English lexical stress perception by Beijing and Changsha listeners at different proficiency levels:

support or hindrance from vowel reduction, duration, and intonation

Zhou Wangqian, You Yilin

Introduction

- The L1 phonological system, which begins to take shape as early as in the first year following birth, may either make it difficult or easier for the listener to distinguish between sounds that are native to other people (Werker & Tees, 1984).

- **Lexical stress** can be defined as the relative prominence given to a syllable within a word (Teschner & Whitley, 2004). A stressed syllable has greater respiratory energy than unstressed ones, which typically results in a higher fundamental frequency (f0), longer duration, and higher intensity; a non-schwa vowel in a stressed syllable can be reduced to a schwa if stress is shifted away, which, for example, serves to contrast the noun CONtrast /ˈkɑːntræst/ and the verb conTRAST /kənˈtræst/ (stress indicated by capitalisation) (Ladefoged & Johnson, 2015).

Do L2 learners of non-stress languages process stress as native speakers do?

- An L1 Chinese speaker often has difficulty producing or perceiving, native-like English lexical stress due to factors like L1 prosodic transfer effects (Qin et al., 2017; Zhang & Francis, 2010; Zhang et al., 2008).
    - Yu and Andruski (2010) have reported that Mandarin listeners primarily attend to f0 (and duration as a secondary cue) but tend to **disregard the vowel quality** differences.
    - Meanwhile, some related studies have shown that **Mandarin listeners rely more heavily on vowel reduction** than f0, intensity, and duration (Chrabaszcz et al., 2014; Zhang, et al., 2010).
    - Zhang, et al. (2010) reported native-like performance by Mandarin listeners on the basis that both English and Mandarin listeners **combined vowel quality with duration** as a cue to the recognition of stress.

An exception?

- Choi et al. (2019): a remarkable Cantonese advantage in English stress discrimination — Cantonese listeners discriminated English stress more accurately than did English listeners in a falling pitch accent pattern where fundamental frequency (f0) was effective for cueing stress. The Cantonese perceptual advantage observed is the result of higher f0 acuity due to lexical tone experience. — **the OPERA hypothesis** (Patel, 2011, 2014)

- However,

  - In Choi (2021), while the Cantonese listeners outperformed English native speakers, the scores of musical pitch perception did not differ significantly between the two groups;

  - Hwang et al. (2022) found no similar advantage for Chinese listeners in the interpretation of English contrastive prosody.

Vowel reduction

- Choi, et al. (2019): neglect of important phonological, segmental cues such as vowel reduction, which is argued by many (eg. Sluijter & Van Heuven, 1996) to be a stronger cue relative to f0 for native speakers and even prioritised by Cutler (1986) over all other prosodic information in differentiation of English stress.

- In other words, the non-native advantage might be an occasional phenomenon found only in contrived experimental settings, and is thus of little use in real-life conversations where the vowel quality difference is often present.

Intonation

- A follow-up study, Choi (2022):
  - the two groups were comparable in discerning lexical stress with a falling intonation pattern (such as a statement intonation);
  - Cantonese listeners were **adversely affected by the rising intonation** pattern (such as a yes/no question intonation), while English listeners were not subject to its influence;
  - the full/reduced vowel distinction facilitated Cantonese listeners' discrimination between differently stressed English words only in the rising pitch accent context.
- Ou's (2016): Mandarin speakers performed similarly to native speakers in contexts with falling pitch contours but worse than English speakers in contexts with rising pitch contours.

Beijing Mandarin

- Li and Grigos (2018): compared with Cantonese speakers, Mandarin speakers may benefit from their use of **neutral tone** and demonstrate a better ability perceiving and producing non-native lexical stress.
    - According to Duanmu (2004, 2007), syllables with neutral tones, commonly seen in two or more syllable structures, are shorter, audibly less loud, and less intense. These two metrical structures are used to distinguish minimal pairs like dōng-xī (full-full, meaning 'East-West') and dōng-xi (full-reduced, meaning ‘something’).
- Changsha stress: the significant impact of syllable duration on perceiving the metrical structure in Changshanese, with f0 being secondarily helpful, intensity being of least importance, and vowel reduction being absent in Changshanese.

