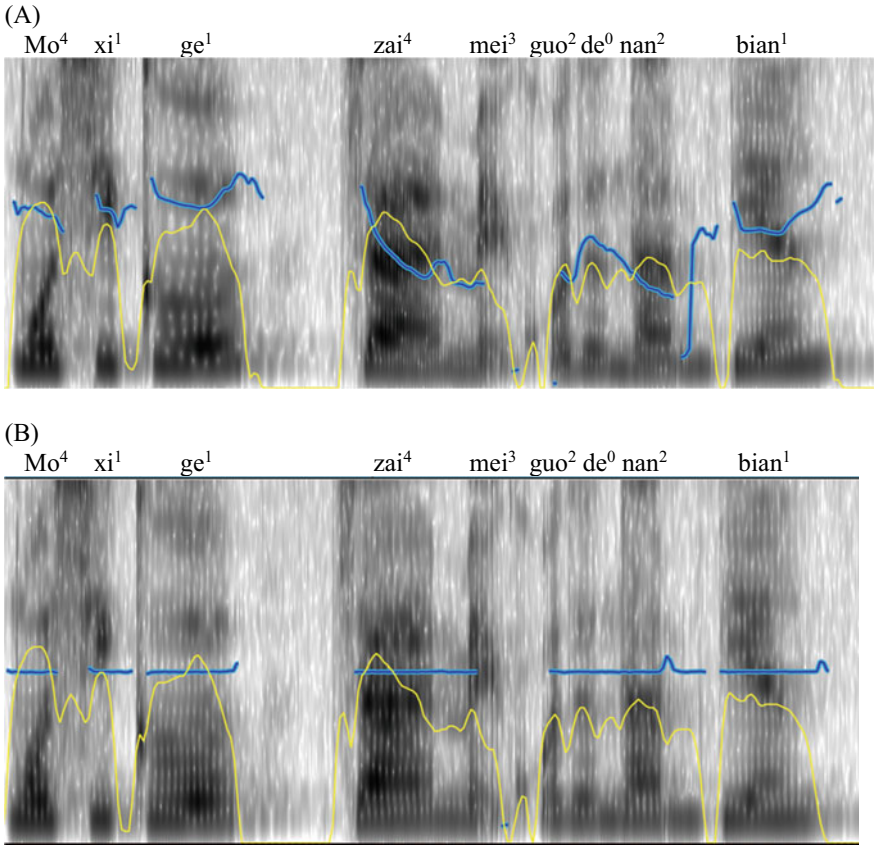
### 2.2 Stimuli

18 Chinese sentences were created by the first author and read by a female Beijing Mandarin speaker in her 30s. All vocabulary and grammar were taken from the following Chinese textbook: *Integrated Chinese (volume 1 and 2)* (Liu, Yao, Bi, Ge and Shi 2016), *Basic Mandarin Chinese* (Kubler 2017) and *Intermediate Spoken Chinese* (Kubler 2013). Appendix 1 presents the whole list of these sentences in Chinese characters and their English translations. To help participants become familiar with the task, two practice sentences were prepared. Additionally, five filler sentences were inserted among the target sentences intermittently to alleviate the impact from cognitive confounding variables such as attention.

Praat (Boersma and Weenink 2018) and Praat vocal toolkit (Corretge 2012–2020) were used to manipulate the stimuli. Specifically, monotones were created by flattening the F0 contour of each sentence at the sentence's mean F0 (Fig. 1). In this sense, pitch-flattened sentence neutralized the intonations and lexical tones while keeping other syllabic and sub-syllabic acoustic information (such as intensity and duration) intact. White noise at + 65 SNR level was added. After manipulations, there were four conditions for each sentence: natural tone, natural tone + 65 dB noise, flat tone, and flat tone + 65 dB noise. Altogether there are 100 sentences (25 sentences  $\times$  2 F0 conditions  $\times$  2 noise conditions). All 100 sentences were amplitude normalized using Praat. Then the sentences were randomized into four blocks, equally distributed across the F0 conditions and noise conditions. Each block has all 25 sentences from the sentence list but in different F0 and noise conditions, all counterbalanced.

