Open Library of Humanities

Juncture prosody across languages: Similar production but dissimilar perception

Martin Ho Kwan Ip, The MARCS Institute, Western Sydney University, Australia; ARC Centre of Excellence for the Dynamics of Language, Australia; Integrated Language Sciences and Technology (ILST), MindCORE, University of Pennyslvania, US, mhkip@sas.upenn.edu

Anne Cutler, The MARCS Institute, Western Sydney University, Australia; ARC Centre of Excellence for the Dynamics of Language, Australia, a.cutler@westernsydney.edu.au

How do speakers of languages with different intonation systems produce and perceive prosodic junctures in sentences with identical structural ambiguity? Native speakers of English and of Mandarin produced potentially ambiguous sentences with a prosodic juncture either earlier in the utterance (e.g., "He gave her # dog biscuits," "他给她 # 狗饼干"), or later (e.g., "He gave her dog # biscuits," "他给她狗 # 饼干"). These production data showed that prosodic disambiguation is realized very similarly in the two languages, despite some differences in the degree to which individual juncture cues (e.g., pausing) were favoured. In perception experiments with a new disambiguation task, requiring speeded responses to select the correct meaning for structurally ambiguous sentences, language differences in disambiguation response time appeared: Mandarin speakers correctly disambiguated sentences with earlier juncture faster than those with later juncture, while English speakers showed the reverse. Mandarin speakers also showed higher levels of accuracy in disambiguation compared to English speakers, indicating language-specific differences in the extent to which prosodic cues are used. However, Mandarin, but not English, speakers showed a decrease in accuracy when pausing cues were removed. Thus even with high similarity in both structural ambiguity and production cues, prosodic juncture perception across languages can differ.

1. Introduction

In any language, a vast repository of words and an infinite range of sentences are based on just a handful of phonemes and syntactic rules. Spoken language, furthermore, is never produced in discrete chunks. Instead, it often contains ambiguity; words can appear embedded within other words, and sentences can carry more than one distinct meaning (consider "He gave her son glasses" versus "He gave her sunglasses"). Yet in everyday conversations, we all produce and understand most ambiguous utterances without much effort. How do we as talkers signal our meaning, and how do we as listeners deduce it? The present study addresses these questions by comparing how speakers of English and of Mandarin Chinese use prosodic cues to resolve syntactic ambiguity.

The use of prosody to signal phrasal junctures has been argued to be a universal feature of all languages (Bolinger, 1978). Formal language theory also suggests that prosody is itself a hierarchical structure that is organized in a similar way across languages (Beckman & Pierrehumbert, 1986). Different levels of prosodic constituents can govern the prominence relations and intonational, rhythmic, and pausing patterns in the speech signal (e.g., Beckman, 1996; Ladd, 1986; Liberman & Prince, 1977; Selkirk, 2003), and from birth, language learners can attend to the prosodic cues that correspond to these levels to detect relevant boundaries (Johnson, 2016). In this respect, prosodic cues to juncture can be seen as a skeletal foundation for integrating different aspects of speech during the early stages of sentence processing (Frazier, Carlson, & Clifton, 2006).

The production of prosodic juncture has been widely researched over the past decades, with remarkable similarity appearing across an impressive number of differing languages, in both tonal and temporal domains (see **Table 1** for a non-exhaustive sample). However, it is still an empirical question whether the cross-language similarities observed in production are also relevant for perception. Certainly, overall juncture cues and the way prosodic structure is organized are highly similar, even across typologically distinct languages (e.g., English and Japanese: Liberman & Pierrehumbert, 1984), but how exactly these cues are realized in phonetic effects can vary due to differences in phonological structure. For example, domain-initial articulations of voiceless aspirated stops in English, German, and Korean are more likely to be produced with longer Voice Onset Time (VOT) (Cho & Jun, 2000; Kuzla & Ernestus, 2011; Pierrehumbert & Talkin, 1992), while voiced stops in Dutch undergo VOT shortening to enhance prevoicing (Cho & McQueen, 2005). Similarly, postboundary nasals receive greater linguopalatal contact and reduced nasal airflow in French and slower lip movements and reduced nasal energy in English (Byrd & Saltzman, 1998; Cho & Keating, 2009; Fougeron & Keating, 1996), but only durational lengthening in Tamil (Byrd, Narayanan, Kaun, & Saltzman, 1997). Thus an important challenge is to examine how universal and language-specific factors interact. By adopting a cross-language approach, the present study will examine the extent to which strategies in juncture processing are shared across languages.

F0 Cues	Language	Reference(s)
• Preboundary F_0 lowering • Boundary tones • Postboundary F_0 reset	Catalan	Frota, D'Imperio, Elodieta, Prieto, and Vigáro (2007)
	Danish	Thorsen (1985)
	Dutch	Gussenhoven and Rietveld (1988); Swerts (1997)
	English	Arvaniti and Godjevac (2003); Ladd (1988); Liberman and Pierrehumbert (1984); O'Brien, Jackson, and Gardner (2014); Price, Ostendorf, Shattuck-Hufnagel, and Fong (1991); Streeter (1978)
	European Portuguese	Frota, D'Imperio, Elodieta, Prieto, and Vigáro (2007)
	French	Vaissière (1983)
	German	Grabe (1998)
	Greek	Arvaniti and Godjevac (2003)
	Italian	Frota, D'Imperio, Elodieta, Prieto, and Vigáro (2007)
	Japanese	Beckman and Pierrehumbert (1986)
	Kikuyu	Clements and Ford (1981)
	Kipare	Herman (1996)
	Korean	Jun (1998); Jun, Kim, Lee, and Jun (2004); Kim (2019)
	Mandarin Chinese	Shen (1993); Shih (2000); Xu and Wang (2001); Yuan and Liberman (2014)
	Spanish	Frota, D'Imperio, Elodieta, Prieto, and Vigáro (2007); Prieto, Shih, and Nibert (1996)
	Swedish	Swerts, Strangert, and Heldner (1996)
	Taiwanese	Peng (1997)
	Yorùbá	Laniran (1992)
Duration Cues	**Language**	**Reference(s)**
• Prosodic break/Pausing • Preboundary deceleration • Postboundary lengthening	English	Beckman and Edwards (1990); Byrd, Krivokapić, and Lee (2006); Campbell and Isard (1991); Cooper and Paccia-Cooper (1980); Goldman-Eisler (1972); Grosjean and Deschamps (1975); Grosjean, Grosjean, and Lane (1979); Harris and Umeda (1974); Hawkins (1971); Klatt (1976); Krivokapić (2007); Lehiste (1972); Price, Ostendorf, Shattuck-Hufnagel, and Fong (1991); Shattuck-Hufnagel and Turk (1998); Streeter (1978); Turk and Shattuck-Hufnagel (2007); Wightman, Shattuck-Hufnagel, Ostendorf, and Price (1992)
	Dutch	Cambier-Langeveld (1997); Quené (1992); Swerts (1997)

(Contd.)

	Estonian	Krull (1997)
	Finnish	Nakai et al. (2012)
	French	Grosjean and Deschamps (1975); Michelas and D'Imperio (2012)
	German	Kohler (1983); Männel and Friederici (2009); Männel, Schipke, and Friederici (2013); Silverman (1990)
	Greek	Katsika (2009, 2016)
	Hebrew	Berkovits (1993)
	Hungarian	Hockey and Zsuzsanna (1998)
	Mandarin Chinese	Kuang (2010); Shen (1993)
	Japanese	Liberman and Pierrehumbert (1984); Shepard (2008); Takeda, Sagisaka, and Kuwabara (1989)
	Korean	Jun (1998, 2003)
	Swedish	Lindblom and Rapp (1973); Lyberg (1977)
	Taiwanese	Peng (1997); Wang and Fon (2012)
Prosodically Conditioned Segmental Cues	**Language (Segment)**	**Reference(s)**
• Domain-initial segmental strengthening	English (Consonant Clusters; Fricatives; Nasals; Plosives; Vowels)	Byrd and Choi (2010); Cho, Lee, and Kim (2011); Cooper (1991); Dilley, Shattuck-Hufnagel and Ostendorf (1996); Fougeron and Keating (1997); Pierrehumbert and Talkin (1992)
	Dholuo (Affricates)	Degenshein and Chitoran (2004)
	Djambarrpuyŋu (Nasals; Plosives)	Jepson, Fletcher, and Stoakes (2019)
	Estonian (Nasals)	Gordon (1996)
	French (Fricatives; Nasals; Plosives; Trills; Vowels)	Christophe, Peperkamp, Pallier, Block, and Mehler (2004); Fougeron (1999); Georgeton and Fougeron (2014); Georgeton, Antolik, and Fougeron (2016); Spinelli, McQueen, and Cutler, 2003; Tabain (2003)
	German (Consonant Clusters; Fricatives; Plosives)	Bombien, Mooshammer, Hoole, Rathcke, and Kühnert (2007); Kuzla and Ernestus (2011); Kuzla, Cho, and Ernestus (2007)

(Contd.)

	Japanese (Plosives; Nasals)	Onaka (2003); Onaka, Watson, Palethorpe, and Harrington (2003)
	Korean (Plosives; Nasals)	Cho and Jun (2000); Cho and Keating (2001)
	Taiwanese (Plosives)	Hayashi, Hsu, and Keating (1999)
	Turkish (Vowels)	Barnes (2002)

Table 1: Some examples of crosslanguage consistency in prosodic juncture production.

1.1. Universal versus language-specific juncture processing

Evidence for language-universal juncture processing comes from experiments on listeners' use of prosodic cues in an unfamiliar language. For example, Carlson, Hirschberg, and Swerts (2005) asked native speakers of Swedish, American English, and Mandarin Chinese to listen to single and multi-word fragments of natural Swedish speech extracted from a radio interview. Listeners were asked to evaluate whether each fragment had been followed by a major or minor prosodic break or no break at all. Despite no knowledge of Swedish, the American participants' judgements during both single and multi-word fragments were as accurate as those of the Swedish participants. Mandarin speakers also showed comparable performance, although only in the multiword stimuli. Acoustic analyses of the stimuli revealed that boundary strength in F_0 and glottalization were correlated with judgement accuracy.

Similarly, Endress and Hauser (2010) showed that listeners can use prosodic cues to parse nonnative speech with an intonational system different from their native language. In their study, native speakers of English (a language with mostly word-initial stress) were asked to identify word boundaries in low-pass filtered speech samples produced in Turkish (a language with word-final stress). Listeners could extract words from speech at both the end and middle of intonational phrases even though they had had no prior exposure to the test language. As prosody was the only cue available, listeners must have employed a universally accessible prelexical mechanism to segment the speech input.

However, even if there is a common universal substrate that dictates the way we process prosodic junctures (thus, in both a native and nonnative language), this substrate might, over the course of development, be gradually shaped by the structure of our mother tongue, leading to strategies that are optimized for the native language. For example, languages can differ in the degree to which various juncture cues are interrelated in production. Consider the case of German and Mandarin Chinese. In German, intonational phrase boundaries are always marked by both preboundary lengthening (66.2%) and F_0 reset (74%), but rarely by pauses (38.3%)

(see Kohler, Peters, & Scheffers, 2017 for results from the Kiel corpus). This has implications for perception, and both ERP and behavioural data show that German listeners can only detect prosodic boundaries when pitch cues and preboundary lengthening co-occur (Holzgrefe-Lang et al., 2016). German listeners show a brain signature associated with boundary detection (a so-called Closure Positive Shift) even when pause duration is made uninformative, suggesting that pausing is not a crucial cue (e.g., Steinhauer, Alter, & Friederici, 1999; Männel & Friederici, 2009; Männel, Schipke, & Friederici, 2013). In addition, there is a developmental trend whereby German-learning infants lose their sensitivity to pausing cues after eight months of age (for a similar case in English, see Seidl & Cristià, 2008).

In Mandarin, in contrast, pausing is a more frequent cue to phrase boundaries (97.2%) than preboundary lengthening (less marked; Wang, Xu, & Zhang, 2019) or boundary-related pitch rises and falls (less predictable due to the presence of contour tones; Yu & Tao, 2005). Mandarin listeners are correspondingly better at detecting prosodic boundaries in sentences that only contain pausing cues, compared to sentences with only preboundary lengthening and postboundary F_0 reset (Yang, Shen, Li, & Yang, 2014). Whether only pausing or both pausing and other boundary-related cues are present does not affect Mandarin listeners' boundary detection, suggesting that pausing is the most reliable cue in Mandarin (e.g., for a similar case in Dutch and Swedish, see Sanderman & Collier, 1997; Horne, Strangert, & Heldner, 1995). Therefore, even when all juncture cues exist across a language pair (so, boundary-related pausing, pitch, and lengthening cues can all be found in German and Mandarin), listeners have developed processing preferences for different cues.

Another line of evidence for language differences comes from studies that have used sentences with ambiguous complex noun phrases and relative clauses (e.g., "Someone shot the servant of the actress who was on the balcony"), where the relative clause (RC) could be construed as modifying the NP headed by either the first noun (i.e., *servant*) or the second (i.e., *actress*). Across languages, listeners adopt different attachment bias due to variation in default prosodic phrasing (Fodor, 1998). High attachment of the RC to the NP1 is favoured in languages where speakers tend to produce a weak boundary between NP1 and NP2 and a strong boundary before the RC (e.g., French, Spanish: Cuetos & Mitchell, 1988; Zagar, Pynte, & Rativeau, 1997). Low attachment is favoured in languages where speakers tend to place a boundary after the NP1 (e.g., English, Mandarin: Kuang, 2010; Jun, 2003). Again, these findings suggest that listeners can differ due to variation in heard input. Languages vary not only in the degree to which different juncture cues are used, but also in the location of these cues.

Interestingly, however, listeners' language-specific attachment preferences can be modulated (Fernández, 2007; Teira & Igoa, 2007) or even reversed (Fromont, Soto-Faraco, & Biau, 2017) if the location of the prosodic boundary in the speech stimuli was manipulated to favour a different interpretation. Moreover, foreign language learners can adopt native-like parsing strategies in

their L2 even when these are different from their native language (L1); English learners of French, for instance, have been shown to use the appropriate French strategy (i.e., high attachment) to disambiguate RC attachment ambiguities even after learning the language for just a few semesters (Dekydtspotter, Donaldson, Edmonds, Liljestrand, & Petrush, 2008). Similarly, English learners of German and German learners of English can both produce and attend to the prosodic cues in their L2 (O'Brien, Jackson, & Gardner, 2014). There is thus certainly flexibility in the processing system for prosodic juncture.

1.2. The present study: General overview

The present study addresses the processing of prosodic juncture in a systematic manner. Although, as is clear from the above brief review, data on prosodic juncture processing has been gathered from many languages, most studies have concerned one language at a time. Even in the handful of cross-language studies, the languages under investigation may involve different prosodic realization of boundary cues (e.g., in O'Brien et al., English disambiguation involved only pitch accent, while the German disambiguation involved both pitch accent and F_0 rise), and the languages compared are often from closely related language families with similar prosodic systems.

The experiments we report here, in contrast, compare English and Mandarin, two languages that are typologically distant and have different intonation systems. Crucially, despite this, both languages allow the same kind of structural ambiguity. Consider the following examples:

(a) 爷爷 /　　给 /　　她 # 婴儿奶粉 /　　　　喝
    ye2ye5 /　gei3 /　ta1 # yi1ner2nai2fen3 / he1
    Grandpa / gave /　her # baby formula /　　to drink

(b) 爷爷 /　　给 /　　她 /　婴儿 # 奶粉 /　　　喝
    ye2ye5 /　gei3 /　ta1 /　yi1ner2 # nai2fen3 / he1
    Grandpa / gave / her /　baby # formula /　　to drink

The two sentences differ in the direct object, and as a consequence, differ in juncture location. In (a), the juncture (#) is realized earlier on in the utterance, giving a sentence with a feminine personal pronoun as the indirect object and a compound noun as the direct object. In (b), the same (segmentally identical) sentence is produced with a later boundary, after "*baby,*" so that in this case "*her*" is functioning as a possessive determiner. This ambiguity can occur in English because "*her*" can be either a possessive or an indirect object. It can also occur in Mandarin because speakers ignore the alienable versus inalienable distinction in everyday speech where the possessive particle -*de* can be omitted (Haiman 1983, 1985; Hsu, 2009). In fact, according to a large database of informal written and spoken Mandarin, almost half (45%) of associative noun phrases in Mandarin are produced without the particle (Chappell & Thompson, 1992).