Proficiency level

- Few studies have examined the relationship between English proficiency and stress perception of L2 English learners. Existent studies either use bilinguals with advanced English proficiency (e.g., Choi, 2022; Choi et al., 2019), or with participants of similar proficiency level (e.g., Guo & Chen, 2022; Lin et al., 2014; Wang, 2008).
    - Some studies have found that learners with high proficiency in L2 can still have difficulty in recognizing stress (Dupoux et al., 2008; Wang, 2008).

Research objectives

- The current study aimed to test the effects of three acoustic/segmental correlates, f0, duration, and vowel reduction on non-native perception of English stress. Intensity was not chosen because it is the least salient cue to stress in BJ and CS dialect (Guo & Chen, 2022), and therefore L1 transfer in L2 perception, if any, may be difficult to detect.

- Also of concern were the effect of proficiency level and cross-dialectal (Beijing and Changsha) variation in terms of stress cue deployment.

- The current study hypothesised that advanced learners performed alike while less proficient listeners differ in their use of cues across dialectal groups.

Test stimuli

	Trochaic (noun)	Lambic (verb)
No vowel reduction	PERmit /'pəˠmɪt/	perMIT /pəˠ'mɪt/
	INsert /'ɪnsɜˠt/	inSERT /ɪn'sɜˠt/
	SUSpect /'səspekt/	susPECT /səs'pekt/
	IMport /'ɪmport/	imPORT /ɪm'port/
	INsult /'ɪnsʌlt/	inSULT /ɪn'sʌlt/
	INcrease /'ɪnkris/	inCREASE /ɪn'kris/
Vowel reduction involved	CONduct /'kɑndʌkt/	conDUCT /kən'dʌkt/
	OBject /'ɑbdʒɪkt/	obJECT /əb'dʒekt/
	PROject /'prɑdʒekt/	proJECT /prə'dʒekt/
	REcord /'rekəˠd/	reCORD /rɪ'kɔrd/
	CONflict /'kɑnflɪkt/	conFLICT /kən'flɪkt/
	CONtest /'kɑntest/	conTEST /kən'test/

Two intonation patterns

- 'She said conFLICT.'
- 'She said <u>conFLICT</u>?')

Two male speakers

- Speaker A, aged 22, Toronto, Canada.
- Speaker B, aged 27, Tennessee, US.

Duration manipulation

- Equalised syllable duration
- Naturally occurring duration normalised to 600 ms/word

Participants

- 33 native Chinese listeners (mean age=22.4 years; SD=1.87 year) were recruited in the perception task. Among them, there are 17 native speakers of Beijing dialect (5 males and 12 females; mean age=22.0 years; SD=2.06 years) and 16 of Changsha dialect (9 males and 7 females; mean age=22.8 years; SD=1.55 year)

- For those who provided a TOEFL or IELTS score, we grouped those with more than TOEFL 98 and IELTS 7.0 (included) as high-level achievers; for those with a CET-4 or CET-6 score, we grouped those with more than CET-6 596 as **high-level proficiency group**, and those with only CET-4 score or are below CET-6 596 as **low-level group**.

    - This categorization is based on the Common European Framework of Reference for Languages (CEFR) and China's Standards of English Language Ability (CSE), where we categorized those with B2 level or below into the low proficiency group, and those with C1 level or above into high proficiency group. Standards for converting different standardized test scores into CEFR scale are extracted from test official websites (ETS, 2023; IELTS, 2023) as well as from Jin et al. (2022).

AX discrimination task

- In each trial, the pair of stimuli was automatically played in sequence with an inter-stimulus silence of 400 ms. The Canadian speaker's audio was always paired with an American speaker's audio, and vice versa.
- After hearing the tokens, participants were asked to determine if the word sequence they heard was composed of two differently stressed words or not. They responded by pressing the 'F' button on the keyboard to indicate 'same' or pressing the 'J' button to indicate 'different'.
- A total of **384 formal trials** were carried out (12 pairs * 4 conditions * 8 combinations), from which the correct responses, incorrect responses, and response time were collected.
- Four **attention check trials** were embedded in the four sections of the stress discrimination (AX) task. In these four trials, identical tokens produced by the same speaker were paired, so the participants should not fail in these trials. No participant was excluded from the data for analysis as they all achieved above-chance (>50%) performance in the attention check.
- Defining a hit as 'the correct discrimination of a minimal pair' and a false alarm as 'the designation of an identical pair as different', hit rates and alarm rates were calculated respectively. Raw data was then converted into the **sensitivity index (d')** before analysis since from the signal detection theoretical viewpoint d' cancels the effect of bias and indexes perceptual sensitivity only (MacMillan & Creelman, 2005; Choi et al., 2019).