### 2.3 Procedure

Participants were recruited through the help of the instructors of various classes. When participants came to the study, they were given the consent forms first and were



**Fig. 1** Acoustic features of sample speech stimuli. Broadband spectrograms in black, intensity contours in yellow, and F0 contours in blue. Panel A: normal (natural F0) sentence. Panel B: F0-flattened counterpart

asked to read and sign before starting the task. Participants were tested individually in a quiet classroom while facing a Mac Pro. They heard each sentence from the speaker of the laptop at a comfortable level. The participants were asked to write down the sentences they heard in either Chinese characters or *pinyin* Romanization. The verbal instruction was in Chinese only, due to the “only Chinese” language pledge signed by all the students in the summer program. To ensure that the participants understood the Chinese instructions, written English instructions were also provided. The progression of the task was controlled by the first author. After listening to one sentence, the participants wrote down the sentence and then translated it into English on the answer sheet. Then the first author would proceed to the next sentence. Each participant listened to each stimulus only once.

The whole task took around 10–15 min. To avoid learning effect and confounding factors, such as attention and fatigue, each participant only listened to one block of



stimuli and the other three participants from the same level listened to the rest of the three different blocks.

## 2.4 The Measurement of Intelligibility

Following Lane (1963), Munro and Derwing (1999), Derwing and Munro (2005), and Yang (2016), intelligibility was measured by the proportions of correct syllables over the total syllable numbers in a sentence. Only when consonant(s) (if any), vowel, and tone were all correct, was a syllable considered to be correct. Because we adopted both narrow and broad measures of intelligibility, apart from word and utterance recognition, understanding of the sentence was also our concern. English translation was used to test whether participants understood the sentences correctly. If their translation was wrong, they did not really comprehend the meanings of the sentence. Therefore, as long as the participants did not translate the sentence correctly, even though they had correctly transcribed syllables, the syllables were not taken as correct. Correct intelligibility should include both correctly transcribed syllables and correct English translations. The intelligibility score was calculated for each sentence in different F0 and noise conditions.

Table 1 gives two examples of how intelligibility scores were calculated. This correct sentence, “Wǒ fēi cháng xǐ huān běi jīng dòng wù yuán,” was the baseline and we calculated how many syllables each participant transcribed and translated correctly. Participant A did not write the last five syllables (běi jīng dòng wù yuán) correctly, but instead wrote “zhōng guó rén” and accordingly translated it as “Chinese people” wrongly. This participant only transcribed and translated the first five syllabus correctly. Thus,  $5/10 = 0.5$  is participant A’s intelligibility score for this sentence. Participant B listened to the same sentence but in different conditions (natural tone without noise). A mixture of Chinese characters and *pinyin* were given in the answer. This participant transcribed two syllables (dōng wú) wrongly. Thus, two correct syllables were missing in the transcription. For the syllable “yuán,”

**Table 1** Samples of intelligibility scoring

Target sentence and its translation (condition)	Transcription and translation (participant)	Intelligibility score
wǒ fēi cháng xǐ huān běi jīng dòng wù yuán. (natural tone in noise) I really like Beijing Zoo	wǒ fēi cháng xǐ huān zhōng guó rén. (Participant A) I really like Chines people	0.5
wǒ fēi cháng xǐ huān běi jīng dòng wù yuán. (natural tone without noise) I really like Beijing Zoo	我非常喜欢北京的 dōng wú yuán. (Participant B) I really like Beijing’s pet stores	0.7

although it was transcribed correctly, the translation was wrong. As a result, only the first 7 syllables got credits, and participant B’s intelligibility score for this sentence is “7/10 = 0.7”.

### 3 Data Analysis and Results

We used a mixed-effect model with proficiency level, flat tone, and noise as fixed variables and sentence number as a random variable. In this case, semantic context was held constant when testing other variables. The model can be written as:

$$y_i = \text{Noise} \times \beta_1 + \text{Proficiency Level} \times \beta_2 + \text{Flat Tone} \times \beta_3 + \text{Sentence Number} \times u + \epsilon$$

where  $y_i$  represents each intelligibility score.

First, we looked at the two-way and three-way interactions, and no significant interactions were found between/among any variables, as shown in Table 2.

Since there were no interactions, we excluded the interactions from our model. Table 3 presents the main effects of all three predictors, and tables in the appendix (see appendix 2) show the estimated marginal means for noise and flat tone from models without interactions. It can be seen that different proficiency levels predict

**Table 2** Analysis of variance with interactions

Variable	F	Degrees of freedom	Degrees of freedom of the residues	P value	Eta_sq
ProficiencyLevel	32.11	4	323	< 0.0001	0.237
FlatTone	31.41	1	323.52	< 0.0001	0.059
Noise	47.62	1	323.51	< 0.0001	0.087
ProficiencyLevel:FlatTone	0.25	4	323.02	0.91	0.002
ProficiencyLevel:Noise	0.44	4	323.04	0.78	0.003
FlatTone:Noise	3.48	1	323.02	0.06	0.006
ProficiencyLevel:FlatTone:Noise	0.53	4	323.02	0.71	0.004

