Universal and Language-specific Processing: The Case of Prosody

Martin Ho Kwan Ip

BPsycSc (Hons I with University Medal), UQ

DipLang (French), UQ

A thesis submitted for the degree of Doctor of Philosophy

The MARCS Institute for Brain, Behaviour and Development
The Australian Research Council Centre of Excellence for the Dynamics of Language
Western Sydney University

January 2019

Supervisory Committee

Distinguished Professor Anne Cutler (Primary Supervisor)

Assistant Professor Jason Anthony Shaw (Co-supervisor 1)

Doctor Mark Antoniou (Co-supervisor 2)

Abstract

A key question in the science of language is how speech processing can be influenced by both language-universal and language-specific mechanisms (Cutler, Klein, & Levinson, 2005). My graduate research aimed to address this question by adopting a crosslanguage approach to compare languages with different phonological systems. Of all components of linguistic structure, prosody is often considered to be one of the most language-specific dimensions of speech. This can have significant implications for our understanding of language use, because much of speech processing is specifically tailored to the structure and requirements of the native language. However, it is still unclear whether prosody may also play a universal role across languages, and very little comparative attempts have been made to explore this possibility.

In this thesis, I examined both the production and perception of prosodic cues to prominence and phrasing in native speakers of English and Mandarin Chinese. In focus production, our research revealed that English and Mandarin speakers were alike in how they used prosody to encode prominence, but there were also systematic language-specific differences in the exact degree to which they enhanced the different prosodic cues (**Chapter 2**). This, however, was not the case in focus perception, where English and Mandarin listeners were alike in the degree to which they used prosody to predict upcoming prominence, even though the precise cues in the preceding prosody could differ (**Chapter 3**). Further experiments examining prosodic focus prediction in the speech of different talkers have demonstrated functional cue equivalence in prosodic focus detection (**Chapter 4**). Likewise, our experiments have also revealed both crosslanguage similarities and differences in the production and perception of juncture cues (**Chapter 5**). Overall, prosodic processing is the result of a complex but subtle interplay of universal and language-specific structure.

Statement of Authentication

I hereby declare that this thesis is composed of my original work, and contains no research works conducted and published by another person except where due reference is made in the text. The content of my thesis is the result of the research experiments I have carried out since the commencement of my graduate degree by research candidature. I have not submitted this material, either in part or in full, to qualify for the award of any other degree or diploma at Western Sydney University or at any other university or tertiary institution.

[redacted]

Martin Ho Kwan Ip (葉晧鈞)

26 January 2019

Publications

Journals

Ip, M. H. K., Imuta, K., & Slaughter, V. (2018). Which button will I press? Preference for correctly ordered counting sequences in 18-month-olds. *Developmental Psychology, 54*, 1199-1207. doi: 10.1037/dev0000515

Ip, M. H. K., Shaw, J. A., & Cutler, A. (submitted). Prosodic strategies of focus expression across languages

Ip, M. H. K., & Cutler, A. (under review). Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages

Ip, M. H. K., & Cutler, A. (under review). In search of salience: Focus detection in the speech of different talkers

Ip, M. H. K., & Cutler, A. (in prep). Prosodic cues to juncture in production and perception: A crosslanguage perspective

Conference Abstracts and Proceedings

* = Oral Presentation, † = Proceedings

*†**Ip, M. H. K.,** & Cutler, A. (2018). Cue equivalence in prosodic entrainment for focus detection. In J. Epps, J. Wolfe, J. Smith, & C. Jones (Eds.), *Proceedings of the 17th Australasian International Conference on Speech Science and Technology* (pp. 153-156), Sydney, Australia: ASSTA.

*†**Ip, M. H. K.,** & Cutler, A. (2018). Asymmetric efficiency of juncture perception in L1 and L2. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 289-296). Poznań, Poland: ISCA. doi:10.21437/SpeechProsody.2018-59

Ip, M. H. K., & Cutler, A. (2017). Crosslanguage experiments on the production and perception of prosody. *Architectures and Mechanisms for Language Processing*, Lancaster University, UK, September.

*†**Ip, M. H. K.,** & Cutler, A. (2017). Intonation facilitates prediction of focus even in the presence of lexical tones. In F. Lacerda, S. Strombergsson, M. Wlodarczak, M. Heldner, J. Gustafson, & D. House (Eds.), *Proceedings of the 18th Annual Conference of the International Speech Communication Association* (pp. 1218-1222). Stockholm, Sweden: ISCA. doi:10.21437/Interspeech.2017-264

Ip, M. H. K., & Cutler, A. (2017). Prosodic strategies of information structure. *ARC Centre of Excellence Summer Meeting 2017*, Mooloolaba, Australia, February.

Ip, M. H. K., & Cutler, A. (2016). Language-specificity in speakers' strategies of focus expression. *Abstracts of Laboratory Phonology 15*, Cornell University, Ithaca, NY, July.

*†**Ip, M. H. K.,** & Cutler, A. (2016). Cross-language data on five types of prosodic focus. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of the 8th International Conference on Speech Prosody* (pp. 330-334). Boston, USA: ISCA. doi: 10.21437/SpeechProsody.2016-68

Ip, M. H. K., & Cutler, A. (2016). Prosodic strategies of focus. *ARC Centre of Excellence Summer Meeting 2016*, Western Sydney University, Sydney, Australia, February.

Ip, M. H. K., & Slaughter, V. (2013). Infants' understanding of counting. Biennial Conference of the Australian Human Development Association, Gold Coast, Australia, June.

*Slaughter, V., & **Ip, M. H. K.** (2013). Infants' understanding of counting. Biennial Meeting of the Society of Research in Child Development, Seattle, USA, April.

***Ip, M. H. K.,** & Slaughter, V. (2012). Infants' understanding of counting. UQ Undergraduate Research Conference, The University of Queensland, St Lucia, Australia, September.

Research Ethics

Ethical approval and amendments for this research (Project No. H11216) were granted by the Western Sydney University Human Research Ethics Committee. All experiments were conducted at the MARCS Institute, Western Sydney University. A copy of the approval letter is provided in Appendix A.

Financial Support

I would not have been able to conduct my graduate research without the generous research funding from my primary supervisor and the financial support provided by the MARCS Institute. This research was supported by both the MARCS Institute and the Australian Research Council Centre of Excellence for the Dynamics of Language [CE140100041]. The MARCS Institute and the Australian Research Council have also funded all of my international and domestic conference travels, conference meals, registration fees, and summer coursework (Canberra in 2015; Boston, New York, and Melbourne in 2016; Mooloolaba, Stockholm, Edinburgh, and Lancaster in 2017; Poznań and Sydney in 2018). I would also like to thank the administrative staff at the MARCS Institute for managing the paperwork, particularly Darlene Williams, Craig Purcell, and Siobhain O'Leary. Also very importantly, I would like to thank the Australasian Speech Science and Technology Association (ASSTA) for providing me with the ASSTA Travel Award in 2017 and 2018. As for my living expenses, I am very grateful to be funded by both the Australian Postgraduate Award (APA) from the Australian Federal Government and the Top-Up Award from Western Sydney University.

Acknowledgements

This thesis would not have been possible without all the people who have supported me throughout my research training. First of all, I owe my deepest gratitude to my primary supervisor, Anne Cutler. As a mentor, Anne has been a source of inspiration, both academically and personally. I would like to thank Anne for opening this new and fascinating research journey for me, for giving me the freedom to pursue my own interests, for constantly giving me her honest advice, for helping me realise my mistakes, for helping me make better talks (and give better answers), answering my emails (even during the weekends and holidays), and for many countless things. But most importantly, I want to thank Anne for being my mentor. Working with a scientist like Anne has been one of my greatest joys as a student and it is a memory that I will always cherish.

At the same time, I would also like to thank my co-supervisors, Jason Shaw and Mark Antoniou. Jason has been my co-supervisor when I first started my degree at MARCS and was still part of my supervisory committee after he moved to the US. As a student with no undergraduate linguistics training, I benefited greatly from Jason's expert supervision. I am very grateful to Jason for helping me become acquainted with the field and for sharing all his ideas and constructive feedback. I really enjoyed discussing my research results with Jason. Mark became my co-supervisor during my second year at MARCS, but even before then, he was already like an unofficial co-supervisor. My study space was right outside his office, and whenever I approached him, he was always free to help me out with my acoustic analyses or E-Prime set-ups. And thanks to his technical advice, I was able to find the best equipment for our experiments. I cannot ask for a better supervisory panel!

I would also like to take the opportunity here to thank Paola Escudero and Chris Carignan. Paola was briefly on my panel before her maternity leave, and she was also the first person I met on my first day at MARCS. Paola was very kind to show me around and introduce me to all the technical team and students. Chris had on more than a few occasions offered his help in very testing situations when I couldn't find anyone else (like when there was a sudden program crash just before my experiments).

Throughout my candidature, the technical team has also been a source of support. In particular, I wish to thank Colin Schoknecht, Steven Fazio, and Johnson Chen, for providing me with the best facilities and advice. I would also like to thank Ma Jiayi from the humanities department. I have only met Ma Jiayi as a total stranger outside the Face and Voice Lab where I conducted my experiments. She agreed to translate my English written instructions into Chinese and managed to make them look really professional and still mean exactly the same thing. Even as native speakers, translating all these instructions was not easy. Her help was invaluable.

Speaking of help from strangers, I would also like to acknowledge all the people who made it physically possible for me to conduct my experiments – my research participants. I wish to thank all of my participants who embraced the research process with patience, enthusiasm, and willingness! I would also like to thank many of my Mandarin participants who introduced my research to other people, in particular, Zhang Yong and Cheng Cheng. Their kind and generous support has helped me recruit many people within an extremely short timeframe. Also, I would like to thank Han Chong, also from the humanities department, for allowing me to hand out my recruitment flyers at the start of her lecture. And finally, I would like to thank the Western Sydney University Student Clubs Officer Matthew Stansfield, for giving me the permission to set up the "Speech Science Research" club. Without this club, I would not have met so many people. I will never forget the wonderful time I had during Clubs Week and the chocolate fountains at our end-of-year party!

I would like to express my heartfelt thanks to all of my fellow students for their supportive network. I should also note that many of the people who kindly volunteered in my research (e.g., pilot studies) were my fellow students, and I am also very grateful to the people who spent over an hour recording the sentences and instruction voiceovers for my research stimuli. These people were Sarah Wright, Sarah Fenwick, Stacey Sherwood, Sandra Garrido (post-doc), and Jiang Nan. They have all sacrificed their time outside of their working hours to help me at the recording booth and I sometimes made them repeat the same sentences several times. I am very lucky to have volunteers who are so patient. And I couldn't ask for a better set of stimuli!

I would also like to thank Cathi Best for agreeing to chair my candidature confirmation and Elizabeth Johnson for agreeing to be the external reviewer. I would like to thank them for being part of my confirmation panel at a time when we were desperate to begin our research.

Finally, I would like to acknowledge all the people outside of MARCS who provided constructive criticisms both before and after my data collection. In particular, I would very much wish to thank Bob Ladd, for spending two days giving me his feedback on all the data I collected during my candidature. During my first year, Michael Yip gave me helpful suggestions. I really enjoyed talking with him. Also when I started my degree, I met Janet Wiles at the CoEDL summer meeting. I wish to thank Janet for giving me important tips on how to formulate my research questions as a scientist. Very importantly, I would like to express my gratitude to my Honours supervisor, Virginia Slaughter. When I was an undergrad at UQ, it was Virginia who first inspired me to do research.

Above all, I am eternally indebted to my mother, 明美玲, and my father, 葉偉仁.

I would like to dedicate this thesis to my mother, 明美玲.

TABLE OF CONTENTS

List of Figures

Chapter 2: Prosodic Strategies of Focus Expression Across Languages

Chapter 3: Universals of Listening

List of Figures (cont.)

Chapter 4: In Search of Salience

List of Figures (cont.)

Chapter 5: Prosodic Cues to Juncture

List of Figures (cont.)

Chapter 5: Prosodic Cues to Juncture

List of Tables

Chapter 2: Prosodic Strategies of Focus Expression Across Languages

Chapter 3: Universals of Listening

List of Tables (cont.)

Chapter 4: In Search of Salience

List of Tables (cont.)

Chapter 5: Prosodic Cues to Juncture

List of Symbols and Abbreviations

# – prosodic juncture

ANOVA – analysis of variance

CI – confidence interval

ERP – event-related potential

F_0 – fundamental frequency

H – high tone

H* – single-tone high pitch accent type

Hz – hertz

IPA – International Phonetic Alphabet

L – low tone

L* – single-tone low pitch accent type

L+H* – bitonal rising pitch accent type

L1 – first language; native language

L2 – second language; non-native language

ms – milliseconds

RMS – root mean square

RT – response time

T1 – high-level tone

T2 – high rising tone

T3 – low-dippng tone

T4 – high-falling tone

T5 – neutral tone

VOT – voice onset time

CHAPTER 1

– General Overview –

– General Overview –

1.1. “Around the Edge of Language”

Human language is a system of astounding complexity. With only a meagre set of phonemes, a limited repertoire of articulatory gestures, and a finite grammar, hundreds of thousands of words can be constructed and combined to generate an infinite range of expressions. For this reason, the speech signal is never immediately transparent. Almost every spoken utterance we encounter in our conversations will be a new utterance, and almost every word will resemble or occur embedded within another word. At the same time, speech in all languages is fast, continuous, transitory, and highly variable. In the face of so much uncertainty, how do listeners convert such a messy and complex string of sounds into meaningful words and sentences?

Over the past decades, there has been an explosion of experimental discoveries on the way speech is decoded (Cutler, 2012). Perception of spoken language involves a formidable array of processing operations, including mental tasks where listeners must distinguish speech from other auditory inputs (e.g., Bregman, 1990; Darwin, 1984; 2007), detect boundaries between clauses, words, syllables, and morae (e.g., Cutler & Norris, 1988; Hirsh-Pasek et al., 1987; Norris, McQueen, & Cutler, 1997; Otake, Hatano, Cutler, & Mehler, 1993), make statistical inferences from the structure of the native lexicon (e.g., Cutler, Otake, & Bruggeman, 2012; Mirman, Magnusun, Graf Estes, & Dixon, 2008), entertain multiple hypotheses about possible word candidates (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986; McQueen, Norris, & Cutler, 1999), use coarticulatory information to anticipate upcoming sound forms (e.g., Gow & McMurray, 2007; Salverda, Kleinschmidt, & Tanenhaus, 2014), adapt to variations in the acoustic-phonetic productions of different speakers (e.g., Mullennix & Pisoni, 1990; Sjerps, Mitterer, & McQueen, 2011), predict syntactic structures (e.g., Arai, van Gompel, & Scheepers, 2007; Kazanina, 2017; Pickering & Ferreira, 2008), and relate utterances to the wider discourse (e.g., van Berkum, Zwitserlood, Hagoort, & Brown, 2003). Many of these tasks can be achieved by adopting both language-universal constraints based on syllabic structures (e.g., Sonority Sequencing Principle: Gómez et al., 2014) or patterning of vowels and consonants (e.g., Possible-Word-Constraint: Cutler, Demuth, & McQueen, 2002; Norris, McQueen, Cutler, & Butterfield, 1997), as well as strategies relevant to specific features of the native language, such as coarticulatory word-onset variations (e.g., Davis, Marslen-Wilson, & Gaskell, 2002), phonotactic or allophonic regularities (e.g., Christiansen, Allen, & Seidenberg, 1998; Juscyzk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; McQueen, 1998; Vitevitch & Luce, 1999), and transitional probabilities between syllables (e.g., Saffran, Aslin, & Newport, 1996). Likewise, knowledge-based processing from higher-level domains such as syntax (e.g., verb argument structure) and semantics (e.g., word frequency, lexical neighbourhood) has also been shown to play various roles in speech perception, including phoneme restoration (e.g., Samuel, 2001), word segmentation (e.g., Gaskell & Marslen-Wilson, 1997; Mattys, Melhorn, & White, 2007), and lexical selection and disambiguation (e.g., Altmann & Kamide, 1999; Seidenberg, Tanenhaus, Leiman, & Bienkowsky, 1982; Garlock, Walley, & Metsala, 2001; Goldinger, Luce, & Pisoni, 1989; Storkel, Armbruster & Hogan, 2006).

However, beyond the segmental level, much less research has focused on the role of prosody*. Very few attempts have been made to uncover the possible universal and language-specific mechanisms that may define the way language users exploit prosody in speaking and listening. This is because there has been a lack of emphasis on comparing prosodic processing in speakers of different languages, despite growing appreciation of

* Prosody is the linguistic structure expressed in the suprasegmental features that convey word-level (Jun, 2014) or postlexical/sentence-level meaning (Ladd, 2008). Linguistic tone, the use of pitch to distinguish lexical or grammatical meaning, is therefore not an expression of prosody.

comparative approaches in the segmental literature. Even in the handful of crosslanguage studies on prosodic processing, the data so far have largely been restricted to speakers of closely related languages with similar intonation systems (e.g., Akan and Ga: Genzel, Renans, Kügler, 2018; Bengali and Hindi: Choudhury & Kaiser, 2016; Dutch and English: Akker & Cutler, 2003; German and English: O'Brien, Jackson, & Gardner, 2014; Taiwan Mandarin and Beijing Mandarin: Xu, Chen, & Wang, 2012). To address these shortcomings, my graduate research aims to compare prosodic strategies in speakers of typologically distinct languages with different intonation systems.

The role of prosody can be seen from two very different standpoints. For most of its history, speech prosody has been neglected as a trivial feature, largely owing to the view that communication without prosody is possible, albeit more challenging (e.g., as in text messages or monotone speech). It is also a phonetic dimension of which language users are least aware. Few languages choose to incorporate prosodic features in their writing system, and perhaps for this reason, prosody is almost never explicitly taught in first (L1) or second (L2) language learning (Lengeris, 2012). Indeed, the great linguist Dwight Bolinger even referred to intonation as a part of speech that is "around the edge of language" (1964). At the same time, the lack of research attention on prosodic processing could also be due to difficulties in adopting suprasegmental features as discrete parameters in current frameworks of spoken language recognition (Cutler, 2012). No models of language processing have succeeded at incorporating prosody into speech perception, and the two automatic speech recognition models that have attempted to use suprasegmental information (e.g., lexical stress) for word identification have so far failed (Sholicar & Fallside, 1988; van Kuijk & Boves, 1999). Certainly, prosody appears more fine-grained and elusive than the segmental.

However, this is not to say that prosody is a random and trivial component. Prosody has an “organisational structure” (Beckman, 1996), and like the hierarchical structures embodied in syntactic trees, there are also different levels of prosodic constituents that govern prominence relations and intonational, rhythmic, and pausing patterns across different languages. Subsequent to the introduction of this phonological hierarchy (e.g., Beckman & Pierrehumbert, 1986; Ladd, 1986; Liberman & Prince, 1977; Nespor & Vogel, 1986; Selkirk, 1986; 2003), evidence from language learning research suggests that prosodic structure is intricately intertwined with segmental phonology (e.g., Ulbrich & Mennen, 2015) and can impinge on higher levels of linguistic representations. This can be seen in morphological development, where constraints arising from prosodic structure can support young children’s production of grammatical morphemes (e.g., Demuth & Tremblay, 2007; Demuth, McCullough, & Adamo, 2007), or in word learning and syntactic processing, where attention to phrase-level prosodic cues can help preverbal infants detect syntactic boundaries and map auditory word forms onto visual referents (e.g., Gleitman & Wanner, 1982; Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Shukla, White, & Aslin, 2011; Soderstrom, Seidl, Nelson, & Jusczyk, 2003), or in discourse processing, where patterns of intonation and prominence can be used to express the speaker’s affect, pragmatic intent, and illocutionary force (e.g., Austin, 1962; Krifka, 2006; Pierrehumbert & Hirschberg, 1990). In this respect, prosody can be seen as the skeletal foundation of language. Examining how prosody facilitates speech perception can therefore provide part of the answer to the “binding” problem of how listeners integrate and unify different domains of language in online processing (Frazier, Carlson, & Clifton, 2006).

Importantly, no theory of speech processing can be complete without taking prosody into account. This is because all components of utterances, even at the phonetic segment, have a certain duration, fundamental frequency (F_0), and amplitude (Lehiste, 1970). Since these physical manifestations of prosody are, themselves, intrinsic determinants of speech, any intervention to promote language education or speech synthesis must take into account how speakers and listeners represent prosodically determined variation. Being able to learn and process these prosodic patterns is an extraordinary cognitive achievement. In ontogeny, speech prosody is most likely the first acoustic cues that prelinguistic infants acquire before any other domains in language development (e.g., Gleitman & Wanner, 1982), and even before birth, the foetus can already implicitly learn the intonational and rhythmic patterns of the outside language (Mampe, Friederici, Christophe, & Wermke, 2009; Mehler et al., 1988; Ramus, 2002). In phylogeny, the capacity to process prosody as a domain-general ability may also be the first to have appeared in language evolution; both vocal (e.g., zebra finches and budgerigars) and non-vocal (e.g., rats) learning species can use some aspects of speech prosody (e.g., basic stress patterns) to disambiguate words or syllables in human speech (e.g., Hoeschele & Fitch, 2016; Spierings & ten Cate, 2014; Toro & Hoeschele, 2017). Yet, it is fascinating that no scholars could formulate a coherent model to capture the sheer intricacy of prosody. So from both a philosophical and paedagogical viewpoint, a better understanding of prosodic processing can provide us with fundamental hints about the human mind and its machineries.

1.2. Two Prosodic Universals: Processing Salience and Junctures

In this thesis, I examine how the production and perception of prosody can be influenced by both language-universal and language-specific mechanisms in speech processing. There are as yet no recent proposals on how prosody may play a universal role in language processing. The only source of information about possible universal functions comes from a proposal more than four decades ago by Bolinger (1978), who identified two aspects of prosody that all speakers and listeners may use to process spoken languages. While the segmental level of speech conveys messages, prosody, expressed at the suprasegmental level, conveys the import of messages within their context. First, prosodic structure is the component of the grammar that may be used across all languages to encode the utterance information structure and mark prominence as signals to semantic salience. Second, prosodic structure is the component of the grammar that may be used across all languages to mark boundaries and organise speech into linguistically significant cognitive units. From this point of view, prosody serves a universal role in the encoding of information structure (e.g., focus) and in the marking of syntactic junctures. From a crosslanguage point of view, what is the nature of prosodic processing? To what extent is this processing universal across languages? To what extent is this processing influenced by speakers' experience with their native language? To what extent do language-general and language-specific factors interact? How is prosodic structure related to discourse interactions and structural disambiguation? In the rest of this chapter, I will briefly discuss some of the previous findings in relation to Bolinger's proposal of the two prosodic universals. Bolinger's ideas, and the research questions that stem from it, will form the theoretical foundations for my thesis.

There are three hypotheses for how prosody may be exploited in language processing. The first possibility is that Bolinger may be right in that prosodic processing of focus and junctures is universal across all languages. However, although there are no explicit models on how humans exploit prosody, from the current literature, it is reasonable to claim that most researchers would maintain that prosodic processing is either a purely language-specific phenomenon or a complex interaction between both universal and language-specific mechanisms. From the existing record of the various sound systems in different languages, prosody is widely acknowledged to be one of the most language-specific, and even dialect-specific, dimensions of speech (see Cruttenden, 2006; Himmelmann & Ladd, 2008). For instance, languages differ in whether the rhythm is based on stress (e.g., English, Arabic), syllables (e.g., Italian, Yorùbá), or the mora (e.g., Bengali, Tokyo Japanese). Likewise, even closely related languages (e.g., the Slavic family) can differ in prosodic structure, such as whether the fixed stress locations are initial (e.g., Slovak, Sorbian), antepenultimate (e.g., Macedonian), penultimate (e.g., Polish), or non-existent, i.e., there is no fixed location at all (e.g., Bulgarian, Russian, Slovenian). At a much broader level, the prosodic systems of different languages can also vary in terms of whether the macro-rhythmic structure is characterised by regular (strong) or irregular (weak) alternations of high and low pitch targets (Jun, 2014).

On this position, cues to prosodic structure would be language-specific. In the domain of prosodic focus (Bolinger's first prosodic universal), there are a variety of language-specific suprasegmental features that are assigned to focused constituents in different languages (Selkirk, 2004), including pitch accents (e.g., English and German: Selkirk, 1984), tonal morphemes (e.g., Bengali: Hayes & Lahiri, 1991), demarcation of prosodic phrase edge (e.g., Chichewa: Downing, Mtenje, & Pompino-Marschall, 2004), pitch range expansion (e.g., Shanghai Chinese: Selkirk & Shen, 1990), and vowel duration increase (e.g., European Portuguese: Frota, 2000). Similarly, in prosodic junctures (Bolinger's second prosodic universal), languages can differ in prosodic marking of phonological phrase boundaries depending on the left versus right boundary of their corresponding morphosyntactic category (e.g., Selkirk's End-based theory, 1986). Given that experience with the structure of the mother tongue induces a language-peculiar pattern of processing (e.g., Best, 1994; Best & McRoberts, 2003; Werker & Tees, 1984), it can be concluded from the high level of linguistic diversity in prosodic systems that there are no language universals dictating how prosody is processed.

However, a perspective solely founded on the current documentation of different languages may be problematic. This is because many of these and other languages have not yet been fully analysed in sufficient depth to answer whether there are potential underlying universals in prosody (see Hirst & Di Cristo, 1998). Further, in order to gain a better understanding of language *processing*, it is important to adopt rigorous experimental procedures to assess how different prosodic structures are actually processed by the language user. Even in cases where processing strategies appear language-specific across languages, there may often be some common underlying mechanisms that are responsible for these differences. Consider the case of lexical stress perception in English and Dutch. Native Dutch listeners have been found to use word-internal stress cues to distinguish segmentally ambiguous words (e.g., *VOORnaam "first name"* from *voorNAAM "respectable"*, Cutler & van Donselaar, 2001), but the same cues are not exploited by English listeners (Cooper, Cutler, & Wales, 2002). This is because in English, as opposed to Dutch, the distribution of suprasegmental stressed and unstressed syllables nearly always corresponds with segmental distinctions in vowel quality (Fear, Cutler, & Butterfield, 1994). Making use of suprasegmental cues to stress would therefore be redundant in English but necessary in Dutch. In another line of experiments comparing pitch accent versus non-pitch accent varieties of Japanese (e.g., Tokyo Japanese vs. Fukushima Japanese), it is pitch contrast that serves a useful cue in lexical selection (e.g., recognising *nagashi* LHH or *nagasa* HLL after hearing the initial segment *naga* in LH or HL: Cutler & Otake, 1999; Otake & Higuchi, 2008). On this interpretation, all listeners possess a common strategy in that they process only the most relevant input that is specified by the phonology of their native language. The different listening strategies (pitch accents in Japanese, stress in Dutch, vowel quality in English) reflect a common strategy shared across all three languages: that listeners ignore redundant information (Cutler, 2012). From crosslanguage experiments, we can reveal how different processing across languages can be based on a complex interplay between language-specific structures and the underlying crosslinguistically shared strategies.

1.3. Thesis Outline

As mentioned earlier in the chapter, little is known about the possible universal and language-specific prosodic mechanisms in speech processing because very few attempts have been made to compare speakers of different languages. To address these issues, my graduate research adopts an explicit crosslanguage approach to examine both the production and perception of prosodic salience and junctures in native speakers of English and Mandarin Chinese. In many of my experiments, I also explore prosodic processing across both native (L1) and non-native (L2) contexts. The ultimate goal of this thesis is to explore how language-universal/crosslinguistic and language-specific processes interact in the production and perception of prosody across speakers of typologically distinct languages with very different intonation systems.

Comparisons between diverse languages, in particular between tone and non-tone languages, are rare. From a methodological point of view, it is a challenge to examine speakers of these languages using comparable materials and procedures, but such comparisons can provide useful insights in two very important ways. First, linguistic tone is a feature that exists in more than half of the world's languages (Hyman, 2001; Yip, 2002), so a crosslanguage assessment with two typologically representative languages can bring us closer to a better understanding of the universal/crosslinguistic influences in speech processing. Second, a comparison between English and Mandarin Chinese, as two exemplary cases of non-tone versus tone languages (Duanmu, 2004), can help us address questions about the sound patterns of natural language with respect to the relationship between tone and intonation system. It has often been claimed (see Nolan, 2006 and Pierrehumbert, 1999) that languages with a complex tone system (e.g., Mandarin Chinese, Cantonese Chinese) have, in general, simpler intonation systems than non-tone languages (e.g., English) and languages with pitch accents (e.g., Japanese) or less stable lexical tones (e.g., Shanghai Mandarin), presumably because much of the F_0 contour is exhausted in the phonetic expression of complex tones.

The first question that I address is how native speakers of English ($n = 24$) and Mandarin Chinese ($n = 24$) use prosody to encode information structure in production (**Chapter 2**). More specifically, our crosslanguage production experiment examines whether English and Mandarin are alike in the ways in which speakers use the different prosodic parameters (e.g., duration, F_0) to produce focus. Previous studies suggest that speakers in both languages realise prosodic focus in very similar ways; both languages effect focus with lengthened duration, greater intensity, boosted spectral energy, heightened F_0, and greater F_0 range expansion, followed by postfocal F_0 and intensity compression (e.g., Breen, Fedorenko, Wagner, & Gibson, 2010; Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; Xu, 1999; Xu & Xu, 2005; Xu et al., 2012). However, it is not possible to reach any definite conclusions about the crosslanguage similarities, because the experimental designs and the structure of the materials used in the existing studies are often quite different. Some of the past experiments that examined focus production also involved procedures where participants were given explicit instructions to produce focus using sentences that were rather ecologically unnatural (e.g., "*māomī mō māomī*" "*Kitty touches Kitty"*: Xu, 1999). To address these issues, our production experiment aims to elicit focus production by using structured dialogues where the discourse context is written in a way that can inform the native speaker's prosodic choices. By eliminating any explicit instructions, our experiment can determine not only how speakers of different languages naturally use prosody to highlight salience, but also the *degree* to which they increase the different aspects of prosody in focus production.

The experiments described in the next chapter (**Chapter 3**) aim to understand whether English and Mandarin listeners perceive prosodic focus in a similar way despite the language-specific differences in intonation systems. Previous experiments adopting the phoneme detection paradigm demonstrate that native speakers of Germanic languages (e.g., American or British English and Dutch) can entrain to prosodic contours to predict where focus will fall in an utterance (Cutler, 1976; Cutler & Darwin, 1981; Akker & Cutler, 2003). Using the same phoneme detection methodology, our crosslanguage focus perception experiments investigates whether prosodic entrainment is also found in Mandarin, a language with a complex tone system where, in principle, the use of pitch primarily for lexical identity may take precedence over the use of pitch cues to salience. Experiment 1a examines prosodic entrainment by listeners of Australian English (*n* = 23) and Mandarin (*n* = 23) in their native language. In Experiment 1b, we also examine whether native Mandarin listeners (*n* = 36) can entrain to prosody in a second language.

A related question is whether listeners within a given language variety (Australian English) can also engage in prosodic entrainment to the speech of different speakers (**Chapter 4**). Previous research suggests that listeners' prediction of upcoming speech forms can be influenced by a variety of distal cues from the preceding prosody, including speech rate (e.g., Dilley & Pitt, 2010), pausing (e.g., Gee & Grosjean, 1984), and rhythmic patterns in pitch and timing (e.g., Dilley, & McAuley, 2008; Morrill, Dilley, McAuley, & Pitt, 2014). However, no research to date has examined whether production of these preceding cues may vary across individual speakers. Moreover, no experiments have used unsynthesised speech stimuli to investigate the role of different types of preceding cues in prosodic entrainment. Across five experiments (*N* = 113), participants listen to a series of series of sentences produced by one of four native female speakers of Australian English from Sydney, Australia. Our experiments aim to explore three important questions. First, we are interested in whether native speakers of a given language variety produce different preceding cues to prosodic focus. Second, to the extent that there is substantial talker variability, we address whether listeners are capable of engaging in an entrainment strategy in any speaker-specific situations. Finally, we examine the relative contributions of different preceding cues in focus detection.