Statistical analyses

- To examine the effects of different acoustic cues on stress discrimination ability by different groups of participants, a **five-way mixed ANOVA** was conducted on **sensitivity index (d')** with **pitch accent** (rising and falling), **vowel reduction** (full and reduced), and **duration** (equalised and naturally-occurring) as the within-subject factors, **L1 group** (BJ and CS) and **proficiency** (intermediate and high) as the between-subject factors.

- To supplement the main findings, a **five-way mixed ANOVA** was conducted on response time with **pitch accent** (rising and falling), **vowel reduction** (full and reduced), and **duration** (equalised and naturally-occurred) as the within-subject factors, **L1 group** (BJ and CS) and **proficiency** (intermediate and high) as the between-subjects factors.

    - The result shows no significant finding in any variable or interaction ($p>0.05$).

Results

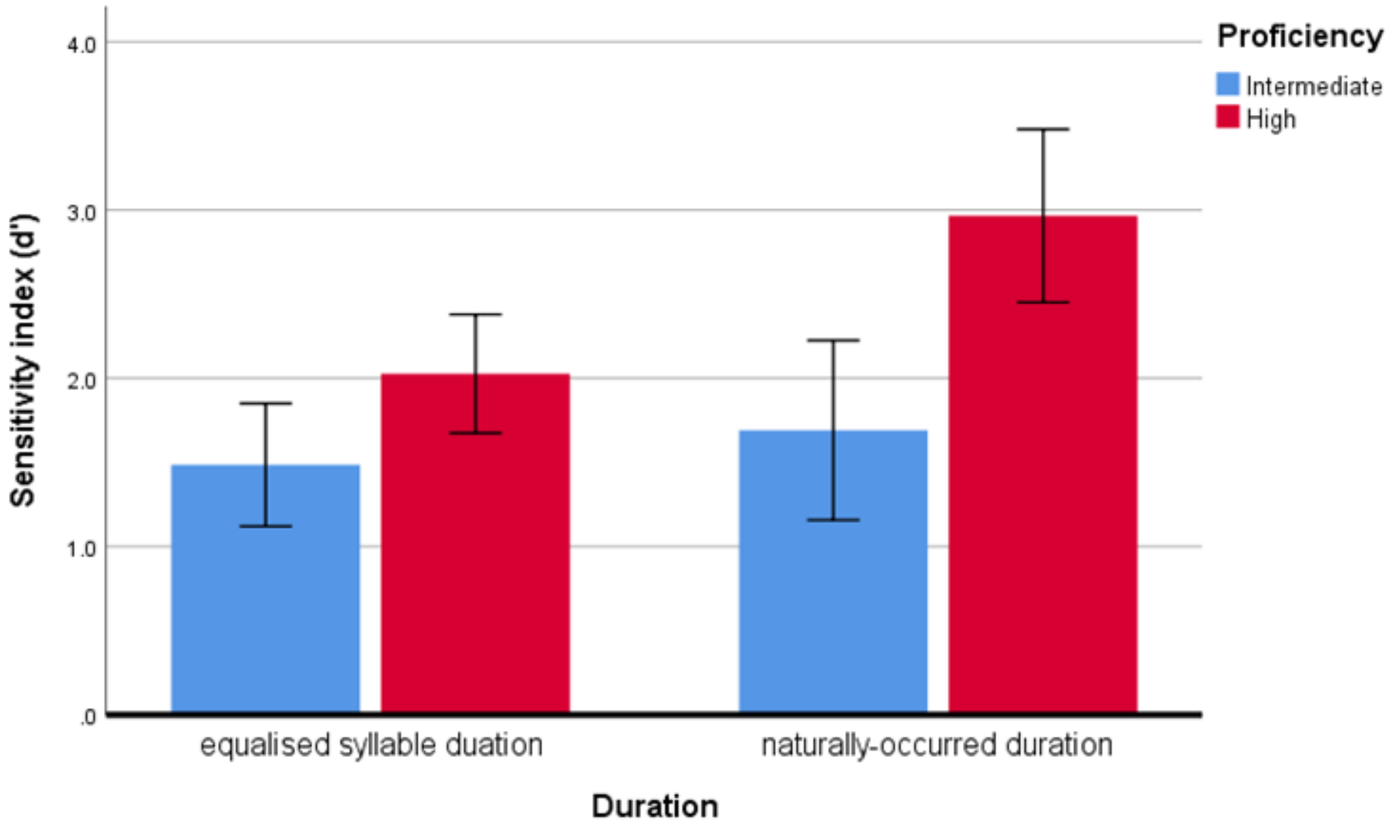
- In terms of findings in between-subject factors, **BJ and CS** participants have no significantly different performance on English lexical stress discrimination task, $F(1,29)=0.04$, $p=0.84>0.05$.