**Table 3** Analysis of variance without interactions

Variable	F	Degrees of freedom	Degrees of freedom of the residues	P value	Eta_sq
ProficiencyLevel	32.57	4	336	< 0.0001	0.24
FlatTone	31.83	1	336.53	< 0.0001	0.06
Noise	48.24	1	336.5	< 0.0001	0.09

different intelligibility scores ( $p < 0.0001, \eta^2 = 0.24$ ) when controlling for noise and flat tones; compared to flat tones ( $M = 0.51$ ), natural tones ( $M = 0.64$ ) predict a higher intelligibility score ( $p < 0.0001, \eta^2 = 0.06$ ) when taking proficiency level and noise into account; compared to noise condition ( $M = 0.49$ ), no noise condition ( $M = 0.66$ ) predicts a higher intelligibility score ( $p < 0.0001, \eta^2 = 0.09$ ) over proficiency level and flat tones. In addition, the effect size of proficiency level is large, accounting for 24% of the variance of sentence scores; the effect size of flat tone is medium, explaining 5.9% of the variance; the effect size of noise is medium, with 8.6% of variance in sentence scores explained. The main effects of the three variables can also be observed in Fig. 2.

Finally, to investigate how specific proficiency level predicts the intelligibility score, we compared each level to a reference level. The reference level here is proficiency level 1. Tables 4 and 5 present the estimated marginal means for each proficiency level and the pairwise differences across levels. We can see from the tables that there are no significant differences between proficiency Level 2 and Level 2.5 and between Level 3 and Level 4. For the rest of the comparisons, they are significantly different in predicting intelligibility scores when adjusted for noise and flat tone. Specifically, Level 2 ( $M = 0.49$ ), Level 2.5 ( $M = 0.56$ ), Level 3 ( $M = 0.68$ ) and Level 4 ( $M = 0.76$ ) significantly predict higher intelligibility scores than Level 1 ( $M = 0.38$ ),  $p_{2-1} < 0.05, p_{2.5-1} < 0.05, p_{3-1} < 0.05, p_{4-1} < 0.05$ ; Level 3 ( $M = 0.68$ )

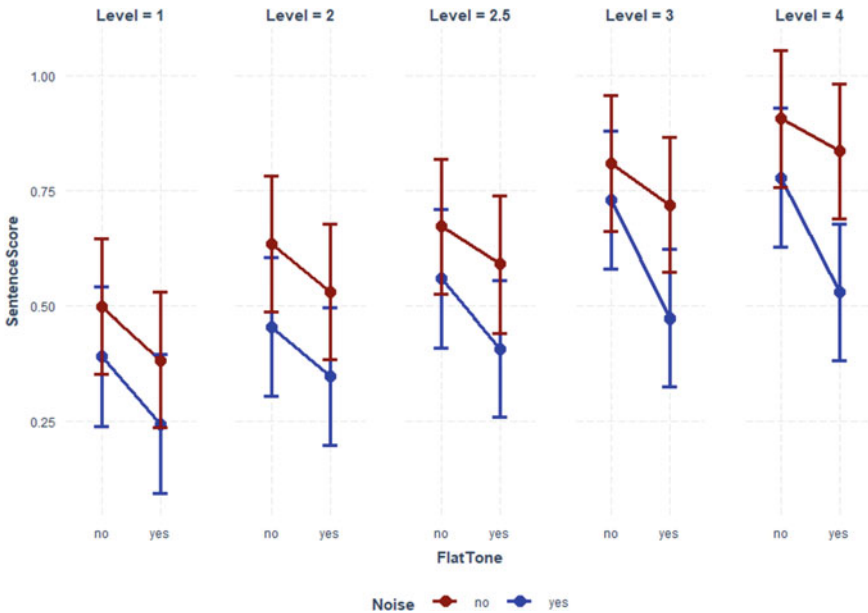


Fig. 2 Relationships of proficiency level, noise, flat tone, and sentence score

**Table 4** Estimated marginal means for proficiency level

<i>Level</i>	<i>Emmean</i>	<i>SE</i>	<i>Df</i>	<i>Lower.CL</i>	<i>Upper.CL</i>
1	0.38	0.06	24.20	0.25	0.50
2	0.49	0.06	24.19	0.37	0.61
2.5	0.56	0.06	24.19	0.43	0.68
3	0.68	0.06	24.19	0.56	0.80
4	0.76	0.06	24.19	0.64	0.89