The present study comprehensively examines juncture processing in these two languages, in both production and perception. For production, we address the following questions:

1. Do English and Mandarin speakers use the same prosodic cues to signal juncture in these near-identical structures?
2. To the extent that they do, are there differences in the degree to which specific juncture cues are deployed?

For perception, we ask:

1. Do English and Mandarin listeners differ in their perceptual processing of juncture in these same structures?

If juncture processing were to be universal across languages, then both English and Mandarin speakers would presumably use prosodic cues in the same way to process the intended meaning of the ambiguous utterances. However, since prior literature has reported some cross-language differences, our study may reveal differences even in this case where the syntactic structure is closely similar.

2. Experiment 1: Production

2.1. Method

2.1.1. Participants

Our participants were 24 native speakers of Australian English (M_{age} = 21.50 years; 21 females) and 24 native speakers of Mandarin Chinese (M_{age} = 27.56 years; 19 females). The English speakers were born and raised in Australia, while the Mandarin speakers were born in Mainland China and had been living in Australia for an average of two years and 10 months (range: 2 months – 9 years). We excluded additional data from one English speaker who had some disfluency in oral reading and three Mandarin speakers who grew up in Chinese-speaking communities outside of Mainland China (e.g., Taiwan). All participants were naïve to the specific purpose of the experiment.

2.1.2. Reading Passages

Our materials were three pairs of short reading passages written in English and Simplified Chinese (see **Table 2**). Each passage pair contained the same target ambiguous sentence as the last sentence in the passage. The target sentences were manipulated to have different meaning by virtue of the different storylines provided by the preceding sentences. In one version, the context would elicit production of the target ambiguous sentence with an Early Juncture, where the boundary occurred earlier in the sentence (e.g., "He gave her # dog biscuits"). In another

English	Mandarin
"He gave her dog biscuits" /hi: gæɪv hɜ: dɔg bɪskəts/	**"他给她狗饼干"** /t^{h}a$_1$ kei$_2$ t^{h}a$_1$ kou$_3$ pin$_3$kan$_1$/
***	***
Early Juncture: "He gave her # dog biscuits" /hi: gæɪv hɜ: # dɔg bɪskəts/	**Early Juncture: "他给她 # 狗饼干"** /t^{h}a$_1$ kei$_2$ t^{h}a$_1$ # kou$_3$ pin$_3$kan$_1$/
Joe's new neighbour is a little girl named Amy who lives with her grandma. Every time he walks past Amy's home, Amy would greet him and ask him for some biscuits. Usually, Joe offers her a few Danish cookies. But today, he gave her dog biscuits.	小周的新邻居住着一位小女孩叫爱玲。她和奶奶一起住。每次小周走路经过爱玲的家时，爱玲都向着他问好，还跟他要饼干吃。通常，小周都会给爱玲一些丹麦奶油饼干。可是今天，他给她狗饼干。
Late Juncture: "He gave her dog # biscuits" /hi: gæɪv hɜ: dɔg # bɪskəts/	**Late Juncture: "他给她狗 # 饼干"** /t^{h}a$_1$ kei$_2$ t^{h}a$_1$ kou$_3$ # pin$_3$kan$_1$/
Adam has just moved to Sydney from Melbourne. His new neighbour is an old lady named Gertrude. Gertrude has been living with her dog in Sydney for over ten years. Every time Adam walks past their front yard, Gertrude's dog would run towards the gate and bark at him. Usually, Adam would ignore Gertrude's dog and continue walking. But today, he gave her dog biscuits.	阿德刚从墨尔本搬到悉尼。他隔壁是一位老奶奶。老奶奶和她的狗住在悉尼已经超过十年了。每次阿德路过老奶奶的前院，老奶奶的狗就跑到门前冲着他嗷嗷叫。通常，阿德都不理老奶奶的狗就继续往前走。可是今天，他给她狗饼干。
"He saw her duck under the chair" /hi: sɔ: hɜ: dʌk ʌndɐ ðə tʃeː/	**"他看见她猫在凳子底下"** /t^{h}a$_1$ k^{h}an$_4$ tɕjɛn$_4$ t^{h}a$_1$ mau$_1$ tsai$_4$ təŋ$_4$tsɿ$_5$ ti$_3$ ɕja$_4$/
***	***
Early Juncture: "He saw her # duck under the chair" /hi: sɔ: hɜ: # dʌk ʌndɐ ðə tʃeː/	**Early Juncture: "他看见她 # 猫在凳子底下"** /t^{h}a$_1$ k^{h}an$_4$ tɕjɛn$_4$ t^{h}a$_1$ # mau$_1$ tsai$_4$ təŋ$_4$tsɿ$_5$ ti$_3$ ɕja$_4$/
Ethan and Maria go to the same primary school and they love to play hide and seek. Ethan loves to duck under tables and Maria loves to duck under chairs. The first time they played hide and seek was in the classroom. Maria was too slow to hide and Ethan quickly found out what she was doing. He saw her duck under her chair.	叶生和玛利亚是小学同学。他们喜欢玩捉迷藏。叶生喜欢猫在桌子底下。玛利亚喜欢猫在凳子底下。他们第一次玩捉迷藏是在教室里玩。玛利亚藏得太慢，叶生很快就发现她藏在哪里。他看见她猫在凳子底下。

English	Mandarin
Late Juncture: "He saw her duck # under the chair" /hiː soː hɜː dʌk # ʌndɐ ðə tʃeː/	**Late Juncture: "他看见她猫 # 在凳子底下"** /tʰa₁ kʰan₄ tɕjɛn₄ tʰa₁ mau₁ # tsai₄ təŋ₄tsɨ₅ ti₃ ɕja₄/
Lily loves her pet duck very, very much. One day, she brought her pet duck to primary school. Lily knew that it is forbidden to bring pets to school. Before her teacher, Mr. Johnson, arrived, Lily quickly hid her duck under her chair. But Mr. Johnson saw Lily's pet duck. He saw her duck under her chair.	莉莉很喜欢她的小猫。有一天，她带着她小猫一起去上学。莉莉知道学校不让带宠物去上学。在左老师到达之前，莉莉很快把小猫藏在凳子底下。可是左老师马上发现莉莉带了小猫来到教室。他看见她猫在凳子底下。
"He gave her baby milk" /hiː gæɪv hɜː bæɪbɪ mɪlk/	**"他给她婴儿奶粉"** /tʰa₁ kei₃ tʰa₁ jiŋ₁ɚ₂ nai₃fən₃/
***	***
Early Juncture: "He gave her # baby milk" /hiː gæɪv hɜː # bæɪbɪ mɪlk/	**Early Juncture: "他给她 # 婴儿奶粉"** /tʰa₁ kei₃ tʰa₁ # jiŋ₁ɚ₂ nai₃fən₃/
Sally is a self-confessed alcoholic and loves to go to the pub. One night, at her favourite pub, she was very drunk. What's more, Sally was behaving very badly. As she was asking for more beer, the bartender decided not to give her more alcohol. Instead of beer, the bartender poured baby milk in the beer bottle and hoped Sally was too drunk to notice. Indeed, Sally didn't notice at all. So he gave her baby milk.	李三丽小姐是个酒迷。她喜欢去酒巴。有一天晚上，三丽喝醉了。而且，三丽的行为很出丑。她还要继续喝酒，可是调酒师不想再给她更多酒了。调酒师把婴儿奶粉倒进酒瓶里。调酒师发现三丽没有看见酒瓶里有婴儿奶粉。所以，他给她婴儿奶粉。
Late Juncture: "He gave her baby # milk" /hiː gæɪv hɜː bæɪbɪ # mɪlk/	**Late Juncture: "他给她婴儿 # 奶粉"** /tʰa₁ kei₃ tʰa₁ jiŋ₁ɚ₂ # nai₃fən₃/
David is a teenager who works as a nanny for his neighbour, Mrs. Berry, who has a baby boy called Bob. One night, Mrs. Berry went out and left Bob in David's care. Before she went out, Mrs. Berry told David to feed Bob some porridge before he went to bed. But David later found out that there was no porridge in the cupboard. He didn't want Mrs. Berry's baby boy to go hungry. David found a carton of milk in Mrs. Berry's kitchen. So he gave her baby milk.	小伙子大伟有时候帮邻居薄阿姨看孩子。薄阿姨有个婴儿是男孩叫薄海。有一天晚上，薄阿姨要出门，让大伟照顾小薄海。薄阿姨出门前告诉大伟给小薄海睡觉前吃粥。可是大伟发现锅里已经没粥了。他不想让薄阿姨的婴儿埃饿。大伟看见在薄阿姨的厨房里有奶粉 。所以，他给她婴儿奶粉。

Table 2: Reading passages (Early versus Late Juncture versions) in English and Mandarin with IPA transcriptions. Transcriptions in English were based on the Harrington-Cox-Evans (1997) system for Australian English.

version, the same target sentence was manipulated to elicit production of Late Juncture, where the boundary occurred later in the sentence (e.g., "He gave her dog # biscuits").

The English and Chinese reading passages were highly comparable in three important ways. First, the English and Chinese ambiguous sentences, as well as the storylines, were identical in meaning, except for one minor deviation in translation in the second passage where the ambiguous sentence in English was "he saw her duck under the chair" and the sentence in Chinese was "他看见她猫在坐凳子底下" "he saw her cat/hide under the chair" (n.b., 猫 can mean either "*cat*" or "*hide*"). Second, both the English and Chinese sentences involved the same ambiguity. The Early Juncture sentences involved a feminine personal pronoun (i.e., *her*/她) before the juncture, followed by a postboundary compound noun or verb and preposition (e.g., *dog biscuit*/狗饼干; *duck under*/猫在), while in the Late Juncture sentences, the compound noun or verb became a simple noun (e.g., dog/狗; *duck*/猫) and the personal pronoun became a possessive determiner. Third, we selected target sentences involving pre- or postboundary consonant onsets that were, as far as possible, highly comparable in their manner of articulation (e.g., /d̠ɔg # b̠ɪskəts/ versus /k̠ou # p̠inkan/; /b̠æɪbɪ # m̠ɪlk/ versus /jiŋɚ # n̠aifən/).

2.1.3. Recording procedures

All participants were tested by the first author, a fluent speaker of both English and Standard Mandarin. Recordings were made inside a sound-attenuated booth at the MARCS Institute, using a Shure SM10A-CN headset microphone connected to a laptop via a Roland Quad-Capture USB audio interface. Recording sessions for each reading passage lasted for approximately five minutes and were performed individually by the participant in the presence of the experimenter. Before each session, all participants spent a few minutes reading through each of the passages by themselves to prepare. To ensure successful elicitation, the experimenter asked participants to pay careful attention to how they chose to speak in each passage, and encouraged them to speak in a way that would "really flesh out the meaning of the entire passage." Participants were also told that the study aimed to examine how speakers produce speech in everyday contexts, and they were told to try to be "as normal as possible." However, the experimenter did not give any explicit instructions to produce the relevant juncture cues in the target ambiguous sentences. Furthermore, the passages were written in plain text without any markers (such as hashtags) between phrases that could signal the designated boundaries.

After reading each passage, participants were asked a series of follow-up questions to test their comprehension of the passage (see **Table 3**). This was done to confirm that they understood the ambiguous sentences. If participants did not know the answers or answered incorrectly, they were encouraged to read the passage by themselves again, and were the given another chance to produce the passage. In such cases, only data from the latest recordings were included in our final analyses. Every participant produced all the reading passages. No participant had to reread a passage more than twice.

English	Mandarin
Early Juncture: "He gave her # dog biscuits" /hi: gæɪv hɜ: # dɔg bɪskəts/	**Early Juncture: "他给她 # 狗饼干"** /tʰa$_1$ kei$_2$ tʰa$_1$ # kou$_2$ pin$_3$kan$_1$/
Questions about Joe and Amy 1. What kind of biscuit did Joe give her today? 2. Did he give Amy some Danish biscuits? 3. Did he give Amy's dog some dog biscuits?	关于小周和爱玲的问题 1. 小周今天给她什么饼干？ 2. 小周有没有给她丹麦奶油饼干？ 3. 小周有没有给爱玲的狗吃狗饼干？
Late Juncture: "He gave her dog # biscuits" /hi: gæɪv hɜ: dɔg # bɪskəts/	**Late Juncture: "他给她狗 # 饼干"** /tʰa$_1$ kei$_2$ tʰa$_1$ kou$_3$ # pin$_3$kan$_1$/
Questions about Adam and Gertrude's dog 1.What did Adam give her dog today? 2. Did he give Gertrude any biscuits?	关于阿德和老奶奶的狗的问题 1. 阿德今天给她狗什么？ 2. 阿德有没有给老奶奶吃饼干？
Early Juncture: "He saw her # duck under the chair" /hi: so: hɜ: # dʌk ʌndɚ ðə tʃeː/	**Early Juncture: "他看见她 # 猫在凳子底下"** /tʰa$_1$ kʰan$_4$ tɕjɛn$_4$ tʰa$_1$ # mau$_1$ tsai$_4$ təŋ$_4$tsɨ$_5$ ti$_3$ ɕja$_4$/
Questions about Maria 1. Where did Maria hide? 2. Was Maria under the stairs?	关于玛利亚的问题 1. 玛利亚藏在哪里？ 2. 玛利亚有没有藏在桌子底下？
Late Juncture: "He saw her duck # under the chair" /hi: so: hɜ: dʌk # ʌndɚ ðə tʃeː/	**Late Juncture: "他看见她猫 # 在凳子底下"** /tʰa$_1$ kʰan$_4$ tɕjɛn$_4$ tʰa$_1$ mau$_1$ # tsai$_4$ təŋ$_4$tsɨ$_5$ ti$_3$ ɕja$_4$/
Questions about Lily's duck 1. Who does the duck belong to? 2. Where did Mr. Johnson see her duck?	关于莉莉的小猫的问题 1. 这只小猫是谁的猫？ 2. 左师在哪里看见莉莉的猫？
Early Juncture: "He gave her # baby milk" /hi: gæɪv hɜ: # bæɪbɪ mɪlk/	**Early Juncture: "他给她 # 婴儿奶粉"** /tʰa$_1$ kei$_3$ tʰa$_1$ # jiŋ$_1$ɚ$_2$ nai$_2$fən$_3$/
Questions about Mrs. Berry's baby 1. What is the drunken woman's name? 2. What did the bartender give Sally to drink? 3. Did the bartender give her beer with the baby milk?	关于莉莉的小猫的问题 1. 喝醉酒的女人叫什么名字？ 2. 调酒师给李三丽喝什么？ 3. 调酒师有没有给李三丽酒和婴儿奶粉？
Late Juncture: "He gave her baby # milk" /hi: gæɪv hɜ: bæɪbɪ # mɪlk/	**Late Juncture: "他给她婴儿 # 奶粉"** /tʰa$_1$ kei$_3$ tʰa$_1$ jiŋ$_1$ɚ$_2$ # nai$_3$fən$_3$/
Questions about Mrs. Berry's baby 1. What is Mrs. Berry's baby's name? 2. What did Bob drink? 3. Did Bob get any porridge？	关于薄阿姨的孩子的问题 1. 薄阿姨的儿子叫什么名字？ 2. 薄海喝了什么？ 3. 薄海有没有吃粥？

Table 3: Examples of follow-up questions in English and Mandarin Chinese.

2.1.4. Acoustic analyses

Four potential disambiguation cues were analyzed in Praat (Boersma & Weenink, 2018). These were (1) pausing, (2) pre- and postboundary vowel lengthening, (3) F_0 modification, and (4) domain-initial/postboundary segmental strengthening (see **Figure 1** for an example sentence pair in English).

For pausing, we measured the duration of each potential juncture pause, i.e., the one that would indicate the early juncture in Early Juncture sentences (P1), and the one that would indicate the late juncture in Late Juncture sentences (P2). This was done for all sentences, so both Early Juncture and Late Juncture sentences had two measures; for the sentence "He gave her dog biscuits," for example, we measured durations between "*her*" and "*dog*" and between "*dog*" and "*biscuits.*" We then compared the duration at each designated juncture across the two juncture versions. If the spectrogram showed no visible pause at any designated juncture, then a rating of zero was given.