In the last experimental chapter (**Chapter 5**), I will introduce a series of experiments where we compare how native speakers of English and Mandarin process prosodic cues to junctures in both production and perception. Although prosodic disambiguation has already been studied in both languages, most of the previous studies have looked at structural ambiguities that are expressed quite differently in English and Mandarin (e.g., relative clause structures). Adopting a crosslanguage perspective, we investigate juncture processing using structural ambiguities that are identical in both English and Mandarin. Across our production and perception experiments, we use pairs of segmentally identical ambiguous sentences that differ in meaning as a function of the timing and location of the prosodic junctures (e.g., "*Simon gave her # dog biscuits*" vs. "*Simon gave her dog # biscuits*"; "*大卫不小心给她 # 狗饼干*" vs. "*大卫不小心给她狗 # 饼干*"). These sentences are segmentally identical and display surface (syntactic) ambiguity that can only be disambiguated through the use of prosody. In the production experiment, we compare English (*n* = 24) and Mandarin (*n* = 24) speakers across four types of disambiguation juncture marking strategies: pausing, pre- and postboundary lengthening, and F_0 modification, and domain-initial strengthening. We explore whether English and Mandarin speakers are alike in the ways in which they use prosody to produce junctures, and also whether there are differences in the degree to which they use the different disambiguation strategies.

In our perception experiments, we design a new disambiguation task where participants are required to make speeded responses to select the correct meaning for a series of structurally ambiguous sentences. Our first perception experiment explores whether English (*n* = 40) and Mandarin (*n* = 40) listeners differ in (1) their perception of sentences with different juncture location, and (2) whether there is also language variation in the degree to which prosody is used for ambiguity resolution. In the second perception experiment, we investigate whether English (*n* = 12) and Mandarin (*n* = 19) listeners can disambiguate sentences when the pause duration of the junctures is rendered uninformative. In the third perception experiment, we explore how Mandarin speakers (*n* = 29) use non-native cues to prosodic junctures when the sentences were in English.

Together, these crosslanguage experiments address the two prosodic universals proposed by Bolinger in terms of both production and perception. A universalist account predicts that native speakers of different languages will produce and perceive prosodic structure for the same functions in the same manner. A purely language-specific account predicts that there will be no association between the functions that are realised by prosody in one language versus another. Where the same function happens to be realised by prosody, there may still be language differences in the exact ways through which certain prosodic information is exploited. On the other hand, there may be an interplay of universal function and language-specific realisation. For instance, different languages may differ in how prosody is realised, but their effect on speech perception is the same.

CHAPTER 2

– Prosodic Strategies of Focus Expression Across Languages –

A version of this chapter was submitted to the *Journal of Phonetics*: Ip, M. H. K., Shaw, J. A., & Cutler, A. (submitted). Prosodic strategies of focus expression across languages.

Portions of the data presented in this chapter were presented at the 8th International Conference on Speech Prosody: Ip, M. H. K., & Cutler, A. (2016). Cross-language data on five types of prosodic focus. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of the 8th International Conference on Speech Prosody* (pp. 330-334). Boston, USA: ISCA. doi: 10.21437/SpeechProsody.2016-68

2.0. Abstract

To examine the relative roles of crosslanguage and language-specific mechanisms in the production of prosodic focus, we compared production of five different types of focus by native speakers of English and of Mandarin Chinese. Structured dialogue scripts were constructed for each language, with the same words appearing in focused and unfocused position; 48 speakers recorded five dialogues each in their respective native language. Duration, F_0 (mean, maximum, range), and RMS-intensity (mean, maximum, range) of all critical word and phrase tokens were measured. In total, the present experiment compiled prosodic data from 34,944 measurements. Overall, English and Mandarin speakers were alike in the ways in which they used prosody to effect focus. However, there were also some cross-language differences: Mandarin speakers produced greater increases in mean and maximum F_0 and F_0 range, while English speakers tended to produce focused words with higher increases in mean and maximum intensity and intensity range. Further, the pattern of language-specific differences also varied across different dialogues and focus types. Our findings provide evidence of language-specificity in prosodic processing and show that production of information structure can differ even when the same prosodic resources are employed in the same manner.

– Prosodic Strategies of Focus Expression Across Languages –

2.1. Introduction

2.1.1. *Universal versus Language-specific*

Information structure is a linguistic universal. As long as speech is used for communication between people, utterances will concern some things that are, in one sense or another (Halliday, 1967; Krifka, 2006), more important, and some that are less important. All speakers thus have the option to convey this structure in the way they speak, and they may use prosody to do it. Indeed, as Dwight Bolinger (1978) noted some four decades ago, the highlighting of more important elements may be one of only two prosodic universals in human language.

However, increasing evidence from different languages has cast doubt on whether prosody plays a universal role in the transmission of discourse information. Firstly, how this phonetic highlighting – expression of focus – is achieved by means of prosody can vary depending on the intonational phonology of the language (Ladd, 2008; Jun, 2014). For instance, edge-prominence languages such as Korean and Japanese mark focus by employing local pitch range expansion on the phrase edge through boundary tones (Venditti, Jun, & Beckman, 1996), while head-prominence languages such as English, German, and Dutch mark focus on the phrase head through the use of intonationally determined pitch accents (Chen, 2012; Féry & Kügler, 2008; Gussenhoven, 2004; Jun, 2011). Other variation in prosodic production can also be seen in languages that express focus through assignments of tonal morphemes (e.g., Bengali: Hayes & Lahiri, 1991; Kinande: Hyman, 1990), durational lengthening (e.g., Cantonese: Fung & Mok, 2018; European Portuguese: Frota, 2000; German: Hay, Sato, Coren, Moran, & Diehl, 2006), or implementations of prosodic breaks to induce tonal changes or blocking of elision (e.g., Chichewa: Downing, Mtenje, & Pompino-Marshall, 2004; Kwa languages of Côte d'Ivoire: Leben & Ahoua, 2006; Shanghai Chinese: Selkirk & Shen, 1990). Even within language groups similar in prosodic structure, cross-linguistic variation could exist in how different prosodic parameters are used to express focus due to differences in rhythmic structures (e.g., variation in the degree of regular alternations between high and low pitch targets: Burdin et al., 2015).

Secondly, the relation between accentuation and relative semantic weight may not be consistent across languages. For example, while speakers of American or British English tend to use prosody to highlight focused words and deaccent repeated information, there exist languages (e.g., Romance languages such as Italian, Spanish, and Romanian: Avesani & Vayra, 2005; Cruttenden, 1993; Ladd, 1990a; 2008) and even regional dialects (e.g., Indian and Caribbean English: Gumperz, 1982; Tunisian and Cairene Arabic: Cruttenden, 2006; Hellmuth, 2005) in which speakers are less likely to deaccent repeated words, or compress the post-focal region of the utterance (e.g., Taiwanese Mandarin: Xu, Chen, & Wang, 2012; Yi: Wang, Wang, & Qadir, 2011). In addition, there are also reports of languages where speakers do not use prosody for focus marking (e.g., Ambonese Malay: Maskikit-Essed & Gussenhoven, 2016; Northern Sotho: Zerbian, 2006; Yucatec Maya: Kügler & Skopeteas, 2007) or where it is only optional (e.g., Hausa: Hartmann & Zimmermann, 2007; Hungarian: Mády, 2015; Chichewa: Downing & Pompino-Marschall, 2013).

Finally, the extent to which speakers use prosody to highlight focus could also be constrained by the morpho-syntactic structure of the language. Thus in Wolof (Rialland & Roberts, 2001), morphological markers are available, and speakers do not redundantly use intonation; in languages with greater word order flexibility such as Czech, Catalan, and Italian (e.g., Dubĕda & Mády, 2010; Ladd, 2008; Vallduví, 1991;

1992; Zubizarreta, 1994; 1998), speakers tend to move the focused elements to the default utterance position that bears strong prominence, such that pitch accents may provide less discourse structure information (Swerts, Krahmer, & Avesani, 2002); and likewise in Indonesian, syntax has been reported to be the only means for focus marking due to fixed word stress in phrase-final positions (Goedemans & van Zanten, 2007). Given this variation and the panoply of resources and means that speakers can exploit, there may be no universal manner of focus expression.

Nonetheless, it is still an empirical question whether language-specific differences in prosodic production arise from a common underlying processing mechanism. Acoustically, words that are stressed tend to be produced with greater intensity, higher pitch, greater and more rapid changes in pitch range, lengthened syllables, and vowels articulated with spectral characteristics that are closer to their citation form (e.g., Cooper, Eady, & Mueller, 1985; Dahan & Bernard, 1996; Heldner & Strangert, 2001; Klatt, 1976; Lehiste, 1970; Lehiste & Peterson, 1959; Hart, Collier & Cohen, 1990; Heldner, 2003; Sluijter & van Heuven, 1995). Moreover, several of these cues may be functionally equivalent across languages (Vaissière, 2005). Therefore, accentuation may contribute to greater perceptual clarity and benefit listeners by attracting their attention to the most informative part of the utterance (e.g., Kristensen, Wang, Petersson, & Hagoort, 2013; Ladd & Cutler, 1983; Lieberman, 1963). And indeed, across various languages, words or syllables that are prosodically prominent are better retained in memory (e.g., Fraundorf, Watson, & Benjamin, 2010; Kember, Choi, Yu & Cutler, submitted), recognised more rapidly and accurately (e.g., Cutler & Foss, 1977; Lee, Chiu, & Xu, 2016; McAllister, 1991), processed more deeply in lexical activation (Blutner & Sommer, 1988; Brunellière, Auran, & Delrue, in press; Li & Ren, 2012; Norris, Cutler, McQueen, & Butterfield, 2006), and are more likely to direct listeners' attention to new elements of the discourse structure (e.g., Dahan, Tanenhaus, & Chambers, 2002; Fowler & Housum, 1987; Hsu, Evans, & Lee, 2015). Moreover, prosodic communication of emphasis may also have some universal processing properties because it may be related to its prelinguistic use as a signal to the speaker's emotional state, a notion that has gained support from studies involving prosodic communication of emotions occurring independently of verbal comprehension (Kitayama, & Ishii, 2002) and also in crosslanguage comparisons (Thompson & Balkwill, 2006). On this interpretation, prosodic focus may have originated from an innate physiological mechanism where the most 'interesting' or 'informative' part of an utterance is associated with heightened arousal, greater respiratory effort, dramatic pitch change, and more energetic movements (e.g., Bolinger, 1978; 1986). Consistent with this view, some research examining the developmental origins of prosody shows that young children can produce prosodic cues to focus, presumably as an automatic physiological reflex, before they start to understand them as markers of information structure (Baltaxe, 1984; Cutler & Swinney, 1987).

For these reasons, prosody may, from a processing standpoint, be universally available as a resource that all speakers can use to highlight focus – from languages where prosody is largely ignored for this purpose, to languages where it is the main way focus is expressed, with many differences in the precise ways in which various parameters are used for this purpose. Certainly it is clear that there are languages where prosodic cues coexist with other means to signal focus. For example, unrelated languages such as English, Korean, Japanese, and Mandarin Chinese all allow focused words to be marked via non-prosodic means (e.g., by particles, phrasing, or word order), but speakers nonetheless convey focus using prosodic parameters such as pitch and duration (e.g., Chen & Gussenhoven, 2008; Jun & Lee, 1998; Maekawa, 1997; Pierrehumbert & Beckman, 1988; Xu, 1999; Xu & Xu, 2005). Further, in contrast to the Wolof, Czech, Italian, and Indonesian cases, there are also counterexamples of languages where focus is still realised prosodically despite optional or even obligatory syntactic constructions (e.g., Finnish: Kaiser, 2006, Arnold, 2016; Hungarian: Genzel, Ishihara, & Surányi, 2015), morphological markers (e.g., Chickasaw: Gordon, 2004; Ewe: Fiedler & Jannedy, 2013; Ga: Genzel, Renan, & Kügler, 2018), word order flexibility (e.g., Bulgarian: Andreeva, Koreman, & Barry, 2016), or fixed word stress (e.g., Polish: Hamlaoui, Zygis, Engelmann, & Wagner, 2018). It is therefore not entirely clear how the production differences in focus strategies reflect language-particular arbitrary choices or some principled differences in certain linguistic properties. For this reason, a research focus on processing mechanisms in focus production may provide a useful perspective about these crosslanguage similarities and differences.

2.1.2. ***Prosodic Focus in English and Mandarin Chinese***

Although the production of focus cues has been examined in many languages, the existing studies are difficult to compare due to differences in experimental design and materials. One way to address the question of universality versus language-specificity in prosodic processing is to adopt an explicit comparative approach by examining speakers of typologically distinct languages with different intonational systems that are nonetheless similar in their strategies for focus construction. Based on this approach, the present study examined prosodic production of focus in English and Standard Mandarin Chinese. A large body of research on information structure in the two languages has revealed striking similarities in speakers' prosodic encoding of focus. Both languages exhibit focus with lengthened duration, greater intensity, boosted spectral energy, heightened F_0 and wider F_0 range, followed by a post-focal compression of intensity and F_0 range (e.g., Breen, Fedorenko, Wagner, & Gibson, 2010; Cao, 2012; Chao, 1968; Chen, 2006; Chen & Gussenhoven, 2008; Chen, Xu, & Guion-Anderson, 2014; Cooper, Eady, & Mueller, 1985; Ito & Speer, 2006; de Jong, 2004; Jin, 1996; Ladd, 2008; Lee et al., 2015; Lieberman, 1960; Liu & Xu, 2005; Ouyang & Kaiser, 2015; Pierrehumbert & Beckman, 1988; Shih, 1988; Xu, 1999; Xu & Xu, 2005; Xu, Chen, & Wang, 2012; Wang & Xu, 2011; Yuan, 2004). Importantly, Mandarin speakers manage to employ prosody for focus construction in ways that do not interfere with the identity of the lexical tones (i.e., their F_0 shapes), such that focused elements have greater F_0 range expansion for contour tones, higher high-level tones, and lower low-level tones. These effects may, in principle, be analogous to the lowering or rising of the low tone (L) and the rising of the high tone (H*) in L+H* accents found in English and Dutch (e.g., Arvaniti & Garding, 2007; Gussenhoven & Rietveld, 2000; Liberman & Pierrehumbert, 1984). Moreover, prosodic cues in both English and Mandarin can co-occur with other means of focus expressions (e.g., cleft structures) and these cues are marked independently of other linguistic functions (e.g., boundary marking) in Mandarin (Wang, Xu, Ding, 2017). In addition, both languages may also have other similar characteristics. For example, Duanmu (2000) posited that Mandarin speakers also obey the same compound stress rule and nuclear stress rule as in English (Chomsky & Halle, 1968), a proposal that has partly been supported in research where speakers produced final syllable lengthening and wider F_0 range to disambiguate between a modifier-head compound and a verb-object phrase (e.g., *"傳言/chuan-yan"* "*hearsay*" vs. "*to pass a message*": Shen, Vaissière, & Isel, 2013).

An open question is whether there is still cross-language variation in the exact degree to which English and Mandarin speakers use each of the different prosodic parameters to express emphasis. In earlier works, G. T. Chen (1972) compared the production of read words and sentences in speakers of Mandarin and Midwestern American English and found that Mandarin speakers (particularly female speakers) produced greater changes in average F_0 and in F_0 range to show emphasis, compared to the English speakers. More recently, in the literature on speaking fundamental frequency characteristics, Xue, Hagstrom, and Hao (2002) found that Mandarin-English bilinguals tend to produce higher average F_0 and greater F_0 range when speaking Mandarin compared to English, suggesting that differences in F_0 are learned on a language-specific basis and are not simply due to talker-specific strategies or to physiological differences between speakers of different languages (see also Mang, 2001 and Graham, 2015 for similar findings in English-Cantonese and -Japanese bilinguals respectively). Further, Yuan and Liberman (2014) revealed in a corpus of broadcast news that Mandarin has wider F_0 range and F_0 fluctuations compared to English. Similarly, an experiment by Keating and Kuo (2012) found Mandarin speakers used higher mean F_0 and F_0 range than English speakers when producing isolated words in excited versus normal pitch (e.g., "*示!*"/[ʂi] or "*Sure, Sure, SURE!*; n.b., only mean F_0 was higher when a prose passage was spoken). All these data suggest Mandarin production involves greater use of F_0 cues.

In English, on the other hand, there is less consensus with respect to which parameter is most relied on. Some argue that F_0 is the strongest cue for stress (e.g., Cooper et al., 1985; Fry, 1958; Lieberman, 1960), while others have noted that duration is more reliable than intensity (e.g., Fry, 1955; Turk & Sawusch, 1996). More recently, however, Kochanski, Grabe, Coleman and Rosner (2005) demonstrated that, at least at a syllabic level, prominence is not always reliably signaled by F_0. In a large corpus involving three styles of speech (sentence lists, story paragraphs, retelling of a story) from seven dialects of the British Isles, Kochanski and colleagues examined the extent to which various suprasegmental properties can separate prominent from non-prominent syllables. Prominent syllables were best predicted by greater loudness, followed by longer duration, while F_0 cues lent little support for prominence detection (see Silipo & Greenberg, 2000 for similar findings). In fact, many non-prominent syllables have also been found to have high pitch levels that were comparable to prominent ones. These findings are also in line with listening experiments suggesting that English listeners can still entrain with the utterance intonation to anticipate an upcoming prosodically focused word when the F_0 is monotonised (Cutler & Darwin, 1981). On the other hand, intensity may generally only have a secondary effect in Mandarin production. For example, phonetic data show that Mandarin speakers do not necessarily produce lower intensity for the destressed neutral tone compared to the full tones (e.g., Cao, 1992), and the neutral tone tends to have a higher intensity after a low-dipping tone (Lin, 2006). However, Mandarin speakers may show greater use of intensity range than English speakers, arguably due to the presence of lexical tones (S. Chen, 2005).

Overall, these data suggest that English and Mandarin speakers may differ in their use of prosodic cues, even if the way focus is produced in the two languages is highly similar. This could be due to a number of reasons. For instance, Vaissière (1983) proposed that prosodic production across languages can vary due to differences in timing, in the relationship between the prosodic parameters, or in the order of priorities. From the studies so far mentioned, we formulate two competing hypotheses. On the one hand, speakers may produce focus by enhancing the cues that are already present rather than introduce new phonetic cues. On this view, Mandarin speakers may be more likely to rely on F_0 and English speakers on intensity cues.

On the other hand, from a functional point of view, it is unlikely that speakers can use the same prosodic parameter to exactly the same extent for two different purposes. For instance, given that English has both prosodic focus and lexical stress occurring on a word's most prominent syllable, the findings from Kochanski and colleagues (2005) that intensity provides the most reliable cue to lexical stress may indicate that English speakers may be more restricted in their manipulation of intensity during focus production compared to Mandarin speakers. Likewise, Mandarin speakers may be less likely than English speakers to rely on F_0 cues, firstly because the higher speaking F_0 in Mandarin may place a ceiling effect on the degree to which F_0 cues can be maximally exaggerated within the constraints of its intonational structure, and secondly, because the presence of lexical tones may preempt the use of F_0 cues for focus expression (Yip, 2002). Supporting this view, Lee, Wang, and Liberman (2016) investigated the role of the different Mandarin tones in focus production and found that the low-dipping third tone showed the smallest effect of F_0 range because of its documented smaller capacity for pitch range expansion. Further, research from other languages has also compared the production of phrase-final focused words in tonal versus non-tonal dialects of Kammu (a Mon-Khmer language) and found that the tonal dialect had a more narrow and less varied pitch range during focus production (particularly when the lexical tone was low), despite having an almost identical intonational system to the non-tonal dialect (House, Karlsson, Svantesson, & Tayanin, 2009). Similarly, data from Triqui, an Otomanguean language with nine lexical tones, also show that speakers only use duration to produce focus (e.g., DiCanio & Hatcher, 2018).

2.1.3. *Role of Focus Types Across Languages*

Another unresolved issue in information structure research concerns how different pragmatic contexts can affect the prosodic realisation of focus across languages. On the one hand, Krifka (2006) outlined at least four pragmatic uses of focus; focus markings to highlight the part of the answer in response to a wh-question, focus used for correcting previously-conveyed information, focus used for confirmation, and focus used for highlighting parallels. According to the central tenet of Alternative Semantics (Krifka; Rooth, 1992), all of these focus constructions “indicate the presence of alternatives that are relevant for the interpretation of linguistic expressions”. On this view, there is no principled difference between the different uses of focus, on the grounds that all expressions of focus evoke two semantic representations: the actual meaning of a focused expression and a set of alternatives. Nevertheless, Krifka proposed that there can still be different prosodic realisations for different uses of focus, since different ways of focus marking can still indicate different ways alternatives are expressed. On the other hand, others (e.g., Kiss, 1998; Rochemont, 1986) have hinted that there are two separate types of focus, one from a contrastive standpoint and another conveying new nonpresupposed information without expressing exhaustive identification. Incorporating both these views, the present study examines from a crosslanguage standpoint how prosody is realised in different pragmatic contexts.

In past studies, some experimental evidence has revealed different acoustic correlates of focus across different contexts. For instance, in Mandarin, experiments by Ouyang and Kaiser (2015) found prosodic differences where words denoting new information had longer duration and greater F_0 range expansion compared to given information, while focused words in corrective contexts also had greater intensity ranges in addition to longer duration and pitch expansion. In another study, Chen and Braun (2005) show that focus under corrective contexts had a larger F_0 range compared to focus production in wh-question contexts. Similarly, Greif (2010) looked at subtypes of corrective focus and found that Mandarin speakers produced more robust cues when correcting the entire presupposed background information of a preceding wh-question compared to correcting just the focused part corresponding to the wh-question.

In English, Katz and Selkirk (2011) compared focus used for correction and new information within the same utterance and found that production of focus in corrective contexts has relatively more prominent duration, F_0 profiles, and intensity compared to focus in discourse-new contexts. Other studies within the auto-segmental metrical phonology framework show that English speakers are more likely to produce rising bitonal (L+H*) pitch accents (i.e., greater use of F_0 range) to encode discourse elements that signal contrastive/corrective contexts, while elements indicating new information are more likely to be marked by simple high (H*) pitch accents (e.g., Ito, Speer, & Beckman, 2004; Pierrehumbert & Hirschberg, 1990), although this is still tentative (c.f. Hedberg, & Sosa, 2008). Meanwhile, experiments in speech perception reveal that listeners are more likely to interpret an L+H* accent as contrastive, but an H* accent can be perceived as indicating both correction and new information (Watson, Tanenhaus, Gunlogson, 2008). In another study, Breen et al. (2010) conducted a series of production and perception experiments and found that speakers can distinguish corrective from new-information focus, though only when they were made aware of the prosodic ambiguity, while listeners could not identify the different types of focus even when they were presented with reliable cues to focus types.

Finally, how focus type affects prosodic production may also be language-specific. In a recent crosslanguage study, Choudhury and Kaiser (2016) looked at the production and perception of corrective versus new-information focus in speakers of Bengali and Hindi (two closely-related languages) and found that the relation between focus type and certain prosodic signals is language-specific. For instance, only Bengali use F_0 to distinguish between the two focus types (e.g., higher F_0 for corrective focus on objects). At the same time, although speakers of both languages produced corrective focus with longer duration, only Hindi speakers reliably used it to distinguish between focus types in perception. However, given the paucity of crosslanguage research and only evidence from two levels of focus type, it remains uncertain if differences exist across focus types and languages in a systematic manner.

2.1.3. *General Overview of Production Experiment*

The present experiment adopts a cross-language approach to investigate the production of prosodic focus in native speakers of English and Mandarin Chinese. For both languages, we compiled a database of focused and unfocused realisations of the same words uttered by multiple speakers, using contexts that were both relatively realistic and closely matched across the two languages. Twenty-four speakers from each language were recruited and five structured dialogues were created to elicit production of five types of focus (comparable Chinese and English versions were made for each dialogue). Each dialogue contained pairs of words occurring in a focused and unfocused context, and production of each pair of focused and unfocused words was measured across 7 prosodic parameters (duration, mean F_0, maximum F_0, F_0 range, root-mean-square (RMS) mean intensity, maximum intensity, and intensity range). Although the dialogues are thus not fully spontaneous speech, they served as a controlled and structured means of eliciting natural production of prosodic focus. All participant performed the five dialogues with the same experimenter who was bilingual in both English and Mandarin.

The present experiment has two major aims. First, we seek to (1a) confirm whether speakers of English and Mandarin produce prosodic focus in the same way, and (1b) to the extent that they do, whether there is still cross-language variation in the degree to which speakers use the different prosodic parameters. Second, we address (2a) whether five focus types are conveyed by different prosodic realisation, and if so, (2b) whether the pattern of such difference is uniform across languages.

2.2. Method

2.2.1. *Participants*

Productions were obtained from 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). All of the English speakers reported that they were born and raised in Australia, while the Mandarin speakers were born in Mainland China and had been living in Australia for less than ten years (*M*= 2.84 years; range: 2 months – 9 years).

Given the prosodic differences between the Mandarin spoken in Mainland China and other parts of the Sinophone world (e.g., Xu et al., 2012), additional data from three further Mandarin speakers who grew up in communities outside of Mainland China (e.g., Taiwan) were excluded from final analysis. We also excluded data from one further English-speaking participant who appeared to have some disfluency in oral reading (e.g., occasional unintended pauses between words). Participants were naïve to the purpose of the experiment and had no self-reported hearing or speech impairment. The English speakers were recruited via an undergraduate subject pool and the Mandarin speakers were recruited using advertisements around the university campus. All participants were university students at the time of the experiment.

2.2.2. *Dialogue Scripts*

Four of the five types of focus were based on Krifka's (2006) proposal of the various pragmatic functions of focus. These were: focus used in response to wh-questions (wh-focus), in correction (corrective focus), in confirmation (confirmatory focus), and in parallel constructions (parallel focus). A fifth type of focus was that involving introduction of new information (new-information focus: e.g., Halliday, 1967; Jackendoff, 1972). New-information focus refers to discourse-new non-presupposed information that is unpredictable (i.e., pragmatically non-recoverable, Lambrecht, 1994) from the preceding utterances (see Table 1 for examples from each focus type in English and Mandarin).

Dialogues written in casual Australian English and Standard Mandarin were constructed to elicit participants' production of prosodic focus (see Appendices B and C). The English and the Mandarin versions of the dialogues all went through several iterations to perfect them. In each language, we used five dialogues, where each dialogue contained pairs involving the same word or phrase tokens in a focused versus unfocused realisation. For each of the focused and unfocused tokens, we measured 7 prosodic parameters: duration, mean F_0, maximum F_0, F_0 range, mean RMS-intensity, maximum RMS-intensity, and RMS-intensity range. Different focus types appeared throughout all five dialogues, although not equally often. Unfocused tokens were defined as presupposed/given information in the information-structural contexts of the dialogues, and in most cases, were words or phrases that had already been made salient as focused tokens earlier in the dialogues. Both within and across each dialogue, there were cases where some token pairs occurred more than once (see Table 2).

In total, each language has 52 pairs of focused and unfocused tokens (48 words; 4 phrases), with 12 pairs in the first dialogue (4 new-information, 2 wh-question, 2 corrective, 2 confirmatory, 2 parallel), 9 pairs in the second dialogue (4 new-information, 4 corrective, 1 confirmatory), 12 pairs in the third dialogue (3 new-information, 2 wh-question, 6 corrective, 1 confirmatory), 12 pairs in the fourth dialogue (6 new-information, 5 corrective, 1 confirmatory), and 7 pairs in the fifth dialogue (5 corrective, 2 parallel). In consequence, we compiled data from a total of 34,944 measurements (2 languages × 24 participants × 52 pairs × 2 focus levels × 7 prosodic parameters).

Since we were relying on reading materials, we have taken extra steps to exclude participants who produced mostly read speech. The naturalness of participants' focus production can be reflected by the fact that they did not randomly emphasise words in the dialogue scripts. We would exclude participants who randomly emphasised words that were not meant to be emphasised, or failed to mark focus on a large number of designated focused words. However, none of our participants had to be excluded for this reason.

Table 1. *Examples of focused and unfocused tokens in English and Mandarin. Underlined words indicate unfocused/given tokens.*

English	Mandarin
New-information focus **Vendor (Experimenter)**: hmm…What about 80 dollars for each sweater? **Buyer (Participant)**: 80 dollars is still too much… (*looking at the green* sweater) Oh look! There's a **[STAIN]** on the green sweater. Maybe you can reduce your price a bit since there is a stain on one of your sweaters.	***New-information focus*** **小贩 (实验者)**: 那么…每件 80 块，行不行? **顾客(受试者)**: 唉呀! 80 块还是太贵了… (*正在看着绿毛衣*) 哎, 你看! 你看这绿毛衣这块 **[脏了]** … 能不能再便宜点，既然绿毛衣都脏了。
Wh-focus **Police** (**Experimenter**): Who did you give the ring to? **John** (**Participant**): I gave it to [**MARY**]. …(*3 turns later*)… I only showed Mary the….	***Wh-focus*** **警察 (实验者)**: 你把这戒指交给谁了? **温西山(受试者)**: 我把它交给了**[马小姐]**。 …(*3 turns later*)… 我只给马小姐看过我的...
Corrective focus **Inspector** (**Experimenter**): ... you heard two books dropped? **Student** (**Participant**): No, I heard two [**GUNSHOTS**]. …(*1 turn later*)… Yes that's right, I heard two gunshots…	***Corrective focus*** **警察 (实验者)**: …你突然间听到两声炮响? **学生 (受试者)**: 不是，我听到 **[枪响]。** …(*1 turn later*)…是啊没错，我听到两声枪响…
Confirmatory focus **Police** (**Experimenter**): The ruby ring is for your fiancée? **John** (**Participant**): Yes, the ruby ring is for my [**FIANCÉE**] …(*9 turns later*)… I now have nothing to give to my fiancée.	***Confirmatory focus*** **警察 (实验者)**: 这红宝石戒指, 是给你未婚妻的吗? **温西山(受试者)**: 是啊, 这红宝石戒指是给我**[未婚妻]**的..(*9 turns later*)..我没有订婚戒指给我未婚妻了。
Parallel focus **Buyer** (**Participant**): … a [**GREEN**] sweater for my friend and a [**RED**] sweater for my sister. …(*3 turns later*)…I'd be happy to pay fifty dollars for the green sweater and another fifty dollars for the red sweater.	***Parallel focus*** **顾客(受试者)**: 我要买件**[绿色]**的毛衣给我朋友，**[蓝色]**的毛衣给我弟弟。 …(*3 turns later*)…我愿意花五十块买一件绿色毛衣，再加五十块买一件蓝色的毛衣。

Table 2. *Target tokens used in the experiment with rough IPA transcriptions in Mandarin. *= Occurred twice.*

	English	Mandarin		Dialogue
New-information	Green	绿色	/ly4 sɤ4/	1
	Two	两	/ljaŋ3/	1
	Fifty	五十	/wu3 ʂɨ2/	1
	Stain	脏了	/tsaŋ1 lɤ5/	1
	Gunshots	枪响	/tɕʰjaŋ1 ɕjaŋ3/	2
	Whispering	叽叽喳喳	/tɕi1 tɕi1 ʈʂa1 ʈʂa1/	2
	Argument	争论	/ʈʂəŋ1 lɤn4/	2
	Read	读过	/tu2 gwo4/	2
	Engagement Ring*	订婚戒指*	/tiŋ4 xwən1 tɕje4 ʈʂɨ3/	3
	Missing	不见	/pu2 tɕjɛn4/	3
	Geology	地理学	/ti4 li3 ɕɥe2/	4
	Volcano	火山	/xwo3 ʂan1/	4
	Mt Wilson	无功山	/wu2 gʊŋ1 ʂan1/	4
	Hundred and Fifty Metres Thick	百五十米厚	/pai3 wu3 ʂɨ2 mi3 xou4/	4
	Fourteen Million Years	十四亿年	/ʂɨ2 sɨ4 ji4 njɛn2/	4
	Blue Mountains	玉林县	/y4 lin2 ɕjɛn4/	4
Wh-question	Sweater	毛衣	/mau2 ji1/	1
	Blue	红色	/xʊŋ2 sɤ4/	1
	Mary	马小姐	/ma3 ɕjau2 tɕje3/	3
	Counter	柜台	/kwei4 tʰai2/	3
Corrective	Sweater	毛衣	/mau2 ji1/	1
	Blue	红色	/xʊŋ2 sɤ4/	1
	Library	图书馆	/tʰu2 ʂu1 kwan3/	2
	Reading	读书	/tu2 ʂu1/	2
	Gunshots	枪响	/ tɕʰjaŋ1 ɕjaŋ3/	2
	Book	书	/ʂu1/	2
	Second	第二次	/di1 ɚ4 tsʰɨ4/	3
	Ruby*	红宝石*	/xʊŋ2 pau3 ʂɨ2/	3
	Mary	马小姐	/ma3 ɕjau2 tɕje3/	3
	Return	还给我	/xwan2 kei2 wo3/	3
	In My Bag	手提包里	/ʂou3 tʰi2 pau1 li3/	3
	Geology	地理学	/ti4 li3 ɕɥe2/	4
	Below	下面	/ɕja4 mjɛn4/	4
	Mt Wilson	无功山	/wu2 gʊŋ1 ʂan1/	4
	Sydney	沈阳	/ʂən3 jaŋ2/	4
	West	西	/ɕi1/	4
	Reporter	记者	/tɕi4 ʈʂɤ3/	5
	National	国立	/kwo2 li4/	5
	Confirm What Has Happened	到底发生了什么事	/tau4ti3fa1ʂəŋ1lɤ5 ʂɤn3mo2ʂʰɨ4/	5
	Full	全部	/ʈʂʰwaŋ2 pu4/	5
	Detective	侦探	/ʈʂən1 tʰan4/	5
Confirmatory	Blue	红色	/xʊŋ2 sɤ4/	1
	Two	两	/ljaŋ3/	1
	Two	两声	/ljaŋ3 ʂəŋ1/	2
	Fiancée	未婚妻	/wei4 xəŋ1 tɕi1/	3
	Hundred and Fifty Metres Thick	百五十米厚	/pai3 wu3 ʂɨ2 mi3 xwo4/	4
Parallel	Green	绿色	/ly4 sɤ4/	1
	Red	蓝色	/lan2 sɤ4/	1
	Local	当地	/taŋ1 ti4/	5
	National	国际	/kwo2 tɕi4/	5

The English and Mandarin dialogues were comparable (close translations), with only small deviations. Occasional minor deviation in translation can be found in some adjectives and nouns (e.g., whether the colour of the sweater was “*red*” or “*blue*”; whether it was “*national*” or “*international*”; or whether the lava was “*fourteen million*” or a “*hundred million*” years old), as we attempted to maintain phonological similarity across the two language versions (e.g., by using words with similar vowel frontness and/or openness or with similar number of syllables). For example, we changed the colour of the sweater from “*red*” in the English dialogue to *blue* “*蓝*” /lan2/ in Chinese because the vowel in the Chinese word for “blue” is closer to the /ɛ/ in the English “red” in terms of both vowel frontness and height than the vowel in the Chinese word for “red” “*红*” /xʊŋ1/. Further, both sets of dialogues involved the same focused and unfocused tokens within the same discourse contexts, and with a few exceptions, most of the focused tokens do not co-occur with other means of focus expression (e.g., syntax, focus-sensitive particles). At the same time, to optimise comparability between the focused and unfocused tokens, we ensured that each focused token and its unfocused counterpart occurred in similar utterance positions. Further, the utterance positions of the focused and unfocused tokens for most pairs were the same across both languages.