- However, **English proficiency** has significant effect on stress discrimination: participants with high-level English achievement significantly outperform the ones with intermediate-level, $F(1,29)=11.31$, $p<0.01$, $\eta p2=0.28$.

- For within-subject factors, the ANOVA showed the significant main effects of **duration**, $F(1,29)=13.82$, $p=0.001$, $\eta p2 =0.32$, Wilks' Lambda=0.68, **vowel reduction**, $F(1,29)=70.26$, $p<0.001$, $\eta p2 =0.71$, Wilks' Lambda=0.29, and **falling/rising pitch accent**, $F(1,29)=61.90$, $p<0.001$, $\eta p2 =0.68$, Wilks' Lambda=0.32.

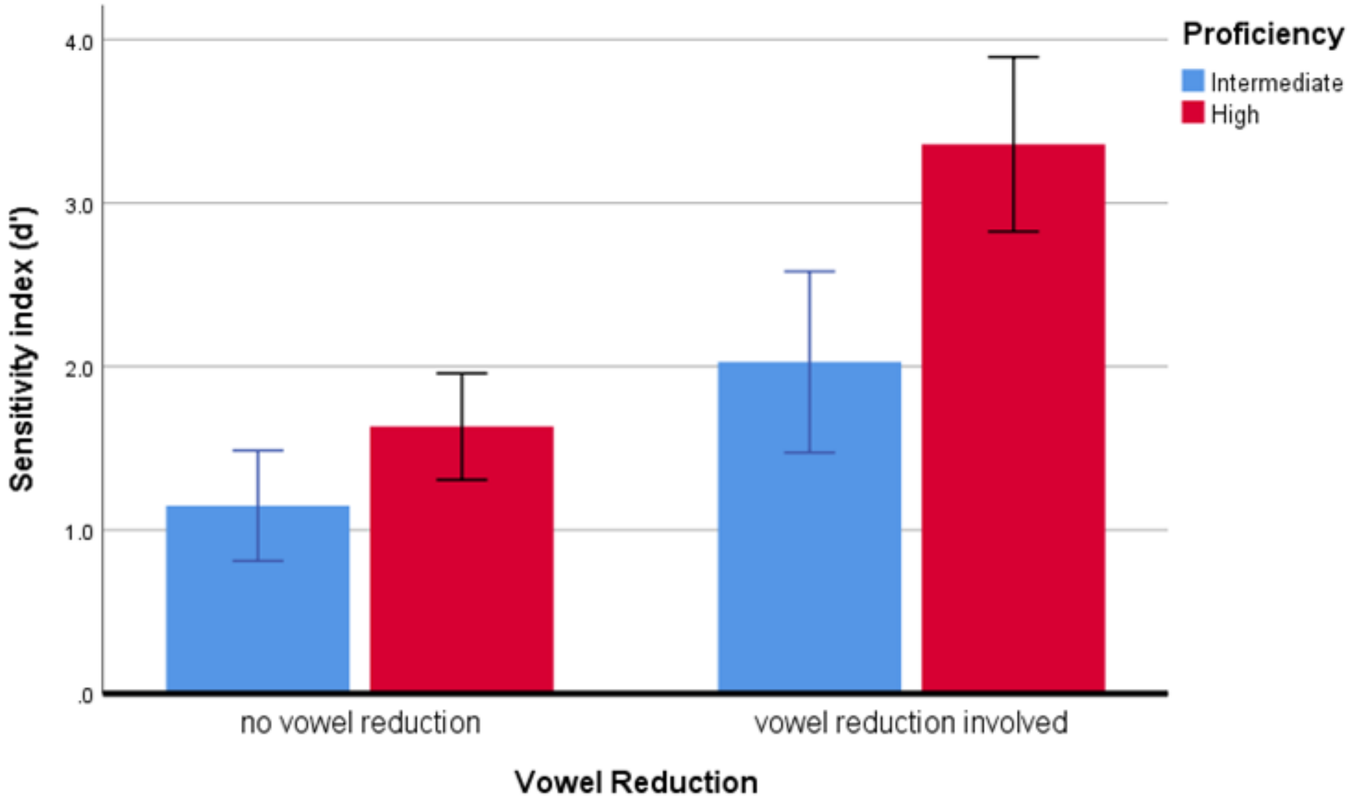
    - Listeners performed better when duration cue is presented, when the target word contains vowel reduction in the unstressed syllable, and when the word is presented in rising intonation.