For boundary lengthening, we compared the pre- and postboundary vowel duration of the words preceding and following the two designated junctures. There were three measures of vowel duration per sentence: of the word before the designated early juncture boundary (V1), of the word before the designated late juncture boundary (V2), and of the word after the designated late juncture boundary (V3).

For F_0, we analyzed the mean, minimum, and maximum F_0 as well as F_0 range of the same three pre- and postboundary vowels. For domain-initial segmental strengthening, we measured the durations of the voice onset time (VOT) and the nasal and affricate or fricative onsets of the words in the potential postboundary location. In English, these segmental duration measures involved one postboundary nasal (i.e., /bæɪbɪ # mɪlk/) and two VOT measures (i.e., /hɜː # dɔg # bɪskəts/ and /hɜː dɔg # bɪskəts/). In Mandarin, we measured one affricate (i.e., /tʰa mau # tsai/), one VOT (i.e., /tʰa # kou/), and two nasals (i.e., /tʰa # mau tsai/ and /jiŋ₁ɚ₂ # nai₃fən₃/).

This led to a total of 5232 measurements across the three sentence pairs in each language ([6 pause duration × 2 languages × 2 juncture versions × 24 speakers] + [9 vowel duration × 2 languages × 2 juncture versions × 24 speakers] + [36 F_0 × 2 languages × 2 juncture versions × 24 speakers] + [3 English segments × 2 juncture versions × 24 speakers] + [4 Mandarin segments × 2 juncture versions × 24 speakers]).

2.2. Results

2.2.1. General overview

The original dataset for all of the experiments in the present study is available on this open access site: https://upenn.box.com/s/n5r5ww7t47dqnywakm580axvujh03amk. Acoustic results for each prosodic cue in Experiment 1 were analyzed. For each prosodic cue, we examined
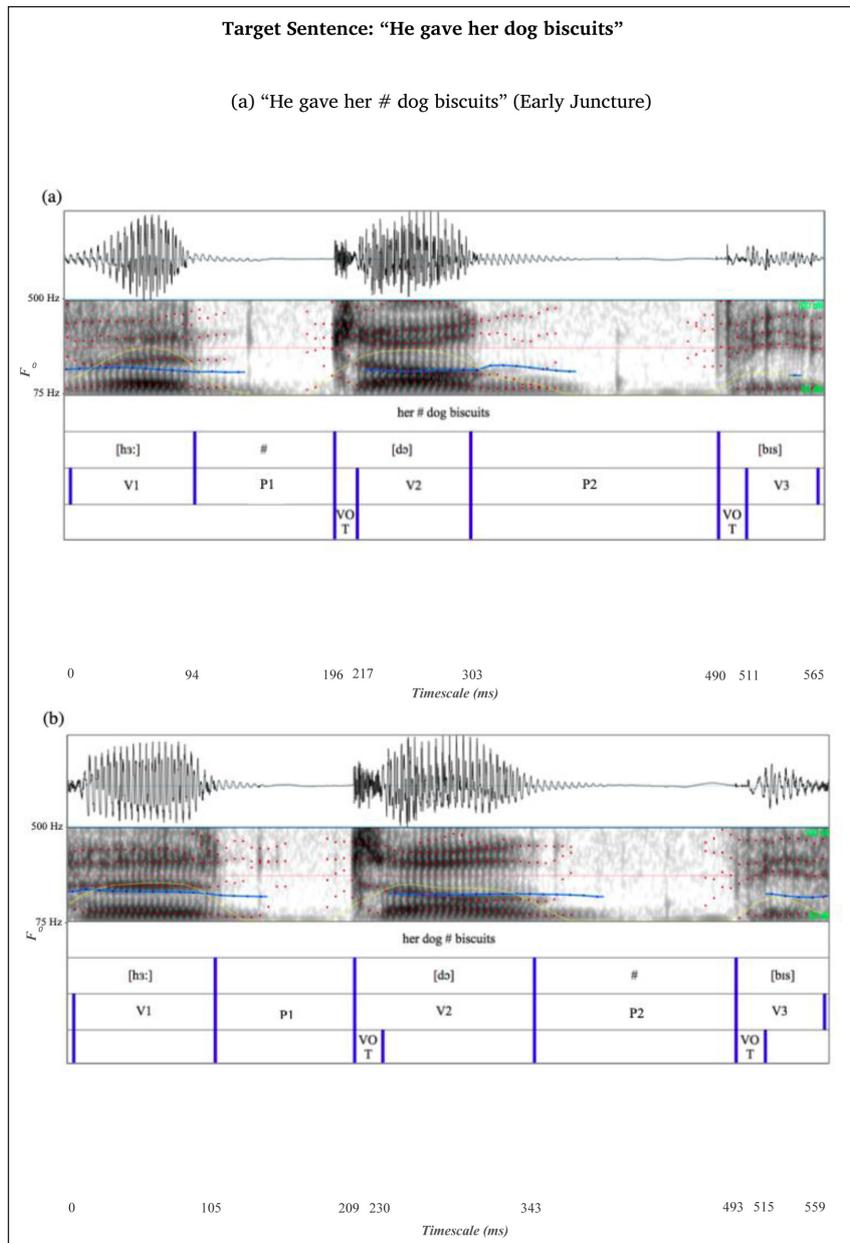
Figure 1: Waveforms and pitch and amplitude contours of an example sentence pair in **(a)** Early Juncture and **(b)** Late Juncture versions. For both versions, we measured the pause duration of the juncture locations that would indicate the designated early juncture (P1) and the designated late juncture (P2). Pre- and postboundary vowel durations (V1, V2, and V3) were also measured. As revealed in the annotations, V1 indicates the preboundary vowel duration before the designated early juncture, while V2 indicates the preboundary vowel duration before the late juncture. V3 is the postboundary vowel duration after the designated late juncture. F_0 measures (mean, minimum, maximum, and range) were calculated from the three pre- and postboundary vowels. Acoustic measures of domain-initial segmental strengthening (i.e., VOT, nasal, or fricative duration) were measured wherever a postboundary word began with a consonant word onset.

whether both languages showed similar patterns of production difference between the Early and Late Juncture sentences. For F_0, a small proportion of the utterances (7.25% of the English data and 2.47% of the Mandarin data) had to be excluded due to octave errors arising from creaky voice production. Using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015, version 1.1-7), Linear Mixed Effects regression (LMER) models were constructed with the maximal random effects justified by the data (Barr, Levy, Scheepers, & Tily, 2013). Prior to the analyses, data skewness of each prosodic cue was first examined based on visual inspections of quartile-quartile plots. We observed that the pause duration after the designated early juncture (i.e., P1) formed a skewed distribution. Skewed distributions were also revealed in the preboundary vowel durations of the designated early junctures (i.e., V1), as well as in the mean F_0 data for the pre- and postboundary vowels of the designated late junctures. These data were therefore transformed prior to the analyses.

It is important to note that for both our production and perception experiments, we chose not to apply log or square root transformations because some of the raw data contained zero or negative values; the zero values in the production data came from instances where participants did not produce a pause, and negative values can be found in subsequent perception experiments where participants correctly disambiguated the sentence before the sentence offset, which occurred in less than 10% of the total correct responses. A common practice is to add a constant value to the datapoints before transforming the data, so that all datapoints become non-zero positive values. However, we refrain from adopting this practice; albeit common, this practice is arbitrary and problematic because it can inflate both Type I and Type II errors as a function of the added constant value (see Feng et al., 2014 for evidence from simulation studies). All of the skewed raw data from our study were therefore transformed using the Yeo-Johnson transformation procedure (Yeo & Johnson, 2000). This was done using the yeojohnson function from the recently updated bestNormalize package in R (version 1.6.1, June, 2020; Peterson & Cavanagh, 2019). Importantly, the Yeo-Johnson procedure is an extension of the Box-Cox inverse transformation procedure in that it handles both positive as well as zero and negative values. Compared to other methods of data transformation (e.g., log transformation), transformations based on the Box-Cox procedure have been argued to be better suited for psycholinguistic data (Lo & Andrews, 2015), and provide a better approximation to normal-distribution and homoscedasticity assumption for linear models (Balota, Aschenbrenner, & Yap, 2013). For the readers' convenience, all of the means, standard deviations, fixed effects estimates (β), and standard errors reported in the main text and figures will be raw values (e.g., in milliseconds).

For each juncture cue, we used as the starting point a baseline model that included by-participant and by-item random intercepts as well as by-participant and by-item random slopes for the effect of juncture version. Juncture version, language, and language by juncture version interaction were added as fixed effects predictors in a step-wise fashion and these

models were compared with the baseline model. Model fit was determined using chi-squared tests of model log-likelihood based on the *p*-values of the chi-squared tests and/or differences in the Akaike Information Criterion (AIC), with the latter being more useful in cases where the complexity of the model cannot be justified by the additional variance explained (see Shaw et al., 2018). Predictors that did not yield significant improvement in the model comparisons were dropped before additional predictors were added. Leave-one-out comparisons were used to ensure that each predictor yielded a significant gain in log likelihood with all other predictors in the model. Planned comparisons following significant interaction effects were carried out using the emmeans package with Tukey-adjusted *p*-values (Lenth, 2020). All fixed effects were coded with mean-centred contrast codes.

2.2.2. Prosodic cues to juncture

Early Pause Duration (P1). We first analyzed the pause duration of the designated early juncture across the two juncture versions. The addition of juncture version showed only a weak/marginally significant improvement in model fit log likelihood (χ^2 (1) = 3.81, p = .051; β = –61.41, SE = 23.23, t = −2.19); speakers produced a significantly longer pause at the designated early juncture (P1) in Early Juncture sentences (M = 71.86 ms, SD = 67.74 ms) compared to the same cue in the Late Juncture sentences (M = 54.90 ms, SD = 41.65 ms). However, there was also a significant main effect of language (χ^2 (1) = 55.06, p < .001; β = –66.01, SE = 11.11, t = −9.67); English speakers produced longer pause after the designated early juncture (M = 81.02 ms, SD = 25.81 ms) compared to Mandarin speakers (M = 45.68 ms, SD = 72.10 ms).

In addition, there was a significant interaction between language and juncture version (χ^2 (1) = 52.98, p < .001; β = –43.24, SE = 6.91, t = −8.93). Follow-up planned comparisons for the significant interaction were again conducted using emmeans with Tukey *p*-value adjustments (Lenth, 2020), and they revealed crosslanguage differences in the degree to which pause duration production was affected by the different juncture versions (see **Figure 2**). In English, speakers produced longer pauses after the designated early junctures in Early Juncture sentences (M = 86.72 ms, SD = 29.02 ms; EMM = 16.33), compared to the Late Juncture sentences (M = 75.32 ms, SD = 20.82 ms; EMM = 14.67), β = 1.66, SE = 0.17, p < .001. In Mandarin, the effect of juncture version on pause duration was also in the same direction and was even stronger; Mandarin speakers produced an even longer pause after the designated early junctures in Early Juncture sentences (M = 57.00 ms, SD = 89.18 ms; EMM = 11.36), compared to the Late Juncture sentences (M = 34.20 ms, SD = 47.07 ms; EMM = 8.05), β = 3.31, SE = 0.35, p < .001.

Late Pause Duration (P2). For the pause duration of the designated late juncture region, adding the main effect of juncture version significantly improved model fit (χ^2 (1) = 6.33, p = .012; β = 218.23, SE = 61.45, t = 3.55); across both languages, speakers produced a longer pause duration at the designated late juncture in Late Juncture sentences (M = 99.28 ms,
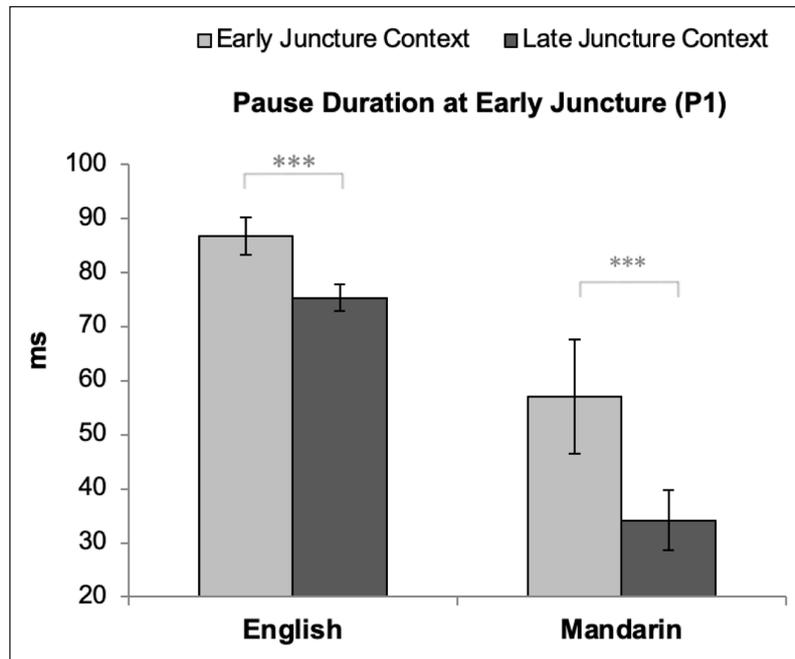
Figure 2: Crosslanguage differences in pause duration (in ms) at the designated Early Juncture (P1) marking as a function of Early Juncture and Late Juncture contexts. Error bars indicate standard error of the mean. ***$p \leq .001$.

$SD = 127.47$ ms) than in Early Juncture sentences ($M = 37.66$ ms, $SD = 47.70$ ms). Unlike the pause duration (P1), however, there was no main effect of language (χ^2 (1) $= 1.41$, $p = 0.234$; $\beta = -21.87$, $SE = 17.83$, $t = -1.23$); mean pause duration at the designated late junctures was similar across English ($M = 67.33$ ms, $SD = 80.82$ ms) and Mandarin ($M = 69.60$ ms, $SD = 117.90$ ms). Our model comparisons also did not reveal a significant interaction between language and juncture version (χ^2 (1) $= 0.42$, $p = .517$; $\beta = -13.10$, $SE = 19.10$, $t = -.69$).

Early Preboundary Vowel Duration (V1). For calculating all the pre- and postboundary vowel durations, we have also taken into account monophthong/diphthong status, vowel height, and vowel frontness. The V1 vowels in English (/ɜ:/) and Mandarin (/a/) were highly comparable; in both languages, they were monophthongal and open to open-mid in vowel height. However, they differed in vowel frontness; Mandarin V1 was a front vowel and English V1 as a central vowel. Prior to analyzing the effect of juncture version and language, we examined whether this difference played any role in speakers' duration production, and compared to the baseline model, adding vowel frontness as a predictor did not improve model fit log likelihood (χ^2 (1) $= 0.53$, $p = .465$). The baseline model therefore did not include this parameter.

Our model comparisons did not reveal a significant main effect of juncture version (χ^2 (1) $= 0.79$, $p = .376$; $\beta = -8.18$, $SE = 6.94$, $t = -0.79$); the preboundary duration of the designated early juncture was similar across the Early Juncture ($M = 99.60$ ms, $SD = 53.31$ ms) and Late

Juncture sentences (M = 91.42 ms, SD = 43.77 ms). Likewise, adding language as a main effect into the model did not improve model fit (χ^2 (1) = 2.35, p = .125; β = 28.59, SE = 14.30, t = 1.53); both the English and Mandarin speakers produced comparable vowel duration before the designated early juncture (English: M = 84.95 ms, SD = 32.23 ms; Mandarin: M = 106.07 ms, SD = 59.39 ms). There was also no significant interaction between juncture version and language (χ^2 (1) = 0.64, p = .423; β = 14.27, SE = 12.25, t = 0.84).

Late Preboundary Vowel Duration (V2). Across languages, there were again differences in vowel quality of V2. For instance, all the V2 tokens were monophthongs in English and most were diphthongs in Mandarin (V2 of the last sentence in Mandarin was rhoticized). Adding whether the vowels were monophthongs or diphthongs did show a significant main effect compared to baseline (χ^2 (1) = 14.44, p < .001; β = 30.49, SE = 7.69, t = 3.97). There was no significant effect of vowel frontness (χ^2 (1) = 0.04, p = .848; β = 4.77, SE = 12.16, t = .39), but there was also a significant main effect of vowel height (χ^2 (1) = 19.65, p < .001; β = –34.45, SE = 7.52, t = –4.58). The bestfitting baseline model therefore included vowel height and vowel monophthongalness.