There were a variety of different discourse contexts among the five dialogues. The first dialogue involved a conversation between a buyer (participant) and a street vendor (experimenter). In the second dialogue, the participant played a high-school student who was being questioned by a police inspector (experimenter). The policeman (experimenter) in the third dialogue was enquiring about a missing ring that belonged to a wealthy customer (participant) at a jewellery store. The fourth dialogue was between a primary school teacher (experimenter) and a student (participant), and the fifth dialogue was based on a job interview conducted by a news company employer (experimenter), with a university graduate (participant).

2.2.3. *Recording Procedures*

All recordings were made in a sound-attenuated booth at the MARCS Institute, Western Sydney University, using a Shure SM10A-CN headset microphone connected to a laptop via a Roland Quad-Capture USB-based audio interface. All dialogues were performed by an individual participant with the experimenter. Recording sessions for each dialogue lasted for approximately five to six minutes, and both roles had equal numbers of turns in each dialogue, except for the fourth dialogue where the participant's role had an extra turn in the beginning.

Participants sat opposite the experimenter and were asked to spend a few minutes reading through each of the dialogues by themselves to prepare for their role before each recording session. Participants still had access to the dialogue scripts during the recordings. To ensure successful elicitation of focus, participants were encouraged to immerse themselves in their roles and be "as natural and genuine as possible". In addition, the experimenter asked all participants to pay careful attention to how they chose to speak in each dialogue. However, as aforementioned, the experimenter gave no explicit instructions to emphasise the focused tokens, and the dialogues were presented in plain text without any typefaces (e.g., boldface, italics) that could indicate the discourse status of focused versus unfocused tokens. All participants were tested by the same experimenter (the first author, who is fluent in both English and Standard Mandarin).

2.2.4. *Data Analyses*

All focused and unfocused tokens in each dialogue were manually segmented and annotated based on inspection of the waveform and the spectrogram in Praat

(Boersma & Weenink, 2018). All tokens were measured for duration (in milliseconds), mean F_0, maximum F_0, F_0 range (maximum F_0 minus minimum F_0), mean Root-Mean-Square (RMS) intensity, maximum intensity, and intensity range (maximum intensity minus minimum intensity). F_0 measures were in Hertz (Hz; note that our analyses were based on within-speaker prosodic differences between focused and unfocused tokens). The English and Mandarin samples also had almost equal proportions of male and female speakers.

In English, focus cues occur on the lexically stressed syllable (e.g., de Jong, 2004), but in Mandarin, cues can occur on any one or all syllables, depending on the word's stress pattern and semantic structure (Gu, Mori, & Kasuya, 2003). To optimise comparability in annotation, prosodic data in both languages were compiled from the entire word or phrase token, except for one particular word in English (i.e., "*Sweater*" /swɛtə/) where only /wɛ/ was segmented because many participants were creaky in their production of the second syllable. Further, in many cases, there were multiple instances of the same unfocused words in each dialogue. Data from each unfocused token were compiled from the same location, but in cases of missing data due to creakiness, a different unfocused token of the same word was used.

In keeping with the descriptive purpose of the paper, we have endeavoured to present the data extensively and with simple statistics focusing on the main parameters of interest. In each dialogue, data for every prosodic parameter for each token were first averaged across the 24 speakers of each language, producing language-specific estimates of each parameter by item. Item estimates were then averaged according to their focus type, which were then averaged across the five dialogues. For each parameter, a series of two-tailed pairwise *t*-tests was conducted to examine whether both languages showed similar patterns of production difference between the unfocused and the focused tokens. We further performed a series of mixed two-way 2 (English vs. Mandarin) × 2 (focused vs. unfocused) ANOVAs to reveal whether there were any cross-language differences in the degree to which speakers increased the different parameters for the focused tokens relative to the unfocused tokens. Significant threshold ($\alpha = .05$) for follow-up *t*-tests was adjusted using the Benjamini-Hochberg (1995) false discovery rate control procedure.

2.3. Results

The results will be presented in three parts. In Sections 2.3.1, we present the crosslanguage differences in the degree to which speakers use the different prosodic parameters to differentiate between focused and unfocused tokens in each of the five focus types averaged across the dialogues. In Section 2.3.2, we present the results of the acoustic analyses concerning the acoustic correlates of prosodic focus in English and Mandarin. In Section 2.3.3, we provide further analyses of the crosslanguage differences within each of the five dialogues.

2.3.1. *Crosslanguage Differences*

We conducted analyses to reveal whether there were crosslanguage differences in the degree to which speakers produced the increases on the different prosodic parameters. Thereby, we conducted a series of mixed two-way 2 (English vs. Mandarin) × 2 (focused vs. unfocused) ANOVAs on the seven prosodic parameters. Thirty-five (5 focus types × 7 parameters) ANOVAs were conducted. Data from each prosodic parameter in each focus type were averaged across all the dialogues. Significant crosslanguage differences from the analyses (i.e., significant interaction effects) are presented in Figures 1 and 2.

Analyses of duration measures averaged across dialogues revealed no crosslanguage differences between English and Mandarin speakers for any of the focus types. For F_0 measures, crosslanguage differences were found for new-information focus in the degree to which English and Mandarin speakers increased their mean F_0, $F(1, 46) = 7.96$, $p = .007$. Simple effects of focus for English and Mandarin revealed that the increase in mean F_0 was greater in Mandarin ($p < .001$), although it was also significant in English speakers ($p < .001$). Similarly, there were also crosslanguage differences in the production of new-information focus for both maximum F_0, $F(1, 46) = 8.50$, $p = .005$, and F_0 range, $F(1, 46) = 13.30$, $p = .001$, where in both cases, Mandarin speakers exhibited a greater increase (all p-values < .001) than English (p-value < .001 for maximum F_0; p-value = .001 for F_0 range). Moreover, there were also crosslanguage differences in corrective focus in maximum F_0, $F(1, 46) = 8.34$, $p = .006$, where Mandarin speakers (p-values < .001) showed a greater increase than English speakers ($p = .024$), and in F_0 range, $F(1, 46) = 12.29$, $p = .001$, where only Mandarin speakers showed greater F_0 range expansion ($p < .001$).

For the RMS-intensity measures, there was a significant crosslanguage difference in mean intensity in the production of new information focus, $F(1, 46) = 4.97$, $p = .031$, where English speakers showed a greater increase than Mandarin speakers (both p-values < .001). In wh-focus, significant crosslanguage differences occurred for all intensity measures; mean intensity, $F(1, 46) = 7.39$, $p = .009$; maximum intensity, $F(1, 46) = 8.99$, $p = .004$; intensity range, $F(1, 46) = 7.38$, $p = .009$. For the mean intensity, only English speakers displayed a significant difference ($p < .001$), whereas for maximum intensity and intensity range, both the English and Mandarin data showed a significant difference, with both cases showing English speakers producing a greater degree of increase (all p-values < .001) than Mandarin speakers (for

maximum intensity, p = .010; intensity range, p < .001. Finally, there was crosslanguage difference in parallel focus for maximum intensity, $F(1, 46) = 5.25$, p = .027, where only English speakers produced a significant difference in focus production (p = .001).

In contrast to the above consistency of a greater increase in intensity for English compared to Mandarin speakers, in confirmatory focus a significant difference in intensity range, $F(1, 46) = 4.71$, p = .035, was due to Mandarin speakers only (p = .003).

To summarise, there were no crosslanguage differences in duration for any of the focus types. For F_0, there were crosslanguage differences in new-information (mean, maximum, range) and corrective focus (maximum, range), where in all cases, Mandarin showed a greater degree of production increase than English. For intensity, wh-focus showed crosslanguage differences on mean, maximum, and range. There were also crosslanguage differences in new-information, parallel, and confirmatory focus on mean intensity, maximum intensity, and intensity range respectively. All cases of intensity differences showed a greater degree of production increase by English speakers, except for the crosslanguage difference in parallel focus.

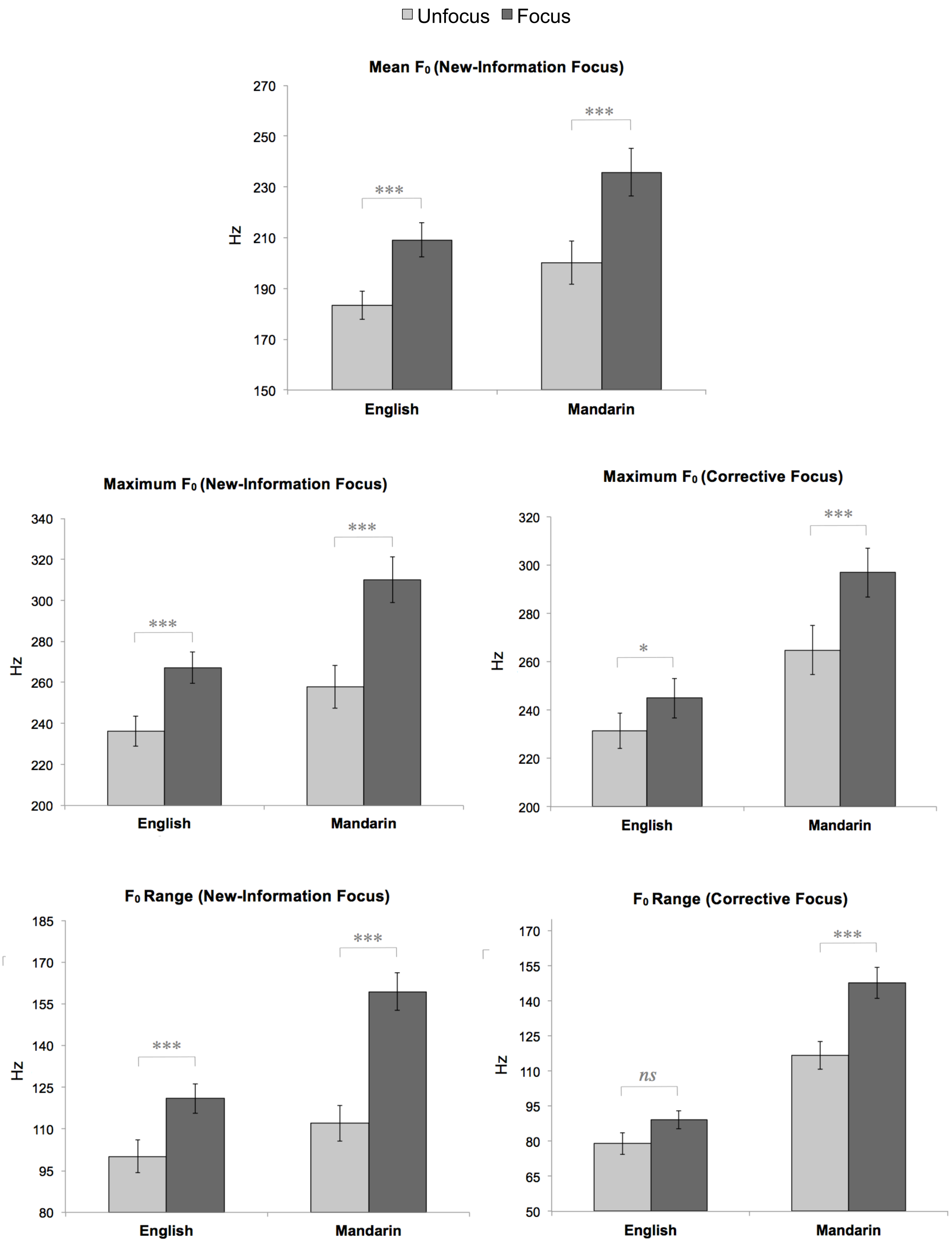


Figure 1. Significant crosslanguage F$_0$ differences in production of focused (black) and unfocused (light grey) tokens. Error bars indicate standard error of the mean. *p ≤ .05, **p ≤ .01, ***p ≤ .001.

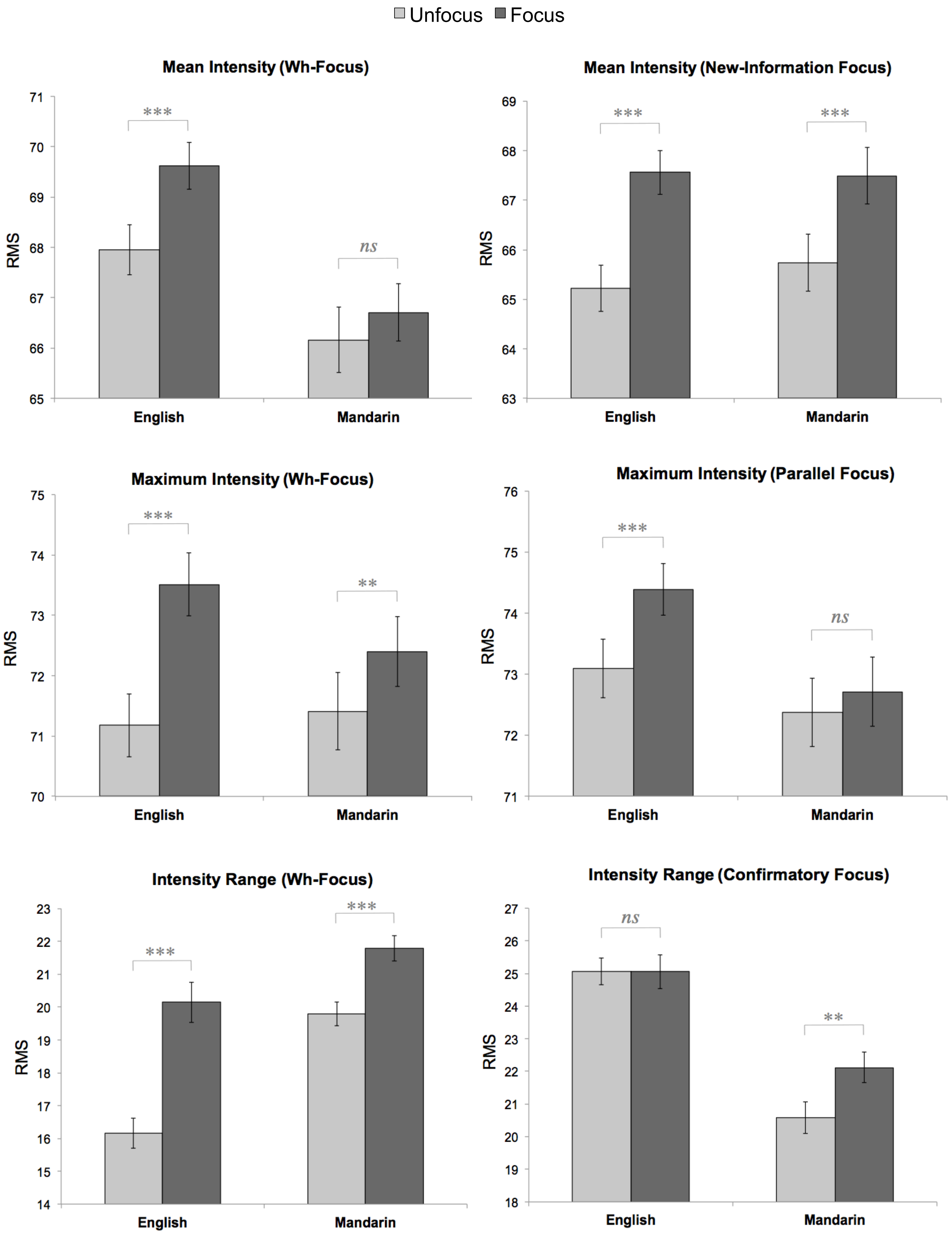


Figure 2. Significant crosslanguage intensity differences in production of focused (black) and unfocused (light grey) tokens. Error bars indicate standard error of the mean. **p ≤ .05, **p ≤ .01, ***p ≤ .001*.

2.3.2. *Acoustic Correlates of Focus*

Results for English and Mandarin speakers' production of each prosodic parameter in the five different focus types, averaged across the five dialogues, are presented in Tables 3 to 7. Overall, a series of pairwise *t*-tests showed similar patterns of production increase from unfocused to focused tokens. However, some variation across different focus types was observed.

For new-information focus, speakers from both language groups showed a significant difference on all prosodic parameters. Compared to unfocused tokens, production of focused tokens in both languages show a lengthened duration, higher average F_0 and maximum F_0, and greater F_0 range expansion. Further, both English and Mandarin speakers produced the focused tokens with greater mean and maximum intensity and intensity range.

For wh-focus, the pattern of production increase was also the same across both languages, except for one intensity measure (i.e., mean intensity), where only English speakers produced a significant difference. Apart from this difference, both groups of speakers produced greater increases on all of the other prosodic parameters (i.e., longer duration, higher mean and maximum F_0, greater F_0 range, maximum intensity, and intensity range). Similarly, for corrective focus, both languages revealed significant production increases on all parameters except for maximum intensity, where only English speakers produced a significant difference in maximum intensity.

For confirmatory focus, English speakers did not show any production prosodic differences between focused and unfocused tokens. In Mandarin, speakers only produced confirmatory focus with a difference in duration, mean F_0, and in intensity range. For parallel focus, English speakers only showed a difference in mean and maximum intensity, while Mandarin speakers did not show any difference on any parameters.

Table 3. *Prosodic differences between focused and unfocused tokens in new-information focus*

	English						Mandarin					
				95% CI						95% CI		
	Unfocused	Focused	SEM	Lower	Upper	t	Unfocused	Focused	SEM	Lower	Upper	t
Duration (ms)	549	604	7.530	38.31	69.47	7.16***	502	547	5.902	33.15	57.57	7.69***
Mean F_0	183.42	209.13	2.374	20.80	30.61	10.83***	200.12	235.81	2.626	30.26	41.12	13.59***
Maximum F_0	236.25	267.27	4.999	20.68	41.36	6.21***	257.91	310.16	5.299	41.29	63.21	9.86***
F_0 Range	100.07	120.98	5.187	10.18	31.64	4.03***	112.09	159.40	5.051	36.86	57.76	9.37***
Mean Intensity	65.22	67.56	0.167	2.00	2.69	14.01***	65.74	67.49	0.202	1.34	2.17	8.67***
Maximum Intensity	70.94	73.81	0.228	2.41	3.35	12.61***	71.41	74.37	0.837	1.23	4.69	3.54**
Intensity Range	23.58	27.70	0.331	3.44	4.81	12.47***	22.34	25.05	0.926	0.80	4.63	2.93**

$^{**}p \le .01$, $^{***}p \le .001$.

Table 4. *Prosodic differences between focused and unfocused tokens in wh-focus*

	English						Mandarin					
				95% CI						95% CI		
	Unfocused	Focused	SEM	Lower	Upper	t	Unfocused	Focused	SEM	Lower	Upper	t
Duration (ms)	257	349	11.336	68.49	115.39	8.11***	444	528	7.793	68.06	100.30	10.80***
Mean F_0	188.46	207.07	5.420	7.40	29.82	3.43**	194.13	213.77	3.250	12.91	26.36	6.04***
Maximum F_0	210.09	244.71	7.408	19.29	49.94	4.67***	265.31	303.97	8.829	20.40	56.92	4.38***
F_0 Range	45.38	82.96	6.493	24.15	51.01	5.79***	120.06	159.76	10.923	17.11	62.30	3.64***
Mean Intensity	67.95	69.62	0.236	1.18	2.16	7.07***	66.16	66.70	0.341	-0.16	1.25	1.59
Maximum Intensity	71.18	73.51	0.279	1.76	2.91	8.39***	71.41	72.40	0.352	0.26	1.72	2.81**
Intensity Range	16.15	20.15	0.611	2.74	5.26	6.55***	19.79	21.79	0.412	1.15	2.85	4.85***

$^{**}p \le .01$, $^{***}p \le .001$.

Table 5. *Prosodic differences between focused and unfocused tokens in corrective focus*

	English						Mandarin					
				95% CI						95% CI		
	Unfocused	Focused	SEM	Lower	Upper	t	Unfocused	Focused	SEM	Lower	Upper	t
Duration (ms)	388	444	4.733	45.84	65.43	11.75***	471	541	7.113	55.55	84.98	9.88***
Mean F_0	191.28	203.56	3.045	5.99	18.59	4.04***	200.12	218.36	1.915	14.28	22.20	9.53***
Maximum F_0	231.45	244.90	5.587	1.90	25.01	2.41*	264.82	296.89	3.219	25.42	38.73	9.96***
F_0 Range	78.98	89.19	4.804	0.27	20.15	2.13	116.69	147.75	3.505	23.81	38.31	8.86***
Mean Intensity	66.29	67.89	0.226	1.13	2.06	7.06***	64.43	65.60	0.186	0.79	1.56	6.31***
Maximum Intensity	71.52	73.44	0.543	0.80	3.04	3.53**	70.34	71.39	0.573	-0.14	2.23	1.83
Intensity Range	22.10	24.59	0.378	1.71	3.27	6.58***	20.67	22.60	0.644	0.60	3.26	3.00**

$*p \leq .05$, $**p \leq .01$, $***p \leq .001$.

Table 6. *Prosodic differences between focused and unfocused tokens in confirmatory focus*

	English						Mandarin					
				95% CI						95% CI		
	Unfocused	Focused	SEM	Lower	Upper	t	Unfocused	Focused	SEM	Lower	Upper	t
Duration (ms)	662	683	16.600	-13.32	55.56	1.27	537	559	9.455	3.35	42.47	2.42*
Mean F_0	204.59	211.63	3.954	-1.14	15.22	1.78	200.60	207.73	2.647	1.65	12.61	2.69*
Maximum F_0	249.01	253.01	7.53	-11.58	19.59	0.53	258.73	271.06	7.675	-3.54	28.21	1.61
F_0 Range	89.44	87.81	8.656	-19.53	16.28	-0.19	111.49	119.29	8.557	-9.91	25.50	0.91
Mean Intensity	65.15	65.55	0.285	-0.19	0.99	1.40	66.52	66.88	0.308	-0.278	1.00	1.17
Maximum Intensity	70.47	70.82	0.343	-3.61	1.06	1.02	72.13	72.78	0.356	-0.08	1.39	1.83
Intensity Range	25.07	25.06	0.536	-1.12	1.10	-0.02	20.58	22.12	0.471	0.57	2.51	3.27**

$*p \leq .05$, $**p \leq .01$.

Table 7. *Prosodic differences between focused and unfocused tokens in parallel focus*

	English						Mandarin					
				95% CI						95% CI		
	Unfocused	Focused	SEM	Lower	Upper	t	Unfocused	Focused	SEM	Lower	Upper	t
Duration (ms)	311	323	8.403	-5.09	29.67	1.46	348	354	7.973	-10.21	22.77	0.79
Mean F_0	199.57	206.94	4.768	-2.50	17.22	1.54	228.83	232.49	3.112	-2.78	10.10	1.18
Maximum F_0	230.31	237.03	7.127	-8.02	21.46	0.94	295.44	301.01	5.099	-4.98	16.12	1.09
F_0 Range	60.03	61.82	5.791	-10.19	13.77	0.31	124.28	126.33	4.826	-7.93	12.03	0.43
Mean Intensity	68.60	69.73	0.348	0.41	1.85	3.24**	67.48	67.88	0.237	-0.09	0.89	1.70
Maximum Intensity	73.09	74.39	0.341	0.60	2.01	3.82***	72.37	72.71	0.244	-0.16	0.85	1.40
Intensity Range	23.76	24.26	0.626	-0.79	1.80	0.81	19.33	20.00	0.518	-0.40	1.75	1.30

$^{**}p \leq .01$, $^{***}p \leq .001$

We have also conducted Linear Mixed Effect (LME) models to further explore the acoustic correlates of focus across English and Mandarin Chinese. This was done in R using the lme4 package (Bates, Maechler, & Dai, 2010). An LME analysis approach is suitable to the present research because it allows the effects of crossed and nested subjects and item factors to be taken into account within a single analysis.

We performed an LME regression to obtain the best fitting model predicting phoneme detection RT. It is important to note that, unlike our ANOVA and t-test results, the LME results for the unfocused condition are based on all unfocused tokens. As a starting point, we first used a baseline model with subject, token, and various other extraneous variables including participant gender, number of syllables, syllabic onset and coda structure, vowel height/frontness, tonal features, and whether the token pair was a word or phrase. Fixed effects parameters (i.e., Focus and Language) were then added in a step-wise fashion to determine which predictors significantly improved model fit. Consistent with our t-test results, our LME results (see Table 8) revealed a significant effect of Focus on all of the prosodic parameters.

Table 8.

Fixed effects results for Focus from the linear mixed-effect modelling analyses (values mapped on the intercept) based on the model with Focus and Language as added predictors.

Prosodic Variable	Fixed Effect for Focus		
	β	SE (β)	t
Duration	5.69e+02	6.12+00	9.29***
Mean F_0	12.32	1.38	8.93***
Maximum F_0	20.50	3.44	5.96***
F_0 Range	2.23e+01	3.75e+01	5.95***
Mean Intensity	1.01	0.14	7.16***
Maximum Intensity	1.55	0.46	3.34***
Intensity Range	1.11e+01	4.52e+00	2.26**

$^{**}p \leq .01$, $^{***}p \leq .001$ (two-tailed).

2.3.3. *Dialogue Differences*

We also conducted a series of mixed ANOVAs within each of the dialogues to examine whether there was variation in the pattern of crosslanguage differences across the five different dialogues. Note that not all focus types were present across the different dialogues (see Section 2.2). The crosslanguage differences within each dialogue are illustrated in Figures 3 to 7.

Dialogue 1 (Street Vendor). Analyses revealed a significant cross-language difference in the degree to which English and Mandarin speakers increased their duration for wh-focus, $F(1, 46) = 14.34$, $p < .001$. Simple effects of focus for the English and Mandarin speakers revealed that the increase in duration for wh-focus was longer in Mandarin ($p < .001$), although it was also significant in English speakers ($p = .001$). Similarly, there were also crosslanguage effects of duration in corrective focus, $F(1, 46) = 5.27$, $p = .026$, and in confirmatory focus, $F(1, 46) = 5.01$, $p = .030$, where in both cases Mandarin speakers (= $p < .001$ in each) produced a greater increase than English speakers (for corrective focus, $p = .003$; nonsignificant for confirmatory focus).

For F_0, there were cross-language differences in mean F_0 for new-information focus, $F(1, 46) = 7.53$, $p = .009$, and for parallel focus, $F(1, 46) = 5.10$, $p = .029$, where in both cases, English speakers (for new-information focus, $p < .001$; for parallel focus, $p = .003$) produced a greater increase than Mandarin speakers (for new-information focus, $p < .001$; nonsignificant for parallel focus). There were also significant cross-language differences in F_0 range for new-information focus, $F(1, 46) = 7.17$, $p = .010$, and for wh-focus, $F(1, 46) = 7.55$, $p = .009$. In new-information focus, both speakers significantly expanded their F_0 range, with higher increase in speakers of Mandarin than English (both p-values < .001). In wh-focus, only Mandarin speakers significantly expanded their F_0 range, ($p = .007$).

For mean and maximum RMS-intensity, significant cross-language differences occurred only in the production of new-information focus; for mean intensity, $F(1, 46) = 13.19$, $p = .001$; for maximum intensity, $F(1, 46) = 12.42$, $p = .001$. In both these cases, English speakers produced greater increases than Mandarin speakers (all p-values < .001). For intensity range, crosslanguage effects were found for both new-information focus, $F(1, 46) = 57.75$, $p < .001$, and confirmatory focus, $F(1, 46) = 15.27$, $p < .001$. For new-information focus, the increase in intensity range was greater in English than in Mandarin (all p-values < .001). However, for confirmatory focus, only Mandarin speakers showed a significant difference ($p < .001$).

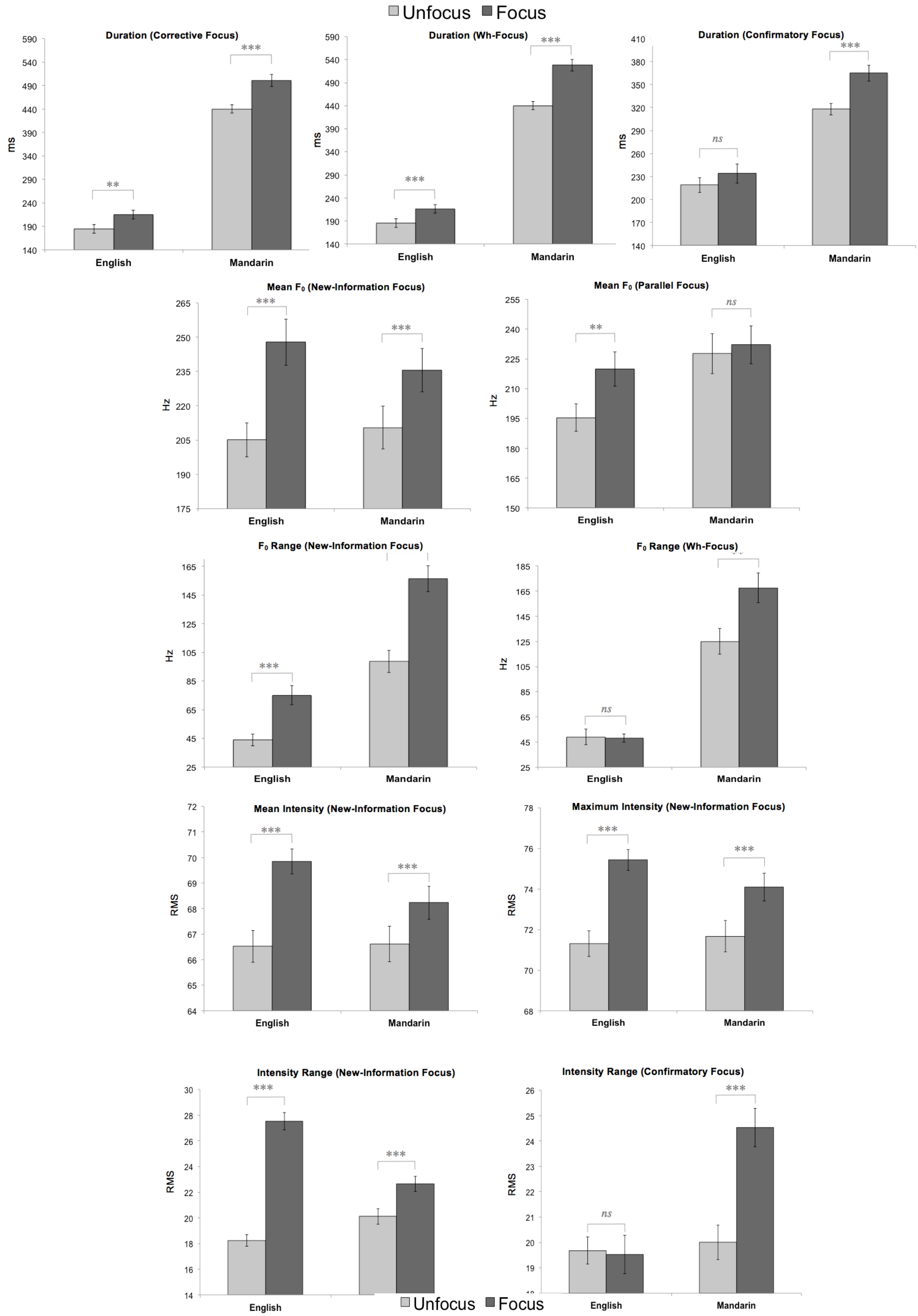


Figure 3. Significant crosslanguage differences in Dialogue 1. Error bars indicate standard error of the mean. *$p \le .05$, **$p \le .01$, ***$p \le .001$.