**Table 5** Differences between estimated marginal means across proficiency level

<i>Contrast</i>	<i>Estimate</i>	<i>SE</i>	<i>Df</i>	<i>T.ratio</i>	<i>p.value</i>
1 - 2	-0.1153	0.0379	336.0000	-3.0448	0.0210
1 - 2.5	-0.1795	0.0379	336.0000	-4.7412	0.0000
1 - 3	-0.3056	0.0379	336.0000	-8.0699	0.0000
1 - 4	-0.3863	0.0379	336.0000	-10.2000	0.0000
2 - 2.5	-0.0642	0.0379	336.0000	-1.6963	0.4376
2 - 3	-0.1903	0.0379	336.0000	-5.0255	0.0000
2 - 4	-0.2710	0.0379	336.0000	-7.1558	0.0000
2.5 - 3	-0.1261	0.0379	336.0000	-3.3291	0.0085
2.5 - 4	-0.2067	0.0379	336.0000	-5.4592	0.0000
3 - 4	-0.0807	0.0379	336.0000	-2.1303	0.2097

and Level 4 ( $M = 0.76$ ) significantly predict higher sentence scores than Level 2 ( $M = 0.49$ ),  $p_{3-2} < 0.05$ ,  $p_{4-2} < 0.05$ ; Level 3 ( $M = 0.68$ ) and Level 4 ( $M = 0.76$ ) significantly predict higher sentence scores than Level 2.5 ( $M = 0.56$ ),  $p_{3-2.5} < 0.05$ ,  $p_{4-2.5} < 0.05$ .

## 4 Discussions

This study investigated the role of F0, listening environment (with or without noise), and proficiency level on the intelligibility of Mandarin Chinese by L2 Mandarin learners. The semantic context in target sentences was held constant (sentence number as a random factor) when testing other variables; in this case, the finding can be generalized to any Mandarin sentence in any semantic context. The three variables, F0 contour, listening environment, and proficiency level, were all found to affect Mandarin intelligibility by L2 Mandarin learners. That is to say, the lack of natural F0 contour, the presence of noise, and the lower proficiency level, would all predict reduction in intelligibility. The relationship of different proficiency levels and intelligibility was also confirmed.

Although the effects of F0 contour and noise are consistent with previous studies (Patel et al. 2010; Wang et al. 2013; Zhou et al. 2017), no significant interactions we hypothesized were found. Looking at Fig. 2, we can see that in a quiet environment, the intelligibility of flat-F0 sentences is lower than that of natural F0 sentences across all proficiency levels. In a noise condition, the pattern is similar, across all proficiency levels. The non-significant interactions of F0 contour and background noise are inconsistent with previous researches on native Mandarin speakers (Patel et al. 2010; Wang et al. 2013; Xu et al. 2013; Chen, et al. 2014; Zhou et al. 2017). These studies have found that the difference of intelligibility of flat F0 and natural F0 sentences depends on the listening environment; namely, flat F0 speech in a quiet environment is as intelligible as natural F0 speech, but in a noise environment, flat F0 dramatically reduced the intelligibility compared to the mild decrease for natural F0 sentences. It was argued that such finding highlighted “the importance of natural F0 contour for sentence intelligibility in noise” (Wang et al. 2013) and “the robustness and flexibility of spoken Mandarin comprehension” (Patel et al. 2010).

We argue that the inconsistent findings on the interactions of factors affecting Mandarin intelligibility are likely due to the change of subjects from native Mandarin listeners to L2 Mandarin listeners. Studies on L2 Mandarin suprasegmentals (Wiener 2017; Yang 2016) have showcased that there are either real-time developments on tone perceptions after classroom learning or various tone and intonation perceptions of L2 learners at different proficiency levels. For example, Yang (2016) found that with respect to the identification of intonation of statements, particularly for those ending with tone 2, native speakers were far more accurate than first-year L2 learners, second-year L2 learners and advanced L2 learners. Yang (2016) interpreted that as L2 learners’ proficiency improved over time, their perception of statement intonation also improved. Furthermore, Yang (2016) proposed that native and L2 listeners may be attending to different cues in perceiving intonation types: native listeners attend to both “global and localized F0 cues” in identifying intonations while L2 listeners primarily depend on “localized terminal F0 cues (mainly the tone of the last syllable).” The difference of mechanism in intonation identification of native and L2 listeners may help explain the different findings on Mandarin intelligibility to some extent. That is to say, L2 listeners tended to focus more on individual words when transcribing and translating, rather than focus on the entire sentence. Yang (2016) also discovered the difference in tone identification between native and L2 listeners: both native and advanced L2 listeners performed much better than first- and second-year L2 listeners. Results also showed a path of improvement from first year to advanced L2 learners in tone perception. Given the aforementioned findings of native and L2 difference in perception of tones and intonation, we assume that if participants in our current study only listen to natural F0 contour sentences, native and L2 listeners will perform differently in intelligibility task. However, for the flattened F0 contour sentences, it would be expected that L2 listeners would not be any worse than native speakers since neither group had tonal and intonational cues to rely on for intelligibility. In other words, L2 listeners in our study were supposed to be similar to native listeners in previous studies in terms of the intelligibility of flat-F0 sentences in a quiet environment. However, the interaction of F0 and the listening environment was