Nevertheless, model comparisons with the bestfitting model revealed that adding juncture version as a predictor significantly improved model fit log likelihood (χ^2 (1) = 4.59, p = .032; β = 158.43, SE = 60.77, t = 2.61); speakers overall produced longer vowels before the designated late juncture in Late Juncture sentences (M = 220.93 ms, SD = 117.15 ms) compared to Early Juncture sentences (M = 176.20 ms, SD = 82.85 ms). There was also a main effect of language (χ^2 (1) = 27.51, p < .001; β = 75.70, SE = 11.66, t = 6.49); there was a trend for longer preboundary vowel duration in Mandarin (M = 226.70 ms, SD = 115.06 ms) than in English (M = 170.43 ms, SD = 82.25 ms). In addition, our model comparisons revealed a significant interaction between language and juncture version (χ^2 (1) = 38.83, p < .001; β = 94.20, SE = 12.32, t = 7.65).

Following upon this significant interaction, planned comparisons with emmeans (Lenth, 2020) revealed greater duration increase in late preboundary vowels in Late Juncture sentences compared to Early Juncture sentences in both languages. In Mandarin, speakers produced longer late preboundary vowels in Late Juncture sentences (M = 259.60 ms, SD = 127.77 ms; EMM = 172.30) compared to Early Juncture sentences (M = 193.80 ms, SD = 90.24 ms; EMM = 141.30), β = 31.00, SE = 4.01, p < .001. In English, speakers also tended to produce longer late preboundary vowels in Late Juncture (M = 182.26 ms, SD 91.02 ms; EMM = 110.20) compared to Early Juncture sentences (M = 158.60 ms, SD = 71.11 ms; EMM = 94.70), β = 15.50, SE = 2.00, p < .001, but the significant interaction stemmed from the fact that the effect of juncture versions was stronger in Mandarin than in English (see **Figure 3**).
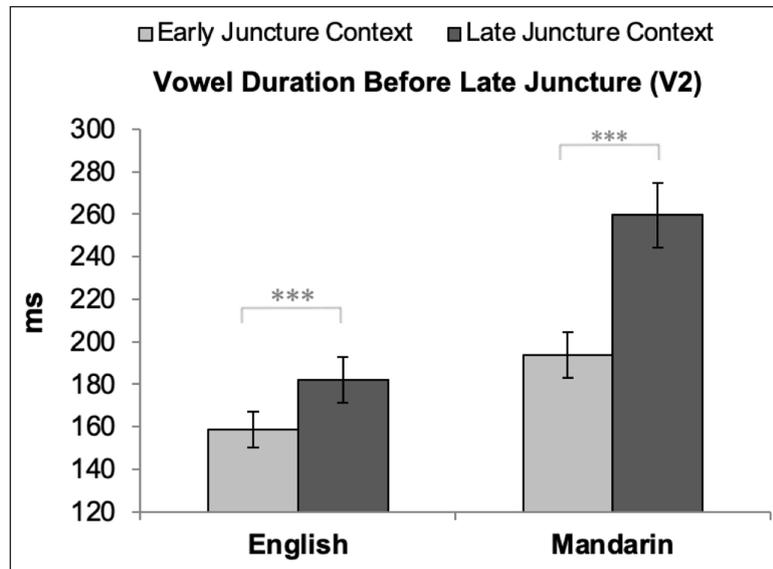
Figure 3: Crosslanguage differences in preboundary vowel duration (in ms) before the designated Late Juncture (V2) marking as a function of Early Juncture and Late Juncture contexts. Error bars indicate standard error of the mean. ***$p \leq .001$.

Late Postboundary Vowel Duration (V3). As in the V1 and V2 tokens, the V3 tokens across English and Mandarin differed in monophthong/diphthong status, vowel height, and vowel frontness. Results revealed that all three parameters played a role in the duration of V3 (monophthong versus diphthong: χ^2 (1) = 24.06, $p < .001$, $\beta = 48.52$, $SE = 9.29$, $t = -5.22$; vowel height: χ^2 (1) = 17.78, $p < .001$, $\beta = 55.26$, $SE = 12.21$, $t = 4.53$; vowel frontness: χ^2 (1) = 13.53, $p < .001$, $\beta = -41.89$, $SE = 10.74$, $t = 3.86$). The baseline model therefore included these parameters.

There was no main effect of juncture version (χ^2 (1) = 1.15, $p = .283$, $\beta = 28.37$, $SE = 28.41$, $t = 0.99$). Mandarin speakers tended to produce longer postboundary vowels after the designated late juncture ($M = 121.16$ ms, $SD = 64.65$ ms) compared to English speakers ($M = 91.92$ ms, $SD = 36.30$ ms), and with the inclusion of the different vowel characteristics, adding language as a fixed predictor also did not improve model fit (χ^2 (1) = 0.18, $p = .674$, $\beta = 6.83$, $SE = 13.97$, $t = 0.49$), suggesting that any cross-language difference should be attributed to these vowel differences. For postboundary vowel, there was also a non-significant interaction between juncture version and language (χ^2 (1) = 1.66, $p = .197$, $\beta = 17.87$, $SE = 12.80$, $t = 1.40$).

Mean F_0. For the preboundary vowels before the designated early juncture, the addition of mean F_0 to the baseline model revealed no significant main effect of juncture version (χ^2 (1) = 2.19,

$p = .139$; $\beta = 22.60$, $SE = 14.58$, $t = 1.55$). There was also no main effect of language (χ^2 (1) = 0.20, $p = .653$; $\beta = -1.43$, $SE = 21.22$, $t = -0.07$), and no significant interaction (χ^2 (1) = 0.26, $p = .612$; $\beta = 12.47$, $SE = 12.12$, $t = 1.03$).

For the preboundary vowel mean F_0 in the late juncture regions, there was also no significant main effect of juncture version (χ^2 (1) = 1.28, $p = .257$; $\beta = -29.67$, $SE = 20.22$, $t = -1.47$), and no interaction between language and juncture version (χ^2 (1) = 0.29, $p = .592$; $\beta = 24.17$, $SE = 15.46$, $t = 1.56$), although there was a marginally significant improvement for the main effect of language (χ^2 (1) = 3.66, $p = .056$; $\beta = 42.02$, $SE = 17.68$, $t = 2.38$); Mandarin speakers tended to produce higher mean F_0 at the designated late juncture regions ($M = 205.78$ Hz, $SD = 64.57$ Hz) compared to English speakers ($M = 179.31$ Hz, $SD = 48.02$ Hz).

For the postboundary vowel after the designated late junctures, there were no main effects of juncture version (χ^2 (1) = 0.13, $p = .724$; $\beta = -11.07$, $SE = 42.69$, $t = -0.26$). Nevertheless, like the preboundary vowels before the designated late junctures, there was a significant main effect of language in terms of mean F_0 (χ^2 (1) = 10.81, $p = .001$; $\beta = 54.56$, $SE = 16.39$, $t = 3.33$), where again Mandarin speakers produced higher mean F_0 ($M = 190.31$ Hz, $SD = 61.36$ Hz) compared to the English speakers ($M = 155.29$ Hz, $SD = 55.94$ Hz). Finally, there was a marginally significant effect of the language by juncture version interaction (χ^2 (1) = 3.10, $p = .078$; $\beta = 9.81$, $SE = 11.64$, $t = 1.96$).

Minimum F_0. For the minimum F_0 of the preboundary vowels in the early juncture regions, there were no significant main effects of juncture version (χ^2 (1) = 2.57, $p = .109$; $\beta = 25.59$, $SE = 15.15$, $t = 1.69$) or of language (χ^2 (1) < .001, $p = .987$; $\beta = 0.30$, $SE = 20.96$, $t = -0.014$), and no language by juncture version interaction (χ^2 (1) = 1.54, $p = .215$; $\beta = 16.90$, $SE = 12.18$, $t = 1.39$). For the min F_0 values in the late juncture regions, there were also no main effects of juncture version (χ^2 (1) = 2.30, $p = .139$; $\beta = -32.02$, $SE = 21.17$, $t = -1.51$) or language (χ^2 (1) = 3.05, $p = .081$; $\beta = 28.59$, $SE = 16.53$, $t = 1.73$), and no interaction (χ^2 (1) = 0.32, $p = .596$; $\beta = 9.14$, $SE = 14.43$, $t = 0.63$). Finally, for the postboundary minimum F_0 after the designated late junctures, there was no main effect of juncture version (χ^2 (1) = 0.51, $p = .474$; $\beta = -16.41$, $SE = 25.18$, $t = -0.65$), but adding the main effect of language did improve model fit (χ^2 (1) = 7.00, $p = .008$; $\beta = 42.55$, $SE = 14.98$, $t = 2.84$), where again, the Mandarin speakers produced higher minimum F_0 ($M = 162.74$ Hz, $SD = 58.00$ Hz), compared to the English speakers ($M = 138.28$ Hz, $SD = 53.05$ Hz). However, there was no significant interaction between language and juncture version (χ^2 (1) = 2.21, $p = .138$; $\beta = 24.25$, $SE = 13.71$, $t = 1.78$).

Maximum F_0. There was no main effect of juncture version for the maximum F_0 values of the preboundary early juncture region (χ^2 (1) = 1.40, $p = .236$; $\beta = 21.25$, $SE = 18.98$, $t = 1.12$). There was also no main effect of language (χ^2 (1) = 0.04, $p = .847$; $\beta = -4.04$, $SE = 21.48$, $t = -0.19$), and no interaction (χ^2 (1) = 0.38, $p = .537$; $\beta = 8.49$, $SE = 13.90$, $t = 0.61$). For preboundary maximum F_0 in the late juncture regions, there was also no main effect of juncture

version (χ^2 (1) = 0.49, p = .481; β = –18.21, SE = 28.67, t = –0.64), and no juncture version by language interaction (χ^2 (1) = 3.47, p = .063; β = 35.27, SE = 16.75, t = 2.11), though there was a main effect of language (χ^2 (1) = 6.76, p = .009; β = 52.26, SE = 19.73, t = 2.65), with Mandarin speakers consistently producing higher preboundary maximum F_0 (M = 233.63 Hz, SD = 70.90 Hz) than English speakers (M = 200.26 Hz, SD = 51.94 Hz). A similar pattern of results was observed for the maximum F_0 values of the postboundary vowels in the late juncture regions, where there was also a significant main effect of language (χ^2 (1) = 14.09, p < .001; β = 86.74, SE = 20.13, t = 4.31); again, Mandarin speakers produced higher postboundary maximum F_0 (M = 225.63 Hz, SD = 70.79 Hz) compared to English speakers (M = 171.34 Hz, SD = 66.82 Hz). However, as in the preboundary vowels, there were no main effects of juncture version (χ^2 (1) = 0.002, p = .969; β = 1.74, SE = 58.56, t = 0.03) and no interaction between juncture version and language interaction (χ^2 (1) = 2.20, p = .138; β = –109.83, SE = 57.26, t = –1.92).

F_0 Range. For the preboundary vowels before the designated early juncture, there were no main effects of juncture version (χ^2 (1) = 0.21, p = .648; β = –4.31, SE = 9.96, t = –0.43) or language (χ^2 (1) = 1.32, p = .251; β = –5.54, SE = 4.85, t = –1.14), and no interaction (χ^2 (1) = 1.59, p = .207; β = –5.44, SE = 4.22, t = –1.29).

For the pre- and postboundary vowels of the late juncture regions, there were also no main effects of juncture version (preboundary: χ^2 (1) = 0.46, p = .498; β = 13.35, SE = 23.75, t = 0.56; postboundary: χ^2 (1) = 0.42, p = .517; β = 17.92, SE = 32.49, t = 0.55). However, there was a main effect of language for both the pre- and postboundary late juncture vowels (preboundary: χ^2 (1) = 5.76, p = .016; β = 24.50, SE = 10.03, t = 2.44; postboundary: χ^2 (1) = 15.99, p < .001; β = 49.28, SE = 11.38, t = 4.33). In both cases, Mandarin speakers produced greater F_0 range (preboundary: M = 54.61 Hz, SD = 55.48 Hz; postboundary: M = 62.89 Hz, SD = 41.22 Hz), compared to English (preboundary: M = 38.90 Hz, SD = 38.67 Hz; postboundary: M = 33.06 Hz, SD = 56.12 Hz). Further, both the pre- and postboundary vowels of the designated late junctures showed an interaction between juncture version and language (preboundary: χ^2 (1) = 6.38, p = .012; β = 23.80, SE = 9.22, t = 2.58; postboundary: χ^2 (1) = 12.95, p < .001; β = 43.95, SE = 11.02, t = 3.99). However, follow-up Tukey-adjusted planned comparisons for preboundary F0 range showed that the effect was only marginally significant (p = .060).

Domain-initial Segmental Effects. Due to different segments used across languages, separate analyses were conducted on the English and Mandarin datasets to examine segmental strengthening effects in the two languages. In English, there were no main effects of juncture version for any of the boundary-related segmental measures. In Mandarin, there was a significant main effect of juncture version for the segmental effect of the preboundary word before the designated early junctures (i.e., the VOT of /tʰ/ in /tʰa/ 她 "her") (χ^2 (1) = 5.08, p = .0242;

$\beta = 13.531$, $SE = 6.25$, $t = 2.17$). However, the effect was not in the predicted direction (see **Figure 4**); the VOT of the preboundary word onset was longer in Late ($M = 70.96$ ms, $SD = 32.74$ ms) than in Early Juncture sentences ($M = 62.43$ ms, $SD = 27.57$ ms).
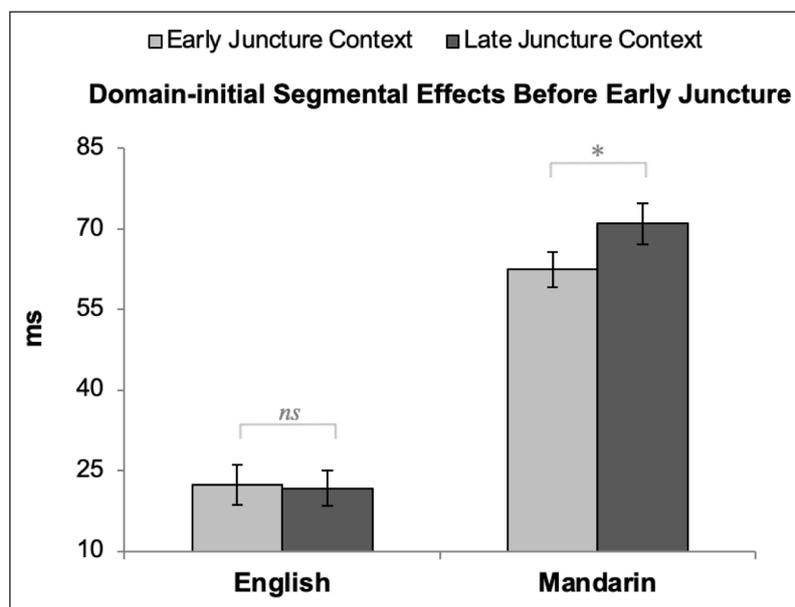
Figure 4: Crosslanguage differences in segmental effects (in ms) before the designated Early Juncture, revealed in the VOT of /tʰ/ from "她" /tʰa/ ("her") in Mandarin, compared to the duration of /h/ from "her" /hɜ/ in English. Error bars indicate standard error of the mean. *$p \leq .05$.

Summary. Both English and Mandarin speakers produced significantly longer pauses at the relevant junctures in both early and late juncture contexts. In early juncture contexts, the pause duration was even longer in Mandarin than in English. In neither language did speakers produce longer preboundary vowels in early juncture contexts, nor did they produce significantly longer postboundary vowels (as measured in the vowels after the designated juncture) in late juncture contexts. Both English and Mandarin speakers produced preboundary duration cues, but the increase in preboundary duration was longer in Mandarin. As for F_0, juncture version played no role at all, though the Mandarin speakers overall produced higher F_0 than the English speakers. For segmental modification, only a longer preboundary VOT in Mandarin occurred, in the unpredicted direction, and hence presumably a chance effect.