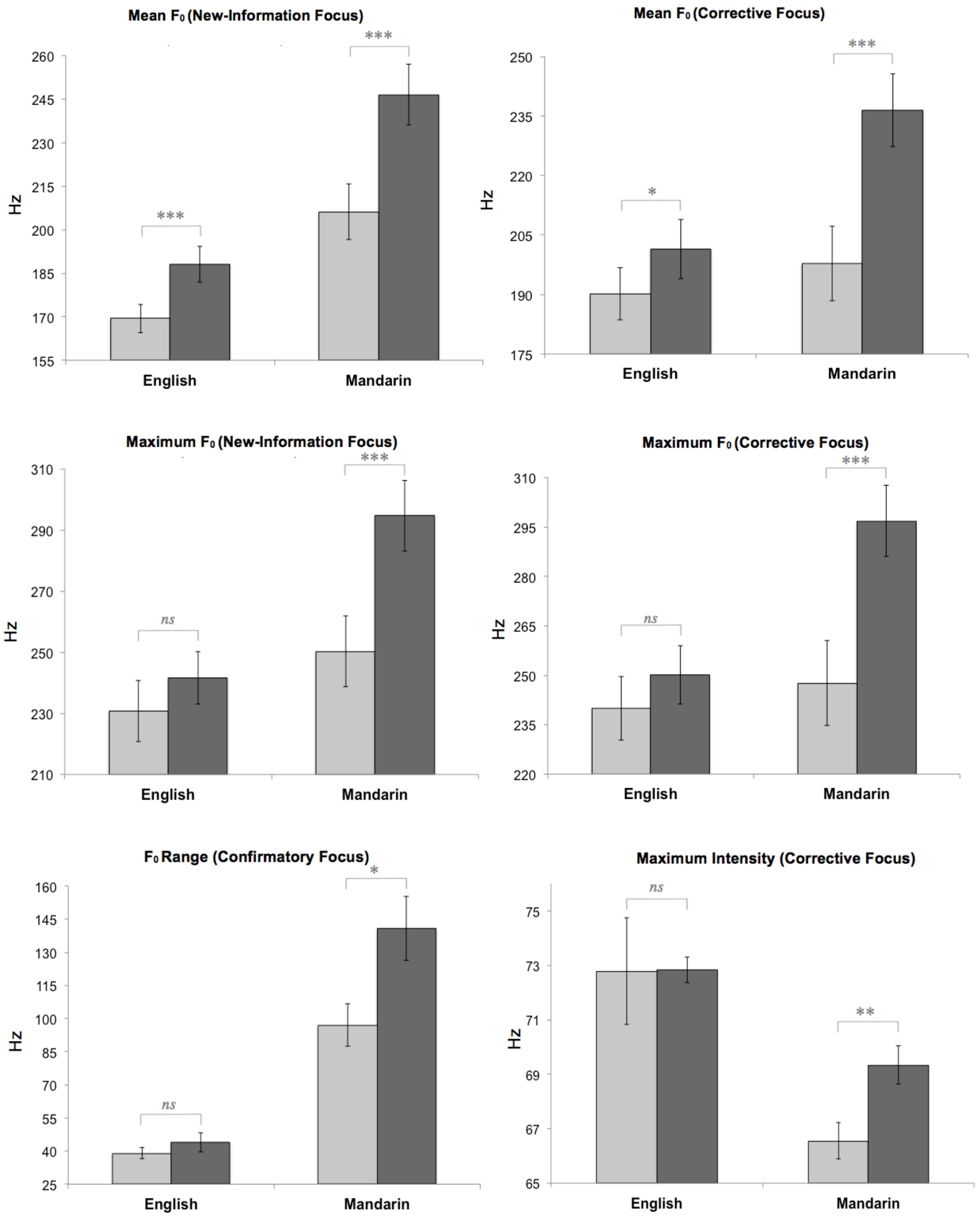


Figure 4. Significant crosslanguage differences in Dialogue 2. Error bars indicate standard error of the mean. *$*p \leq .05$, $**p \leq .01$, $***p \leq .001$*.

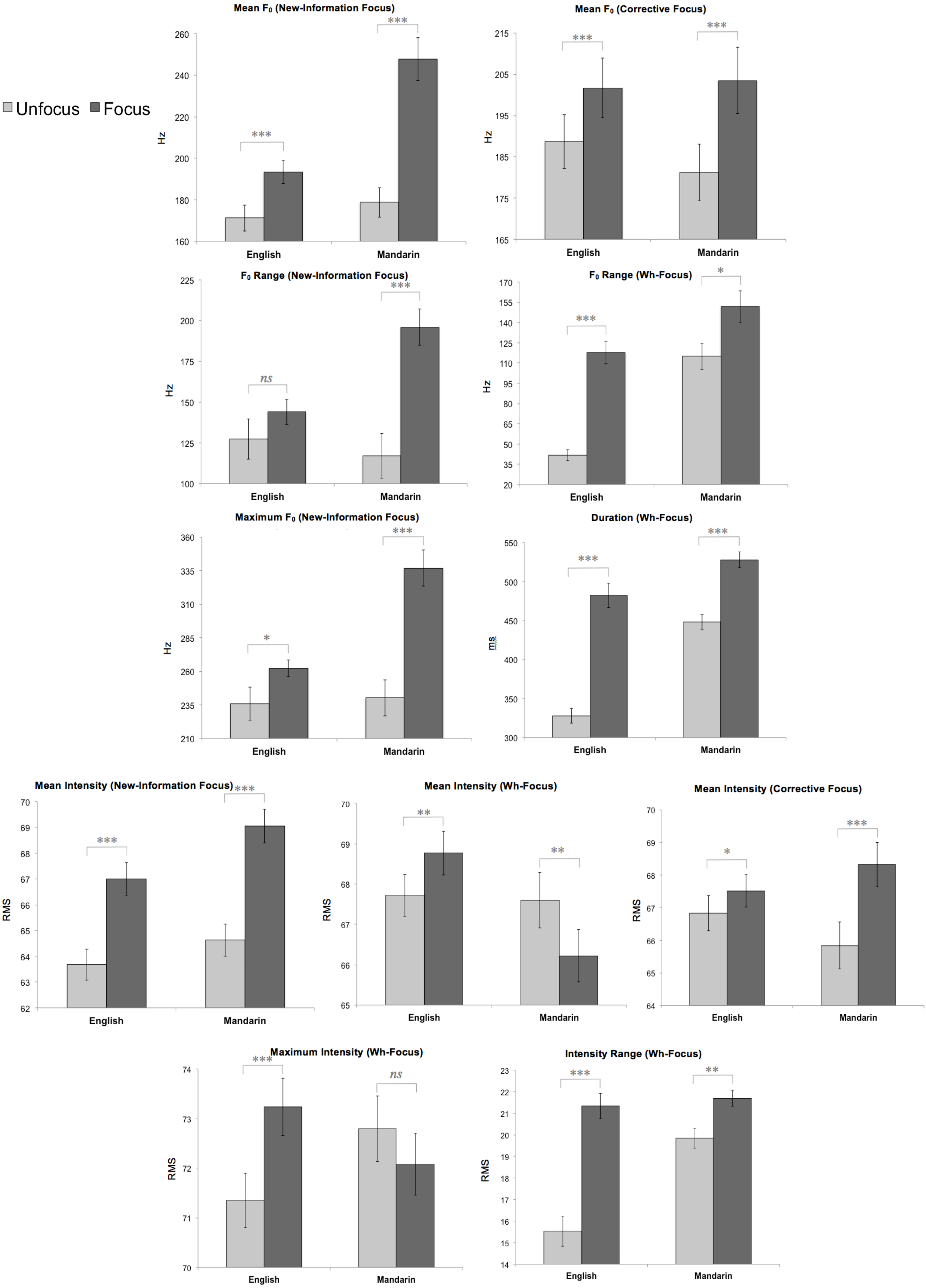


Figure 5. Significant crosslanguage differences in Dialogue 3. Error bars indicate standard error of the mean. **p ≤ .05, **p ≤ .01, ***p ≤ .001*.

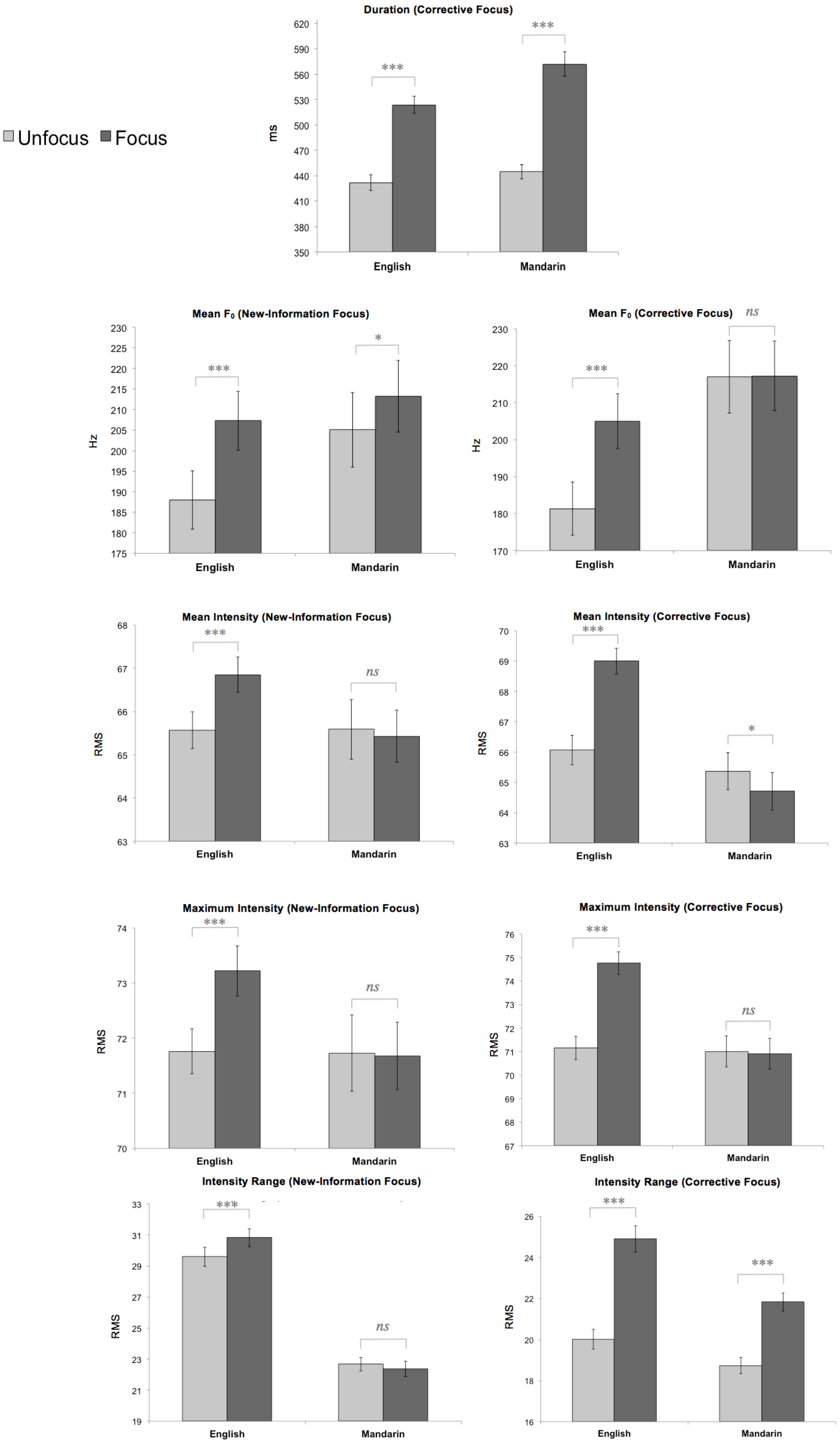


Figure 6. Significant crosslanguage differences in Dialogue 4. Error bars indicate standard error of the mean. **p ≤ .05, **p ≤ .01, ***p ≤ .001*.

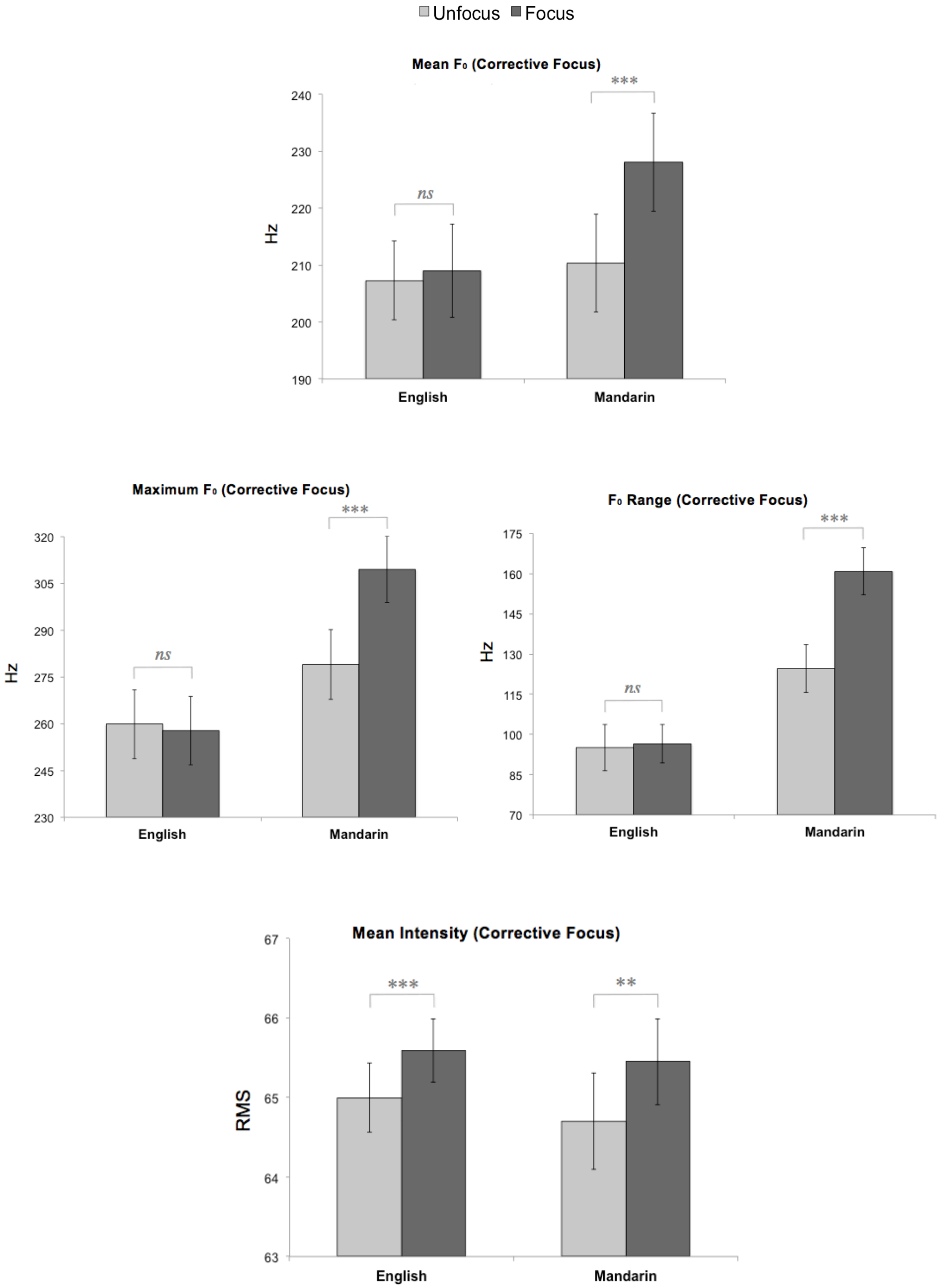


Figure 7. Significant crosslanguage differences in Dialogue 5. Error bars indicate standard error of the mean. *$*p \leq .05$, $**p \leq .01$, $***p \leq .001$*.

Dialogue 2 (Criminal Investigation). There were no significant cross-language differences for duration. For mean F_0, results revealed a significant cross-language difference for new-information focus, $F(1, 46) = 9.30$, p = .004, such that Mandarin speakers produced a greater increase (p < .001) than English speakers ($p = .001$). We also found a significant difference in mean F_0 for corrective focus, $F(1, 46) = 23.77, p < .001$, in which Mandarin speakers showed greater increase ($p < .001$) than English speakers ($p = .014$). For maximum F_0, crosslanguage differences were found for new-information focus, $F(1, 46) = 6.22, p = .016$, and corrective focus, $F(1, 46) = 15. 27, p < .001$, where in both cases only Mandarin speakers showed significant difference ($p < .001$). For F_0 range, there was a language effect for confirmatory focus, $F(1, 46) = 5.05, p = .029$, where only Mandarin speakers showed a significant difference ($p = .014$). For intensity, there was language difference in maximum intensity for corrective focus, $F(1, 46) = 4.37$, $p = .042$, in which only Mandarin speakers showed a significant difference ($p = .004$).

Dialogue 3 (Where Is My Ring?). Significant cross-language difference in duration was only observed for wh-focus, $F(1, 46) = 10.36, p = .002$, in which English speakers showed greater increase than Mandarin speakers (all p-values < .001). For F_0, there were significant cross-language effects for new-information focus; mean F_0, $F(1, 46) = 46.18$, $p < .001$; maximum F_0, $F(1, 46) = 10.81, p = .002$; F_0 range, $F(1, 46) = 8.34, p = .006$. In all these cases, Mandarin speakers produced a greater increase (all p-values < .001) compared to English (for mean F_0, $p < .001$; for maximum F_0, $p = .029$; not significant for F_0 range). Similarly, Mandarin speakers also produced a greater increase in the cross-language mean F_0 difference observed for corrective focus, $F(1, 46) = 4.79, p = .034$, compared to English (all p-values < .001). However, for wh-focus, English speakers produced greater increase in F_0 range, $F(1, 46) = 4.72, p = .035$ (for Mandarin, $p = .024$; for English, $p < .001$).

For mean intensity, various focus types showed significant crosslanguage differences; new-information focus, $F(1, 46) = 4.30$, $p = .044$; wh-focus, $F(1, 46) = 24.85$, $p < .001$; corrective focus, $F(1, 46) = 16.10$, $p < .001$. In all these cases, Mandarin speakers showed greater increase than English speakers (for Mandarin new-information focus and corrective focus, $p < .001$; for Mandarin wh-focus, $p = .002$ (in the opposite direction); for English new-information focus, $p < .001$; English wh-focus, $p = .002$; for English corrective focus, $p = .039$). However for intensity range, English speakers ($p < .001$) showed a greater increase for wh-focus, $F(1, 46) = 22.83$, $p < .001$, although the increase was also significant in Mandarin ($p = .005$). And for maximum intensity, wh-focus also showed a crosslanguage effect, $F(1, 46) = 14.23$, $p < .001$, and this time, only English speakers showed a significant difference ($p < .001$).

Dialogue 4 (Teacher and Student). For duration, only corrective focus showed a crosslanguage difference, $F(1, 46) = 4.20$, $p = .046$, such that Mandarin speakers produced a greater degree of increase than English speakers (all p-values < .001). For mean F_0, there were significant crosslanguage differences in new-information focus, $F(1, 46) = 7.39$, $p = .009$, where English speakers ($p < .001$) produced a greater increase than Mandarin speakers ($p = .011$). There was also a crosslanguage mean F_0 difference for corrective focus, $F(1, 46) = 21.73$, $p < .001$, where only English speakers showed a significant difference ($p < .001$). Similarly, new-information focus also showed crosslanguage effects in all of the intensity measures; mean intensity, $F(1, 46) = 16.78$, $p < .001$; maximum intensity, $F(1, 46) = 11.37$, $p = .002$; intensity range, $F(1, 46) = 4.91$, $p = .032$. In all cases of new-information focus intensity differences, only English speakers produced a significant difference; for mean and maximum intensity, $p < .001$; for intensity range, $p = .037$). A similar trend was also found for corrective focus; mean intensity, $F(1, 46) = 80.87$, $p < .001$; maximum intensity, $F(1, 46) = 70.12$, $p < .001$;

intensity range, $F(1, 46) = 5.78$, $p = .020$. And all these effects showed a significantly greater increase in English (all p-values < .001) compared to Mandarin; mean intensity, $p = .023$; intensity range, $p < .001$; not significant for maximum intensity).

Dialogue 5 (The Job Interview). No significant differences were observed for duration. For corrective focus, only Mandarin speakers (all p-values < .001) showed a significant difference on all F_0 measures; mean F_0, $F(1, 46) = 12.16$, $p = .001$; maximum F_0, $F(1, 46) = 8.80$, $p = .005$; F_0 range, $F(1, 46) = 8.38$, $p = .006$. In the crosslanguage difference for corrective focus in mean intensity, $F(1, 46) = 5.01$, $p = .030$, Mandarin speakers ($p = .005$) showed a greater degree of increase, although it was significant in English speakers ($p < .001$).

2.4. General Discussion

The present experiment sheds new light on the production of prosodic focus in general, and also on the language-particular strategies that underlie speakers' use of prosody. In line with results from previous research (e.g., Chen & Gussenhoven, 2008; Xu, 1999; Xu & Xu, 2005), our experiment shows that native speakers of English and Mandarin are alike in their tendency to express focus by manipulation of duration, F_0, and intensity. However, our results also reveal cases where the two languages did not pattern similarly in the degree to which speakers employed the various prosodic parameters. Based on the prosodic data that were averaged across all the dialogues, there was a systematic trend in cross-language variation where English and Mandarin speakers differed significantly in their production of intensity and F_0. For intensity, English speakers consistently produced greater degrees of increase in mean and maximum intensity. By contrast, in all cases of language-specific differences involving F_0, Mandarin speakers were more likely to produce a greater increase in mean and maximum F_0 as well as F_0 range.

Together, these findings provide evidence that there can still be language-specific differences in focus expression despite similar phonetic cues used to indicate focus across the two languages. These differences indicate subtle variation in the use of the same prosodic resources that are available and used by speakers in both languages. It is also important to note that this crosslanguage variation in realisation of F_0 and intensity also happened to correspond with previous work on English and Mandarin speakers' use of these cues in signalling lexical contrasts. In Mandarin, F_0 contour and height contrasts are the primary acoustic parameters that determine lexical tone identity (Jongman, Wang, Moore, & Sereno, 2006). Likewise, in English, a lexical stress language where focus falls on the primary stressed syllable, intensity has previously been found to be the most reliable cue to syllabic prominence across many dialects (Kochanski et al., 2006).

On the one hand, the fact that speakers would highlight focus with greater increases on the parameters that also signal lexical items is somewhat surprising. For instance, Chen and Gussenhoven (2008) analysed Mandarin speakers' duration and F_0 range expansion under various degrees of emphasis in corrective focus (i.e., emphasis vs. more emphasis) and found that duration was relied on more consistently than F_0, arguably because speakers were restricted in their manipulation of cues that already serve another purpose. Similarly, previous research shows that certain tones (e.g., the low-dipping tone) have a lower degree of freedom for F_0 expansion (e.g., Lee et al., 2016). On this interpretation, one would have predicted a trading relation in language production, where the processing weight of different prosodic dimensions of focus might have depended on their functional load in conveying other linguistic information.

On the other hand, the greater use of F_0 in Mandarin and intensity in English also reinforces the idea that different prosodic parameters can be highly flexible and multifunctional. Firstly, our findings on F_0 and intensity production increase provide a useful insight into how prosodic dimensions play a dual role in encoding both information structure and lexical contrast. As already observed in previous studies from Mandarin, focus production in a tone language involves enhanced distinctiveness of the tonal contrasts where lexical tones are encoded in the shape of F_0 contours while information structure is conveyed through span expansion (F_0 range) and level raising (mean and maximum F_0). The analogous case of intensity increases found in English demonstrates that speakers implement phonetic adjustments on those parameters that would facilitate better detection of the focused constituent, whether it be through enhancement of tonal contrasts or the lexically stressed syllable. Therefore, the greater increase in F_0 in Mandarin and intensity cues in English may even play a complementary role in supporting the speaker's lexical processing during focus production. Through our cross-language findings, we have further illustrated how the production system can make use of its fine-grained ability to implement the same prosodic cues for different linguistic functions.

A possible explanation for the language-specific difference in F_0 and intensity could be that speakers of different languages vary in the level of attention they pay to each prosodic parameter when signalling focus. When speakers choose for some reason to speak carefully, they tend to modify their output in ways that are similar to prosodic focus (e.g., articulating more slowly and loudly; Picheny, Durlach, & Braida, 1986). Since F_0 in Mandarin and loudness in English also play a lexical role, speakers may need to pay more careful attention to the realisation of these parameters to convey their lexical information. This in turn may lead to more exaggerated increase during focus production, so that elements of both focus and lexical contrast are perceptible enough for processing.

The mechanisms that are responsible for production of prosodic focus can thus be seen as a form of hyperarticulation. Consistent with our findings in prosodic focus, past research shows that speakers tend to hyperarticulate cues in ways that are related to the phonological structure of their native language. For example, studies on vowel duration in clear speech produced by Croatian and English speakers revealed an enhanced durational difference for Croatian short and long vowels and English vowels before voiced and voiceless coda stops, but not for English tense and lax vowels (Smiljanic & Bradlow, 2008). This reflects the fact that Croatian has phonemic vowel length contrast and English has "voice induced lengthening" (de Jong, 2004), while English tense and lax vowels differ primarily in their spectral characteristics. Similarly, in Korean, a language with three-way stop distinctions and neither lexical stress nor pitch accent, speakers have been found to produce clear speech with enhanced domain-initial strengthening cues (e.g., marked VOT differences), but without the use either of local F_0 cues to enhance a particular syllable or of global F_0 cues to enhance the overall intelligibility of the utterance (Cho, Lee, & Kim, 2011). On the other hand, English speakers and younger Korean speakers (i.e., those born after a sound change in aspirated stops with shorter VOT) are more likely to use F_0 differences to produce clear speech (Kang & Guion, 2008). Consistent with these findings, our study contributes evidence that cue enhancement strategies may tend to involve greater phonological distinctiveness in the phonetic categories most likely to carry lexical information. We therefore speculate that the very nature of focus is to enhance perception of the focused constituent by strengthening the cues that distinguish the lexical items from others. This could explain why Mandarin speakers produce focused lexical tones with exaggerated F_0 contours and English speakers enhance the F_0 rise for focused high (H*) tones but attenuate the F_0 information in focused low (L*) tones (e.g., Chen & Gussenhoven, 2008; Liberman & Pierrehumbert, 1984).

At the same time, our findings may also suggest that speakers of different languages vary in the degree to which speakers are sensitive to the different prosodic parameters. From a statistical learning standpoint, it may be useful to develop sensitivity to a prosodic parameter that has lower baseline variability. In Mandarin, F_0 information is tightly specified at the syllabic domain where it is controlled on a syllable-by-syllable basis for each syllable carrying a particular tonal target. Having F_0 specification for every single syllable thereby reduces the baseline variability for the prediction of F_0 targets. Therefore, F_0 may be a particularly informative cue for detecting focus because any deviation of this low baseline level of variability is going to signal additional linguistic information beyond tonal identity. Conversely, English speakers may be more sensitive to deviation in intensity cues due to the presence of lexical stress and less sensitive to focal pitch because F_0 in intonational pitch accents is sparsely specified across many syllables. Consistent with this view, data from both speech and music perception show that native speakers of tone languages (e.g., Mandarin, Thai) are more likely than speakers of non-tone languages (e.g., English) to have absolute pitch (Deutsch, Henthorn, Marvin, & Xu, 2006) and be able to discriminate musical and speech stimuli on the basis of F_0 contours (Stevens, Keller, & Tyler, 2011).

It is important to note that the language-specific differences in F_0 and intensity production did not occur in all instances of focus. Firstly, how prosodic focus was realised and whether there was any crosslanguage variation at all also depended on the discourse-pragmatic contexts and whether each specific type of focus occurred in each of the dialogues. Nevertheless, even when the within dialogue differences are taken into account, there is still a systematic trend for English speakers to produce a greater increase in intensity (13 out of 17 cases) and Mandarin speakers to produce a greater degree of increase in F_0 (in 5 out of 19 cases) and duration (2 out of 3 cases), although the latter was too small to reach significance when averaged across all dialogues. For instance, all focus types occurred in the first dialogue and there was at least one case of crosslanguage difference for every type of focus, with Mandarin speakers producing greater increase in three out of four cases of F_0 differences and all cases of durational differences, while English speakers produced greater increase in three out of the four cases of intensity differences. In the second dialogue, by contrast, there were only four words with new-information focus and corrective focus and one word with confirmatory focus, but crosslanguage differences still occurred; these mostly involved Mandarin speakers producing greater increase in F_0, while there were no cases of greater increase in intensity by English speakers. And different again, in the fifth dialogue, where only corrective and parallel focus were represented, crosslanguage differences occurred for all F_0 measures. The different findings may indeed indicate that our participants engage enthusiastically in their role-playing task!

These differences in our results for the different pragmatic expressions of focus have potential implications for how focus is modelled in linguistic theory. Even though it may be more parsimonious to view focus as a unitary construct in information structure theory (e.g., Rooth, 1992; Krifka, 2006), our crosslanguage findings across different dialogues suggest that speakers may prefer their precise prosodic realisation of focus to differ from one pragmatic function to another. Certainly, we found that different kinds of focus have, to a certain extent, their own specific acoustic properties that are unique to a particular type of focus. Specificity of this kind has been reported previously, but in fact the present patterns were not fully in line with those shown in previous studies. For instance, for Mandarin, Ouyang and Kaiser (2015) found new information focus to be produced with longer duration and greater F_0 range, while corrective focus had greater intensity ranges in addition to these duration and F_0 cues. Similarly, Chen and Braun

(2005) suggest that corrective focus has larger F_0 range compared to wh-focus. For English, Katz & Selkirk (2011) and Ito and colleagues (2004) suggest that new information is produced mostly with heightened pitch (i.e., H* pitch accent) while corrective focus is more likely to be associated with increased F_0 range expansion (i.e., L+H*). Contrary to these proposals, our data show that English speakers tended to produce corrective focus with heightened mean and maximum F_0, but no not with F_0 range expansion (see Table 3), while for new-information and wh-focus, both English and Mandarin speakers reliably produced prosodic increases on all F_0 and intensity measures (see Tables 1 and 2). Meanwhile, for corrective focus, our data suggest that both English and Mandarin appeared to be similar in all aspects of prosodic focus except for the F_0 range increase that was only found in Mandarin and the increase in maximum intensity only found in English.

An important question that warrants further research is why certain crosslanguage differences were more likely to occur with certain focus types. We speculate that part of this variation across focus types could be due to differences in lexical tones. Different tones have been documented to have different degrees of F_0 expansion (e.g., Lee et al., 2016), and since we did not control for the tone of the focused token pairs in Mandarin, there is a possibility that variation across focus types is a result of different tones (see Table 8). However, this does not mean that differences in discourse contexts and dialogues do not play a role in the variation across focus types. In the crosslanguage differences that were averaged across the five dialogues (Figures 1 and 2), four out of six of the intensity differences were from new-information focus and wh-focus, and all of the F_0 differences were from cases of new-information focus and corrective focus. This is particularly interesting in that both contrastiveness versus noncontrastiveness and newness versus background are at the centre of decade-long debates concerning the distinction between focus and givenness. Interestingly, the focus types that were most similar in the prosodic dimensions that were used to produce focus (i.e., new-information, wh-question, and corrective focus) were also the very focus types that had the most crosslanguage differences in the degree of prosodic increases across the seven parameters.