A significant two-way interaction of duration and proficiency



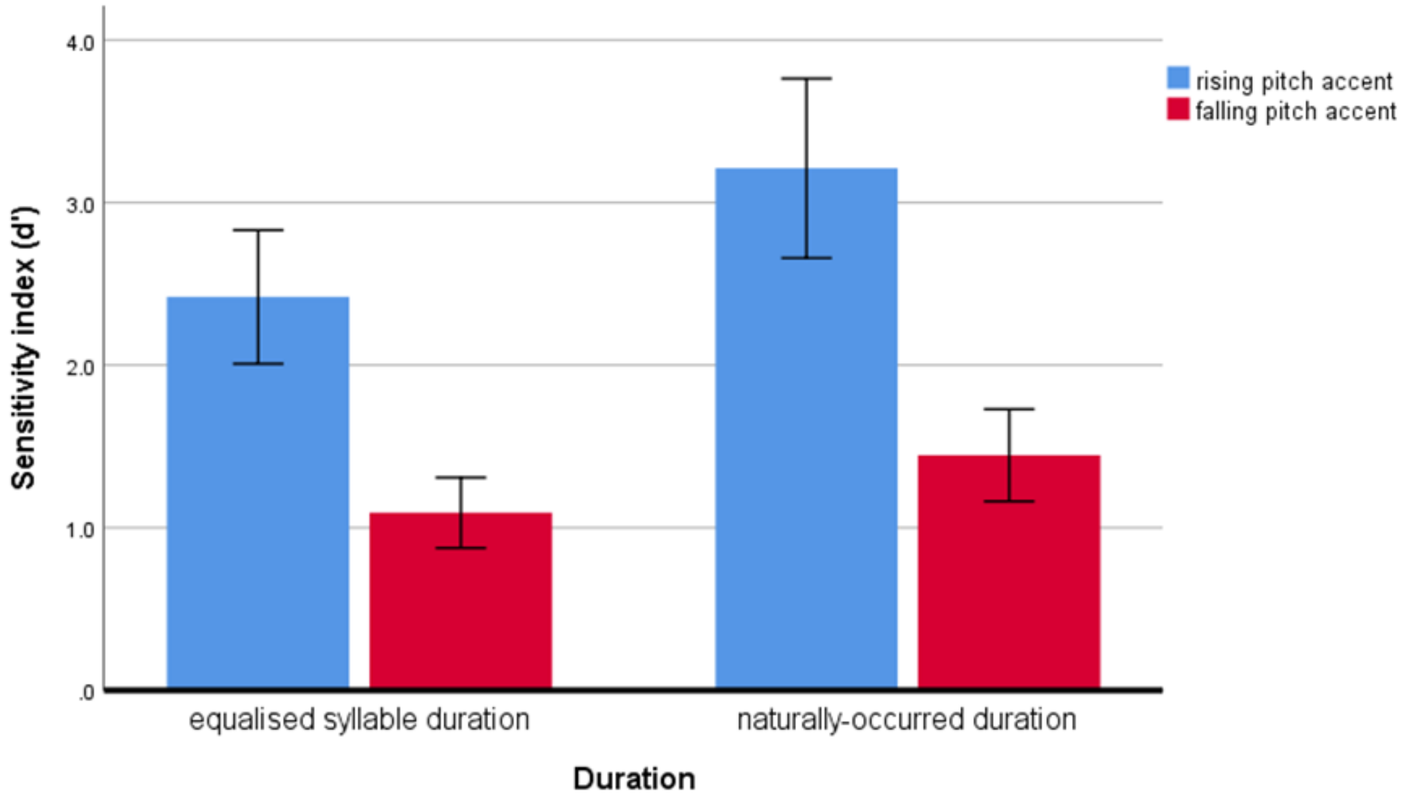
- High-level listeners performed much better when duration cue is presented, while the low-level listeners' performance didn't change much in either condition.
- This suggests that participants with higher levels of English achievement make better use of the duration cue in discriminating lexical stress.