2.3. Discussion

Our production data suggest that English and Mandarin speakers were alike in their use of duration to mark juncture. Particularly in the late juncture sentences, longer pauses were produced at the designated juncture. Similarly, both groups of speakers produced longer preboundary vowels before the designated juncture in Late Juncture sentences.

However, there were language-specific differences in the degree to which different prosodic features were produced across the different juncture versions. Thus the difference in pause duration at the juncture position in Early Juncture sentences was greater in Mandarin, and Mandarin speakers also produced a significantly greater increase in preboundary duration in Late Juncture sentences. We therefore conclude that while English and Mandarin speakers are similar in how they produce duration cues (i.e., pausing, preboundary lengthening), there can still be cross-language differences in where they produce them; in both cases of durational differences, Mandarin speakers were more likely to mark longer pauses in early juncture contexts and longer preboundary vowels in late juncture contexts.

Note also that neither language group produced all the boundary-related cues we measured. For example, neither group produced postboundary or F_0 cues. At the same time, we also observed a case where juncture cues mismatched the prosodic context: Mandarin speakers produced pre-junctural VOT lengthening with late rather than, as might have been expected, with early juncture.

Note that speakers' juncture cue choices could have been influenced by our task, which involved reading passages where the storyline already provided the referential context necessary for effective disambiguation. Reading tasks may be less likely to elicit juncture production, particularly if speakers are unaware of the ambiguity (e.g., Allbritton, McKoon, & Ratcliff, 1996). Disambiguation can also be influenced by many other factors, including lexical bias, situation-specific contextual information, listeners' knowledge of the speaker, as well as speaker awareness of the ambiguity (e.g., Albritton et al., 1996; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Crain & Steedman, 1985; Kim, Stephens, & Pitt, 2012; Snedeker & Trueswell, 2003; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

It is worth noting that all the above studies and proposals have concerned native speakers of English or other Germanic languages (e.g., Dutch, German). Here, we compared languages with very different prosodic systems, but cases where we could adopt a structured approach involving identical storylines and sentences with identical syntactic ambiguity and very similar boundary-related segments. Contrary to previous findings from reading tasks (e.g., Albritton et al., 1996), and even without explicit instructions but with contexts that effectively made the use of prosody redundant, we found that speakers did produce prosodic cues to juncture. Importantly, the choice of cue types was similar across English and Mandarin. The speaker groups varied in the degree to which each type was engaged, however. Thus there appear to be language differences in production preferences across the different early versus late juncture versions.

We now move on to explore the cross-language patterns to be found in perception. The following two perception experiments again exploit the similar ambiguous structure of the Early and Late Juncture parses across English and Mandarin. These experiments involve a novel disambiguation task, in which participants from each language group listen to the ambiguous

sentences without contextual cues, and press a button to choose the correct interpretation; both their response time and their accuracy are measured. With this method we can ascertain whether the cross-language symmetries and asymmetries we have observed in production are also reflected in listening behaviour.

There are two possibilities. On one hand, the cross-language perception results across early and late juncture versions may mimic the language-specific differences in our production data, particularly with respect to duration. As already mentioned, Mandarin speakers were more likely to mark stronger duration cues compared to English. Such cross-language duration differences may reflect processing difficulties across languages; speakers may be more likely to mark prosodic cues to disambiguate a sentence when they deem the sentence difficult to understand (see Kraljic & Brennan, 2005). For this reason, Mandarin speakers' disambiguation perception may be affected when certain cues are rendered uninformative.

On the other hand, perception strategies may be separate from production. Unlike production, where languages may vary in the degree to which different cues are used, listeners in both English and Mandarin may still use whatever cues are available in the signal. For example, in prosodic entrainment, where listeners entrain to prosodic contours to rapidly locate an upcoming prosodic focus, we know that listeners across different languages do not use any one single cue to predict the prosodic forms of upcoming words (e.g., Cutler, 1987; Cutler & Darwin, 1981; Ip & Cutler, 2020). Likewise, the realization of a sentence's prosodic structure may be a blend of different prosodic cues (e.g., duration, F_0) that all listeners may exploit (Cutler & Isard, 1980), and listeners might accordingly exploit whatever cues are available for disambiguation. This could result in no significant relationship between listening behavior and individual disambiguation cues.

We explore these possibilities in the following two perception experiments. In Experiment 2, our first perception experiment, we ask whether English and Mandarin listeners show differences in disambiguation when all prosodic cues are present. We will also analyze if there is a relationship between individual cues and listeners' disambiguation response time and accuracy rates. In our second perception experiment, we ask whether listeners' disambiguation response time and accuracy differ across languages when a primary disambiguation cue (pause duration) is rendered uninformative.

3. Experiment 2: Perception (with all cues)

3.1. Method

3.1.1. Participants

Forty native speakers of Australian English (M_{age} = 22.50 years, SD = 7.70 years, range: 17.89–53.50 years; 30 females) and 40 native speakers of Mandarin Chinese (M_{age} = 25.12 years, SD = 3.61 years, range: 18.75–38.30 years; 21 females) took part. Using a conservative cut-off

value at 64% accuracy, data from a further seven English listeners and two Mandarin listeners were excluded. All Mandarin speakers were born in Mainland China and had been living in Australia for an average of 1.86 years ($SD = 2.27$ years, range: 8 days – 10 years). No participant reported any hearing or reading impairment.

3.1.2. Materials

All the materials used in the present research can be accessed from the following URL link: https://upenn.box.com/s/u72whjp3buhtvwhv9b6adhmdhme7vf71. Twenty-two syntactically ambiguous experimental sentences in English and Mandarin were constructed (see Appendices A and B), each having two different interpretations resulting from different juncture placement. For each language, the sentences were recorded in their two versions by a female native speaker at a natural fast-normal rate. As in the production experiment, the two juncture versions differed in the timing and location of the boundary (i.e., Early Juncture versus Late Juncture). In Early Juncture versions, a boundary occurred earlier in the utterance (e.g., "Larry accidentally gave her # rat poison"; "刘波不小心给她 #老鼠药吃") while in segmentally identical Late Juncture versions, a boundary occurred later in the utterance (e.g., "Larry accidentally gave her rat # poison"; "刘波不小心给她老鼠 # 药吃"). For each experimental sentence, the speaker also produced a pair of interpretation sentences that corresponded to the intended meaning of the Early and Late Juncture versions (e.g., "Larry gave rat poison to Hannah" versus "Larry gave rat poison to Hannah's pet rat Rohan"; "刘波把老鼠药给珍妮" versus "刘波把老鼠药给珍妮的宠物鼠").

Unlike the production experiment on natural usage, where speakers were not explicitly told to disambiguate the sentences, the perception experiment aims to examine whether listeners across languages can use informative juncture cues to disambiguate their understanding of sentences. To manipulate the Early and Late Juncture versions, the English and Mandarin speakers who recorded the stimuli for the perception study were made aware of the ambiguity and were asked to produce each version of the experimental sentences in a way that would match its corresponding interpretation sentence. But as in the production experiment, nonetheless, they were not given any explicit instruction on how they should accomplish the disambiguation. In both languages, the Early and Late Juncture versions for each stimulus sentence pair were segmentally identical, and the two language sets were highly comparable in terms of their syntactic ambiguity (see Appendix C for interlinear morpheme-by-morpheme glosses of the Mandarin sentences).

In each language, 12 additional filler sentences and their corresponding pair of interpretation sentences were also recorded. These filler sentences involved other types of ambiguity that were either easier than the experimental sentences (e.g., homonyms) or more difficult (e.g., sentences with relative clause attachment ambiguity). There were two counterbalanced experimental conditions, each containing one juncture version of each of the 22 experimental sentences, plus the additional 12 filler sentences. The experimental and filler items were pseudo-randomized

(such that the counterbalanced orders contained no more than two consecutive instances of an experimental sentence).

3.1.3. Procedures

The disambiguation task was administered using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) on a laptop computer and a Chronos® USB response device for button pressing. All instructions were given in the form of a pre-recorded voiceover script made by the same speaker who produced the stimuli. Written instructions were also displayed on the screen as the voiceover instructions were being played (see Appendices D and E). Because of the greater distinction between formal/written and spoken language in Mandarin, we added an extra line in our Mandarin instructions to explicitly inform participants that they would hear everyday normal sentences, (and thus should not expect sentences in a formal/written style). All participants were given three practice trials and feedback before starting the actual experiment. Note that we did not give instructions on how to disambiguate the sentences.

From the start of each trial, participants saw on their screen two interpretation sentences that corresponded to the left and right buttons in front of them. Participants heard the test sentences, and were asked to "pay careful attention to the meaning of each sentence" and to choose for each sentence its intended meaning by pressing the button that matched the correct interpretation sentence. Up to five seconds were available for pressing the button before the next trial commenced; the interpretation sentences remained on the screen for the full five seconds after the offset of the sentence. Nevertheless, participants were instructed to choose the correct button "as soon as they understood the sentence" and were told that they could press the button at any time during the trial while the sentence was being played. They were further informed that they would be tested on both their accuracy and on their speed of comprehension. Whether the correct button was the left or right button was counterbalanced across participants.

We recorded participants' response times and number of correct responses. Participants who made errors of disambiguation on one-third or more of the experimental sentences were excluded from the analysis. An absence of button press was also considered an 'incorrect response,' because a failure to press the button, even during the five seconds after the sentence was finished, was interpreted as indicating that the participant was still trying to process the meaning of the ambiguous sentence. No participant failed to respond on more than two occasions during the experimental trials.

At the end, all participants completed a recognition test where they were presented with a list of 22 sentences and were asked to judge whether each of these sentences were from the experiment (see Appendices F and G). Half of these test sentences were indeed from the experiment. All participants scored above 14 out of 22 (64%) on the recognition test (In English,

$M = 88.64\%$, $SD = 9.14\%$, range: 64–100%; In Mandarin, $M = 90.68\%$, $SD = 8.17\%$, range: 73–100%). English and Mandarin listeners did not differ significantly in their recognition scores.

3.2. Results

3.2.1. General Overview

More than 90% of participants' correct responses in both languages were made by pressing the button after the offset of the sentences. Therefore, we measured response time (RT) as the duration between experimental sentence offset and participant button presses. Only data for correct disambiguations were included in our analyses.

The main aims of our statistical analyses were to investigate (1) whether RT differed across juncture version (i.e., Early versus Late Juncture) and (2) whether the pattern of this RT difference varied across languages and experimental trials. To address (2), we performed statistical tests on the RT data to examine whether there was an interaction between language groups and juncture version, and to address (1), we performed separate analyses for the English and Mandarin datasets. We also performed acoustic analyses of all of the experimental sentences to examine whether there were differences in duration cues across the Early and Late Juncture sentences. We measured the following prosodic disambiguation cues in Praat (Boersma & Weenink, 2018): (1) pausing and (2) pre- and postboundary vowel lengthening. As in the previous production study, we measured the pause duration and boundary lengthening of both the designated early and late juncture locations in all sentences. All analyses were conducted using mixed effects models.

3.2.2. Response Time

As in the production experiment, we constructed LMER models to obtain the best fitting model predicting listeners' RT. Visual displays of the quartile-quartile plots revealed that the raw RT data formed a skewed distribution, so we transformed the data using the Yeo-Johnson procedure. Leave-one-out model comparisons were conducted to ensure that each predictor yielded a significant gain in log likelihood with all other predictors in the model. Predictors were added in a step-wise fashion and all fixed effects (i.e., juncture version, language) were coded with mean-centred contrast codes. The baseline model we used as the starting point included by-participant and by-item random intercepts as well as by-participant and by-item random slopes for the effect of juncture version. The effect of trial sequence order as a continuous variable was also tested at the outset. Trial sequence order was included in the model as a continuous predictor, where each level was labelled (i.e., from 1 to 34) according to its trial order across the 34 (22 experimental, 12 filler) trials. Because of its large eigenvalue, we rescaled this variable by centering the trial order levels into numeric values from 0 to 1. Leave-one-out comparisons show that the addition of trial order was significant (χ^2 (1) = 6.22, $p = .013$; $\beta = -255.91$, $SE = 111.64$, $t = -2.68$).

The updated baseline model therefore also consists of trial order as fixed effect, as well as the by-participant and by-item random slopes for the effect of trial order.

Listeners' average RT across languages and juncture versions are shown in **Figure 6**. We examined the effect of juncture version (Early versus Late Juncture), language (English versus Mandarin), and language by juncture version interaction. The addition of language did not significantly increase model log likelihood (χ^2 (1) = 0.20, p = .657; β = –216.85, SE = 184.59, $t = -0.55$). There was also no main effect of juncture version in the whole sample of English and Mandarin listeners together (χ^2 (1) = 1.66, p = .197; β = 204.18, SE = 318.85, t = 1.29); however, there was a significant interaction between language and juncture version (χ^2 (3) = 28.84, p < .001; β = 2628.98, SE = 499.83, t = 5.40).

As in Experiment 1, we followed up the significant interaction using *emmeans* with Tukey adjustments for *p*-values (Lenth, 2020). Juncture version influenced RT in both English and Mandarin, but there was an inverse interaction between language and juncture version (see **Figure 5**): English listeners were faster at disambiguating Late Juncture sentences (M = 1097.23 ms, SD = 982.12 ms; EMM = 1143) compared to Early Juncture sentences (M = 1308.33 ms, SD = 1022.48 ms; EMM = 1338), β = 320, SE = 122, t = 2.63, p = .049, whereas Mandarin listeners were faster to disambiguate the Early Juncture sentences (M = 1022.32 ms, SD = 850.33 ms; EMM = 1018) than the Late Juncture sentences (M = 1258.30 ms, SD = 1082.35 ms; EMM = 1292), β = –274, SE = 96.10, t = –2.86, p = .037.
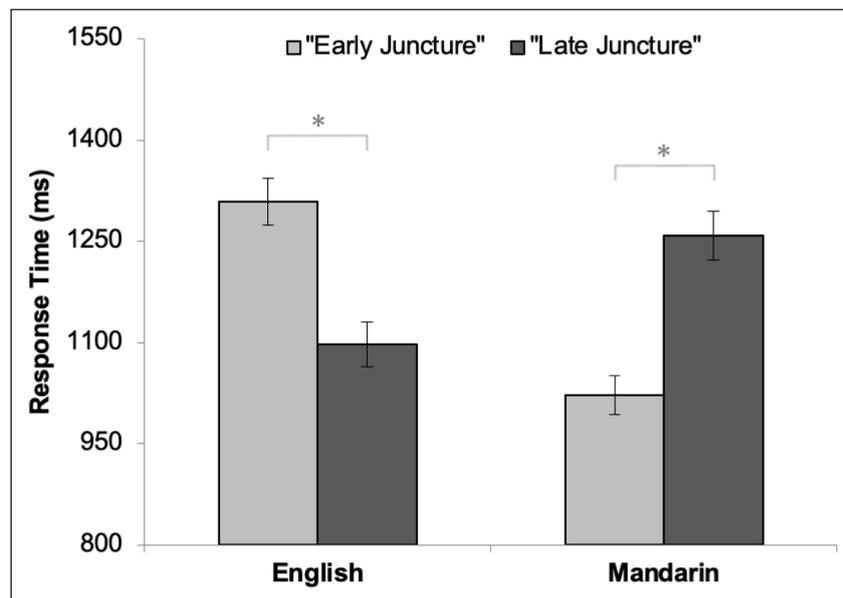
Figure 5: Significant crosslanguage difference in disambiguation response time (in ms) as a function of Early Juncture and Late Juncture contexts. Error bars indicate standard error of the mean. *p ≤ .05.*

3.2.3. Accuracy

On average, Mandarin-speaking participants made 3.3 incorrect disambiguation responses ($SD = 1.82$) throughout the 22 experimental sentences, whereas the English-speaking participants in contrast averaged 5.6 incorrect disambiguations ($SD = 2.1$) across the 22 experimental sentences (see **Table 4**). We used the *glmer* function from the lme4 package (Bates et al., 2015, version 1.1-7) to build Generalized Linear Mixed-Effects regression (GLMER) models to examine whether there were accuracy differences across languages and juncture versions. GLMER models were used because they enabled us to assess the influence of language background and juncture version on accuracy as a categorical dependent measure (i.e., a binary distribution) while also accounting for individual patterns across participants and sentence items (Bolker et al., 2008). The accuracy data were coded as either "1" for correct disambiguation or "2" for incorrect responses. Specifically, we were interested in whether sentences with early versus late junctures had an effect on listeners' ability to correctly disambiguate the sentences, and whether this varied across the languages. As in the RT analyses, by-participant and by-item random slopes for the effect of juncture version, as well as by-participant and by-item random intercepts, were added as random effects, and juncture version, language, and language by juncture version interaction were entered as fixed effects.