Table 9. *Distribution of lexical tones as a function of focus types in Mandarin in terms of Tone 1 (high-levelled), Tone 2 (rising), Tone 3 (low-dipping), Tone 4 (high-falling), and Tone 5 (neutral).*

	Focus Types [%]				
	New	Wh	Correct	Confirm	Parallel
Tone 1	11 [26%]	1 [11%]	13 [24%]	5 [39%]	1 [12.5%]
Tone 2	9 [21%]	4 [44%]	16 [29%]	3 [23%]	2 [25%]
Tone 3	9 [21%]	2 [22%]	13 [24%]	2 [15%]	0 [0%]
Tone 4	13 [30%]	2 [22%]	12 [22%]	3 [23%]	5 [62.5%]
Tone 5	1 [2%]	0 [0%]	1 [1%]	0 [0%]	0 [0%]

Another question that has sometimes been overlooked in information structure research is whether variation in each prosodic parameter may also reflect differences in the degree of deaccenting. Languages can differ substantially in the extent to which speakers deaccent given words (e.g., Cruttenden, 2006), and there may even be a hierarchy of different degrees of givenness in everyday discourse (e.g., Gundel, Hedberg, & Zacharski, 1993). Whether this givenness hierarchy is also language-specific remains largely uncertain, because compared to focus production, the production of deaccenting in different languages has received much less attention. Using the same dialogue paradigm from the present study, future research could examine how languages may differ in the degree to which speakers deaccent given information across different levels of givenness (e.g., first vs. second instance of repeated information).

To the best of our knowledge, the present report is the first to draw on an extensive collection of experimental crosslanguage production data on more than two types of prosodic focus from an unusually large sample of native speakers. It is important to note that our data analyses are based on within-token comparison between each focused token and its unfocused counterpart, in dialogues that were constructed to sound natural despite multiple occurrences of identical words in a focused and unfocused position. For the most part, each pair of focused and unfocused tokens occurred in the same or similar phonetic contexts and utterance positions. From a methodological standpoint, the present study also addressed debates about the dichotomy between "spontaneous" versus "laboratory" speech (Beckman, 1997; Xu, 2010) by using structured dialogues written in everyday casual speech and adopting procedures where focus production could be naturally elicited. We agree with Xu that systematic experimental procedures are vital to fostering knowledge about language production. However, we have also addressed some problems associated with laboratory speech in previous studies (e.g., experiments where speakers were explicitly instructed to produce focus "*māomī mō māomī*" "*Kitty touches Kitty"*: Xu, 1999). Through the use of structured dialogues written in casual speech, the present experiment provides a novel approach in eliciting a more naturalistic form of speech that was nonetheless produced under controlled laboratory conditions. Note that participants were never instructed to emphasise any of the focused tokens. Therefore, the prosodic focus elicited in the present experiment is likely to be a good reflection of speech production in natural settings.

2.5. Conclusion

Together, our findings provide new insights into how crosslanguage and language-specific mechanisms interact in the speaker's use of prosody to encode information structure. By examining the phonetics of prosodic focus across languages, our experiment demonstrates how speakers can differ in their use of prosodic parameters based on their experience with their native language. Of course, it is still an empirical question how the production differences across languages relate to focus perception. For example, even though languages have different production strategies for focus expression, listeners may still share a common strategy for focus perception. One way in which all listeners may be similar in focus perception is in how they use the cues from the immediate speech stream to anticipate an upcoming focused word. It is possible that languages with different production strategies for focus may share a common perceptual strategy that underlies listeners' ability to search for the discourse-marker that is attached to the focused constituent. However, differences may still exist as a result of speakers' varying sensitivity and attention to different prosodic parameters. If focus production relates to focus perception, then based on our current production findings, it could be the case that English listeners are more sensitive to intensity information while Mandarin listeners attend more to pitch cues. To test this idea, future research could conduct a perceptual task where certain prosodic information is rendered uninformative. For example, previous work has shown that English speakers could still entrain to prosodic structure for locating sentence focus even when F_0 cues were removed by monotonising the sentences (Cutler & Darwin, 1981). Whether this is also the case in other languages is still an open question. Prosody may be universally available for expressing focus, but the means of its employment and its precise realisation may be considerably language-specific.

CHAPTER 3

– Universals of Listening –

A version of this chapter was submitted to *Cognition*: Ip, M. H. K., & Cutler, A. (under review). Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages.

Portions of the data presented in this chapter were presented at the 18th Annual Conference of the International Speech Communication Association (Interspeech 2017): Ip, M. H. K., & Cutler, A. (2017). Intonation facilitates prediction of focus even in the presence of lexical tones. In F. Lacerda, S. Strombergsson, M. Wlodarczak, M. Heldner, J. Gustafson, & D. House (Eds.), *Proceedings of the 18th Annual Conference of the International Speech Communication Association* (pp. 1218-1222). Stockholm, Sweden: ISCA. doi:10.21437/Interspeech.2017-264

3.0. Abstract

In English and Dutch, listeners entrain to prosodic contours to predict where focus will fall in an utterance. Here we ask whether this strategy is universally available, even in languages with different phonological systems. In a phoneme detection experiment, we examined whether prosodic entrainment also occurs in Mandarin Chinese, a tone language, where the use of various suprasegmental cues to lexical identity may take precedence over their use in salience. Consistent with the results from Germanic languages, response times were facilitated when preceding intonation predicted high stress on the target-bearing word, and the lexical tone of the target word (i.e., rising vs. falling) did not affect the Mandarin listeners' response. Further, the extent to which prosodic entrainment was used to detect the target phoneme was the same in both English and Mandarin listeners. However, acoustic analyses of the preceding intonation of the English stimuli revealed greater mean F_0, maximum F_0, F_0 range, overall duration, and pausing before the predicted accent, while the Mandarin stimuli only showed differences in maximum F_0 and F_0 range. Nevertheless, native Mandarin speakers did not adopt an entrainment strategy when the sentences were presented in English. These findings have implications for how universal and language-specific mechanisms interact in the perception of focus structure in everyday discourse.

– Universals of Listening –

3.1. Introduction

The speech stream is a continual cascade of information, from the physical properties of the speech sounds to the sequencing of words and the discourse context. To anticipate the likely continuation, listeners must constantly build up knowledge about the incoming signal by attending to cues from different parts of the language structure (Norris, McQueen, & Cutler, 2000). In the segmental domain, considerable research over the past decades has revealed both universal and language-specific mechanisms in speech perception. For example, across languages with differing phonological structures, there is evidence that listeners can use the same strategies to recognise words by tracking information based on their syllabic structure (e.g., Sonority Sequencing Principle: Gómez, et al., 2014) or patterning of vowels and consonants (e.g., Possible Word Constraint: Brent & Cartwright, 1996; Cutler, Demuth, & McQueen, 2002; Norris, McQueen, Cutler, & Butterfield, 1997). At the same time, it is also well known that listeners are sensitive to language-specific features such as the transitional probabilities between syllables (Saffran, Aslin, & Newport, 1996), coarticulatory word-onset variations (Davis, Marslen-Wilson, & Gaskell, 2002), and phonotactic or allophonic regularities (Christiansen, Allen, & Seidenberg, 1998; Juscyzk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; McQueen, 1998; Vitevitch & Luce, 1999). Likewise, knowledge-based processing from higher-level domains (e.g., syntax, semantics) has also been shown to support perception of word boundaries (Gaskell & Marslen-Wilson, 1997; Mattys, Melhorn, & White, 2007), phoneme restoration (Samuel, 2001), and lexical selection and disambiguation (Altmann & Kamide, 1999; Seidenberg, Tanenhaus, Leiman, & Bienkowsky, 1982).

However, much less research has focused on the role of prosody. In everyday discourse, the entire meaning of an utterance cannot always be conveyed solely by the syntax and the meaning and segmental compositions of the individual words. Importantly, conversations between people can only occur if both speakers and listeners share a common understanding on some information about the world. One way in which prosody can facilitate communication is by conveying the speaker's state of mind through the focus structure, or the "information packaging" (Chafe, 1976), of the utterance. Speakers rarely assign equal acoustic weight to each word in the sentence; words with different discourse status (e.g., focus vs. background) can be produced with different degrees of prosodic prominence to express the utterance semantic structure. In this way, even identical sentences can have different implications depending on how certain words are produced; as illustrated in (1), where "*poodle*" is prosodically highlighted to show that the new information being conveyed is about the Archduke's poodles, and not some other dog breed, compared to (2), where it is deaccented and the prosodic emphasis occurs later in the sentence. Therefore, it is important for listeners to identify both the location and features of different prosodic cues in order to understand the intended message.

(1) I was quite shocked to see the Archduke's POODLES eating truffles for lunch.

(2) I was quite shocked to see the Archduke's poodles eating TRUFFLES for lunch.

Prosodically highlighted words can speed up the sentence comprehension process, in part because the phonetic features of these words play an important role in perception. In English, for instance, where more than 60% of spoken words deviate from their citation form in at least one segment (Johnson, 2004), stressed syllables of focused words are realised with longer vowel duration, higher relative pitch, and greater peak amplitude and spectral clarity (e.g., de Jong, 2004; Lehiste, 1970; Sluijter & van Heuven, 1996). Conversely, unfocused words tend to have shorter duration, more centralised vowels, and lower pitch and intensity. These acoustic differences allow focused words to stand out from the background elements, making them clearer and easier to understand (e.g., Lieberman, 1963; Mattys & Samuel, 2000). Indeed, behavioural and ERP studies from various languages have shown that prosodic focus can provide many listening advantages. Prosodically highlighted words are recognised more rapidly and accurately (Cutler & Foss, 1997; Lee, Chiu, & Xu, 2016; McAllister, 1991) and are processed more deeply in lexical activation (Blutner & Sommer, 1988; Brunellière, Auran, & Delrue, in press; Li & Ren, 2012; Norris, Cutler, McQueen, & Butterfield, 2006). Further, given the intimate relation between prosody and discourse in some languages, prosodically highlighted words can also speed up sentence comprehension (Birch & Clifton, 1995), support processing of contextual alternatives, and help listeners identify different elements of the discourse structure (Braun & Tagliapietra, 2010; Dahan, Tanenhaus, & Chambers, 2002; Fowler & Housum, 1987; Hsu, Evans, & Lee, 2015). In addition, crosslanguage comparisons between typologically unrelated languages (e.g., English and Korean: Kember, Choi, Yu & Cutler, submitted) have revealed better recognition memory for prosodically focused words (see also, Birch & Garnsey, 1995; Fraundorf, Watson, & Benjamin, 2010). All these findings indicate that prosodic focus may have some universal effects on language processing.

What is less clear, however, is whether there is also a common strategy that all listeners can use to forecast the location of a prosodically focused word, even before it is uttered. For Germanic languages (e.g., English and Dutch), Cutler and colleagues have discovered that listeners could anticipate an upcoming accented word by *entraining* to prosodic features in the utterance intonation contour (Akker & Cutler, 2003; Cutler, 1976; Cutler & Darwin, 1981; Cutler & Fodor, 1979). In a phoneme detection task, participants listened to a series of sentences in their native language and responded as fast as they could to words that began with a specified phoneme target (e.g., respond as soon as they hear the sound /d/ in "*duck*"). Listeners responded faster to the target phoneme in sentences where the preceding intonation contour predicted high stress on the target-bearing word, compared to sentences where the intonation predicted low stress. Response times were still faster for sentences with predicted high stress contexts, even when the original target-bearing words in each context were replaced by an acoustically identical neutral version of the same word. Since the only difference was in the preceding intonation, it was concluded that listeners could attend to (different cues in) the preceding prosodic contour to anticipate an upcoming focused word. Through this phoneme-monitoring approach, we can demonstrate that listeners can engage in prosodic entrainment, a strategy where listeners can attend to the prosodic features of the intonation contour that is immediately available in the speech stream and draw along with it to anticipate the prosodic form of an upcoming word.

In word recognition, similar effects of preceding prosody have also been observed in prediction of upcoming lexical forms. For example, Dilley and Pitt (2010) found that listeners can use contextual speech rate cues to predict the presence or absence of heavily coarticulated function words. Dilley and Pitt presented native English listeners with sentences containing a spectrally reduced function word, and manipulated the speech rate of the preceding prosody (e.g., *or* from *minor or* [maɪnɚː] in "*Anyone must be a minor or child…*"). Compared to sentences with normal speech rate, listeners were less likely to detect the function word when the preceding context was slowed, even though the target words were acoustically identical in both contexts. Conversely, speeding the speech rate caused listeners to hallucinate hearing a function word that was never spoken (e.g., *a* in "*The company moved to (a) different…*").

Subsequent experiments have further demonstrated that preceding speech rate can still facilitate listeners' anticipation of upcoming words even when the target words have been made clearer (e.g., by creating various degrees of amplitude dip at the word onset; Heffner, Dilley, McAuley, & Pitt, 2013). According to Dilley and colleagues, one way in which listeners can use such cues to anticipate upcoming word forms is by extracting the statistical (e.g., distributional) properties of the preceding prosody. For example, Baese-Berk and colleagues (Baese-Berk et al., 2014) examined the role of long-term exposure to varying speech rates and found that perceptual learning of contextual prosody can influence word perception. This indicates that human listeners are constantly updating their model of different prosodic cues to enable more accurate predictions about the upcoming signal. Consistent with this view, similar uses of speech rate have been replicated in other languages (e.g., Russian) in both native (L1) and non-native (L2) processing (Dilley, Morrill, & Banzina, 2013; Lai & Dilley, 2016). Further, the role of preceding prosody on lexical recognition has also been found for other types of prosodic cues such as rhythmic patterns (e.g., Breen, Dilley, McAuley, & Sanders, 2014; Brown, Salverda, Dilley, & Tanenhaus, 2011; 2015; Dilley & McAuley, 2008; Dilley, Mattys, & Vinke, 2010; Kuijpers & van Donselaar, 1998; Morrill, Dilley, McAuley, & Pitt, 2014).

However, unlike lexical processing, it is still an empirical question whether the preceding prosody can also facilitate prediction of upcoming prosodic forms in focus perception across languages. Firstly, the existing data on prosodic focus entrainment come from native speakers of English and Dutch. This makes it difficult to reach any conclusions about universality and language-specificity, since the relation between prosody and focus is essentially the same in these two languages (Gussenhoven, 1983).

Secondly, from a production standpoint, there is considerable crosslanguage variation in how different aspects of the suprasegmental structure are used for the expression of information structure. Variation in focus production can occur due to differences in intonational phonology (e.g., Jun, 2014), rhythmic structure (e.g., Burdin et al., 2015), durational lengthening (e.g., Hay, Sato, Coren, Moran, & Diehl, 2016), or contextual predictability (e.g., Turnbull, Burdin, Clopper, & Tonhauser, 2015). At the same time, languages, and even regional dialects, can differ starkly in the degree to which prosodic prominence is related to discourse structure, from languages where speakers consistently use prosody to highlight focus and deaccent background information (e.g., American and British English: Ladd, 2008; Dutch: Caspers, Bosma, Kramm, & Reya, 2012; Swerts, Krahmer, & Avesani, 2002; German: Féry & Kügler, 2008), to languages where it is only optional (e.g., Indian and Caribbean English: Gumperz, 1982; Hausa: Hartmann & Zimmermann, 2007; Romance languages: Avesani & Vayra, 2005; Cruttenden, 1993; Ladd, 1990b), to languages where prosody is never used for this purpose (e.g., Ambonese Malay: Maskikit-Essed & Gussenhoven, 2016; Northern Sotho: Zerbian, 2006; Yucatec Maya: Kügler & Skopeteas, 2007; Wolof: Rialland & Robert, 2001). On a related note, speakers of languages that already have morphological focus markers (e.g., Wolof: Rialland & Robert) or more flexible word orders (Italian or Catalan: Vallduví, 1991; 1992; Zubizarreta, 1994; 1998) may be more likely to use non-prosodic means to produce focus. It is therefore possible that listeners in some languages may not use the preceding prosody to predict upcoming prosodic focus.

One way to pursue this question of universal versus language-specific is to examine whether speakers of different languages can anticipate prosodic focus using similar listening strategies despite differences in production. For example, it is still an open issue whether an entrainment strategy can be found in another language where listening is adapted to a different prosodic system. A crosslanguage investigation with native speakers of English and Mandarin Chinese could provide new insights into prosodic perception. Mandarin has features that are both similar to and different from English. Despite their typological distance, both languages express prosodic focus in fundamentally the same way (i.e., exaggerated pitch range/pitch accents, increased duration and intensity, and post-focal compression). However, recent work in our laboratory has revealed that the two languages can still differ in the degree to which different prosodic cues (e.g., F_0, intensity) are used to highlight focus (Ip & Cutler, 2016).

Further, other differences in phonological systems could prevent Mandarin speakers from showing the same entrainment effect. In English, entrainment to the intonation contour may be useful because it signals postlexical meaning at the level of the sentence. Also, English sentences would typically contain a focused constituent highlighted by a pitch accent. In Mandarin, however, both lexical tones and intonation share the same prosodic features, and to date, there is no consensus on how the two features co-exist. Xu (2005) argues that having a tonal system may not affect the use of pitch for other purposes because tones only require about one half of speakers' natural pitch range. At the same time, intonational effects may be phonetically layered on existing lexical tones and cause shifts in F_0 register or fluctuation of F_0 range (e.g., Mandarin: Xu, 1999; Yoloxóchitl Mixtec: DiCanio et al., 2018). Contrasting with this view is the suggestion that much of the pitch contour would be exhausted in the phonetic expressions of contour tones, thereby resulting in a less elaborate intonational system (Hayes, 1995; Pierrehumbert, 1999) or not having an intonational system at all (Kratochvil, 1998). For example, research across various tone languages show that pitch accents are minimal or absent (e.g., Mambila: Connell, 2017; Yorùbá: Laniran & Clements, 2003), and not all lexical tones can carry boundary tones (e.g., Akan: Kügler, 2017; Tswana: Zerbian, 2017). Further, there are also tonal differences in phonation and intrinsic duration and amplitude, which have been revealed to affect perception (Blicher, Diehl, & Cohen, 1990; Fu, Zeng, Shannon, & Soli, 1998; Liu & Samuel, 2004; Whalen & Xu, 1992). These tonal cues also co-specify lexical identity. Therefore, even if there is the exaggeration of prosodic cues used for focus (e.g., Chen & Gussenhoven, 2008), it may only be localised on the focused word.

Indeed, some production research suggests that Mandarin speakers may not produce the preceding intonation in a way that would support prosodic entrainment. For example, Xu (1999) found that the intonation contour before a Mandarin focused word tends to be acoustically similar to that of a neutrally produced sentence with no prosodic focus. There are also reports of other tone languages, such the Austronesian language Ma'ya (Remijsen, 2002), and some Otomanguean languages (Chávez-Peón, 2010; DiCanio & Hatcher, 2018), in which speakers only use duration to produce stress, due to the documented use of F_0 primarily for tonal contrasts. In addition, comparisons between tonal and non-tonal dialects of a single language (e.g., Kammu) show that intonation can be influenced by the tone combination in the sentence (Karlsson, House, Svantesson, & Tayanin, 2010). Finally, certain tones (e.g., Mandarin low-dipping tone) are more prone to F_0 restriction (e.g., Lee, Wang, & Liberman, 2016).

These results suggest that the presence of lexical tones may have implications for the perception of intonation. For example, competing F_0 contour adjustments by lexical tones and intonation can hinder recognition of different intonational categories (e.g., statements vs. questions; Liu & Xu, 2005; Yuan, 2011). Several experiments comparing tone and non-tone languages have also suggested that native speakers of tone languages are more likely to process pitch at a lexical level and are less sensitive to sentence intonation (e.g., Gandour et al., 2003; Gussenhoven & Chen, 2000). Therefore, even though suprasegmental features may have a dual function in the production of tone and focus, the presence of tones may still place a limit on the degree to which speakers can produce, and listeners can perceive, the preceding cues that support focus prediction.

In the present study, we adopt the phoneme detection paradigm from Cutler and colleagues' experiments to compare English and Mandarin listeners' use of prosody in their anticipation of focus. Based on the phonological differences between English and Mandarin, Mandarin listeners may not have the ability to adopt an entrainment strategy. However, it is also possible that Mandarin listeners may still adopt the same entrainment strategy, but that the extent to which they do so may be limited due to the presence of lexical tones, either because the intonation is less informative for focus detection, or because listeners make less effective use of the intonational cues. A third possibility is that cues signalling focus may still assist Mandarin listeners to the same extent as the English listeners. This third view would suggest that prosodic entrainment may be a universal strategy that all listeners can adopt despite any differences in prosodic systems.

3.2. Experiment 1a

3.2.1. *Method*

Participants. Two participant samples were tested: 23 native speakers of Australian English (M_{age} = 23.96 years, *SD* = 8.64 years; 16 females) and 23 native speakers of Mandarin Chinese (M_{age} = 25.02 years, *SD* = 3.78 years; 13 females). All of the English speakers reported that they were born and raised in Australia. The Mandarin speakers were born in Mainland China and had been living and studying in Australia for an average of one year and 5 months (*SD* = 25.44 months, range: 23 days – 7.96 years). We excluded additional data from one Mandarin speaker who failed a follow-up recognition test. Data from two English speakers were also excluded due to technical issues. None of the participants reported any hearing or speech impairments.

Materials. The English and Mandarin sentences (see Appendices D and E) were each recorded by a female native speaker who did not know the purpose of the experiment. In both languages, 24 unrelated experimental sentences were recorded in three versions. In the first version, the target-bearing word received emphatic stress. In the second version, emphatic stress was instead placed on a word that occurred later in the sentence than the target-bearing word, which, in consequence, received very reduced stress. In the third version, the target-bearing word and the sentence as a whole were produced in a neutral manner. In all of the experimental sentences, the phoneme target was a voiceless aspirated bilabial stop [pʰ] occurring at the start of the target-bearing word's first syllable (e.g., "*peanuts*" [pʰi:nʌts]; "*葡萄*" *grapes* [pʰu2tʌ5]). Further, the phoneme target in English always occurred on the word's lexically stressed syllable. Given the language differences in stop inventories, we only used one phoneme target for all sentence trials. For Mandarin, we also controlled the tone of the target-bearing words, such that half of the sentences had the phoneme target occurring on a high-rising second tone (e.g., "*葡萄*" *grapes* [pʰu2 tʌ5]) and half had the target on a falling fourth tone (e.g., "*骗子*" *swindler* [pʰjɛn4 tsɨ5]).

Using Praat (Boersma & Weenink, 2018), the target-bearing words were spliced at their nearest zero crossing from all three versions of each experimental sentence. The high- and low-stressed target-bearing words from the first and second versions were replaced by an acoustically identical token of the same target word from the neutral version. For both the English and Mandarin stimuli, participants were randomly assigned to one of two experimental conditions, each containing one version of each of the 24 spliced experimental sentences, plus an additional set of 24 filler sentences. The experimental sentences with predicted high versus predicted low stress were counterbalanced across the two conditions (i.e., "Version A" and "Version B").

The English and Mandarin experimental sentences were comparable in length, as measured in terms of the total number of syllables (English, *M* = 17.92, *SD* = 3.92; Mandarin, *M* = 16.75, *SD* = 2.59). Further, the number of syllables between the start of the sentence and the onset of the target-bearing word was comparable across the two languages (English, *M* = 10.00, *SD* = 2.95; Mandarin, *M* = 9.04, *SD* = 2.35), and was also similar to the set of English sentences used in the previous Cutler and Darwin (1981) experiments (*M* = 10.30, *SD* = 3.16). To avoid interference between the sentences, sentence beginnings were varied and semantic content that could be associated with another sentence in the set was avoided. In previous studies by Cutler and colleagues, 16 out of 20 sentences had target words preceded by a determiner, but here, we also varied the syntactic category of the word immediately preceding the target word, so that less than half of the target words were preceded by a determiner (and we used a variety of determiners). In addition, none of the sentences had any additional occurrence of voice or voiceless bilabial stops beyond that in the target-bearing word. All of the sentences were produced at a natural fast-normal rate.

Finally, we conducted acoustic analyses of the stimulus recordings to determine whether the preceding prosody revealed any prosodic differences between the predicted high and low stress sentences. Analyses were performed on the preceding prosodic context (i.e., the part of the sentence before the release burst of the target phoneme). In line with the Dilley and Pitt's (2010) definition of "distal prosody", our analyses of the preceding prosody included words up to four syllables before the target.

Procedures. All tests were conducted in the participant's native language in a sound-attenuated booth at the MARCS Institute, Western Sydney University. The phoneme-detection task was administered using E-Prime software on a laptop computer, with attached to it a set of headphones and a Chronos® response device for button pressing (Schneider, Eschman, Zuccolotto, 2002). Participants were informed that the experiment aimed to examine listeners' memory and language comprehension; they were further told that they would listen to a series of sentences and had two tasks: first, pay careful attention to the meaning of each sentence, and second, press a button as fast and as accurately as they could whenever they heard a word that began with the target sound [p^h]. Participants received two practice trials and feedback before starting the actual experiment. Instructions were written in the participants' native language (see Appendices F and G). The Chinese instructions were translated from the English version by a professional translator who was an instructor at the university's languages and translation department. The instructions contained no mention of sentence prosody.

At the end of the testing session, participants completed a recognition test in which they were asked to judge whether or not each of the 20 sentences in the list were from the experiment (see Appendices H and I). From the instructions at the start of the trial, all participants were told that they would be quizzed on their comprehension of the sentences at the end of the study. The recognition test was conducted to confirm that participants understood the sentences. Data from participants who scored below 50 percent were excluded because such a low score may indicate insufficient attention to the sentences. All participants in the final sample scored 65 percent or above in the test (Mandarin, *M* = 84.13, *SD* = 10.51, range: 65 – 100; English, *M* = 88.48, *SD* = 7.75, range: 70 – 100).

3.2.2. *Results and Discussion*

Response times (RTs) were measured as the duration of the latencies between the release of the target stop consonant and participants' button presses. We compared RT to the target in predicted high stress sentences with RT in predicted low stress sentences. There were no RT shorter than 100 milliseconds. We excluded a further Mandarin speaker who had average RT scores of over 1000 milliseconds.

In line with standard practice, RT datapoints longer than 2500 milliseconds were excluded from final analyses, since such a delayed response may indicate a reprocessing of the sentence (Ratcliff, 1993). Both the predicted high stress and low stress contexts had two excluded datapoints in Mandarin and one excluded datapoint in English. No participant had more than two instances of RT longer than 2500 milliseconds. There was only one speaker from each language group who had two instances of excluded RTs longer than 2500 milliseconds, which occurred once for each prosodic context.

Preliminary analyses. We conducted control analyses to assess whether there was a significant effect of the counterbalanced experimental conditions (Version A vs. Version B). These analyses revealed no significant effect in both languages. In Mandarin, there was also no effect of the tone (high-rising vs. falling tone) of the target. Therefore, the main analyses were conducted without these variables in the model. For accuracy (see Table 1), there were one miss in the predicted high stress context and five in the predicted low stress context for English and four for Mandarin. However, the differences were too small to reach significance. No individual participant had more than one miss.

Table 1.
Detection accuracy in Experiment 1a and 1b.

Experiment	Sample			
		Predicted High Stress	Predicted Low Stress	*p*
Experiment 1a: L1 Phoneme detection	Native speakers of Australian English (N = 23)	1	5	.219
	Native speakers of Mandarin Chinese (N = 23)	1	4	.375
Experiment 1b: L2 Phoneme detection	Native speakers of Mandarin Chinese (N = 36)	14	8	.286

Response time. The RT results for each language group are displayed in Table 2. To evaluate whether or not the English- and Mandarin-speaking participants had faster RT to the phoneme target in predicted high stress contexts, we computed a 2 (Language: English vs. Mandarin) × 2 (Prosodic context: high vs. low stress) mixed-model ANOVA on the dependent variable of RT. This analysis revealed a significant main effect of prosodic context, $F(1, 44) = 16.959, p < .001, \eta_p^2 = .28$, but the interaction was not significant. We followed up the significant main effect with paired two-tailed *t*-tests by language group. These indicated significantly faster RT to the target phoneme in predicted high stress contexts by both English and Mandarin listeners (both *p*-values = .008; see Figure 1).

Table 2.
Response time (in ms) to the target phoneme [p^h] in Experiment 1a and 1b.

Experiment	Sample	Mean Response Time (SD) [Range]		
		Predicted High Stress	Predicted Low Stress	*t*
Experiment 1a: L1 Phoneme detection	Native speakers of Australian English (N = 23)	418.77 (72.43) [340-603]	459.90 (84.77) [362-568]	2.92**
	Native speakers of Mandarin Chinese (N = 23)	492.03 (100.31) [322-715]	537.57 (129.89) [335-836]	2.91**
Experiment 1b: L2 Phoneme detection	Native speakers of Mandarin Chinese (N = 36)	606.34 (156.61) [386-993]	619.74 (140.36) [446-955]	.69

***p ≤ .01.*

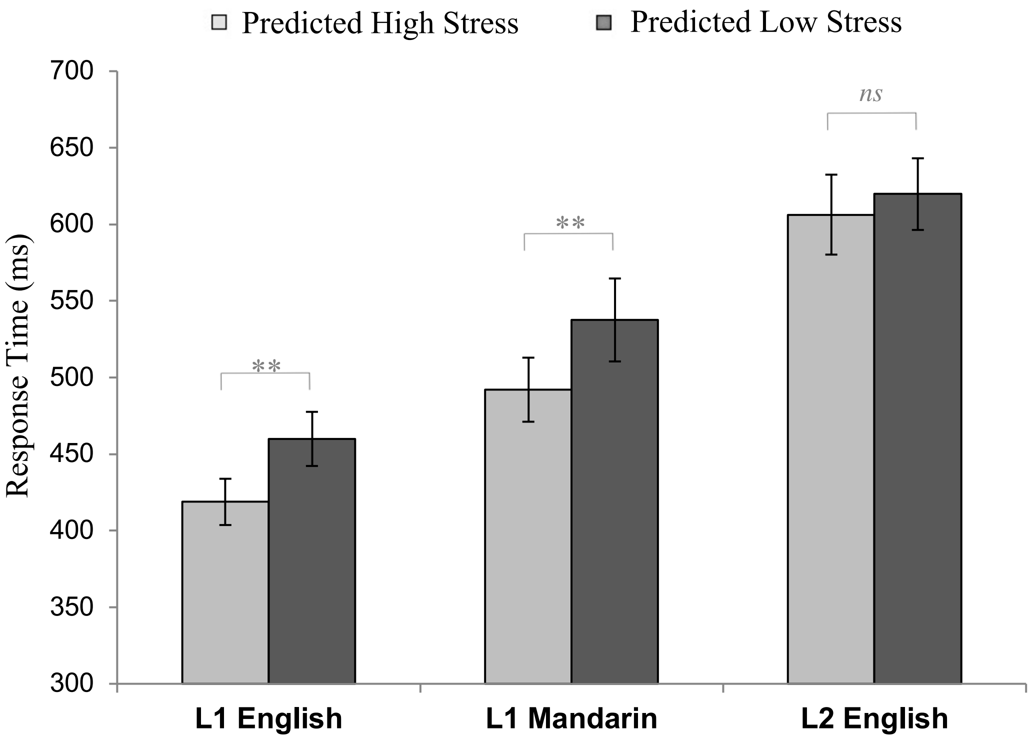


Figure 1.
Response time (in ms) as a function of intonationally predicted high versus low stress in Experiment 1a (L1 English, L1 Mandarin) and 1b (L2 English). Error bars represent standard error of the mean. **$p \leq .01$ (two-tailed).

Response time across sentence trials. Given the similarities in RT across the English and Mandarin speakers, we also examined whether there were language differences in the pattern of listeners' RT across the 24 sentence trials. RT differences (RT in low stress contexts minus RT in high stress contexts) were divided into and averaged across four separate quartiles in time. We conducted a 2 (Language) × 4 (Time) mixed ANOVAs on RT difference as a dependent variable to determine whether there was language variation in listeners' response over the course of the experimental trials. Analyses did not reveal any significant interaction across languages, although there was a significant main effect of time, $F(3, 132) = 3.01, p = .033, \eta_p^2 = .06$ (see Figure 2 and Table 3). Follow-up Bonferroni-adjusted paired *t*-tests only revealed a significant variation in RT difference between Times 1 and 2 ($p = .031$).