A significant two-way interaction of vowel reduction and proficiency



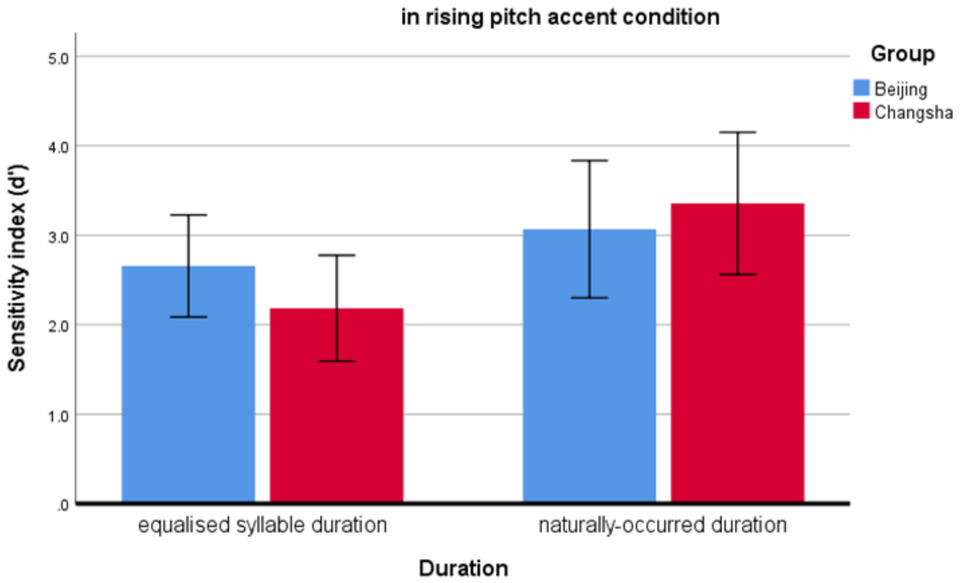
- When vowel reduction is involved, both groups significantly performed better, and the difference of d' between the two groups enlarged.
- This suggests that both intermediate- and high-level listeners are strongly influenced by the presence/absence of vowel reduction cue in the words

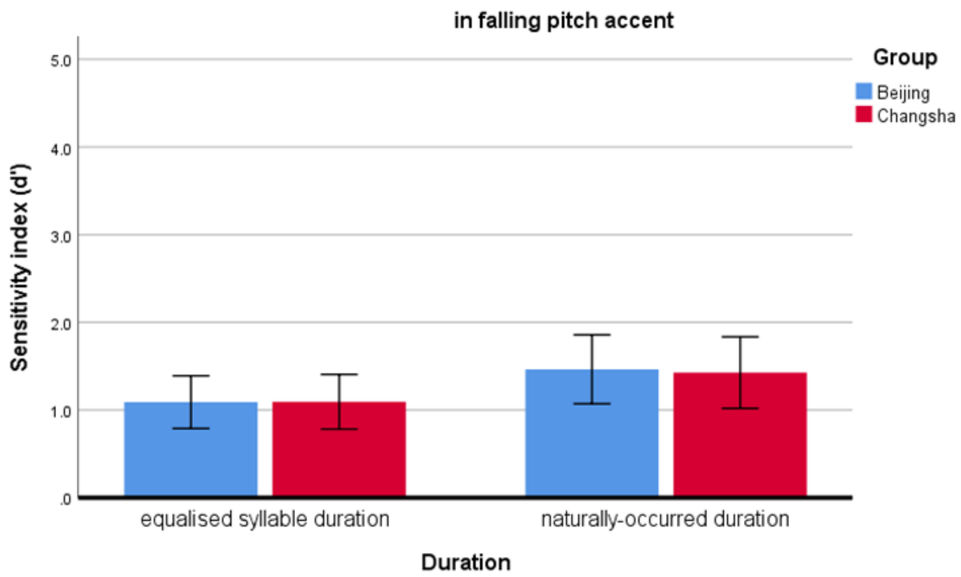
A significant two-way interaction of duration and falling/rising pitch accent



- When duration cue is not presented, rising pitch accent leads to better performance compared with falling pitch accent.
- When duration cue is presented, participants perform better in both falling and rising conditions.
- Pitch accent seems to be a more powerful cue than duration in helping listeners discriminating lexical stress.