Results revealed a main effect of language (χ^2 (1) = 30.25, $p < .001$; $\beta = -1.39$, $SE = 0.24$, Wald $z = -5.83$); English listeners had more errors throughout the 22 experimental sentences ($M = 5.6$ errors, $SD = 2.1$ errors), compared to the Mandarin listeners ($M = 3.3$ errors, $SD = 1.82$). However, there was no main effect of juncture version (χ^2 (1) = 1.64, $p = .200$; $\beta = -0.75$, $SE = 0.57$, Wald $z = -1.32$). Nevertheless, there was a significant interaction between language and juncture version (χ^2 (1) = 18.36, $p < .001$; $\beta = -1.16$, $SE = 0.26$, Wald $z = -4.43$). Model comparisons from the separate English and Mandarin datasets revealed a main effect of juncture version in English, but not in Mandarin; English listeners had significantly more errors in disambiguating Early Juncture sentences ($M = 3.9$ errors, $SD = 1.6$ errors) compared to Late Juncture sentences ($M = 1.7$ errors, $SD = 1.07$) (χ^2 (1) = 6.33, $p = .012$; $\beta = -1.78$, $SE = 0.65$, Wald $z = -2.73$), while Mandarin listeners had similar number of errors across sentences with early junctures ($M = 1.63$ errors, $SD = 1.15$) and sentences with late junctures ($M = 1.68$, $SD = 1.59$) (χ^2 (1) = 0.40, $p = .529$; $\beta = 0.74$, $SE = 1.17$, Wald $z = 0.63$).

	Mean Errors (*SD*)		
	Early Juncture	**Late Juncture**	**Total**
English	3.90 (1.60)	1.70 (1.07)	5.60 (2.10)
Mandarin	1.63 (1.15)	1.68 (1.59)	3.30 (1.82)

Table 4: Number of incorrect responses in English and Mandarin as a function of Early Juncture and Late Juncture contexts in Experiment 2 (With All Cues).

To complement these findings, we also observed the error rates of the English and Mandarin participants who were excluded on the basis of their incorrect responses. In total, we excluded seven English listeners and two Mandarin listeners who failed to correctly disambiguate at least 64% of the experimental sentences. The seven excluded English-speaking participants had an average total of 10.86 incorrect responses (SD = 2.12), with 5.86 errors (SD = 1.77) in the Early Juncture sentences and 5.00 errors (SD = .82) in the Late Juncture sentences. The two excluded Mandarin listeners had a total average of 10 incorrect responses (SD = 1.42), with 4 errors (SD = 1.41) in the Early Juncture sentences and 6 errors (SD = 2.82) in the Late Juncture sentences, a non-significant difference.

3.2.4. Stimuli: Acoustic analyses

We also conducted acoustic analyses of the stimuli sentences in Praat (Broersma & Weenink, 2018) based on inspection of the waveform and the spectrogram. As in the production experiment (see **Figure 1**), we measured, for all stimuli, the pause duration of both the designated Early Juncture (P1) and the designated Late Juncture (P2). We also measured the preboundary and postboundary vowel durations of the designated Early and Late Junctures. Separate LMER models were constructed for the English and Mandarin stimuli sentences, where the random intercept of items was added as a random factor and leave-one-out comparisons were used to ensure that each predictor yielded a significant gain in log likelihood with all other predictors in the model.

In English, there was a main effect of preboundary vowel duration; the preboundary vowels before the designated early junctures were longer in Early Juncture sentences (M = 105 ms, SD = 67 ms) than in Late Juncture sentences (M = 76 ms, SD = 36 ms) (χ^2 (1) = 6.86, p = .009; β = –46.82, SE = 16.89, t = –2.77), and the preboundary vowels before the designated junctures were longer in Late Juncture sentences (M = 216 ms, SD = 95 ms) compared to Early Juncture sentences (M = 155 ms, SD = 71 ms) (χ^2 (1) = 22.29, p < .001; β = 97.13, SE = 16.00, t = 6.07). Adding postboundary vowel duration significantly increased model fit likelihood, such that the postboundary vowels after the designated late junctures (i.e., V3) were longer after Late Juncture sentences (M = 181 ms, SD = 82 ms) than after Early Juncture sentences (M = 162 ms, SD = 79 ms) (χ^2 (1) = 12.56, p < .001; β = 31.84, SE = 7.92, t = 4.02). There was also a significant main effect of pause duration, with the pause duration of the designated early junctures being longer in Early Juncture sentences (M = 102 ms, SD = 93 ms) than in Late Juncture sentences (M = 51 ms, SD = 23 ms) (χ^2 (1) = 5.36, p = .021; β = –81.57, SE = 33.51, t = –2.43), and the pause duration of the designated late junctures being longer in Late Juncture sentences (M = 102 ms, SD = 32 ms) than in Early Juncture sentences (M = 45 ms, SD = 38 ms) (χ^2 (1) = 29.44, p < .001; β = 90.56, SE = 116.20, t = 7.79).

Similarly in Mandarin, there were main effects of both the preboundary vowel durations before the designated early junctures, (χ^2 (1) = 28.48, p < .001; β = –179.35, SE = 24.09, t = –7.45), and the preboundary vowel durations before the designated late junctures, (χ^2 (1) = 19.10, p < .001; β = 129.68, SE = 24.06, t = 5.39). In both these cases, the preboundary vowels were longer before the designated junctures; preboundary vowels before designated early junctures were longer in Early Juncture sentences (M = 254 ms, SD = 91 ms) than in Late Juncture sentences (M = 141 ms, SD = 50 ms), while vowels before designated late junctures were longer in Late Juncture sentences (M = 271 ms, SD = 113 ms) than in Early Juncture sentences (M = 190 ms, SD = 66 ms). Likewise, adding pause durations significantly increased the model log likelihood for both the pause duration of the designated early junctures (χ^2 (1) = 32.67, p < .001; β = –348.24, SE = 51.21, t = –6.80), as well as the pause duration of the designated late junctures (χ^2 (1) = 56.53, p < .001; β = 480.26, SE = 42.88, t = 11.20). The pause duration of the designated early junctures were, on average, longer in Early Juncture sentences (M = 238 ms, SD = 147 ms) than Late Juncture sentences (M = 20 ms, SD = 32 ms), while the pauses of the designated late junctures were longer in the Late Juncture sentences (M = 318 ms, SD = 132 ms), compared to the Early Juncture sentences (M = 17 ms, SD = 30 ms). However, unlike for the English stimuli, there was no main effect of postboundary vowel duration (χ^2 (1) = 0.22, p = .638; β = –22.61, SE = 49.03, t = –0.46); the Mandarin speaker who recorded the stimuli sentences produced similar postboundary duration across the Early Juncture sentences (M = 205 ms, SD = 134 ms) and the Late Juncture sentences (M = 191 ms, SD = 77 ms).

Acoustic Cues and Response Time. As RT differences across the different juncture versions could be attributed to acoustic differences between the two stimulus sets, we constructed separate LMER models for the two datasets and tested for main effects of each acoustic cue as well as for interactions of each acoustic cue with juncture version. In English, the pause duration of the designated early juncture showed a main effect on listeners' RT (χ^2 (1) = 4.03, p = .044; β = 125.59, SE = 61.86, t = 2.03). There was also a significant interaction between postboundary duration of the designated late juncture and juncture version (χ^2 (1) = 6.86, p = .009; β = 313.64, SE = 114.94, t = –2.73). In Mandarin, the duration of the preboundary vowels before the designated early juncture showed a marginally significant main effect on RT (χ^2 (1) = 3.27, p = .071; β = –119,17, SE = 63.53, t = –1.88).

Acoustic Cues and Accuracy. We also constructed leave-one-out GLMER models for the English and Mandarin datasets to examine the effect of each acoustic cue on listeners' disambiguation accuracy. Similar to the RT results, adding the duration of the preboundary vowels before the designated early juncture significantly improved model fit in Mandarin (χ^2 (1) = 15.25, p < .001; β = –2.53, SE = 0.63, Wald z = –4.04). The pause duration of the designated early

juncture also marginally improved model fit (χ^2 (1) = 3.56, p = .059; β = –0.82, SE = 0.44, Wald z = –1.86). No other acoustic cues in Mandarin predicted accuracy. Finally, none of the acoustic features significantly affected English listeners' accuracy scores.

3.3. Discussion

In line with the cross-language asymmetry observed in English and Mandarin speakers' duration production, our perception experiment revealed a similar cross-language difference in RT pattern across the different juncture versions. In English, listeners were significantly faster at disambiguating Late Juncture sentences than Early Juncture sentences. In contrast, Mandarin listeners were faster at disambiguating Early Juncture sentences. The English and Mandarin listeners also differed in interpretation accuracy, with more errors made by English listeners. The results therefore indicate (1) language differences in listeners' sentence processing as a function of different juncture context and (2) language differences in the extent to which listeners use prosody at all to correctly disambiguate an ambiguous sentence.

The perception results could be interpreted in light of the production differences in duration. In our production experiment, we have observed that Mandarin speakers tend to produce sentences with longer preboundary lengthening and pauses. From our analyses, only Mandarin listeners were able to use preboundary duration cues to resolve the ambiguous sentence; the preboundary vowel duration of the designated early juncture, and also to a weaker extent its pause duration, showed a significant effect on improving Mandarin listeners' accuracy rates. In English, although response time was related to the duration of the early juncture pause, there was no relation between listeners' disambiguation accuracy and any of the individual disambiguation cues. These findings thus seem to suggest that listeners do not necessarily exploit all available cue(s) for disambiguation; cues are weighted differently across languages and listeners across languages vary in the cues they rely on to correctly disambiguate a sentence.

In light of this we might expect that Mandarin listeners would pay particular attention to boundary-related duration cues. In a second perception experiment, therefore, we test whether native English and Mandarin speakers would show the same RT pattern and accuracy scores when pause duration was rendered uninformative. If the Mandarin listeners assign more weight to pausing than English listeners, then their accuracy and RT performance would be affected to a greater degree by the removal of the pausing cue. Given that pre- and postboundary lengthening cues were still preserved, a lack of change in disambiguation performance would indicate that Mandarin listeners could attend to boundary-related lengthening to disambiguate the sentences. Likewise, the English listeners' disambiguation performance would be unaffected if they do not rely on pause duration as a cue to prosodic juncture.

4. Experiment 3: Perception (with pause duration removed)

4.1. Method

4.1.1. Participants

The final sample contained 12 native Australian English speakers (M_{age} = 23.46 years, SD = 8.84 years, range: 18.16–49.61 years; 10 females) and 19 native Mandarin speakers (M_{age} = 28.76 years, SD = 8.77 years, range: 19.72–51.45 years; 13 females). None had taken part in the earlier perception experiment. The Mandarin-speaking participants had been living in Australia for an average of 5 years and 2 months (SD = 7.32, range: 41 days to 24 years and 9 months). We excluded additional data from four English listeners and we also excluded one Mandarin listener who failed to correctly disambiguate at least 64% of the experimental sentences. All participants were university students at the time of the experiment and reported no hearing or reading impairment.

4.1.2. Materials and procedures

The procedures were identical to the first perception experiment, except that the pauses were removed from all experimental sentences. Across both the Early and Late Juncture versions of the experimental sentences, we deleted all pauses indicating either the designated early juncture (P1) or the late juncture (P2). As a result, there were no interword silences at all in these two positions, and this lack of pausing became potentially inconsistent with the other durational cues (e.g., preboundary lengthening) that coexisted with it. In the follow-up recognition test, the included listeners had an average respective score of 81.46% (or 17.92 out of 22; SD = 9.59%, range: 68–100%) for the English group, and 92.82% (20.42 of 22; SD = 6.50%, range: 82–100%) for the Mandarin group. These recognition scores did not statistically differ from those in the first perception experiment.

4.2. Results

4.2.1. Response time

As in the first perception experiment, the raw RT data were transformed using Yeo-Johnson procedure and LMER models were constructed to examine the role of language, juncture version, and language by juncture version interaction. Again, we used a baseline model with random intercepts of participants and items and random slopes of participants and items for the effect of juncture version. In our model comparisons, there was no significant main effect of language (χ^2 (1) = 0.02, p = .895; β = –255.08, SE = 406.34, t = –0.15). Adding juncture version also did not improve model fit (χ^2 (1) = 0.43, p = .510; β = 39.41, SE = 198.26, t = 0.65). In addition, there was no significant language by juncture version interaction (χ^2 (3) = 5.05, p = .168; β = 1035.04, SE = 483.73, t = 2.15).

Nevertheless, the direction of the results was the same as in the first perception experiment (see **Figure 6**). The English listeners showed a faster RT in Late Juncture sentences (M = 1175.26, SD = 1123.45) compared to Early Juncture sentences (M = 1410.51, SD = 1179.05). The Mandarin listeners showed a faster RT in Early Juncture sentences (M = 1113.83, SD = 1069.51) compared to Late Juncture sentences (M = 1269.12, SD = 1264.58).
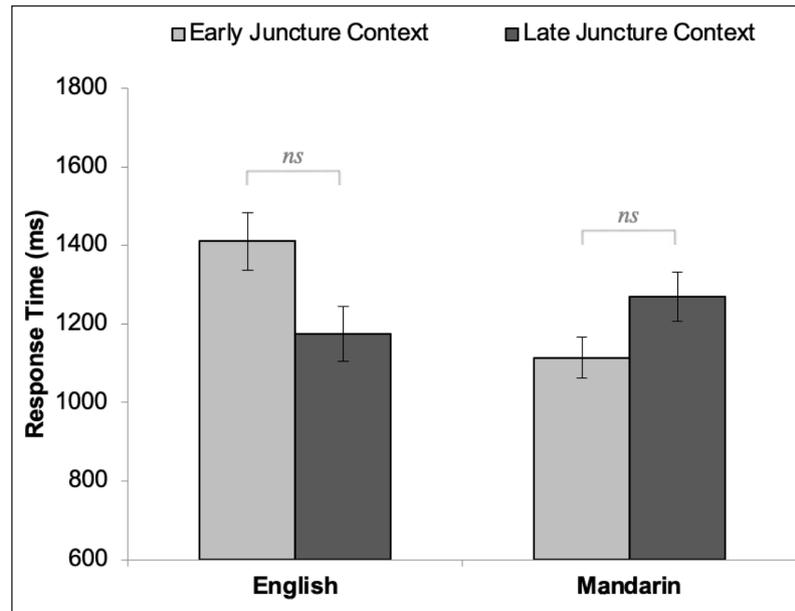
Figure 6: Disambiguation response time (in ms) as a function of Early Juncture and Late Juncture contexts when pause duration cue was removed in Experiment 3. Error bars indicate standard error of the mean.

4.2.2. Accuracy

As in the first perception experiment, we constructed GLMER models to examine listeners' accuracy scores. This time, we found that adding language as a predictor did not significantly improve model fit (χ^2 (1) = 0.24, p = .622; β = –0.18 SE = 0.36, Wald z = –0.50); on average, English listeners averaged 5.75 (SD = 1.49) incorrect responses and Mandarin listeners averaged 5.68 errors (SD = 1.89). It is important to note that Mandarin listeners in this second perception experiment made statistically more disambiguation errors than the Mandarin listeners in the first perception experiment (M = 3.30, SD = 1.82), as revealed in a follow-up pairwise t-test, t(18, 39) = 4.66, p < .001. Mandarin speakers in the present experiment showed an effect of juncture version, with more errors on Early Juncture sentences compared to Late Juncture sentences. There was no overall main effect of juncture version (χ^2 (1) = 1.64, p = .200; β = 0.85, SE = 0.66, Wald z = 1.30) and no significant interaction between language and juncture version (χ^2 (1) = 0.46, p = .500; β = 0.22, SE = 0.32, Wald z = 0.69). The number of detection misses across the juncture version conditions are shown in **Table 5**.