not borne out, implying that there are some other cues native listeners can access to assist intelligibility but L2 listeners cannot.

Besides F0, previous studies have shown that native listeners make use of secondary cues, such as duration, amplitude, or acoustic boundaries/landmarks, when F0 cues are not accessible (Liu and Samuel 2004; Li and Loizou 2008; Patel et al. 2010; Chen et al. 2014). Thus, due to their limited exposure to Mandarin Chinese, L2 learners are not as good as native speakers at making use of these secondary cues when tones and intonations were flattened. Thus, we propose that the constraints of proficiency, specifically the underdeveloped utilization of secondary cues other than tone contours, may lead to the non-significant interactions.

As we stated in the introduction, context is also a big factor influencing L1's intelligibility (Wang et al. 2013; Zhou et al. 2017). This could also be one aspect that L2 listeners lack. Since we have controlled semantic context to be constant in the present study, we could not know how different contexts affect L2's intelligibility. It is possible that L2 learners might still be in the process of developing the sensitivity to semantic context.

When we look at Fig. 2, we could clearly see that as proficiency level improves, the slopes of the red horizontal lines and the blue horizontal lines are progressing toward a converging point, showing their possible tendency to interact with each other and move off the parallels. We argue that two factors may be playing a role here. Firstly, the L2 learners in this study, including the Level 4 learners, are still in the process of developing their proficiency. This is due to their limited exposure to Mandarin Chinese, especially in terms of both phonetic/phonological variations often occurring in actual communication and the phonotactic constraints in the language. In this sense, the Level 4 participants are still not advanced enough, at least not native-like. On the other hand, the small sample size in our study is another factor which may prevent the occurrence of the interaction of flat F0 and noise. Future studies can be expanded to include more advanced L2 learners and increase the sample size of each level to 20 or 30.

Lastly, the measure of intelligibility in this study may lead to the inconsistent findings from previous studies. In this study, we adopted both narrow and broad definition of intelligibility and the measure we used included both the orthographic transcription and English translation. However, previous researchers used various measures of intelligibility, such as orthographic transcription (Patel et al. 2010; Wang et al. 2013), verbal repetition (Zhou et al. 2017), and scale ratings of comprehension (Xu et al. 2013). They are either mere recognition (Patel et al. 2010; Wang et al. 2013; Zhou et al. 2017) or comprehension (Xu et al. 2013). But none of these studies combined transcription/recognition and translation/comprehension in their measurement of intelligibility.

## 5 Pedagogical Implications, Limitations and Future Studies

This study has significant pedagogical implications. The finding of the effect of F0 on intelligibility highlights the importance of tone accuracy in L2 Mandarin teaching and learning. Although monotone sentences can be as equally intelligible as natural F0 sentences for native speakers in a quiet environment, this unfortunately does not apply to L2 learners. L2 listeners' ability to use secondary cues, such as duration and amplitude, is still developing, and their limited experience and exposure do not provide them with phonetic/phonological variations and the (implicit) knowledge of Mandarin phonotactic constraints. Thus, they do not have the resources to rely on to recognize and comprehend utterances when F0 is not available in both quiet and noise environments. To help L2 learners become better listeners, tone accuracy should be emphasized in L2 Chinese classes, not only at the beginning level, but also at the intermediate and advanced levels. More importantly, tone training should be incorporated in meaningful communicative activities or focus-on-form tasks in addition to mechanical drillings (Yang 2016 and 2020). To help L2 learners understand well in undesirable environments, such as in a noisy listening condition, they should be provided with access to different types of linguistic input. For example, L2 learners should listen to both slow speech and fast speech, both standard speech and non-standard, or even accented speech, and both speech by native speakers and speech by non-native speakers. By exposing L2 learners to a diversity of linguistic input and integrating tone and pronunciation training in task-based pronunciation activities in various listening environments, L2 learners will acquire allophonic/allotonic knowledge of Mandarin tones and learn to use secondary acoustic cues (i.e., duration and amplitude).