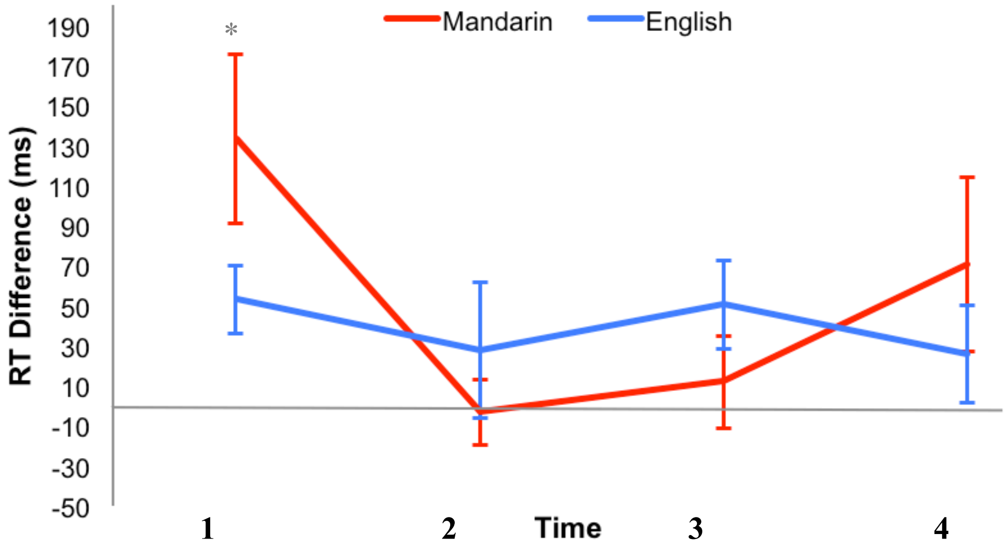


Figure 2.
RT difference (in ms) across trials (divided into Times 1 to 4) in English and Mandarin. Error bars represent standard error of the mean. *$p \leq .05$ (two-tailed).

Table 3.
RT difference (in ms) across trial in native English and Mandarin listeners.

	Mean Response Time Difference (SD) [Range]			
	Time 1	Time 2	Time 3	Time4
English	53.50 (81.94) [-97-210]	27.86 (163.04) [-268-587]	50.41 (107.16) [-77-460]	25.82 (117.61) [-202-302]
Mandarin	133.58 (203.40) [-122-718]	-3.04 (78.83) [-161-125]	12.23 (110.89) [-154-295]	70.85 (208.91) [-127-884]

Acoustic analyses. Analyses of prosodic features of the stimulus recordings were conducted in Praat based on inspection of both the waveform and the spectrogram as well as the pitch tracks and amplitude envelopes (Boersma & Weenink, 2018). In each experimental sentence, we segmented the preceding prosody (i.e., two to four syllables before the onset of the target-bearing word), in which we measured duration, mean F_0, maximum F_0, F_0 range, root-mean-square (RMS) mean intensity, maximum intensity, and intensity range (see Figures 3 for an example in Mandarin). We also measured the pre-target interval, the duration of the silence between the release of the target stop consonant and the offset of the preceding word.

Target: [pʰ]

mei2 jou3 ɹən2 tsai4 ʈʂʊŋ1kwo3 nəŋ2 ɕjɑŋ1 ɕin4 pʰu2 tʌ5 nəŋ2 ʈʂɨ 4 tsau4 ɕjaŋ1 ʂwei3

没 有 人 在 中国 能 相信 葡萄 能 制造 香 水

No one in China believes that grapes can be used to make perfumes

(a) 没有人在中国能相信 [**葡萄**] 能制造香水

(b) 没有人在中国能相信葡萄能制造 [**香水**]

(c) 没有人在中国能相信葡萄能制造香水

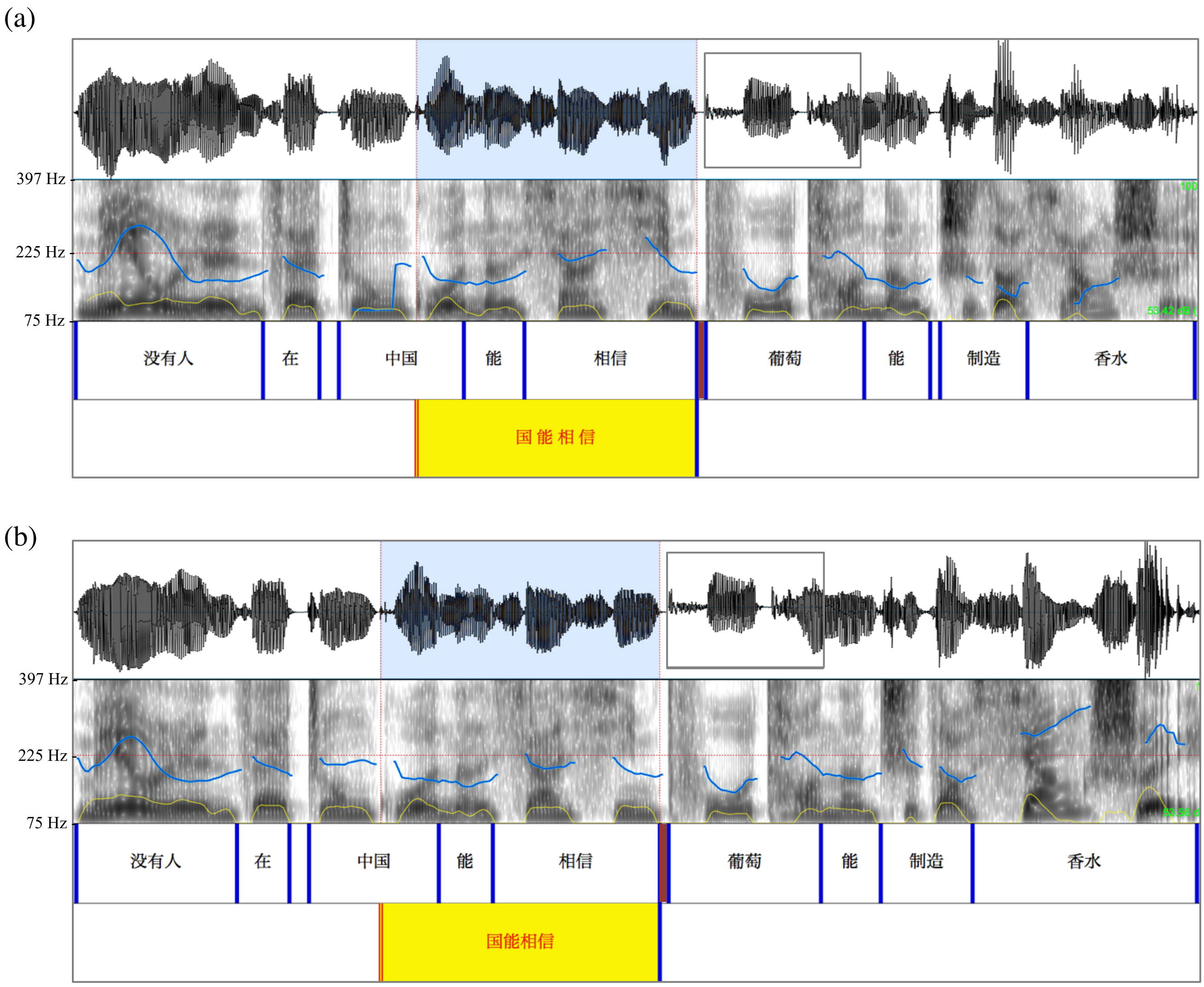


Figure 3.
Waveforms and pitch and amplitude contours of an example experimental sentence in Mandarin predicted high (a) and low (b) stress contexts; text (c) gives the neutral context. Prosodic parameters (i.e., overall duration, mean and maximum F_0, F_0 range, mean and maximum intensity, and intensity range) in the shaded portion – four syllables preceding the target-bearing word (squared) – were measured for our acoustic analyses. The red shaded portion indicates the duration of the pre-target interval.

The acoustic results for the preceding duration, F_0, and intensity are displayed in Tables 4 and 5. Statistical evaluation of the acoustic data for the Mandarin stimuli showed a significant difference in F_0 range between the predicted high and low stress contexts, such that syllables before target-bearing words had greater F_0 range in predicted high stress sentences than in predicted low stress contexts, $t(23) = 3.78$, $p = .001$. Maximum F_0 was also greater in predicted high stress sentences in Mandarin, $t(23) = 2.65$, $p = .014$. There was also a longer pre-target interval for high stress context sentences, $t(23) = 4.99$, $p < .001$. No significant differences were observed for mean F_0, overall duration, or any of the intensity cues. In contrast, in English, the preceding prosody of predicted high stress sentences was produced with higher values on all measures except for intensity range. Compared to predicted low stress contexts, the preceding prosody of English high stress context sentences had higher mean F_0, $t(23) = 2.23$, $p = .036$, higher maximum F_0, $t(23) = 3.78$, $p = .001$, greater F_0 range, $t(23) = 4.61$, $p < .001$, longer overall duration, $t(23) = 2.23$, $p = .036$, longer pause duration, $t(23) = 4.46$, $p < .001$, greater mean intensity, $t(23) = 4.88$, $p < .001$, and greater maximum intensity, $t(23) = 5.30$, $p < .001$.

We also conducted additional 2 (Language: English vs. Mandarin) × 2 (Prosodic context: high vs. low stress) mixed-model ANOVAs for maximum F_0, F_0 range, and pre-target interval duration. This was to examine whether the magnitude of these prosodic differences between high and low stress contexts were different across the English and Mandarin sentences, despite being the parameters that showed significant differences in both languages. However, none of the analyses showed a significant interaction between language and prosodic context. Therefore, there were no crosslanguage differences in the degree to which the English and Mandarin speaker used these parameters to differentiate the high and low stress contexts.

Table 4.

Preceding prosody F_0 (mean, maximum, and range in Hz) and duration (in ms) three or four syllables before target onset in predicted high versus low stress contexts.

Stimuli	Mean Prosodic Variables (SD) [Range]									
	Mean F_0		Maximum F_0		F_0 Range		Overall Duration		Pre-target Interval Duration	
	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress
English (24 sentence pairs)	180.84* (15.43) [161-223]	176.11 (14.60) [154-201]	213.97*** (22.57) [175-286]	203.25 (25.99) [165-255]	58.38*** (20.08) [19-100]	44.67 (20.02) [17-90]	585.04* (159.22) [385-1000]	553.58 (142.91) [317-940]	74.35*** (10.91) [55-95]	61.71 (13.91) [33-89]
Mandarin (24 sentence pairs)	200.97 (22.85) [140-251]	197.36 (19.29) [152-252]	252.62* (22.25) [195-291]	242.42 (17.10) [200-293]	106.43*** (42.04) [23-204]	85.41 (35.61) [37-176]	754.67 (130.83) [500-1101]	755.04 (140.76) [510-1070]	66.67*** (26.09) [14-120]	49.04 (19.10) [4-71]

**p ≤ .05, **p ≤ .01, ***p ≤ .001 significant differences from predicted low stress contexts (two-tailed).*

Table 5.
Preceding prosody intensity (mean, maximum, and range in RMS) three or four syllables before target onset in predicted high versus low stress contexts.

Stimuli	Mean Prosodic Variables (SD) [Range]					
	Mean Intensity		Maximum Intensity		Intensity Range	
	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress
English (24 sentence pairs)	53.63*** (2.09) [50-58]	52.46 (1.99) [48-56]	59.03*** (1.88) [56-62]	57.32 (1.97) [53-62]	26.94*** (7.17) [19-41]	25.63 (6.03) [14-40]
Mandarin (24 sentence pairs)	54.44 (3.60) [51-64]	55.43 (4.36) [51-63]	59.06 (3.85) [56-69]	59.75 (4.37) [55-68]	26.47 (8.37) [15-42]	27.23 (8.61) [14-44]

****p ≤ .001 significant differences from predicted low stress contexts (two-tailed).*

Relation between preceding prosodic cues and response time. We also conducted a series of Pearson's two-tailed correlation analyses to examine whether there was any link between the strength of the different prosodic cues in each sentence item and the degree to which listeners showed a RT difference between high and low stress contexts (see Tables 6 and 7). For each sentence item, we calculated each prosodic parameter's proportional difference (i.e., percentage change) between high and low stress contexts. For each sentence, we also calculated the proportional difference in RT averaged across the participants. In English, there were no significant correlations between RT difference and any of the parameters. This was also the case when we conducted correlation based on absolute differences. In Mandarin, there were only significant negative correlations between proportional differences in RT and mean intensity (r = -.57, p = .004) and maximum intensity (r = -.58, p = .003).

Table 6.

Proportional (% change) differences in English RT, F_0 (mean, maximum, range), overall duration, pre-target interval duration, and intensity (mean, maximum, range) by sentence item (presented according to trial order).

	Proportion Difference								
	RT	Mean F0	Max F0	F0 Range	Duration	Pre-Target	Mean Intensity	Max Intensity	Intensity Range
1. Partner	7.31	1.01	-0.26	-19.56	1.64	24.42	2.69	3.35	-12.51
2. Polish	23.20	3.43	12.91	43.85	24.83	4.35	-2.01	-1.26	32.35
3. Puzzles	-1.52	11.01	6.51	-12.62	-0.44	28.57	1.21	1.28	16.18
4. Pixies	20.63	6.97	7.41	19.28	6.62	20.69	0.99	-0.07	6.41
5. Pendulum	11.83	-4.97	4.76	50.45	-10.71	19.35	6.84	9.15	-15.05
6. Peak	-17.69	-2.94	-0.51	-2.70	-5.76	22.22	0.06	1.31	-29.74
7. Powder	9.44	-8.87	-8.15	19.22	23.37	6.56	0.71	2.62	20.23
8. Panic	32.04	-1.02	4.19	29.05	-6.04	9.76	5.29	6.51	6.26
9. Power	8.04	-7.97	-3.20	19.61	16.29	32.93	0.74	4.40	15.61
10. Porters	-1.70	5.01	11.13	40.07	11.55	-45.45	4.49	6.66	36.67
11. Petrol	1.00	5.38	3.68	4.77	8.83	13.68	6.36	4.23	15.72
12. Picnic	-11.89	0.87	11.76	43.14	3.43	-11.25	1.59	4.65	3.00
13. Pencil	16.75	4.16	19.11	71.29	-4.22	39.71	2.24	4.06	-8.85
14. Poodles	11.81	7.98	-1.31	14.28	6.00	22.22	-0.57	0.14	17.45
15. Parrot	-15.00	6.44	6.63	11.74	-12.08	21.62	3.83	5.50	-9.79
16. Perfect	23.11	9.89	10.83	10.16	3.55	18.18	3.40	0.36	-14.12
17. Pardoned	22.01	6.04	12.38	36.63	0.86	-4.92	4.02	4.94	29.45
18. Pirate	3.87	8.42	7.00	24.98	-19.96	7.79	2.58	0.95	-29.89
19. Park	9.56	-6.05	-5.41	18.85	12.93	31.75	1.14	-0.35	-8.57
20. Party	-0.04	0.88	2.83	46.95	2.58	15.52	1.81	3.42	13.85
21. Punish	1.97	5.22	6.91	20.13	0.87	7.06	-1.14	-0.22	-18.42
22. Pancreas	1.57	5.17	8.87	38.89	25.06	15.49	1.22	2.13	6.65
23. Peanuts	13.56	-0.63	2.02	41.74	15.37	27.95	2.71	4.24	11.05
24. Padlocks	6.13	3.68	1.16	-25.53	6.86	59.26	1.88	0.75	-18.75

Light Green = 0-25 percentile, Light Blue = 25-50 percentile, Medium Blue = 50-75 percentile, Dark Blue = 75-100 percentile.

Table 7.

Proportional (% change) differences in Mandarin RT, F_0 (mean, maximum, range), overall duration, pre-target interval duration, and intensity (mean, maximum, range) by sentence item (presented according to trial order).

	Proportion Difference								
	RT	Mean F0	F0 Max	F0 Range	Duration	Pre-target	Mean Intensity	Max Intensity	Intensity Range
1. 爬山	-8.33	-0.55	-1.83	16.31	3.33	-9.43	-1.62	-5.36	-18.56
2. 票	48.74	5.93	5.78	12.85	0.00	25.00	-12.90	-11.86	5.24
3. 怕	-13.22	-4.11	-6.34	28.14	0.00	46.00	0.67	2.31	2.41
4. 葡萄	3.92	4.79	13.35	32.19	-5.00	33.33	-0.20	-0.46	-35.82
5. 碰撞	13.54	-1.17	-1.28	0.13	2.82	36.36	1.42	1.31	-32.04
6. 便宜	27.87	-3.96	-0.73	38.02	-2.70	16.67	-2.30	-0.09	-20.17
7. 破壞	19.94	3.95	6.99	21.27	-10.13	92.73	0.86	-1.26	-10.38
8. 叛徒	15.48	2.81	0.76	6.94	5.56	2.70	0.71	-2.48	-15.83
9. 脾氣	8.30	4.58	14.92	57.62	-4.00	30.00	1.97	2.53	-5.04
10. 鋪頭	-14.72	1.12	-2.54	-64.79	-1.67	25.00	0.39	2.58	-19.40
11. 騙子	9.31	-8.34	-4.99	-16.23	-6.15	-2.90	-0.47	1.21	1.45
12. 朋友	-22.55	12.12	10.03	-5.43	-2.00	19.40	5.28	6.32	-6.44
13. 牌子	29.27	0.23	6.72	13.55	3.51	5.26	-13.48	-12.58	24.37
14. 漂亮	21.33	8.92	10.02	42.89	9.72	27.63	-0.78	0.81	30.93
15. 盤子	12.01	6.64	0.02	-2.42	1.82	8.00	1.82	4.30	26.40
16. 皮衣	-7.31	5.55	3.58	29.98	0	-21.43	1.47	2.49	22.08
17. 盼望	-26.30	1.07	6.61	20.34	-5.95	53.33	1.92	4.36	12.54
18. 膨脹	14.46	-4.71	-0.67	11.44	2.58	42.50	-18.95	-16.50	-4.46
19. 砲響	-22.91	7.70	7.74	35.69	-4.11	32.43	-0.78	-0.52	-0.79
20. 螃蟹	-22.22	-13.50	16.98	21.31	3.13	37.78	-1.57	-1.13	12.99
21. 蘋果樹	40.84	-1.32	-14.55	-48.95	-14.81	0.00	-12.58	-11.01	-52.50
22. 胖子	-19.89	1.92	10.17	25.83	2.94	14.29	2.75	4.83	-68.17
23. 排骨	-28.19	3.98	4.99	14.86	-4.22	11.29	0.26	-0.46	-1.71
24. 屁股	26.87	1.31	1.33	58.49	-2.60	41.58	0.73	0.31	13.81

Light Yellow = 0-25 percentile, Dark Yellow = 25-50 percentile, Light Orange = 50-75 percentile, Red = 75-100 percentile.

Discussion. Overall, both English and Mandarin listeners responded faster to the target phoneme in sentences where the preceding prosody predicted high stress on the target-bearing word. Further, no significant language-specific difference appeared in the degree to which high stress contexts facilitated RT, despite the acoustic data showing more cues being available in the English stimuli. Thus, this listening strategy appears to be used to equivalent extent in each language. This shows that listeners can exploit whatever cues are available in the speech signal. Also, in the acoustic analyses of the two preceding prosodic measures (maximum F_0 and F_0 range) that were significant in the stimuli of both languages, there were no crosslanguage differences in the degree to which they differentiated the prosodic high and low stress contexts.

However, all of the Mandarin-speaking participants were proficient in English and had been living and studying in an English-speaking country. Exposure to English as an L2 might have helped the Mandarin speakers develop a non-native listening strategy that they could apply when listening to their native language. To test this competing explanation, we conducted Experiment 1b to examine whether Mandarin speakers would also respond faster to phoneme targets due to high stress contexts in the English sentences. The same pattern of response in English by Mandarin speakers may indicate that they have acquired this prediction strategy from their L2 experience with English, but it could also mean that prosodic entrainment is general strategy that all listeners can use in any language that has prosodic cues to upcoming focus.

3.3. Experiment 1b

3.3.1. *Method*

Participants. Participants in Experiment 1b were 36 native Mandarin speakers who were born and raised in Mainland China (M_{age} = 24.94, *SD* = 3.72; 20 females), of whom 19 had also taken part in Experiment 1a. We aimed for a larger sample size in order to capture a wider range of Mandarin speakers with different levels of English proficiency. All participants spoke English as their second language and had been living and studying in Australia for minimally 20 days and maximally just over 10 years (*M* = 2.18 years, *SD* = 2.39 years).

Materials and Procedures. The procedures were identical to those in Experiment 1, except in that the English sentences and recognition test as used for the native English speakers in Experiment 1a were now presented to the native Mandarin speakers. All participants scored at 55 percent or above on the follow-up recognition test (*M* = 72.78, *SD* = 12.16, range: 55 – 100). We excluded additional data from a participant who did not score significantly different from chance (i.e., 55%) on the recognition test and three participants whose average RT scores were over 1000 milliseconds.

3.3.2. *Results and Discussion*

Two RT responses longer than 2500 milliseconds and one RT response shorter than 100 milliseconds were removed from the predicted high stress data set and one such response from the predicted low stress set. Control analyses revealed no significant effect of the counterbalanced conditions. Further, there was no RT difference between the 19 participants who had previously participated in the Mandarin condition of Experiment 1a and the 17 new participants without experience of similar experiments, so the experience factor was ignored and data from all participants were included in the main analyses.

In striking contrast to Experiment 1a, the RTs of Experiment 1b revealed no effect of predicted high versus low stress (see Figure 1 and Table 2). Thus, native Mandarin speakers' phoneme detection in English did not display the entrainment that they had demonstrated in their native language. In accuracy, participants had 14 detection misses in predicted high stress sentences and 8 misses in predicted low stress sentences, which was not statistically different from chance, *p* = .286 (see Table 1).

We have also examined whether RT was related to L2 proficiency. Length of stay abroad was used since it is a reliable indicator of L2 proficiency (e.g., Dwyer, 2004; Félix-Brasdefer, 2004; Ife, Vives, & Meara 2000). Pearson's correlation analyses revealed no significant association between the proportion of RT difference between high and low stress contexts and participants' length of stay in Australia (i.e., date of testing minus date of arrival; *r* = .208, *n* = 38, *p* = .223) or their scores on the recognition test (*r* = .029, *n* = 38, *p* = .869). For the sample of the Mandarin speakers who participated in the Mandarin experiment in Experiment 1a, there was also no significant correlation between their length of stay and the proportion of RT difference between the high and low stress conditions (*r* = -.266, *n* = 23, *p* = .219). With these results taken into account, Mandarin speakers' RT seemed very unlikely to be due to their amount of L2 proficiency.

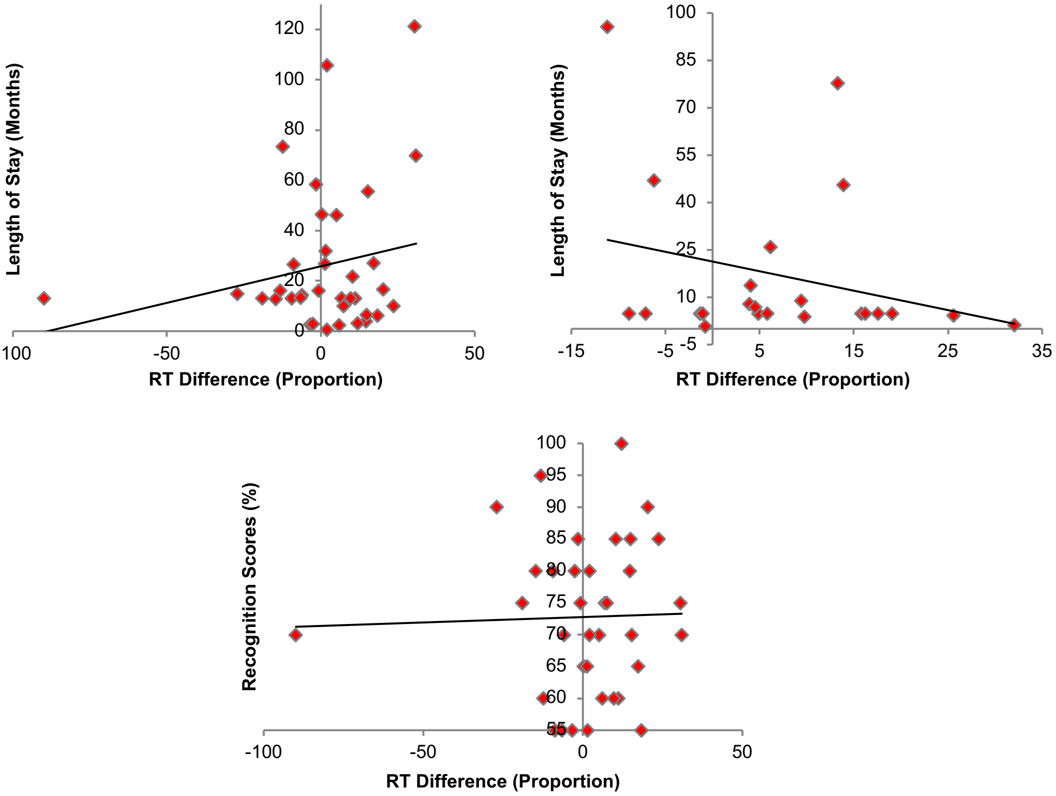


Figure 4
Non-significant correlations between Mandarin-speaking participants' response time difference between high and low stress prosodic contexts (in proportions) and length of stay in Australia (in months) in Experiment 1b (top left) and 1a (top right), and their post-test recognition scores (between 55% to 100%) in Experiment 1b (bottom centred).

3.4. Linear Mixed Effect (LME) Analyses

Using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015), Linear Mixed Effect (LME) models were also tested on all of the significant phoneme detection RT results. This analysis approach is also suitable to the present research because it allows the effects of crossed and nested subjects and item factors to be taken into account within a single analysis. Previous studies have demonstrated that residual effects of stimulus attributes, trial sequence, and stimuli list construction can in some cases explain substantial variance in RT, even when stimuli were carefully matched and counterbalanced (see Baayen, 2008). Performing LME analyses can partial out this variance and thereby eliminate artificial effects and increase power to detect real effects.

We performed an LME regression to obtain the best fitting model predicting phoneme detection RT. A baseline model with subject, sentence item, and counterbalanced experimental version was used as the starting point, and parameters (i.e., Focus and Language*Focus) were then added in a step-wise fashion to determine which predictors significantly improved model fit. Consistent with the mixed ANOVA results from Experiment 1a, our LME results (see Table 8) revealed a significant effect of predicted high stress on detection RT, but there were no language-specific differences (i.e., no Language by Focus interactions).

Table 8.
Results from the linear mixed-effect models for the significant results in Experiment 1a (based on values mapped on the intercept).

Parameters	β	SE (β)	t
Focus	42.53	10.83	3.93***
Language*Focus	6.27	21.65	0.29

****p ≤ .001* (two-tailed).

3.5. General Discussion

The present experiments offer a useful insight into how both language-universal and language-specific mechanisms influence the sentence comprehension process. Consistent with previous findings in English and Dutch (e.g., Akker & Cutler, 2003; Cutler, 1976), native Mandarin listeners can also entrain to the intonation contour to forecast an upcoming focus, despite their language being one where much of the same prosodic information in the speech signal is also used for lexical tone perception. As in the predecessor studies, the entrainment was confirmed by the fact that the original target-bearing words had been replaced by neutrally produced words, so that in both sentence contexts the targets being reacted to were acoustically identical. However, we have also found that Mandarin speakers failed to adopt an entrainment strategy when they were listening to sentences in English. In light of these results, our findings support the view that a common strategy may still exist in listeners' prefocus entrainment to prosody, despite the lack of transference to a non-native language.

The fact that prosodic cues to focus can co-exist in speech with lexical cues to tone is already well known. As demonstrated in previous studies (e.g., Xu, 1999; Chen & Gussenhoven, 2008), prosody can be used for producing focused words in Mandarin in ways that do not interfere with tonal identity (e.g., by exaggeration of pitch register while maintaining pitch contour shape). What is interesting here is the perceptual reflection of this dual role for prosody: Mandarin listeners were as likely as the English listeners to make use of the preceding intonation contour even before they heard the predicted focused word. According to some scholars (e.g., Hayes, 1995; Nolan, 2006; Pierrehumbert, 1999), languages with lexical tones ought to have less scope for a complex intonational system, given that much of the prosodic contour is preempted by its use for distinguishing words. Indeed, previous studies have suggested that Mandarin listeners fail to distinguish between intonational categories if the features in the intonation conflict with the tonal cues (e.g., Liu & Xu, 2005), suggesting that these listeners give processing priority to lexical tones over intonation. Similarly, cross-dialect comparisons of a single language suggest that intonation production as well as the use of prosodic features for focus can be more restricted in the presence of lexical tones (e.g., House, Karlsson, Svantesson, & Tayanin, 2009; Karlsson et al., 2010). On this interpretation, any prosodic cues to focus in a tone language would most likely have to be locally restricted to the focused region of the utterance, and the preceding intonation contour would thus be uninformative. However, our acoustic analyses reveal that, contrary to previous findings (e.g., Xu, 2005), some pitch and duration cues to upcoming focus were present in the preceding intonation, at least in the form of significantly longer pre-target intervals and greater F_0 range expansion and heightened pitch peaks before the onset of the anticipated accent.

This would indicate that if Mandarin listeners were to try to anticipate upcoming focus, the duration of the short pause before focus and the maximum F_0 and F_0 range cues would be available in the signal to enable this. However, correlation analyses did not reveal any significant association between each sentence item's accent effect (measured as listeners' RT proportional difference between high and low stress contexts) and the degree to which each item had hyperarticulation of pitch or duration cues such as maximum F_0 in the preceding prosody (measured as the proportional difference in prefocus cues between high and low stress contexts). The correlations were also non-significant when we examined the link between RT and these cues in absolute values.

These findings appear puzzling, as they suggest that the Mandarin listeners were not exclusively using the pre-target interval duration and maximum F_0 and F_0 range to predict the incoming focus, even though they were more reliable than the other prosodic cues. It is also noteworthy that the extent to which the predicted high stress context supported entrainment was the same in both English and Mandarin, even though acoustic analyses of the English stimuli showed reliable support from not only pre-target duration and maximum F_0 and F_0 range cues, but also cues to mean F_0, overall prefocus duration, and all the intensity cues. Similarly, in previous research such as the Cutler and Darwin (1988) study in British English, listeners did not make use of any single prosodic dimension, since they still responded faster in predicted high stress sentences when the F_0 contour was rendered uninformative (i.e., artificially levelled out).

Therefore, a possible explanation for the apparent discrepancy in the RT results versus the acoustic measures in Mandarin could be that listeners were flexible in their prosodic entrainment and were thus open to whatever cues that could help them anticipate the upcoming focus, even when only the pre-target duration, maximum F_0, and F_0 range cues were the most consistent cues throughout the experimental trials. After all, a cue can only be called a cue if it is used by the listeners. Since no one prosodic dimension was directly related to response time, we speculate that listeners do not rely any one particular cue; efficient processing of the upcoming focus may occur even when there is just one feature present in the preceding prosody. What listeners may be relying on is probably the overall prosodic pattern from a *combination* of features. Our RT difference scores for each sentence trial suggest that listeners may be attending to consistency in the different prosodic patterns. For instance, listeners may be less efficient at using the preceding prosody to predict upcoming focus if some prosodic features are in conflict with each other. As seen in our English RT results across sentence trials (see Table 6), there were five sentence items where the degree of RT difference between high and low stress contexts was on the 25th percentile. These five sentence items still had prosodic features that were in line with the prosodic contexts (i.e., higher values for high stress contexts, lower values for low stress contexts), but they also have at least one prosodic feature with differences in values that were 10 to 45 percent different in the opposite direction (i.e., higher values for low stress contexts). On the other hand, for the sentence items with huge RT differences (i.e., those in the top 25 percent), there were also some conflicting cues, but these were only below 6 percent in the opposite direction. In other words, sentence items that produce greater RT difference between high and low stress context have fewer conflicting preceding cues. Although still an empirical question, this may indicate that the extent to which a particular feature conflicts with other features may hinder listeners' use of preceding cues for focus prediction. The human language comprehension system can generalise and integrate abstract patterns from multiple sources in the speech input.