	Mean Errors (*SD*)		
	Early Juncture	**Late Juncture**	**Total**
English	3.08 (1.51)	2.67 (1.50)	5.75 (1.49)
Mandarin	2.37 (1.17)	3.32 (1.86)	5.68 (1.89)

Table 5: Number of incorrect responses in English and Mandarin as a function of Early Juncture and Late Juncture contexts in Experiment 3 (With Pause Duration Removed).

Consistent with the accuracy data in the final sample, the excluded English listeners averaged 9.5 errors (*SD* = .58), spread almost equally across juncture version (*M* = 4.75, *SD* = 1.71). The excluded Mandarin listener had three errors in Early Juncture sentences and six errors in Late Juncture sentences.

4.3. Discussion

Mandarin listeners' disambiguation accuracy was significantly lower when the pausing cue was rendered uninformative. The English listeners, however, showed no significant increase in errors. Thus removal of pausing cues affected the Mandarin listeners' performance accuracy, but had little effect on the English listeners. However, the RT results were non-significant, presumably as a result of the (unavoidably) lower number of participants. Nevertheless, it is noteworthy that the pattern of RT difference between the two juncture versions remained unchanged.

5. General discussion

The present experiments provide new findings on how native speakers of two phonologically very different languages can differ in their use of prosody in juncture processing. Even when utterances involve the very same structural ambiguity, and even when users of these two languages choose generally the same cues to signal a particular reading, the precise deployment of the disambiguating prosody may vary in several ways. In production, therefore, speakers can differ in the degree to which they enhance the various juncture features. In perception, likewise, listeners' disambiguation accuracy and RT patterns can vary across languages for each prosodic effect.

Prosodic boundaries can be signalled by many cues (pausing, pre- and post-boundary vowel lengthening, prosodically conditioned domain-initial segmental strengthening, and pitch variation such as preboundary F_0 lowering and postboundary F_0 reset). Previous production studies have suggested that English and Mandarin use much the same signals (e.g., Beach, 1991; Cooper & Paccia-Cooper, 1980; Keating, Cho, Fougeron, & Hsu, 2004; Kuang, 2010; Li & Yang, 2009; Liberman & Pierrehumbert, 1984; Shen, 1993; Shih, 1988, 2000). Of course, juncture cues are not always produced; speakers unaware of an ambiguity may initiate no disambiguation, and even aware speakers may ignore the need to distinguish a reading if the context is already

informative (e.g., Allbritton et al., 1996; Snedeker & Trueswell, 2003; Straub, 1997). Note, though, that studies with interactive tasks have shown that speakers may spontaneously produce juncture cues even in an unambiguous context (e.g., Kraljic & Brennan, 2005).

Consistent with these latter findings, our English and Mandarin speakers produced juncture cues even though the referential context we provided had rendered the use of prosody in principle unnecessary. Language-specific differences appeared in the degree to which speakers would optionally deploy the different cues to mark juncture: The temporal cues (e.g., pause duration and boundary-related vowel lengthening) were, overall, used to a greater extent by the Mandarin speakers, and Mandarin speakers tended to produce higher F_0. Speakers of different languages can vary in their prosodic choices.

In perception, we revealed a difference in how native English and Mandarin listeners use the prosodic cues to resolve structural ambiguity. First, we observed differences for juncture locations; Mandarin listeners disambiguated Early Juncture sentences faster than Late Juncture, while English listeners were faster (and also more accurate) at disambiguating Late Juncture than Early Juncture sentences. Second, as shown by the accuracy data in Experiment 2, English and Mandarin listeners differed in the degree to which they successfully disambiguated the sentences. Across both Early and Late Juncture contexts, Mandarin listeners were significantly more accurate at disambiguating the sentences compared to English listeners. Third, reliance on different juncture cues (e.g., pausing) also varied; Mandarin listeners became less accurate when pausing was neutralized in Experiment 3, though English listeners' accuracy was unaffected.

Why were there language differences in RT across the Early and Late Juncture contexts? One reason may well be the frequency of these ambiguous structures across languages, in conjunction with findings from work using structural priming; from the latter it has been long known that multiple auditory presentations of sentences with a particular syntactic structure can facilitate processing of subsequent sentences with the same structure (e.g., Carey, Mehler, & Bever, 1970; Mehler & Carey, 1967). Recall that as noted in the introduction, the Early Juncture structure is more frequent in Mandarin than the Late Juncture. Interpretation of the latter in Mandarin is only possible because speakers can omit the genitive particle -*de*. However, whether -*de* is omitted or not depends on a number of factors. Based on a large database of informal written and spoken Mandarin, Chappel and Thompson (1992) identified a number of reasons for the omission. First, Chappel and Thompson showed that -*de* omitted sentences are almost as frequent (at 45%) as -*de* included ones (55%), and inalienable possessions (e.g., body parts) are not always associated with -*de* omitted sentences. Whether speakers choose to omit or include -*de* depends on the conceptual closeness between the possessor and possessee in a given situation (e.g., economic motivation; see also, Haiman, 1983). The degree to which the particle is used also occurs along a continuum with respect to the inherent semantics of the subject and referent. Likewise, there are also pragmatic factors, including the information structure of a conversation and whether the object attached

to the optional particle is topicalized (see also Hsu, 2009). For example, *-de* is more likely to be omitted in the case of given referents; in everyday conversations, once an association between possessor and possessee is established, there is no need to signal it again through the use of *-de*. At the same time, *-de* constructions are syntactically heavy in that they have various functions beyond indicating possession, so processing sentences that involves such a particle (or even the lack of it) may incur extra processing costs. The Late Juncture sentence structure in Mandarin is thus less straightforward than the Early Juncture structure. In our perception experiments, we removed contextual bias by providing listeners with the two possible interpretations before they heard the test sentence; nonetheless, listeners might still be better at accessing a given version of an ambiguous sentence if the use of its structure for a given interpretation is easier to process.

Why were there language differences in listeners' sensitivity to different juncture cues? Again, we suggest that the frequency and strength of a given cue in production is likely to have influenced whether listeners would use it in perception. English and Mandarin speakers differed in their production preferences; our production data showed Mandarin speakers producing greater increases in pause duration than English speakers. Thus again, as a result of their native language experience, Mandarin listeners would be more used to attending to pausing as a juncture cue than English listeners would. This experience is the most likely source of the better disambiguation accuracy in Mandarin. The results of our second perception experiment (Experiment 3), showing that disambiguation accuracy in Mandarin, but not in English, was significantly degraded when pausing cues were absent, support such an interpretation.

Note that our findings resemble previous data of Yang et al. (2014) in which Mandarin listeners showed better Intonational Phrase boundary detection when only pausing was preserved, compared to conditions where preboundary lengthening or F_0 cues were present. Yang and colleagues focused on a more conscious form of boundary detection by adopting a judgement task where listeners had to respond "Yes" or "No" when asked if they heard a boundary. We have extended their findings by showing that Mandarin listeners relied more on pausing under conditions where prosody was the only source of disambiguation information.

Language-specific preference for a given prosodic cue to boundary placement is not the whole story, however; the details of a cue's realization are also part of the native strategy. There is extensive evidence that even when the same cues (e.g., VOT, domain-initial strengthening) are used across languages, the realization may vary (e.g., Byrd et al., 1997; Cho & McQueen, 2005; Kuzla & Ernestus, 2011; Pierrehumbert & Talkin, 1992). In English, both our perceptual findings and existing ERP data (e.g., Aasland & Baum, 2003) indicate that listeners are less reliant on pausing than on other cues. Interestingly, in language development, English-learning infants undergo a developmental change in cue weighting, from attending to all prosodic boundary cues (i.e., pause, pitch, and vowel duration) at three months, to only pitch cues at six months of age (Seidl, 2007; Seidl & Cristià, 2008; see Männel et al., 2013 for similar findings in German).

It is always possible that languages might use potential cues to juncture less if such cues compete with other functions of the same suprasegmental dimension, such as making lexical distinctions. In this respect, English and Mandarin differ. Mandarin has only 29 phonemes (seven vowels, 22 consonants), while General Australian English has 43 (24 consonants and 19 vowels: pp. xv and xvii respectively in Cox, 2012). At least 12 of the 22 Mandarin consonants involve phonemic distinction based on duration (potentially in combination with aspiration, as for VOT); this number is double that for English. Mandarin also has lexical tones, differing in duration, F_0 and amplitude. The tones alone may diminish the likelihood of suprasegmental cues being useful for non-lexical purposes (for a similar suggestion, see Pierrehumbert, 1999). Probably the strongest asymmetry between English and Mandarin is however lexical ambiguity, which, though common in all languages, is particularly rampant in Mandarin due to the small phoneme inventory and severe restrictions on syllable structure (just as an example: /ʂu1/ with high level tone can represent at least 40 words!). Disambiguation is thus more a feature of Mandarin speech processing than of English, and actually inserting a pause between words or phrases is a way of disambiguating sequences without altering either F_0 or segmental durations. Consistent with this, our native Mandarin listeners showed higher rates of disambiguation accuracy in their native language compared to the English participants. In the disambiguation task used in our perception experiments, only prosody could disambiguate the heard sentences. The fact that there were more interpretation errors in English than Mandarin indicates that English listeners may be less likely overall to rely on prosodic juncture cues for disambiguation.

6. Conclusion

Our findings demonstrate that identical structural ambiguity does not entail identical processing. Cues chosen in production can be similar in type but nevertheless different in degree, and perceptual weighting of cues can also differ. All humans may use prosody to segment speech streams into meaningful units, but even when the prosodic cues and the structural options are the same, the ease and the degree to which speakers and listeners use those cues in disambiguation will still show cross-language variability.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Experimental and filler sentences in English. DOI: https://doi.org/10.5334/labphon.6464.s1
- **Appendix B.** Experimental and filler sentences in Mandarin. DOI: https://doi.org/10.5334/labphon.6464.s2
- **Appendix C.** Experimental sentences in Mandarin with interlinear glosses. DOI: https://doi.org/10.5334/labphon.6464.s3
- **Appendix D.** Juncture perception experiment instructions in English. DOI: https://doi.org/10.5334/labphon.6464.s4
- **Appendix E.** Juncture perception experiment instructions in Simplified Chinese. DOI: https://doi.org/10.5334/labphon.6464.s5
- **Appendix F.** Recognition test in English. DOI: https://doi.org/10.5334/labphon.6464.s6
- **Appendix G.** Recognition test in Mandarin. DOI: https://doi.org/10.5334/labphon.6464.s7

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Aasland, W. A., & Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: A comparison of processing by left- and right-hemisphere brain- damaged individuals. *Brain and Language, 87*, 385–399. DOI: https://doi.org/10.1016/S0093-934X(03)00138-X

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 714–735. DOI: https://doi.org/10.1037/0278-7393.22.3.714

Arvaniti, A., & Godjevac, S. (2003). The origins and scope of final lowering in English and Greek. *Proceedings of the 15th International Congress of the Phonetic Sciences* (pp. 1077–1080).

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1563–1571. DOI: https://doi.org/10.1037/a0032186

Barnes, J. (2002). Domain-initial strengthening and the phonetics and phonology of positional neutralization. *North East Linguistics Society, 32*, Article 2.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Language and Memory, 30*, 644–663. DOI: https://doi.org/10.1016/0749-596X(91)90030-N

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11*, 17–67. DOI: https://doi.org/10.1080/016909696387213

Beckman, M. E., & Edwards, J. (1990). Lengthening and shortening and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 152–214). Cambridge, UK: Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255–309. DOI: https://doi.org/10.1017/S095267570000066X

Berkovits R. (1993). Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics. 21*, 476–489. DOI: https://doi.org/10.1016/S0095-4470(19)30231-1

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*, 774–806. DOI: https://doi.org/10.1006/jmla.1995.1034

Bolinger, D. L. (1978). Intonation across languages. In J. Greenberg (Ed.), *Universals of human language II: Phonology* (pp. 471–524). Palo Alto, CA: Stanford University Press.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., et al. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. Trends in Ecology & Evolution, *24*(3), 127–135. DOI: https://doi.org/10.1016/j.tree.2008.10.008

Bombien, L., Mooshammer, C., Hoole, P., Rathcke, T., & Kuhnert, B. (2007). Articulatory strengthening in initial German /kl/ clusters under prosodic variation. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 457–460). Saarbrücken, Germany.

Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics: The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In C. Fougeron, B. Kuehnert, M. D'Imperio & N. Vallee (Eds.), *Papers in Laboratory Phonology 10.* Mouton de Gruyter.

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America, 120*, 1589–1599. DOI: https://doi.org/10.1121/1.2217135

Byrd, D., Narayanan, S., Kaun, A., & Saltzman, E. (1997). Phrasal signatures in articulation. In *Proceedings of Laboratory Phonology V* (pp. 70–87). Cambridge University Press.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics, 26*, 173–199. DOI: https://doi.org/10.1017/S0952675700001019

Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of Dutch. In J. Coerts & H. de Hoop. (Eds.), *Linguistics in the Netherlands* (pp. 13–24). Amsterdam, The Netherlands: John Benjamins. DOI: https://doi.org/10.1075/avt.14.04cam

Campbell, W. N., & Isard, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics, 19*, 37–47. DOI: https://doi.org/10.1016/S0095-4470(19)30315-8

Carey, P. W., Mehler, J., & Bever, T. (1970). Judging the veracity of ambiguous sentences. *Journal of Verbal Learning and Verbal Behavior, 9*, 243–254. DOI: https://doi.org/10.1016/S0022-5371(70)80058-5

Carlson, R., Hirschberg, J., & Swerts, M. (2005). Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication, 46*, 326–333. DOI: https://doi.org/10.1016/j.specom.2005.02.013

Chappell, H., & Thompson, S. A. (1992). The semantics and pragmatics of associative DE in Mandarin discourse. *Cahiers de linguistique – Asie orientale, 21*, 199–229. DOI: https://doi.org/10.1163/19606028-90000330

Cho, T., & Jun, S. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society, 36*, 31–44.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190. DOI: https://doi.org/10.1006/jpho.2001.0131

Cho, T., & Keating, P. A. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics, 37*, 466–485. DOI: https://doi.org/10.1016/j.wocn.2009.08.001

Cho, T., Lee, Y., & Kim, S. (2011). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics, 46*, 128–146. DOI: https://doi.org/10.1016/j.wocn.2014.06.003

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics, 33*, 121–157. DOI: https://doi.org/10.1016/j.wocn.2005.01.001

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler., J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language, 51*, 523–547. DOI: https://doi.org/10.1016/j.jml.2004.07.001

Clements, G. N., & Ford, K. C. (1981). On the Phonological Status of Downstep in Kikuyu. In D. L. Goyvaerts (Ed.), *Phonology in the 1980's* (pp. 309–357). Ghent, Belgium: Story- Scientia. DOI: https://doi.org/10.1075/ssls.4.14for

Cooper, A. M. (1991). An Articulatory Account of Aspiration in English. Doctoral dissertation, Yale University.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press. DOI: https://doi.org/10.4159/harvard.9780674283947

Cox, F. (2012). *Australian English pronunciation and transcription.* Port Melbourne, Australia: Cambridge University Press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. *Natural Language Parsing,* 320–358. DOI: https://doi.org/10.1017/CBO9780511597855.011

Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30,* 73–105. DOI: https://doi.org/10.1016/0010-0277(88)90004-2

Cutler, A. (1987). Components of prosodic effects in speech recognition. *Proceedings of the Eleventh International Congress of Phonetic Sciences, Tallinn, Estonia, 1,* 84–87.

Cutler, A., & Isard, S. D. (1980). The production of prosody. In B. Butterworth (Ed.), Language production (pp. 245–269). London: Academic Press.

Degenshein, R., & Chitoran, I. (2004). Dholuo interdentals: Fricatives or affricates? Evidence from domain-initial strengthening. *Journal of the Acoustical Society of America, 115,* 2542. DOI: https://doi.org/10.1121/1.4783613

Dekydtspotter, L., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrusch, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English- French learners. *Studies in Second Language Acquisition, 30,* 453–480. DOI: https://doi.org/10.1017/S0272263108080728

Dilley, L. C., & Shattuck-Hufnagel, S. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24,* 423–444. DOI: https://doi.org/10.1006/jpho.1996.0023

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology, 61,* 177–199. DOI: https://doi.org/10.1016/j.cogpsych.2010.05.001

Fernández E. M. (2007). How might a rapid serial visual presentation of text affect the prosody projected implicitly during silent reading? In *Conferências do V Congreso Internacional da Associaçao Brasiliea de Lingüistica Belo Horizonte.*

Fodor, J. D. (1998). Learning to parse. *Journal of Psycholinguistic Research, 27,* 285–319. DOI: https://doi.org/10.1023/A:1023258301588

Fougeron, C. (1999). Articulation of prosodic phrasing in French. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 675–678). San Francisco, USA.