One limitation of current study is that we did not consider individual difference. Cognitive variables, such as attention and working memory, vary from person to person. Fatigue can also be a confounding variable as the first author has witnessed some participants saying "very tired" when coming to testing venue for this study right after their immersion class. The alternative choice of either transcribing in Chinese characters or *Pinyin* is also a limitation in manipulating individual difference. If participants have not formed automatic connection between meaning, sound and form yet, it would cost them more cognitive resources to write characters, which may lower their intelligibility scores compared to *pinyin* users. The first author witnessed a participant stuck with a character and miss the remaining part of a complete sentence. Additionally, individual's attitudes and strategies are different. After missing some words from the recording, some were more "risky" and would try their best to recall and guess what it might be and wrote them down, while others may be very "conservative" (frustrated as well) and gave up the whole sentences. Future studies are expected to take all these individual differences into consideration.

Another limitation or a confounding factor is the way intelligibility is measured. The orthographic transcription measure was carried out in such a way that if the answers were in *pinyin*, only by transcribing all segmental (consonants and vowels) and suprasegmental (tones) components of a syllable correctly, can they be treated

as correct. We observed that some participants did not write tone marks, but all consonants, vowels, and translations were correct. They lose that intelligibility score for doing so. However, we do not know whether it was because they just forgot the tone marks or they did not recognize the tones. Since it is common to see L2 Mandarin learners ignore the tone marks when writing Chinese *pinyin* because of the lack of suprasegmental counterparts in their native language English, it is possible that in this study, they already recognized the tones and understood the sentences, but just forgot to write down tones. If it was the case, could the incorrect tones only be treated as typos, like misspellings in English, and credits should not be deducted. Unfortunately, we had no idea of which scenario out of the two lead to the lack of tone marks in some sentences. As a result, we adopted a more stringent and consistent measure and deducted points for those cases without tone marks. Future studies may require the testers to monitor participants' response and remind them to always include tone marks when transcribing in *pinyin* to avoid the potential ambiguity in intelligibility measurement.

This study expands previous studies on Mandarin intelligibility by focusing on L2 Mandarin learners across proficiency levels. Future studies are warranted to further examine the possible interaction of flat F0 and noise, and the chance of achieving closer intelligibility to native speakers, by including L2 learners of various proficiency levels and increasing the sample size. We could further explore at what advanced proficiency level or threshold L2 learners can recognize and understand the flattened sentences in the quiet environment as native speakers do, namely the issue of ultimate attainment in L2 intelligibility.

As argued in the discussion part, secondary cues like amplitude, duration, and acoustic boundaries may assist listener's intelligibility when sentences are flattened, especially in quiet environment. We have yet to know to what extent L2 learners may utilize these cues and what are their relationships with intelligibility. More studies are needed to explore L2 learners' developing competence of using secondary cues. In terms of semantic context, although we controlled sentence semantic variations and make it constant by statistical measures to reduce total errors, we still do not know how it impacts L2 learners' speech intelligibility in different semantic context. Future studies can examine whether normal sentences and wordlist sentences make a difference to intelligibility judgment.

## 6 Concluding Remarks

This study examined the effects of fundamental frequency, listening environment, and proficiency levels on the intelligibility of Mandarin Chinese by L2 learners. The findings revealed that flattened F0, background noise, and lower proficiency levels all lead to the decrease in intelligibility when holding semantic context constant. However, no interactions were found among the three factors, which is not consistent with previous finding on native Mandarin speakers. The hypothesis on the difference



of the intelligibility of flat F0 speech and natural F0 speech in quiet and noise environments for advanced learners were not borne out. Different from native speakers, L2 Mandarin learners did not understand the flat F0 and natural F0 sentences equally well in the quiet environment. As a matter of fact, the intelligibility of flat F0 sentences was lower for L2 learners across proficiency levels. Several accounts were proposed for the non-significant interactions and discrepancy between native speakers and L2 learners, such as the underdeveloped capability for the utilization of semantic contexts, lack of knowledge of phonetic/phonological variations and phonotactic constraints, and not attending to secondary cues, such as amplitude, duration, and acoustic boundaries.