Interestingly, in our Mandarin data (Table 7), we observed cases where the conflict in prosodic features only seemed to matter mostly when there was a conflict in the different F_0 cues (e.g., where the mean F_0 difference was drastically different from maximum F_0 and F_0 range). Therefore, even if no single feature is relied on, it is still useful for future research to do more crosslanguage comparisons to examine whether certain *group* of prosodic cues (e.g., F_0, duration) may prove more informative in certain languages. One way to address this is to examine the relative contribution of different prosodic cues to focus entrainment in different languages. For example, future research could examine whether listeners of different languages vary in the degree to which they could use the preceding prosody to predict upcoming focus in contexts where the stimuli are manipulated to only have one type of prosodic cue throughout the entire experiment. For one thing, languages may vary in the type of prosodic cues that are most conducive to prosodic entrainment. In Mandarin, F_0 contour shape is most commonly assumed to be the major carrier of lexical tones and it is also, as we have shown in previous studies (e.g., Ip & Cutler, 2016), more likely to be exaggerated in focus production than is the case in English. On this account, it could be that a condition was the only cue is F_0 range can facilitate prosodic entrainment in Mandarin to a greater extent than in English, because of Mandarin listeners' enhanced use of this cue in tone processing and focus production. On the other hand, listeners of languages that mark prosodic focus using only or mainly duration cues (e.g., Chinese Cantonese: Fung & Mok, 2018; Moroccan Arabic: Burdin et al., 2016; some Otomanguean languages: Chávez-Peón, 2010; DiCanio & Hatcher, 2018) may probably benefit more from conditions where only duration cues were made informative in the preceding prosody.

Even if language variation exists in listeners' exploitation of different cues, prosodic entrainment may still be a universal strategy on the view listeners are going to process whatever cue in the preceding prosody that is most useful to them. Prosodic entrainment to locate focus may be justified by its value as a comprehension strategy for everyday social interactions. Irrespective of language or culture, holding a conversation presents mental challenges where listeners are continually presented with differing cues that must be processed quickly and accurately. For example, conversational utterances tend to be fragmentary and elliptical (Garrod & Pickering, 2004), and there is always uncertainty with respect to how a dialogue will unfold. This means that listeners need to repeatedly organise and update their ongoing discourse model. Since speech perception must involve bottom-up processing (Norris, McQueen & Cutler, 2000), entrainment to intonation contours to detect the semantically most central part of the utterance may provide a headstart for listeners in navigating the utterance information structure early on, making it a strategy useful for all listeners for maintaining a socially effective conversation.

On a related note, our acoustic findings are noteworthy in light of a recent production study from our laboratory (Ip, Shaw, & Cutler, submitted), where Mandarin speakers were more likely to produce focused words with greater degrees of increase in F_0 cues (and duration to a certain extent), while English speakers tended to produce greater increases in intensity. Given that salience is fundamentally gradient (Flemming, 2008), it could be the case that Mandarin speakers start to expand their pitch range quite early in the utterance, in preparation for pitch increases and pitch range exaggeration on the upcoming focused words. This may partly even result from an automatic physiological mechanism. As Bolinger (1978) noted almost four decades ago, the semantically most "interesting" or "important" content in an utterance is associated with heightened arousal, greater respiratory effort, dramatic pitch changes, and more energetic movement. Not only speakers' realisation of focus, but also listeners' entrainment to intonation contours and their faster response times in predicted high stress context could thus be due to increasing levels of physiological arousal as an acoustically salient word approaches in the speech stream. On this view, the maximum F_0 and F_0 range difference found in the Mandarin stimuli, although being the most statistically reliable, may only be one of the many prosodic cues that can facilitate entrainment as a result of the increase in physiological arousal. Similarly, increases in pre-target interval duration may be interpreted as a short break before the upcoming focused word. Physiological measurement techniques would however be necessary to test this suggestion.

At the same time, our findings raise an important issue concerning how listeners can process sentence intonation in the presence of lexical tones. From a functional point of view, it has been argued that the same phonetic dimension cannot be used to the same extent for two different purposes (e.g., Liang & van Heuven, 2005; Seddoh, 2002; Vogel, Athanasopoulou, & Pincus, 2016). A difference between our results and those of previous studies is that we did not manipulate the lexical tones of the preceding intonation. This may partly explain why our Mandarin listeners could still entrain with the intonation contour, since Mandarin listeners' perception of pitch contours at the word and sentence level may involve separate processes (e.g., different hemispheric lateralisation; Gandour et al., 2003). Nevertheless, in our phoneme detection test, we manipulated the tone of the syllable that bore the phoneme target as either rising or falling, and this did not show any significant difference in response time.

Likewise, our findings can also support the idea that cues to lexical tones may be redundant if listeners can already use segmental analysis. This is because lexical tones are primarily realised on vowels, and so listeners cannot process tones until the vowel information is available (Tong, Francis, & Gandour, 2008). Supporting this view, several experiments in both Chinese Mandarin and Cantonese found that listeners process tonal information less rapidly and less accurately compared to segmental cues (e.g., Cutler & Chen, 1997; Repp & Lin, 1990; Taft & Chen, 1992; Tsang & Hoosain, 1979). Further, similar to adult lexical processing, vowels maintain primacy over lexical tones in infant word recognition (Ma, Zhou, Singh, & Gao, 2017), and recent results suggest that lexical tones are not fully acquired until late in childhood (Singh & Fu, 2016; Wong & Strange, 2017; Wong, Fu, & Cheung, 2017; although the literature so far has been mixed, see Götz, Yeung, Krasotkina, & Schwarzer, & Höhle, 2018; Yeung, Chen, & Werker, 2013). In addition, in conceptual development, infants can categorise objects using word labels but not using tone sequences (Ferry, Hespos, & Waxman, 2010). Therefore, we suggest that the same suprasegmental cues to lexical tones may not exercise so crucial a role to the extent where tone perception would hinder prosodic entrainment in focus perception.

Moreover, whether or not tone exists in the language of the stimuli does not seem to affect the listeners' response at all. In Experiment 1b, we demonstrate that native

Mandarin speakers failed to use the preceding prosodic cues to anticipate upcoming focus when they were presented with sentences in English. This is interesting for two reasons. Firstly, Mandarin speakers failed to entrain in English even though the English sentences had richer and more consistent prosodic features to support an entrainment strategy than the Mandarin sentences. As revealed in the acoustics, the preceding intonation of the English stimuli showed higher mean and maximum pitch and pitch range as well as longer overall duration and pre-target pause duration for the high stress contexts, while the Mandarin sentences only showed differences in pre-target duration and maximum pitch and pitch range. Again, as already mentioned, whether prosodic features actually support sentence processing depends entirely on how listeners use them.

Secondly, our L2 findings seem to be contrary to the view that the existence of a suprasegmental cue used for one linguistic purpose may actually enhance listeners' use of the same cue for a different purpose. For example, from the perspective of cue-weighting theory in speech perception, recent experiments have found that Mandarin listeners can encode lexical stress in English better than Korean listeners, presumably because of Mandarin listeners' enhanced use of the same suprasegmental cue to process lexical tones in their L1 (Connell et al., 2018; Lin, Wang, Idsardi, & Xu, 2014). Similarly, Tremblay, Broersma, and Coughlin (2016) showed that native listeners of Dutch can transfer their L1 use of F_0 cues for lexical stress to perceive word-final boundaries in French.

One reason for the lack of non-native transference in our study might be thought to be the lower levels of overall English proficiency of the Mandarin speakers. Across both high and low stress contexts, the Mandarin speakers had a slower average RT and lower scores on the recognition test in their L2 compared to their L1 and those of the English speakers. However, lower levels of English proficiency cannot fully explain the lack of non-native transference, as there was no significant correlation between listeners' RT in their L2 and their amount of exposure to English (measured as length of time in Australia) or their recognition scores.

More important is that prosodic processing, and in particular the mapping of prosody to information structure, differs across native and non-native listeners. As a result, differences in native versus non-native prosodic processing could arise from listeners' adaptation to different intonation systems. For example, Pennington and Ellis (2000) assessed native Cantonese speakers' memory of English sentences produced in different prosodic versions (e.g., with or without a prosodically focused word). Participants first heard a set of 24 simple sentences and were later asked to judge whether or not each of the 48 test sentences came from one of the earlier 24 sentences. Even highly proficient non-native speakers were poor at distinguishing between prosodically altered sentences when they were not made aware of the different intonation patterns. Similarly, Vanlancker-Sidtis (2003) found that non-native speakers are less likely to be able to discriminate between idiomatic and literal readings of word sequences in their L2.

We suggest therefore that prosodic entrainment is developed as a listening strategy that is tailored to the specific structure of the mother tongue. As infants or young children, listeners may begin with various universal auditory mechanisms (e.g., for processing basic distinctions between falling and rising contours; Grabe, Rosner, García-Albea, & Zhou, 2003) that are, over the course of development, gradually shaped by experience with a given language. At the same time, because acquisition of non-native prosody is a protracted process (Mennen, 2004), whether listeners can apply their L1 prosodic strategies in their L2 may also depend on how they process the interactions between suprasegmental and segmental information in the non-native language (Lee & Nusbaum, 1993). Future experiments could provide a more in depth look at L1 to L2 transfer by examining listeners' entrainment in English sentences that are acoustically manipulated to have Mandarin intonation. For example, studies looking at the interplay between segments and prosody in foreign accent perception created stimuli where the prosodic information from a recording produced by an individual speaker was extracted and superimposed onto segments produced by a different speaker (e.g., Ulbrich & Mennen, 2015; Winters & O'Brien, 2013). Using this procedure, it would be interesting to examine whether Mandarin listeners could engage in L2 prosodic entrainment if intonation contours of Mandarin stimuli were superimposed upon English sentences.

Another major step for future research would be crosslanguage comparisons from a language learning perspective. It would be interesting to investigate how focus is processed by first and second language learners of languages that reportedly do not use intonation to mark focus (e.g., Ambonese Malay: Maskikit-Essed & Gussenhoven, 2016; Jakartan Indonesian: van Zanten & van Heuven, 1998; Northern Sotho: Zerbian, 2006; Triqui: DiCanio & Hatcher, 2018; Wolof: Rialland & Roberts, 2001; Yucatec Maya: Gussenhoven & Teeuw, 2008). Future research could combine the phoneme-detection methodology with an artificial language learning paradigm where different prosodic cues could facilitate anticipation of upcoming accented words. Speakers of languages that do not use prosody for focus marking may not be able to adapt the prosodic features as efficiently as speakers of languages with prosodic cues to focus, but there may still be some subtle processing similarities For instance, if tested in follow-up recognition tests, all participants may be more likely to be able to remember the target words from sentences with predicted high stress contexts compared to low stress contexts. Similarly, listeners may show greater influence of word priming for target words in predicted high stress contexts, even when these words are acoustically neutral.

3.6. Conclusion

Even though Mandarin has lexical tone, whereby F_0 patterns carry a lexical as well as a sentence-level functional load, Mandarin listeners entrain to preceding intonation across utterances to predict upcoming focus. Consistent with data from speech production in Mandarin, acoustic analyses of the present stimuli revealed longer pre-target interval duration, higher maximum F_0, greater F_0 expansion in the preceding intonation of predicted high stress sentences, while the English stimuli showed a larger variety of prosodic cues (i.e., greater F_0 range, mean F_0, maximum F_0, overall duration, and pause duration). However, Mandarin listeners failed to engage in prosodic entrainment strategy when the sentences were presented in English, suggesting that the listening strategy is developed as a language-specific strategy. Nevertheless, the fact that entrainment can be used to the same extent in both English and Mandarin native processing, despite the acoustic differences, indicates that the strategy operates in a universal manner. In everyday conversations, one of the most crucial of the listener's tasks is to actively search for the semantically most important word in the speaker's message. Attending to whatever cues that are available in the speech stream can help listeners anticipate where this word will occur, even before it is uttered.

CHAPTER 4

– In Search of Salience –

A version of this chapter was submitted to *Language and Speech*: Ip, M. H. K., & Cutler, A. (under review). In search of salience: Focus detection in the speech of different talkers.

Portions of the data presented in this chapter were presented at the 17th Australasian International Conference on Speech Science and Technology (SST 2018): Ip, M. H. K., & Cutler, A. (2018). Cue equivalence in prosodic entrainment for focus detection. In J. Epps, J. Wolfe, J. Smith, & C. Jones (Eds.), *Proceedings of the 17th Australasian International Conference on Speech Science and Technology* (pp. 153-156), Sydney, Australia: ASSTA.

4.0. Abstract

Many different prosodic cues can help listeners predict upcoming speech. However, no research to date has assessed listeners' processing of preceding prosody from different speakers. The present experiments examine (1) whether individual speakers (of the same language variety) are likely to vary in their production of preceding prosody; (2) to the extent that there is talker variability, whether listeners are flexible enough to use any prosodic cues signalled by the individual speaker, and (3) whether types of prosodic cues (e.g., F_0 versus speech rate) vary in informativeness. Using a phoneme detection task, we examined whether listeners can entrain to different combinations of preceding prosodic cues to predict where focus will fall in an utterance. We used unsynthesised sentences recorded by four female native speakers of Australian English who happened to have used different preceding cues to produce sentences with prosodic focus: a combination of prefocus speech rate cues, F_0 and intensity (mean, maximum, range), and longer pre-target interval before the focused word onset (Speaker 1), and only mean F_0 cues, mean and maximum intensity, and longer pre-target interval (Speaker 2), only pre-target interval duration (Speaker 3), and only speech rate and maximum intensity (Speaker 4). Results revealed that listeners could entrain to almost every speaker's cues except for when speech rate and maximum intensity were the only reliable cues. Further, listeners could use whatever cues were available even when one of the cue sources was rendered uninformative. Our findings demonstrate functional equivalence of different preceding cues to prosodic entrainment.

– In Search of Salience –

4.1. Introduction

4.1.1. *Finding the Right Referent*

Holding a conversation can present a number of cognitive challenges. For one thing, listeners not only need to decode the phonetic sequence that determines what words and utterances they hear, but also the pragmatic structure that underlies how various information is expressed in the speaker's message. For another, conversational utterances tend to be elliptical and spontaneous, and there is often much uncertainty with respect to how a dialogue will unfold (Cutler, 1997; Garrod & Pickering, 2004). For example, the speaker may change topics, correct a previous response, or suddenly introduce new information that is not under discussion (e.g., Kiss, 1998; Krifka, 2006; Rochemont, 1986). To maintain a socially effective conversation, all listeners must adopt strategies to organise and update their discourse model with speed and accuracy.

One such strategy is to actively search for the most important word in the utterance. According to Chafe (1987), speakers tend to plan and produce sentences that contain at least one piece of new information. With the arguable exception of a few languages and regional dialects, there is a general tendency for prosodically highlighted words to be associated with the semantically most central portions of the message. This could partly be due to a basic physiological mechanism. For instance, Bolinger (1978) argued that the most "informative" or "interesting" aspects of the message are often associated with a heightened state of arousal, and when expressed in speech, this arousal would induce prosodic focus through greater articulatory efforts and more energetic movements. On the other hand, Gussenhoven (2000; 2002) proposed that speakers can intentionally exploit certain biologically determined conditions (e.g., amount of articulatory energy exerted) to create intonational meanings (e.g., emphasis) through various phonetic implementations (e.g., wider pitch excursion: Wichmann, House, & Rietveld, 1997), even though a minority of languages may exhibit “unnatural” arbitrary form-function relations due to language change. In either case, words marked with prosodic focus are easier to process because of their acoustic clarity and greater spectral balance (Dahan & Bernard, 1996; Hawkins & Warren, 1994; Klatt, 1976; Redford, Stine, & Vatikiotis-Bateson, 2014), and various perceptual advantages have been revealed in different languages, including a deeper processing of focused words in lexical activation (Blutner & Sommer, 1988; Brunellière, Auran, & Delrue, 2018; Li & Ren, 2012; Norris, Cutler, McQueen, & Butterfield, 2006; Wang, Bastiaansen, Yang, & Hagoort, 2011), a faster and more accurate word recognition (Birch & Clifton, 1995; Cutler & Foss, 1997; Lee, Chiu, & Xu, 2016; McAllister, 1991), better retention in memory (Birch & Garnsey, 1995; Kember, Choi, Cutler, 2016; Fraundorf, Watson, & Benjamin, 2010), and better access to contextual alternatives (Braun & Tagliapietra, 2010).

An important question is how listeners can begin their search for focus even when they have not yet received any acoustic signals from the focused word. Cutler (1976) discovered that listeners can direct their attention to upcoming focused words by entraining with cues in the utterance intonation contour. In a phoneme detection task, participants listened to a series of sentences and responded as fast as they could to words that began with a specified phoneme target (e.g., [b] in “*book*”). Results show that listeners responded faster to the target in sentences where the preceding prosodic contour predicted high stress on the target-bearing word (1a), compared to sentences where high stress was predicted to occur a few syllables later (1b).

(1) Target [b]

a. The couple had quarrelled over a BOOK they had read.

b. The couple had quarrelled over a book they hadn’t even READ.

Importantly, there was still a significant response time difference even when the original accented and unaccented target words were replaced by an acoustically identical neutral version of the same words. Since the only difference was in the preceding prosody, it was proposed that listeners can entrain with the prosodic information that is immediately available in the speech stream to predict the timing of future accents.

Subsequent research confirmed that this entrainment strategy operates in the same way as a search for semantic salience. In a crosslanguage study with native listeners of English and Dutch, Akker and Cutler (2003) used a similar phoneme detection task where they manipulated both the prosodic context and the semantic structure of the sentences. The task involved sentences like “*The manager of the dairy will check the bank account*” where accent would either fall on “*dairy*” or “*bank*” and the phoneme target would be [d] or [b]. However, at the start of each trial, participants were also primed with one of two questions that would bias their attention to either the accented target word or the distractor word (e.g., “*which manager…*” or “*which account…*” two seconds before hearing the sentence with predicted high stress on “*dairy*”). With this approach, the experiment revealed that prosodic accent and the question-induced focus interact in the degree to which listeners use prosodic entrainment as a search strategy. As in the original Cutler (1976) study, response times were faster when the preceding prosody predicted high stress on the specified target, but the effect of prosodic accent was significantly reduced when the semantic context also facilitated prediction of the target word. Therefore, the degree to which listeners engage in prosodic entrainment depends on whether other cues (e.g., context) are also available to support their search for salience.

In addition, eye-tracking experiments have provided evidence that listeners integrate prosodic information with the discourse structure during the early stages of focus processing. For example, Dahan, Tanenhaus, and Chambers (2000) examined the effect of accent on lexical competition and found that listeners were able to integrate prosodic focus with discourse cues. Participants were asked to move objects in a display where they first heard an instruction sentence such as "*Put the candle above the triangle*", followed by a second instruction "*Now put the candle above the square*" with accent on either the noun "*candle*" or on the prepositional phrase "*above the square*". When the noun in the second instruction was accented, participants looked more often at the lexical competitor "*candy*", arguably because they misinterpreted the inappropriate pitch accent on the noun "*cand-*" to be discourse-new (see also Ito & Speer, 2008 and Weber, Braun, & Crocker, 2006 for similar results). Consistent with the experiments demonstrating prosodic entrainment, these findings all indicate that prosodic and discourse processing are part of the same strategy to facilitate the listener's search for the focused referent.

4.1.2. *Variation, Flexibility, and Cue Weighting*

In the present study, we address three questions. The first question concerns whether individual speakers vary in their production of different cues in the preceding prosody. Second, if there are considerable individual differences, we examine whether listeners are flexible enough to exploit whatever cues are available in the immediate speech signal to efficiently search for the focused word in the utterance. Third, we examine whether some cues in the preceding prosody are more informative than others (e.g., F_0 vs. duration).

There is to date no data on how individual speakers within a given language would differ in their prosodic production before focus. However, recent research has shown that individual speakers within a given language can differ in on-focus word production (e.g., variation in F_0 shapes and ranges; Ouyang & Kaiser, 2015). Likewise, speakers can also differ in the number of cues used to signal different intonational contrasts, and listeners are sensitive to these differences (e.g., Cangemi, Krüger, Grice, 2015). It is therefore highly likely that individual speakers would also vary in their production of different preceding cues.

Nevertheless, listeners may still be flexible in their prosodic entrainment. For example, Cutler and Darwin (1981) found that listeners can engage in prosodic entrainment to predict accent location even when some of the cues in the preceding prosody were rendered uninformative (e.g., by removing the closure duration of the target stop phoneme or monotonising the F_0 information). In another study, Cutler (1987) transposed the timing patterns and the pitch and intensity contours across sentences and found that listeners still showed a response time advantage in predicted high stress contexts. However, predicted accent no longer had an effect on response time, even with intact pitch and intensity, when only the timing patterns were transposed across the sentences. Therefore, processing would still be affected if the preceding intonation had an unnatural conflicting prosodic structure (e.g., where the pitch and intensity contours signal one pattern while the timing contour signals another).

Similar findings by Dilley and colleagues have also shown that the global rhythmic patterns in pitch and timing can influence the perception of upcoming words in cases of lexical ambiguity. In one study, Dilley and McAuley (2008) asked listeners to report the final words of eight-syllable sequences where the initial five syllables contained two disyllabic trochaic words and a third monosyllabic word (e.g., "*chocolate lyric down…*"), followed by three final syllables that can be processed in more than one way (e.g., "*…town shipwreck*", "*…township wreck*"). For the first five syllables, Dilley and McAuley manipulated the periodic alternation of strong and weak syllables to produce either a predicted monosyllabic context, where the initial syllables contained two strong-weak disyllabic words followed by a lengthened third monosyllabic word, or a disyllabic context, where there were two strong-weak disyllabic words and a shortened monosyllabic word. When the third monosyllabic words were shortened, listeners were more likely to report hearing the sequence final words as disyllabic (e.g., "*shipwreck*"), even though the final words were acoustically identical in both contexts. Presumably, the preceding rhythmic pattern involving the shortened monosyllabic word caused listeners to continue hearing a binary strong-weak grouping of sequence elements (e.g., "*downtown...*" rather than "*down…town*"). Consistent with these findings, later work using eye-tracking methodology has also revealed that listeners can use information from preceding rhythmic patterns to predict upcoming lexical stress (e.g., "*jury*" versus "*giraffe*"; Brown, Salverda, Dilley, & Tanenhaus, 2011; 2015), and studies using the event-related potential (ERP) recordings show that preceding cues can support prediction of word boundaries and later lexical processing and interpretations of what was heard (Breen, Dilley, McAuley, & Sanders, 2014). Further, recent research has also shown that speech rate can also facilitate prediction of upcoming weak syllables (Baese-Berk, Dilley, Henry, Vinke, & Banzina, in press), suggesting that preceding prosodic cues can have a pervasive role in predicting upcoming words.

Importantly, Dilley and McAuley (2008) also varied the cue type. For example, in the F_0 only condition, the preceding prosody only featured F_0 alternations of high and low pitch units, while the temporal characteristics were held constant, whereas in the duration only condition, F_0 was flat and the temporal characteristics remained intact. Although both types of rhythmic patterns supported word disambiguation, the strongest influence of preceding rhythm was still in the original condition where both F_0 and duration cues were presented to listeners. On the other hand, the duration-only condition showed the smallest effect while an intermediate effect was found for the F_0-only condition. These results suggest that the effect of different cues in the preceding prosody is additive, and that preceding pitch cues are more perceptually informative than duration (see also Morrill, Dilley, McAuley, & Pitt, 2014).

However, apart from rhythm, there are two more possible ways in which the temporal properties of the preceding prosody can alter perception of upcoming speech. One is the presence of pausing before a focused word. The tendency to pause (and pause longer) before adding new information has been revealed in a number of production studies across various languages, including English (Gee & Grosjean, 1984; Redford, 2013), French (Dahan & Bernard, 1996), Dutch (Romøren & Chen, 2015), Chinese Cantonese (Gu & Lee, 2007), and Chinese Mandarin (Huang & Liao, 2002). These pauses may take the form of an extra lengthening effect before a prosodically focused word with plosive word onsets. For example, in English, while stop closures may range from 80 to 250 milliseconds (Dalton & Hardcastle, 1977), stop closure duration before the release burst of focused stops tend to range between 130 to 250 milliseconds (Hieke, Kowal, & O'Connell, 1983), which may be robust enough to facilitate anticipation of the focused word (see Dahan & Bernard).

Another temporal feature of the preceding prosody that may influence processing of upcoming speech features is speech rate. For example, Dilley and Pitt (2010) found that listeners can use contextual speech rate cues to predict the presence of heavily coarticulated function words. Dilley and Pitt used sentences containing a spectrally reduced function word and manipulated the speech rate of the preceding prosody (e.g., *or* from *minor or* [maɪnɚ:] in "*Anyone must be a minor or child…*"). Compared to sentences with normal speech rate, listeners were less likely to detect the function word when the preceding context was slowed, even though the target words were acoustically identical in both contexts. Conversely, speeding the speech rate caused listeners to hallucinate hearing a function word that was never spoken (e.g., *a* in "*The company moved to (a) different…*"). In relation to focus processing, these results may have implications for prediction of upcoming prosodic focus. For example, speakers may, in principle, tend to speak slower before producing a lengthened prosodically highlighted word, thereby leading to better anticipation of the upcoming focused word.

Given that many cues in the preceding prosody may influence processing of the upcoming speech signal, what is the perceptual weighting of these cues? In the traditional literature on the acoustic correlates of lexical stress in English and Dutch, many studies suggest that listeners are more sensitive to F_0 cues than duration and least sensitive to intensity cues (e.g., Bolinger, 1958; Fry, 1955; 1958; van Katwijk, 1974; Lehiste, 1970), while others more recently have shown that intensity is the more reliable correlate (e.g., Kochanski, Grabe, Coleman, & Rosner, 2005) and that the relative importance of cues varies across languages (e.g., Chrabaszcz, Winn, Lin, & Idsardi, 2017; Gordon & Roettger, 2017). Likewise, it is an open issue how different cues are weighted in listeners' processing of the preceding prosody. Apart from a handful of studies comparing F_0 and duration cues (e.g., Dilley & McAuley, 2008), no studies have, to the best of our knowledge, compared the roles of all three aspects of prosody (i.e., pitch, duration, intensity) in prosodic entrainment for focus detection. Moreover, all of the previous experiments used synthesised speech materials (e.g., sentences with manipulated flat F_0: Cutler & Darwin, 1981), which may not always reflect natural speech in everyday conversations. Finally, it is still an empirical question whether cue weighting also exists within a particular prosodic dimension (e.g., F_0 range versus mean F_0; pause duration versus speech rate).

4.1.3. *Overview of Experiments*

In the present experiments, we examine the relative role of different preceding cues to prosodic entrainment in the context of speaker variation. Listeners engage in a phoneme detection task where they listen to a set of sentences that were produced by one of four speakers. The four speakers who produced these sentences were all native speakers of Australian English in their late 20s or early 30s. All four speakers produced the same set of sentences and were not given any explicit instructions on what cues to use. Furthermore, in order to use stimuli that would most reflect natural speech, we did not alter any of the preceding prosodic cues in any experiment. We analysed the preceding prosody (i.e., three to four syllables before the target word onset) of predicted high and low stress sentences on the following parameters: speech rate (measured as overall duration of the preceding syllables), pre-focus pausing (duration of pre-target interval before release of the target plosive), mean F_0, maximum F_0, F_0 range, root-mean-square (RMS) mean intensity, maximum intensity, and intensity range. We hypothesise that there will be between-speaker differences in the production of the stimuli sentences, and that listeners' entrainment may benefit most from the speakers who produced the most variety of preceding cues (see Dilley & McAuley, 2008 and Morrill et al., 2014 for the additive effects of prosodic cues). Based on the past studies on acoustic correlates of prominence, we also hypothesise that listeners may be more sensitive to some preceding cues to others.

4.2. Experiment 1

4.2.1. *Method*

Participants. The final sample comprised 23 native speakers of Australian English (M_{age} = 23.96 years, *SD* = 8.64 years; 16 females). Data from two further participants were also excluded, due to technical issues. None of the participants reported any hearing or speech impairments.

Materials. Twenty-four experimental sentences (see Appendix D) were recorded in three versions by a female native speaker (the first speaker in our experiment series, henceforth S1) who did not know the purpose of the experiment. In the first version, the target-bearing word received emphatic stress. In the second version, emphatic stress was instead placed on a word that occurred later in the sentence than the target-bearing word, which as a result, received very reduced stress. In the third version, the target-bearing word and the sentence as a whole were produced in a neutral manner. In all of the experimental sentences, the phoneme target was a voiceless aspirated bilabial stop [pʰ] occurring at the start of the target-bearing word's first syllable (e.g., "*peanuts*" [pʰi:nʌts]).

The number of syllables between the start of the sentence and the onset of the target-bearing word in the present study (English, *M* = 10.00, *SD* = 2.95) was similar to the previous Cutler and Darwin (1981) study (*M* = 10.30, *SD* = 3.16). Using Praat (Boersma & Weenink, 2018), the target-bearing words were excised (at the nearest zero crossing of the initial consonant burst) from all three versions of each experimental sentence. The high- and low-stressed target-bearing words from the first and second versions were replaced by an acoustically identical token of the same target word from the neutral version. Thereby, two experimental conditions were constructed, each containing one version of each of the 24 spliced experimental sentences, plus an additional set of 24 filler sentences. The experimental sentences with predicted high versus predicted low stress were counterbalanced across the two conditions (henceforth called "Version A" and "Version B"). To avoid interference between the sentences, sentence beginnings were varied and semantic content that could be associated with another sentence in the set was avoided. We also varied the syntactic category of the word immediately preceding the target word, so that less than half of the target words were preceded by a determiner (and we used a variety of determiners). In addition, none of the sentences had any additional occurrence of bilabial stops beyond that in the target-bearing word. All of the sentences were produced at a natural fast-normal rate.

Procedures. Participants were tested in their native language in a sound-attenuated booth at the MARCS Institute, Western Sydney University. The phoneme-detection task was administered using E-Prime software on a laptop computer, with attached to it a set of headphones and a Chronos® USB-based response device for button pressing (Schneider, Eschman, Zuccolotto, 2002). Participants were informed that the experiment aimed to examine listeners' memory and language comprehension. They were told that they would listen to a series of sentences and had two tasks: first, pay attention to the meaning of each sentence, and second, press a button as soon as they heard a word that began with the target sound [p^h]. Participants received two practice trials and feedback before starting the actual experiment (see Appendix F). At the end of the testing session, the participants completed a follow-up recognition test in which they were asked to judge whether or not each of the 20 sentences in the list were from the experiment (see Appendix H). All participants in the final sample scored 65 percent or above in the test (*M* = 88.48, *SD* = 7.75, range: 70 – 100).

Acoustic analyses. Acoustic analyses of the stimulus recordings were conducted based on inspection of the waveform and the spectrogram in Praat. In each experimental sentence, the preceding syllables before the target words (i.e., two to four syllables before the onset of the target-bearing word) were annotated, and overall duration (in milliseconds), F_0 (mean, maximum, range), and root-mean-square (RMS) intensity (mean, maximum, range) were measured (see Figure 1 for an example sentence). We also measured the pre-target interval duration, the duration of the brief pause before the release of the target stop consonant (i.e., silent part of the utterance between the onset of the target bearing word and the offset of the word before it).

Statistical evaluation of the results of these analyses show significant differences for all prosodic measurements except for intensity range: speaker produced the preceding syllables of predicted high stress sentences with longer overall duration, $t(23) = 2.23$, $p = .036$, longer pre-target interval duration, $t(23) = 4.46$, $p < .001$, higher mean F_0, $t(23) = 2.23$, $p = .036$, higher maximum F_0, $t(23) = 3.78$, $p = .001$, greater F_0 range, $t(23) = 4.61$, $p < .001$, greater mean intensity, $t(23) = 4.88$, $p < .001$, and greater maximum intensity, $t(23) = 5.30$, $p < .001$.