Fougeron, C., & Keating, P. (1996). Variations in velic and lingual articulation depending on prosodic position: Results for two French speakers. *UCLA Working Papers in Phonetics, 92,* 88–96.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101,* 3728–3740. DOI: https://doi.org/10.1121/1.418332

Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*, 244–249. DOI: https://doi.org/10.1016/j.tics.2006.04.002

Fromont, L. A., Soto-Faraco, S., & Biau, E. (2017). Searching high and low: Prosodic breaks disambiguate relative clauses. *Frontiers in Psychology, 8*, 96. DOI: https://doi.org/10.3389/fpsyg.2017.00096

Frota, S., D'Imperio, M., Elodieta, G., Prieto, P., & Vigáro, M. (2007). The phonetics and phonology of intonational phrasing in Romance. In P. Prieto (Ed.), *Segmental and prosodic issues in Romance Phonology* (pp. 131–154). Amsterdam, the Netherlands: John Benjamins. DOI: https://doi.org/10.1075/cilt.282.10fro

Georgeton, L., Antolik, T. K., & Fougeron C. (2016). Effect of domain initial strengthening on vowel height and backness contrasts in French: Acoustic and ultrasound data. *Journal of Speech, Language and Hearing Research, 59*, 1575–1586. DOI: https://doi.org/10.1044/2016_JSLHR-S-15-0044

Georgeton, L., & Fougeron, C. (2014). Domain initial strengthening of French vowels and phonological contrasts: Evidence from lip articulation and spectral variation. *Journal of Phonetics, 44*, 83–95. DOI: https://doi.org/10.1016/j.wocn.2014.02.006

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech, 15*, 103–113. DOI: https://doi.org/10.1177/002383097201500201

Gordon, M. (1996). Phonetic correlates of stress and the prosodic hierarchy in Estonian. In J. Ross & I. Lehiste (Eds.), *Estonian prosody: Papers from a symposium.* (pp. 100–124). Institute of Estonian language, Tallinn, Estonia.

Grabe, E. (1998). Comparative intonational phonology: English and German. Doctoral dissertation, Universiteit Nijmegen, Nijmegen, The Netherlands.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica, 31*, 144–184. DOI: https://doi.org/10.1159/000259667

Grosjean. F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology, 11*, 58–81. DOI: https://doi.org/10.1016/0010-0285(79)90004-5

Gussenhoven, C., & Rietveld, A. C. M. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics, 16*, 355–369. DOI: https://doi.org/10.1016/S0095-4470(19)30509-1

Harris, M. S., & Umeda, N. (1974). Umeda Effect of speaking mode on temporal factors in speech: Vowel duration. *Journal of the Acoustical Society of America, 56*, 1016–1018. DOI: https://doi.org/10.1121/1.1903366

Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech, 14*, 277–288. DOI: https://doi.org/10.1177/002383097101400308

Hayashi, W., Hsu, C., & Keating, P. (1999). Domain-initial strengthening in Taiwanese: A follow-up study. *UCLA Working Papers in Phonetics, 97*, 152–156.

Herman, H. (1996). Final lowering in Kipare. *Phonology, 13*, 171–196. DOI: https://doi.org/10.1017/S0952675700002098

Hockey, B. A., & Zsuzsanna, F. (1998). Pre-boundary lengthening: Universal or language-specific? The case of Hungarian. *University of Pennsylvania Working Papers in Linguistics 5.1*, 71–82.

Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Räling, R., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2016). How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience, 31*, 904–920. DOI: https://doi.org/10.1080/23273798.2016.1157195

Horne, M., Strangert, E., & Heldner, M. (1995). Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. In K. Elenius & P. Branderud (Eds.), *Proceedings of the International Congress of Phonetic Sciences* (pp. 170–173). Stockholm, Sweden.

Hsu, Y.-Y. (2009). Possessor extraction in Mandarin Chinese. *University of Pennsylvania Working Papers in Linguistics, 15*, 95–104.

Ip, M. H. K., & Cutler, A. (2020). Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages. *Cognition, 202*, 104311. DOI: https://doi.org/10.1016/j.cognition.2020.104311

Jepson, K., Fletcher, J., & Stoakes, H. (2019). Prosodically conditioned consonant duration in Djambarrpuyŋu. *Language and Speech*. DOI: https://doi.org/10.1177/0023830919826607

Johnson, E. K. (2016). Constructing a proto-lexicon: An integrative view of infant language development. *Annual Review of Linguistics, 2*, 391–412. DOI: https://doi.org/10.1146/annurev-linguistics-011415-040616

Jun, J., Kim, J., Lee, H., & Jun, S.-A. (2004). The prosodic structure of Northern Kyungsang Korean. *Proceedings of Speech Prosody.*

Jun, S. A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology, 15*, 189–226. DOI: https://doi.org/10.1017/S0952675798003571

Jun, S.-A. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research, 32*, 219–249. DOI: https://doi.org/10.1023/A:1022452408944

Katsika, A. (2009). Boundary- and prominence-related lengthening and their interaction. *Journal of the Acoustical Society of America, 125*, 2572–2572. DOI: https://doi.org/10.1121/1.4783765

Katsika A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics, 55*, 149–181. DOI: https://doi.org/10.1016/j.wocn.2015.12.003

Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology VI* (pp. 143–161). Cambridge, UK: Cambridge University Press.

Kim, D., Stephens, J. D. W., & Pitt, M. A. (2012). How does context play a part in splitting words apart? Production and perception of word boundaries in casual speech. *Journal of Memory and Language, 66*, 509–529. DOI: https://doi.org/10.1016/j.jml.2011.12.007

Kim, J. E. (2019). Acoustic characteristics of read and spontaneous speech in Seoul Korean with between-age variability. *Korean Linguistics, 85,* 61–76. DOI: https://doi.org/10.20405/kl.2019.11.85.61

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1208–1221. DOI: https://doi.org/10.1121/1.380986

Kohler, K. (1983). Prosodic boundary signals in German. *Phonetica 40*, 89–134. DOI: https://doi.org/10.1159/000261685

Kohler, K. J., Peters, B., & Scheffers, M. (2017). The Kiel Corpus of spoken German: Read and spontaneous speech. Retrieved from https://www.isfas.uni-kiel.de/de/linguistik/forschung/kiel-corpus/the-kiel-corpus-of-spoken-german-read-and-spontaneous-speech on December 12, 2019.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194–231. DOI: https://doi.org/10.1016/j.cogpsych.2004.08.002

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, 162–179. DOI: https://doi.org/10.1016/j.wocn.2006.04.001

Krull, D. (1997). Prepausal Lengthening in Estonian: Evidence from Conversational Speech. In I., Lehiste & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 136–148). Tallinn, Estonia: Institute of Estonian Language.

Kuang, J. (2010). Prosodic grouping and relative clause disambiguation in Mandarin. In T. Kobayashi, K. Hirose & S. Nakamura (Eds.), *Proceedings of INTERSPEECH*. Makuhari, Japan. DOI: https://doi.org/10.21437/Interspeech.2010-501

Kuzla, C., Cho, T., & Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics, 35*, 301–320. DOI: https://doi.org/10.1016/j.wocn.2006.11.001

Kuzla, C., & Ernestus, M. (2011). Prosodic conditioning of phonetic detail in German plosives. *Journal of Phonetics, 39*, 143–155. DOI: https://doi.org/10.1016/j.wocn.2011.01.001

Ladd, D. R. (1986). Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook, 3*, 311–340. DOI: https://doi.org/10.1017/S0952675700000671

Ladd, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America, 84*, 530–544. DOI: https://doi.org/10.1121/1.396830

Laniran, Y. O. (1992). Intonation in tone languages: The phonetic implementation of tones in Yoruba. Doctoral dissertation, Cornell University, USA. DOI: https://doi.org/10.1121/1.1913062

Lehiste, I. (1972). Timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America, 51*, 2018–2024.

Lenth, R. V. (2020). Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.2-1. https://CRAN.R-project.org/package=emmeans

Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience, 158*, 1416–1425. DOI: https://doi.org/10.1016/j.neuroscience.2008.10.065

Liberman, A. M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249–336.

Liberman, M. Y., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle* (pp. 157–233). Cambridge, MA: MIT Press.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Paper of the Linguistic University of Stockholm, 21*, 1–59.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyze reaction time data. *Frontiers in Psychology, 30*, 1171. DOI: https://doi.org/10.3389/fpsyg

Lyberg, B. (1977). Some observations on the timing of Swedish utterances. *Journal of Phonetics, 5*, 49–59. DOI: https://doi.org/10.1016/S0095-4470(19)31113-1

Männel, C., & Friederici, A. D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience, 21*, 1988–2006. DOI: https://doi.org/10.1162/jocn.2009.21221

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience, 5*, 86–94. DOI: https://doi.org/10.1016/j.dcn.2013.01.003

Mehler, J., & Carey, P. W. (1967). The role 'of surface and base structure in the perception of sentences. *Journal of Verbal Learning and Verbal Behavior, 6*, 335–338. DOI: https://doi.org/10.1016/S0022-5371(67)80122-1

Michelas, A., & D'Imperio, M. (2012). When syntax meets prosody: Tonal and duration variability in French Accentual Phrases. *Journal of Phonetics, 40*, 816–829. DOI: https://doi.org/10.1016/j.wocn.2012.08.004

Nakai, S., Turk, A. E., Suomi, K., Granlund, S., Ylitalo, R., & Kunnari, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics, 40*, 796–807. DOI: https://doi.org/10.1016/j.wocn.2012.08.003

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

Onaka, A. (2003). Domain-initial strengthening in Japanese: An acoustic and articulatory study. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 2091–2094). Barcelona, Spain.

Onaka, A., Watson, C., Palethorpe, S., & Harrington, J. (2003). An acoustic analysis of domain-initial strengthening effect in Japanese. In S. Palethorpe & M. Tabain (Eds.), *Proceedings of the sixth international seminar on speech production* (pp. 201–206). Sydney, Australia.

Peng, S.-H. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics, 25*, 371–400. DOI: https://doi.org/10.1006/jpho.1997.0047

Peterson, R. A., & Cavanaugh, J. E. (2019). Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, 1–16. DOI: https://doi.org/10.1080/02664763.2019.1630372

Pierrehumbert, J. (1999). Prosody and intonation. In R. A. Wilson & F. C. Keil (Eds.), *MIT encyclopedia of cognitive science* (pp. 679–682). Cambridge, MA: MIT Press.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Doherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture segment prosody* (pp. 90–117). Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511519918.005

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956–2970. DOI: https://doi.org/10.1121/1.401770

Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish, *Journal of Phonetics, 24*, 445–473. DOI: https://doi.org/10.1006/jpho.1996.0024

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20*, 331–350. DOI: https://doi.org/10.1016/S0095-4470(19)30638-2

Sanderman, A. A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech, 40*, 391–409. DOI: https://doi.org/10.1177/002383099704000405

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 2.0). [Computer software and manual]. Pittsburgh, PA Psychology Software Tools Inc.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57*, 24–48. DOI: https://doi.org/10.1016/j.jml.2006.10.004

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science, 11*, 596–606. DOI: https://doi.org/10.1111/j.1467-7687.2008.00704.x

Selkirk, E. O. (2003). The prosodic structure of function words. In J. McCarthy (Ed.), *Optimality theory in phonology: A reader* (pp. 464–482). Malden, MA: Blackwell Publishing. DOI: https://doi.org/10.1002/9780470756171.ch25

Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of phrase-final lengthening in English. *Journal of the Acoustical Society of America, 103*, 2889–2889. DOI: https://doi.org/10.1121/1.421798

Shaw, J. A., Best, C. T., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., & Mulak, K. E. (2018). Resilience of English vowel across regional accent variation. *Laboratory Phonology, 9*, 1–36. DOI: https://doi.org/10.5334/labphon.87

Shen, X. S. (1993). The use of prosody in disambiguation in Mandarin. *Phonetica, 50*, 261–271. DOI: https://doi.org/10.1159/000261946

Shepard, M. A. (2008). The scope and effects of preboundary prosodic lengthening in Japanese. *USC Working Papers in Linguistics, 4*, 1–14.

Shih, C. (1988). Tone and intonation in mandarin. *Working Papers, Cornell Phonetics Laboratory, 3*, 83–109. Retrieved from https://ci.nii.ac.jp/naid/10022356702

Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation: Analysis modeling and technology* (pp. 243–268). Dordrecht: Kluwer Academic Publishers. DOI: https://doi.org/10.1007/978-94-011-4317-2_11

Silverman, K. (1990). The separation of prosodies: Comments on Kohler's paper. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech* (pp. 139–151) Cambridge, U.K.: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511627736.008

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential contest. *Journal of Memory and Language, 48*, 103–130. DOI: https://doi.org/10.1016/S0749-596X(02)00519-3

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language, 48,* 233–254. DOI: https://doi.org/10.1016/S0749-596X(02)00513-2

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2,* 191–196. DOI: https://doi.org/10.1038/5757

Straub, K. A. (1997). The production of prosodic cues and their role in the comprehension of syntactically ambiguous sentences. Doctoral dissertation, University of Rochester, Rochester, NY, USA.

Streeter, L. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America, 64,* 1582–1592. DOI: https://doi.org/10.1121/1.382142

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America, 101,* 514–521. DOI: https://doi.org/10.1121/1.418114

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in spontaneous and read-aloud speech. *Proceedings of the fourth international conference on spoken language (ICSLP 96)*. DOI: https://doi.org/10.1109/ICSLP.1996.607901

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Acoustic results. *Journal of the Acoustical Society of America, 113,* 516–531. DOI: https://doi.org/10.1121/1.1523390

Takeda, K., Sagisaka, Y., & Kuwabara, H. (1989). On sentence-level factors governing segmental duration in Japanese. *The Journal of the Acoustical Society of America, 86,* 2081. DOI: https://doi.org/10.1121/1.398467

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science, 268,* 1632–1634. DOI: https://doi.org/10.1126/science.7777863

Teira, C., & Igoa, J. M. (2007). Relaciones entre la prosodia y la sintaxis en el procesamiento de oraciones. *Annuario del Psicología, 38,* 45–69.

Thorsen, N. G. (1985). Intonation and text in Standard Danish. *Journal of the Acoustical Society of America, 77,* 1205–1216. DOI: https://doi.org/10.1121/1.392187

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35,* 445–472. DOI: https://doi.org/10.1016/j.wocn.2006.12.001

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53–66). Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-642-69103-4_5

Wang, C., Xu, Y., & Zhang, J. (2019). Mandarin and English use different temporal means to mark major prosodic boundaries. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of 19th International Congress of Phonetic Sciences*. Melbourne, Australia.

Wang, S. F., & Fon, J. (2012). Durational cues at discourse boundaries in Taiwan Southern Min. In *Proceedings of Speech Prosody*. Shanghai, China.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 92,* 1707–1717. DOI: https://doi.org/10.1121/1.402450

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33*, 319–337. DOI: https://doi.org/10.1016/S0167-6393(00)00063-7

Yang, X., Shen, X., Li, W., & Yang, Y. (2014). How listeners weight acoustic cues to intonational phrase boundaries. *PLoS ONE, 9*, e102166. DOI: https://doi.org/10.1371/journal.pone.0102166

Yeo, I.-K., & Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87*, 954–959. DOI: https://doi.org/10.1093/biomet/87.4.954

Yu, J., & Tao. J. (2005). The pause duration prediction for Mandarin text-to-speech system. In *IEEE International Conference on Natural Language Processing and Knowledge Engineering* (pp. 204–208). Wuhan, China.

Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin Broadcast News Speech. *Speech Communication, 65*, 67–74. DOI: https://doi.org/10.1016/j.specom.2014.06.001

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early-closure attachment on first-pass reading times in French. *Quarterly Journal of Experimental Psychology, 50*, 421–438. DOI: https://doi.org/10.1080/713755715