This study contributes to our understanding of intelligibility from the perspective of second language learners of a tonal language and supports the importance of tone accuracy and diversifying L2 learners' linguistic input in Chinese pronunciation teaching and learning. Future studies should incorporate larger sample size and more advanced L2 Mandarin learners to explore the possibility of ultimate attainment in L2 intelligibility.

**Acknowledgements** We acknowledge the help of Shuang Yin and Timothy E. Moore of UConn's Statistical Consulting Services.

## Appendix 1. Sentences List

Sentence type	Chinese sentences	English translations
Practice 1	他是美国人。	He is American
Practice 2	我爸爸是律师。	My father is a lawyer
Filler 1	时间过得太快了!	Time flies!
Filler 2	我上个星期到加拿大去了。	I went to Canada last week
Filler 3	谢谢你百忙之中还抽空来看我。	Thank you for visiting me even when you are very busy
Filler 4	打太极拳的都是中老年人。	Those who play Tai-Chi are all old people
Filler 5	瑞士是个富有的国家。	Switzerland is a weather country
Target 1	我非常喜欢北京动物园。	I really like Beijing Zoo
Target 2	这是很久以前的事了。	This is the issue long time ago
Target 3	中国总共有几百种方言。	China has hundreds of dialects
Target 4	他对民国时期的文学特别感兴趣。	He is very interested in the literature of the Republic China
Target 5	爱情要紧还是面包要紧?	Love matters or bread matters?
Target 6	爱笑的人活得更长。	Who loves laughing lives longer

(continued)

(continued)

Sentence type	Chinese sentences	English translations
Target 7	很多事情说起来容易做起来难。	Many things are easy to say but hard to do
Target 8	请把今天的功课交给老师!	Please give today's homework to the teacher!
Target 9	他的护照被偷走了。	His passport was stolen
Target 10	红烧牛肉很好吃。	Braised beef is very delicious
Target 11	香港和澳门使用繁体字。	Hong Kong and Macau use traditional characters
Target 12	墨西哥在美国的南边。	Mexico is to the south of America
Target 13	法国是世界上最好的香水。	French has world's best perfume
Target 14	马友友是一位非常著名的音乐家。	Yoyo Ma is a well-known musician
Target 15	美国老一代的华人,大部分是从广东来的。	Old generation Chinese American mostly come from Guangdong
Target 16	在中国,孩子一定要听父母的话。	In China, children must heed what their parents say
Target 17	我有很多朋友。	I have a lot of friends
Target 18	今天天气很糟糕。	Today's weather is very terrible

## Appendix 2

Estimated marginal means for noise from model without interactions.

Noise	Emmean	SE	Df	Lower.CL	Upper.CL
No	0.66	0.06	18.60	0.54	0.77
Yes	0.49	0.06	18.83	0.37	0.61

Estimated marginal means for flat tone from model without interactions.

Flat Tone	Emmean	SE	Df	Lower.CL	Upper.CL
No	0.64	0.06	18.77	0.53	0.76
Yes	0.51	0.06	18.65	0.39	0.62

## References

- Benichov, J. C., Cox, L. A., Tun, P., & Wingfield, A. (2012). Word Recognition Within a Linguistic Context: Effects of Age, Hearing Acuity, Verbal Ability, and Cognitive Function. *Ear and Hearing, 33*(2), 250–256.
- Binns, C., & Culling, J. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *the Journal of the Acoustical Society of America, 122*(3), 1765–1776.
- Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from <https://www.praat.org>
- Chen, F., Wong, L., & Hu, Y. (2014). Effects of lexical tone contour on Mandarin sentence intelligibility. *Journal of Speech, Language, and Hearing Research : JSLHR, 57*(1), 338–345.
- Chen, H. C. (2011). Judgments of intelligibility and foreign accent by listeners of different language backgrounds. *Journal of Asia TEFL., 8*, 61–83.
- Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *J. Verbal Learn. Verbal Behav., 19*, 297–315.
- Corrette, R. (2012–2020). Praat Vocal Toolkit. <https://www.praatvocaltoolkit.com>
- Craig, C. H., Kim, B. W., Rhyner, P. M., & Chirillo, T. K. (1993). Effects of Word Predictability, Child Development, and Aging on Time-Gated Speech Recognition Performance. *Journal of Speech and Hearing Research, 36*(4), 832–841.
- Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20*(1), 55–60.
- Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech, 40*(2), 141–201.
- Derwing, T., & Munro, M. (2005). Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly, 39*(3), 379–397.
- Dubno, J., Ahlstrom, J., & Horwitz, A. (2000). Use of context by young and aged adults with normal hearing. *the Journal of the Acoustical Society of America, 107*(1), 538–546.
- Fernald, A. (2001). *Making use of semantic context in early language understanding*. Minneapolis, MN: Paper Presented at the Society for Research in Child Development.
- Jenkins, J. (2000). *The Phonology of English as an International Language: New models, New Norms, New Goals*. Oxford: OUP.
- Kirkpatrick, A., Deterding, D., & Wong, J. (2008). The international intelligibility of Hong Kong English. *World Englishes, 27*(3–4), 359–377.
- Kubler, C. (2013). *Intermediate Spoken Chinese: A Practical Approach to Fluency in Spoken Mandarin*. Vermont: Tuttle Publishing.
- Kubler, C. (2017). *Basic Mandarin Chinese: Speaking & listening*. Vermont: Tuttle Publishing.
- Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35*(4), 451–453.
- Laures, J., & Bunton, K. (2003). Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions. *Journal of Communication Disorders, 36*(6), 449–464.
- Laures, J., & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research: JSLHR, 42*(5), 1148–1156.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass.: M.I.T. Press.
- Levis, J. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation (Cambridge Applied Linguistics) (Cambridge Applied Linguistics)*. Cambridge: Cambridge University Press.
- Li, N., & Loizou, P. (2008). The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise. *the Journal of the Acoustical Society of America, 124*, 3947–3958.
- Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech, 47*(2), 109–138.