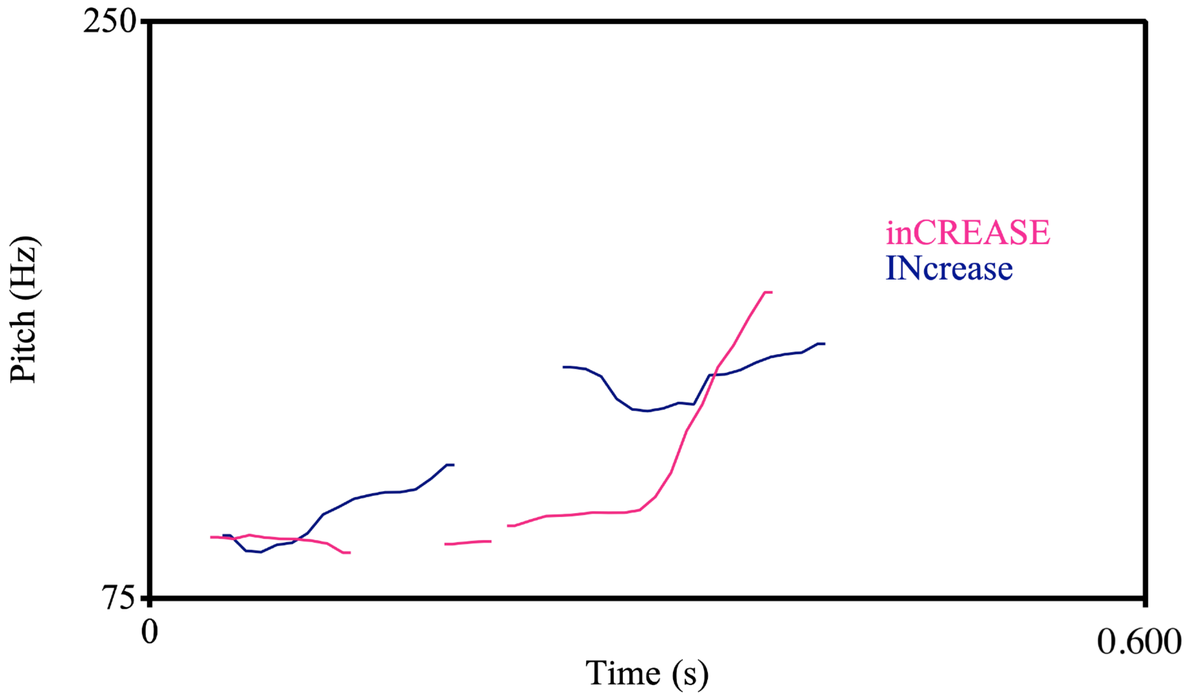
A significant a three-way interaction among duration, falling/rising pitch accent, and L1 group





- Participants generally performed better in rising pitch accent condition and when duration cue is presented.
- BJ and CS groups did not have significant differences in performance under the falling pitch accent condition.
- Under the rising pitch accent condition, the BJ group did marginally worse than the CS group when the duration cue was not removed (BJ: M=3.07, SD=0.38; CS: M=3.36, SD=0.39), but outperformed the CS group when syllabic duration was manipulated as equal (BJ: M=2.66, SD=0.28; CS: M=2.18, SD=0.29).

A facilitative effect of the rising pitch accent



- Contradictory to Ou's (2016) and Choi's (2022) research, which uncovered that the rising intonation pattern handicaps non-native perception.
- English native speakers may impose a perceptually more salient tonic accent L+H* (as compared to L* and H*) to the stressed syllable of a noun in a rising pitch accent, leaving the unstressed syllable a relatively high phrasal tone and a slightly rising boundary tone.
- Chinese-English speakers showed a high sensitivity to the acoustic salience of rising in their two languages (Ortega-Llebaria & Wu, 2021).

- The results disagreed with Yu and Andruski (2010) but supported Zhang and Francis's (2010) stance in the debate over whether or not non-native listeners can make good use of **vowel quality** differences to discern lexical stress patterns.
- **English proficiency** had a significant effect on stress discrimination ability, with high-level achievers outperforming intermediate-level participants; participants with higher level of proficiency make better use of vowel reduction and duration cues.
- The **BJ and CS groups** did not have significantly different performance, suggesting that neither of the group distinctively employed particular acoustic cues that are unique in its dialect.
    - External influence: the promotion of Mandarin in all areas and regions in China; comparable proficiency level.