- Liu, Y., Yao, D., Bi, N., Ge, L., & Shi, Y. (2016). *Integrated Chinese: Zhong wen ting shuo du xie (Volume 1 & 2)*. Boston, MA: Cheng & Tsui.
- Maassen, B., & Povel, D. (1984). The effect of correcting fundamental frequency on the intelligibility of deaf speech and its interaction with temporal aspects. *the Journal of the Acoustical Society of America*, 76(6), 1673–1681.
- Miller, S., Schlauch, R., & Watson, P. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *the Journal of the Acoustical Society of America*, 128(1), 435–443.
- Munro, M., & Derwing, T. (1999). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, 49, 285–310.
- Patel, A. D., Xu, Y., & Wang, B. (2010). The role of F0 variation in the intelligibility of Mandarin sentences. *Proceedings of Speech Prosody 2010*, Chicago, IL
- Pichora-Fuller, M., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *the Journal of the Acoustical Society of America*, 97(1), 593–608.
- Pickering, L. (2006). Current Research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics.*, 26, 219–233.
- Sheldon, S., Pichora-Fuller, M., & Schneider, B. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *the Journal of the Acoustical Society of America.*, 123(1), 489–499.
- Smith, L., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World English*, 4, 33–342.
- Wang, J., Shu, H., Zhang, L., Liu, Z., & Zhang, Y. (2013). The roles of fundamental frequency contours and sentence context in mandarin Chinese speech intelligibility. *the Journal of the Acoustical Society of America*, 134(1), 91–97.
- Wang, W.S.-Y. (1973). The Chinese Language. *Scientific American*, 228(2), 50–60.
- Watson, P., & Schlauch, R. (2008). The Effect of Fundamental Frequency on the Intelligibility of Speech with Flattened Intonation Contours. *American Journal of Speech-Language Pathology*, 17(4), 348–355.
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25–47.
- Wiener, S. (2017). Changes in Early L2 Cue-Weighting of Non-Native Speech: Evidence from Learners of Mandarin Chinese. *INTERSPEECH*.
- Xu, G., Zhang, L., Shu, H., Wang, X., & Li, P. (2013). Access to lexical meaning in pitch-flattened Chinese sentences: An fMRI study. *Neuropsychologia*, 51(3), 550–556.
- Yang, C. (2016). *The acquisition of L2 Mandarin prosody: From experimental studies to pedagogical practice* (Bilingual processing and acquisition; v. 1). Amsterdam: John Benjamins Publishing Company.
- Yang, C. (2020). Teaching Chinese Intonation and Rhythm. In C. Shei, M. E. McLellan Zikpi, & D. Chao (Eds.), *The Routledge handbook of Chinese language teaching* (pp. 180–194). London: Routledge.
- Zhou, H., Li, Y., Liang, M., Guan, C., Zhang, L., Shu, H., & Zhang, Y. (2017). Mandarin-Speaking Children's Speech Recognition: Developmental Changes in the Influence of Semantic context and F0 Contours. *Front. Psychol.*, 8, 1090.