Target: $[p^h]$

(a) The old lady thought she saw three [PIXIES] in her garden

(b) The old lady thought she saw three pixies in her [GARDEN]

(c) The old lady thought she saw three pixies in her garden

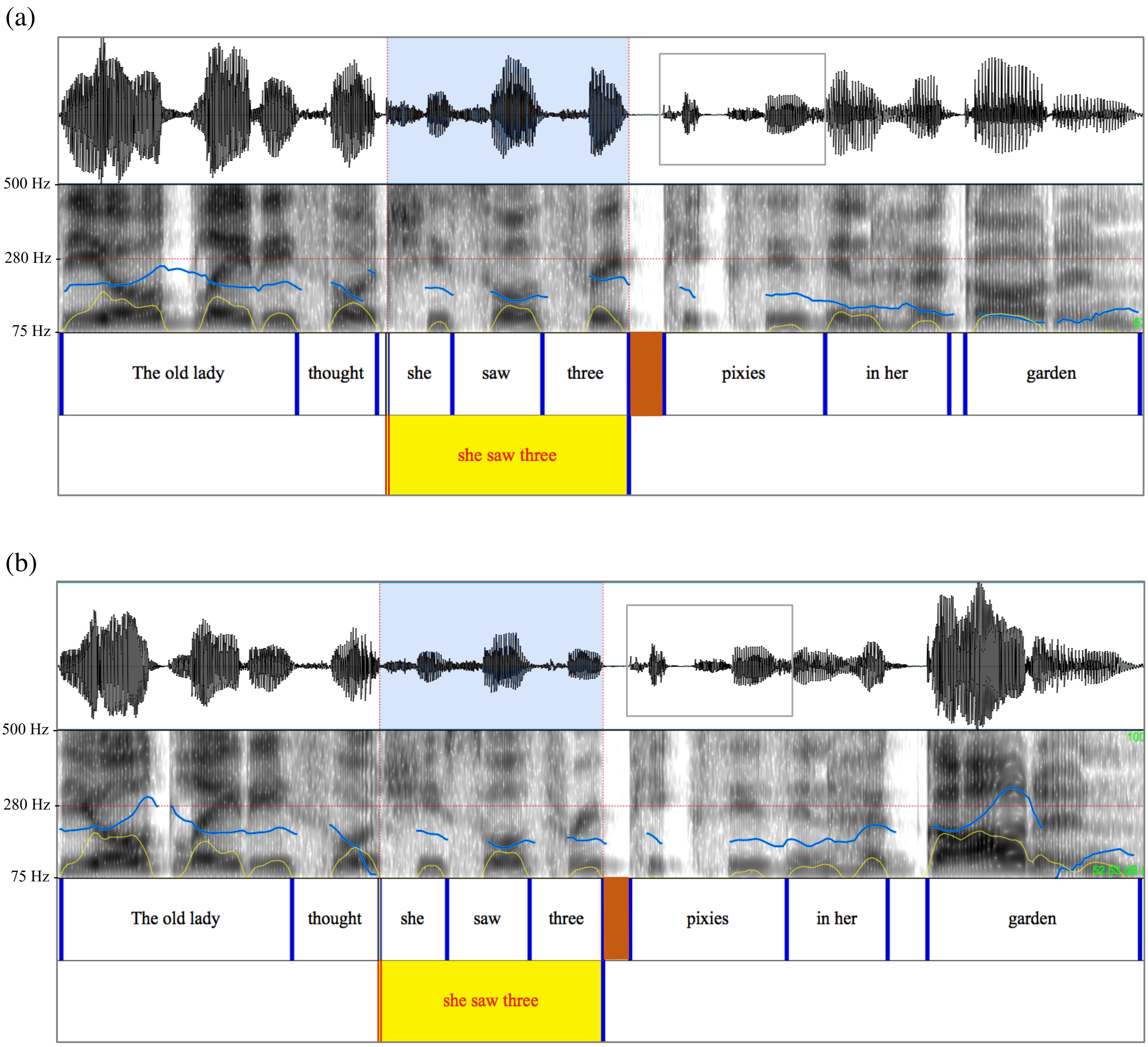


Figure 1.
Waveforms and pitch and amplitude contours of an example experimental sentence in predicted high (a) and low (b) stress contexts from S1 in Experiment 1; text (c) gives the neutral context. Prosodic parameters (i.e., overall duration, mean and maximum F_0, F_0 range, mean and maximum intensity, and intensity range) in the shaded portion – four syllables preceding the target-bearing word (squared) – were measured for our acoustic analyses. The red shaded portion indicates the duration of the pre-target interval.

4.2.2. *Results and Discussion*

Data analyses. For accuracy, we measured the total number of misses participants had in the predicted high and low stress contexts. For response times (RT), we measured the duration of the latencies between the release of the target stop consonant and participants' button presses. We compared participants' RT to the target phoneme in predicted high stress sentences with their RT in predicted low stress sentences. Participants who had an average RT score of less than 100 milliseconds or over 1000 milliseconds would be excluded (Ratcliff, 1993), because an extremely fast response may indicate accidental presses or false alarms (i.e., pressing the button despite not hearing the target) and a delayed response may indicate a reprocessing of the sentence. RT datapoints shorter than 100 milliseconds or longer than 2500 milliseconds were also excluded from the final analyses. In addition, we would exclude participants who had more than two instances of RT longer than 2500 milliseconds.

Response time. No RT datapoints were shorter than 100 milliseconds; two longer than 2500 milliseconds were excluded, one in the high stress context and one in the low stress context. We conducted control analyses on the final sample to assess whether there was a significant effect of the counterbalanced experimental conditions. These analyses revealed no significant effect, so the main analyses were conducted without these variables in the model. To evaluate whether or not the participants had faster RT to the phoneme target in predicted high stress contexts, a two-tailed within-subjects t-test with an alpha threshold of .05 was conducted to assess the difference in RT between the predicted high versus low stress sentences. RTs were significantly faster in predicted high stress sentences (M = 418.77, SD = 72.43) compared to sentences with predicted low stress (M = 459.90, SD = 84.77), t(22) = 2.92, p = .008 (see Table 1 and Figure 2). Similarly, supplementary statistical analyses using Linear Mixed Effect (LME) modelling revealed a significant fixed effect of predicted stress (t = 3.24, β = 39.58, SE = 12.20) (see Table 3).

Accuracy. For accuracy (see Table 2), we performed a two-tailed binomial sign test to determine whether participants were more likely to miss a button press to the phoneme target in sentences with predicted low stress than in predicted high stress. There were a total of six misses, with one miss in the predicted high stress context and five in the predicted low stress context, which was not statistically different from chance, *p* = .219.

Discussion. Consistent with previous studies, listeners responded faster to the target phoneme in sentences where the preceding prosody predicted high stress on the target-bearing word. Australian English speakers can thus entrain to the preceding prosodic contour to forecast an upcoming focused word. However, because the acoustic analyses of the stimuli revealed differences on all preceding cues, it remains unclear as to whether some cues are more informative than others. In the following experiments, we examine listeners' response in contexts involving different prosodic cues.

4.3. Experiment 2

4.3.1. *Method*

Participants. We recruited a new sample of 22 native speakers of Australian English (M_{age} = 20.54 years, *SD* = 3.39 years; all females). We excluded data from five participants who had more than two misses and one participant with average RT scores over 1000 milliseconds.

Materials and procedures. The procedures and recordings produced by S1 were the same as that used in Experiment 1, but the duration of the pre-target interval was rendered uninformative by splicing the target-bearing word at the onset of the closure rather than at the release of the burst (henceforth S1'). Participants scored an average of 84.55 percent on the recognition test (*SD* = 10.90, range: 65 – 100 percent).

4.3.2. *Results and Discussion*

Response time. Two datapoints from the predicted low stress sentences were excluded for being over 2500 milliseconds long. Listeners responded faster to the target in predicted high stress sentences (M = 490.38, SD = 68.72) compared to predicted low stress sentences (M = 542.89, SD = 73.16), t(21) = 3.97, p = .001. Consistent with these ANOVA results, the LMER results showed a significant fixed effect for predicted stress (t = 4.13, β = 55.40, SE = 13.42)

Accuracy. With respect to accuracy, there were 13 misses in the predicted low stress sentences and 4 misses in the predicted high stress sentences, p = .049.

Discussion. Consistent with the results from Experiment 1, the findings from Experiment 2 revealed that listeners can still use other prosodic information in the preceding prosody to efficiently forecast an upcoming focused word even when the brief pause before the target stop release was rendered uninformative. However, it is still uncertain whether listeners may also show the same entrainment strategy using unsynthesised speech from a different speaker. Therefore, in the following experiments, we explore whether different speakers may produce the preceding prosody differently and whether listeners can still use these cues for focus prediction.

4.4. Experiment 3

4.4.1. *Method*

Participants. Another new sample of 23 native speakers of Australian English (M_{age} = 22.16 years, SD = 5.37 years; 17 females) was recruited. Data from one additional participant were excluded for being at an age that was almost 10 standard deviations beyond average age of the mean. We also excluded additional data from one further participant who was born in Australia but grew up in a non-English speaking country.

Materials and procedures. The procedures and sentences were identical to those in the previous experiments, only this time, the sentences were recorded by another female native speaker (S2). Participants scored an average of 81.52 percent on the recognition test (*SD* = 12.38 percent, range: 65 – 100 percent).

Acoustic analyses. Acoustic analyses of the stimuli sentences only showed significant differences in preceding prosody between high and low stress prosodic contexts on pre-target interval duration, $t(23) = 4.61, p < .001$, mean F_0, $t(23) = 3.54, p = .002$, mean intensity, $t(23) = 5.14, p < .001$, and maximum intensity, $t(23) = 5.42, p < .001$. There were no significant differences on other F_0 and intensity measures, and no significant differences on any of the duration measures.

4.4.2. *Results and Discussion*

Response time and accuracy. None of the participants had RT datapoints shorter than 100 milliseconds or longer than 2500 milliseconds. Consistent with the results from the previous two experiments, listeners' RT in Experiment 3 was faster for predicted high stress sentences (*M* = 379.65, *SD* = 68.12) compared to low stress sentences (*M* = 404.52, *SD* = 80.44), $t(22) = 2.54, p = .019$. Likewise, the LMER results showed a significant fixed effect for predicted stress (t = 2.68, β = 25.86, SE = 9.64). In terms of accuracy, there was one miss in the predicted high stress sentences and none in the predicted low stress sentences. There was also one false alarm in each of the sentence stress contexts.

Discussion. With the same sentences, but recorded by a different speaker, the results indicate that listeners are as likely to use the cues from the preceding intonation regardless of whether there is a combination of many different cues (as in the sentences produced by S1 in Experiments 1), whether the closure duration of the target stop from Experiment 1 was made uninformative (Experiment 2), or whether there were only reliable cues from pre-target duration interval, mean F_0 and mean and maximum intensity

(Experiment 3). However, it is an open question whether listeners can still entrain if the most informative cue in the preceding prosody is not pitch-based or intensity-based. It is also at present unclear whether listeners would also engage in prosodic entrainment if only one type of preceding cue was consistently present throughout the sentence trials. The following experiment will use the same set of sentences recorded by speakers who happened to have signaled upcoming focus using mostly duration-based cues.

4.5. Experiment 4

4.5.1. *Method*

Participants. The final sample comprised of 23 native speakers of Australian English (M_{age} = 22.04 years, *SD* = 6.80 years; 19 females). We excluded two participants from the final sample for having average RT scores above 1000 milliseconds. Data from a participant who scored at chance on the recognition test were also excluded.

Materials and procedures. Stimuli sentences identical to the previous experiments were recorded by a third female speaker (S3). The procedures remained the same. Participants scored an average of 82.73 percent on the recognition test (*SD* = 12.70 percent, range: 65 – 100 percent).

Acoustic analyses. Acoustic analyses of the experimental sentences recorded by the third native speaker only revealed significantly longer pre-target interval before the predicted focused word in high stress context, $t(23) = 5.30, p < .001$. There were no significant differences for speech rate or any of the intensity or F_0 measures.

4.5.2. *Results and Discussion*

Response time and accuracy. We excluded two RT datapoints from predicted high stress sentences with RT shorter than 100 milliseconds and one datapoint longer than 2500 milliseconds from predicted low stress sentences. As in the previous experiments, RT was faster for predicted high stress sentences (*M* = 405.19, *SD* = 108.50) compared to low stress sentences (*M* = 445.37, *SD* = 126.41), *t*(22) = 3.96, *p* = .001. Likewise, the LMER results showed a significant fixed effect for predicted stress (t = 3.12, β = 40.80, SE = 13.09). In terms of accuracy, there were only two misses and one false alarm for the predicted low stress sentences.

Discussion. Consistent with the previous experiments, the results demonstrate that listeners could still respond faster to the predicted accented target even when the speaker who recorded the stimuli only consistently produced longer pre-target interval. An interesting follow-up study to these results would be to investigate whether other types of duration cues (e.g., speech rate) could also facilitate the same response.

4.6. Experiment 5

4.6.1. *Method*

Participants. We recruited a new sample of 22 college-aged native speakers of Australian English (16 females). We excluded data from a participant with RT scores beyond 2.5 standard deviations.

Materials and procedures. All participants scored above chance on the recognition test. We used the same procedures and sentences from the previous experiments using stimuli produced by a fourth speaker (S4).

Acoustic analyses. Acoustic analyses of the experimental sentences from this speaker revealed significant differences in overall duration, such that the preceding parts (four to five syllables) of the predicted high stress sentences before the onset of the target-bearing word were longer (i.e., produced slower) than the preceding parts of the low stress sentences, *t*(23) = 4.21, *p* < .001. There were also significant differences in maximum intensity *t*(23) = 3.18, *p* = .004. There were no significant differences in pre-target interval duration, F_0, or in any of the other intensity measures.

Table 1.
Response time (in ms) to the target phoneme [p^h] in Experiments 1 to 5.

Experiment	Mean Response Time (SD) [Range]		
	Predicted High Stress	Predicted Low Stress	*t*
Experiment 1: S1 (*N* = 23)	418.77 (72.43) [340-603]	459.90 (84.77) [362-568]	2.92**
Experiment 2: S1' (*N* = 22)	490.38 (68.72) [381-684]	542.89 (73.16) [404-674]	3.97***
Experiment 3: S2 (*N* = 23)	379.65 (68.12) [275-583]	404.52 (80.44) [292-630]	2.54*
Experiment 4: S3 (*N* = 23)	405.19 (108.50) [267-732]	445.37 (126.41) [312-902]	3.96***
Experiment 5: S5 (*N* = 22)	411.26 (81.47) [287-638]	431.62 (90.45) [267-585]	0.97

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$ (two-tailed).

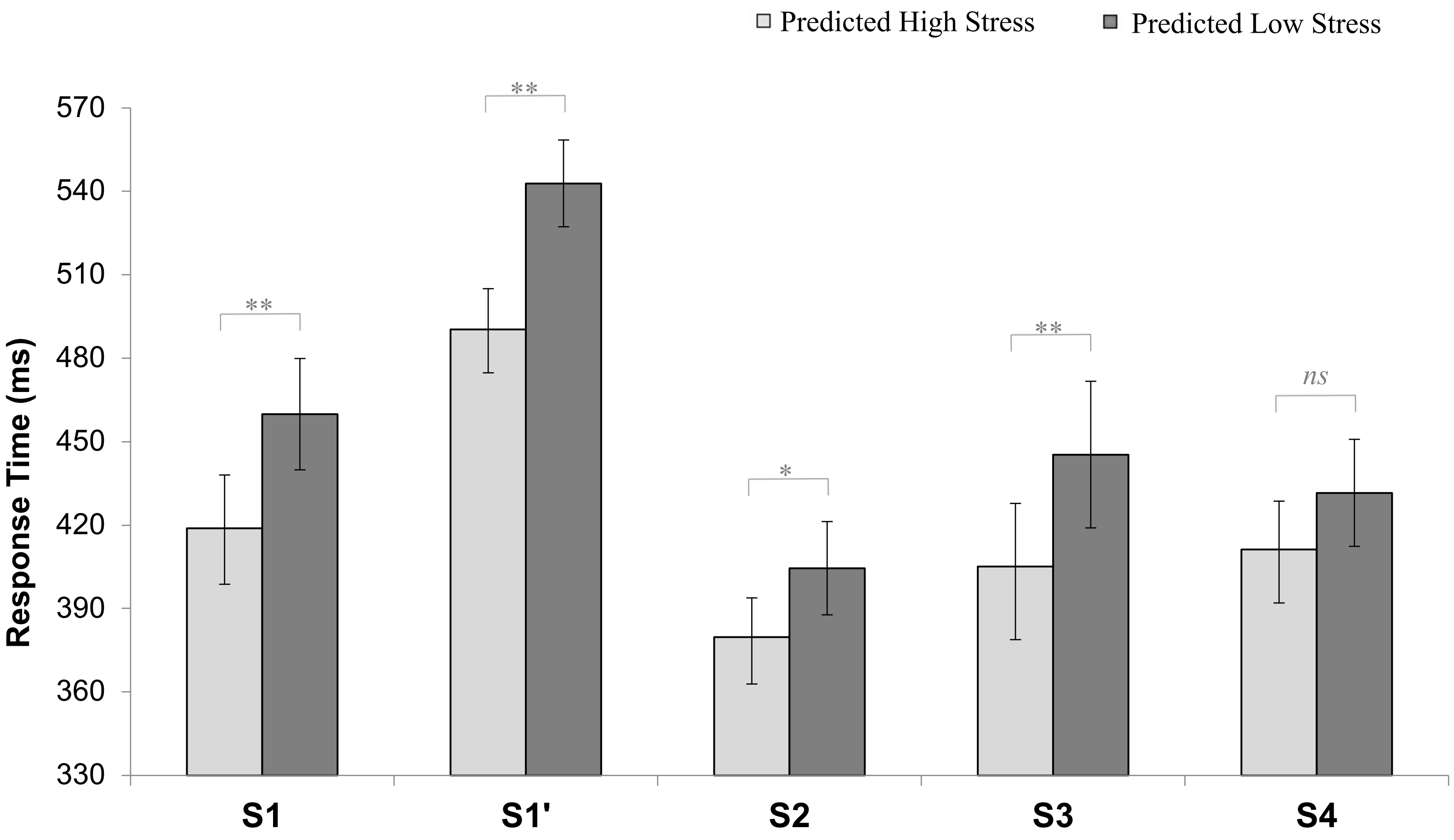


Figure 2. Response time (in ms) as a function of prosodically predicted high versus low stress contexts across different speaker-specific listening conditions in Experiment 1 (S1: with significant acoustic differences in speech rate, pre-target interval duration, mean F_0, maximum F_0, F_0, range, mean intensity, and maximum intensity), Experiment 2 (S1': significant differences in all the preceding cues from S1 except pre-target interval), Experiment 3 (S2: significant differences in mean F_0, mean intensity, maximum intensity, and pre-target interval duration), Experiment 4 (S3: significant difference only in the pre-target intervals), and Experiment 5 (S4: with only significant difference in speech rate). Error bars indicate standard error of the mean. *$*p \leq .05$*, *$**p \leq .01$, $***p \leq .001$* (two-tailed).

Table 2.
Number of detection misses in Experiments 1 to 5.

Experiment	Number of Detection Misses		
	Predicted High Stress	Predicted Low Stress	*p*
Experiment 1: S1 (*N* = 23)	1	5	.219
Experiment 2: S1' (*N* = 22)	4	13	.049*
Experiment 3: S2 (*N* = 23)	1	0	-
Experiment 4: S3 (*N* = 23)	0	2	.500
Experiment 5: S5 (*N* = 22)	4	4	1.000

**p* ≤ *.05* (two-tailed).

Table 3.
Results from the linear mixed-effect models for the results in Experiments 1 to 5 (based on values mapped on the intercept). Baseline model included subject, item, and experimental versions as starting point.

Experiment	Fixed Effect for Predicted High vs. Low Stress		
	β	SE (β)	t
Experiment 1: S1 (*N* = 23)	39.58	12.20	3.24**
Experiment 2: S1' (*N* = 22)	55.40	13.42	4.13***
Experiment 3: S2 (*N* = 23)	25.86	9.64	2.68**
Experiment 4: S3 (*N* = 23)	40.80	13.09	3.12**
Experiment 5: S5 (*N* = 22)	20.28	16.14	1.26

****p ≤ .001, **p ≤ .01, *p ≤ .05* (two-tailed).

However, acoustic analyses of the preceding portions that were one to two syllables (around an average of 350 milliseconds) before the target onset revealed significant prosodic differences in both overall duration, $t(23) = 6.26, p < .001$, and some of the F_0 cues, including mean F_0, $t(23) = 3.08, p = .005$, and maximum F_0, $t(23) = 2.36, p = .027$.

4.6.2. *Results*

Response time and accuracy. One RT datapoint over 2500 milliseconds from the predicted low stress contexts was excluded. In contrast to the previous experiments, there was no significant RT difference between the predicted high ($M = 411.26, SD = 81.47$) versus low stress sentences ($M = 431.62, SD = 90.45$), $t(21) = 0.97, p = .346$. Consistent with these results, our LME results demonstrate no significant fixed effect for predicted stress. For accuracy, the predicted high and low stress conditions each had four misses.

4.7. Cross-speaker Comparisons

4.7.1. *Cross-speaker Differences in On-Focus Production*

Although the actual focused and unfocused target words were replaced by acoustically neutral words in the experiment, it is still helpful to know whether the four female speakers from Experiments 1, 3, and 4 also varied in their production differences between the actual focused and unfocused words from the predicted high and low stress contexts. A series of 2-way (Focus level × Speaker) mixed-model ANOVA was conducted on seven prosodic parameters (word duration, mean F_0, maximum F_0, F_0 range, mean intensity, maximum intensity, intensity range). All speakers produced the focused target words in the same way, with significant increases in duration, mean and maximum F_0, F_0 range, and mean and maximum intensity and intensity ranges (see Tables 3 and 4). However, there were also speaker differences in the degree of production increase on all parameters except for intensity range (see Figures 3 and 4).

Table 4. *Duration (in ms) and F_0 (mean, maximum, range, in Hz) of the actual target-bearing words in focused (high stress context) versus unfocused (low stress context) positions*.

Stimuli	Mean Prosodic Variables (SD) [Range]							
	Duration		Mean F_0		Maximum F_0		F_0 Range	
	Focus	Unfocus	Focus	Unfocus	Focus	Unfocus	Focus	Unfocus
S1 (Experiment 1)	534.58*** (100.62) [328-765]	351.33 (81.68) [193-521]	233.98*** (48.84) [164-317]	197.17 (11.01) [151-197]	316.42*** (84.27) [205-497]	207.20 (74.28) [167-461]	160.55*** (69.95) [61-324]	54.28 (75.66) [13-304]
S2 (Experiment 3)	445.33*** (80.32) [305-604]	336.25 (81.77) [192-540]	230.65*** (19.31) [197-289]	171.95 (10.73) [159-196]	288.81*** (28.67) [251-381]	206.67 (36.06) [173-359]	129.60*** (44.07) [54-210]	63.30 (43.29) [20-199]
S3 (Experiment 4)	455.33*** (79.00) [304-615]	310.17 (78.35) [182-477]	213.38*** (18.23) [181-251]	174.26 (21.24) [108-195]	256.58*** (15.24) [227-293]	209.81 (16.27) [174-240]	98.08*** (41.42) [35-190]	59.16 (35.92) [9-132]
S4 (Experiment 3)	531.38*** (78.29) [398-668]	362.54 (88.49) [154-496]	237.97*** (23.58) [190-278]	177.20 (18.12) [151-224]	310.98*** (35.30) [248-406]	203.65 (17.81) [177-249]	157.18*** (48.38) [77-267]	51.05 (19.89) [20-96]

***$p \leq .001$ (two-tailed).

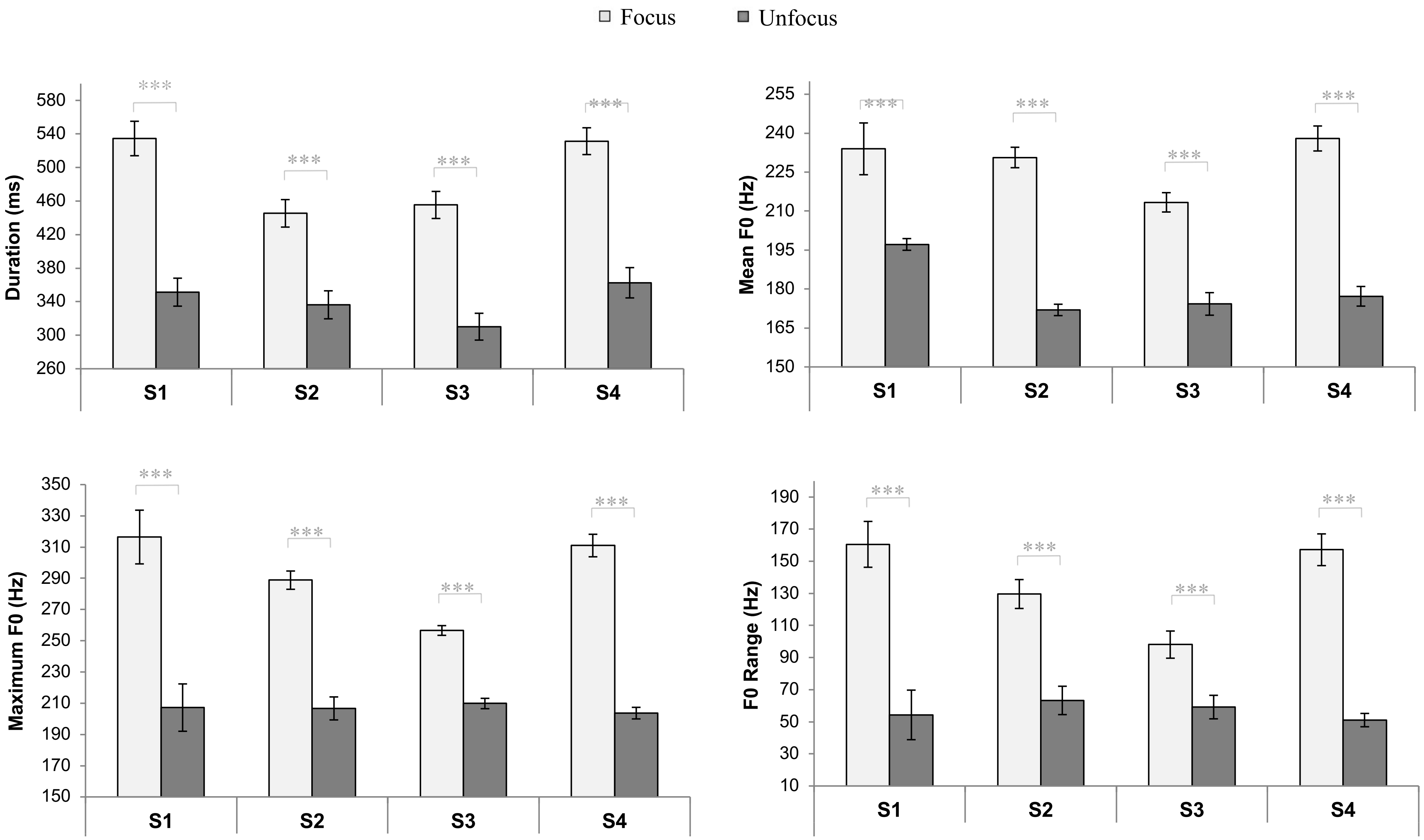


Figure 3. Prosodic realisations of the actual target-bearing words as a function of focused (high stress context) versus unfocused (low stress context) positions in S1 (Experiments 1 and 2), S2 (Experiment 3), S3 (Experiment 4), and S4 (Experiment 5), measured on duration (top left), mean F_0 (top right), maximum F_0 (bottom left), and F_0 range (bottom right). Error bars indicate standard error of the mean.
****p ≤ .001* (two-tailed).

Table 5. *Intensity (mean, maximum, and range, in RMS) of the actual target-bearing words in focused (high stress context) versus unfocused (low stress context) positions*.

Stimuli	Mean Prosodic Variables (SD) [Range]					
	Mean Intensity		Maximum Intensity		Intensity Range	
	Focus	Unfocus	Focus	Unfocus	Focus	Unfocus
S1 (Experiment 1)	57.20*** (3.33) [51-63]	51.42 (2.47) [47-55]	63.72*** (3.29) [58-68]	56.19 (2.43) [52-60]	32.57*** (6.62) [22-47]	26.66 (4.03) [20-37]
S2 (Experiment 3)	58.68*** (2.35) [53-64]	55.41 (2.17) [51-60]	64.58*** (3.46) [60-73]	60.41 (2.56) [56-67]	28.37** (8.75) [13-47]	23.92 (7.17) [13-41]
S3 (Experiment 4)	52.14*** (1.81) [48-56]	46.92 (1.95) [44-52]	57.91*** (1.65) [55-61]	51.78 (1.87) [49-57]	29.50** (9.49) [16-49]	24.45 (6.79) [15-37]
S4 (Experiment 5)	51.40*** (2.54) [46-58]	47.56 (2.03) [45-51]	57.11*** (2.43) [52-63]	51.96 (2.34) [47-57]	27.78*** (7.01) [17-40]	21.74 (6.15) [4-33]

***p ≤ .01, ***p ≤ .001* (two-tailed).

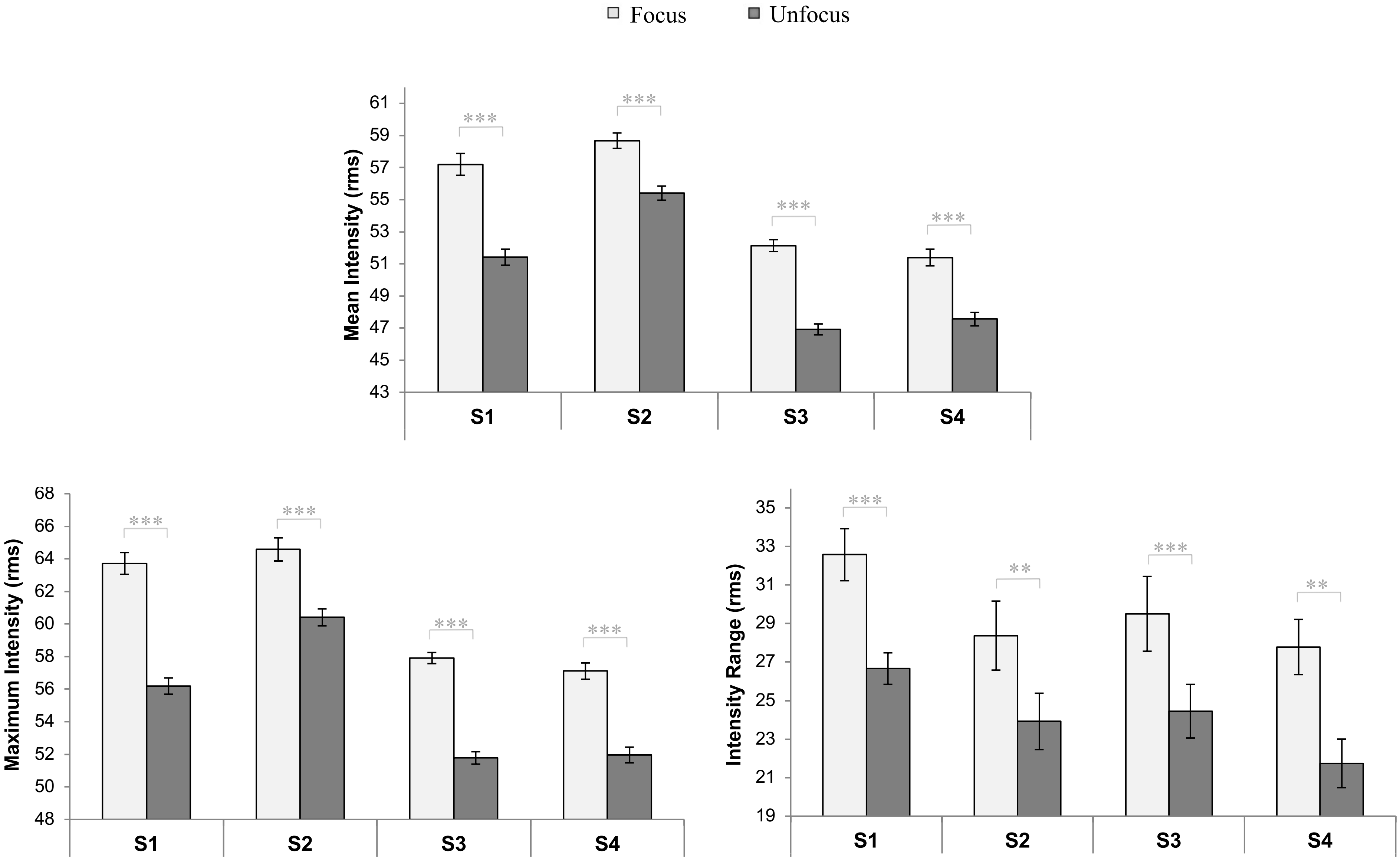


Figure 4. Prosodic realisations of the actual target-bearing words as a function of focused (high stress context) versus unfocused (low stress context) positions in S1 (Experiments 1 and 2), S2 (Experiment 3), S3 (Experiment 4), and S4 (Experiment 5), measured on mean intensity (top), maximum intensity (bottom left), and intensity range (bottom right). Error bars indicate standard error of the mean.
***p ≤ .01, ***p ≤ .001* (two-tailed).

4.7.2. *Cross-speaker Differences in Preceding Prosody*

We conducted a series of 2-way (Prosodic context X Speaker) mixed-model ANOVAs on all the prosodic parameters. This was to examine whether the magnitude of the prosodic differences in the preceding prosody of high and low stress context sentences were different across the four speakers. In other words, we looked at the whether the analyses revealed any significant interactions indicating that the four speakers differed in the degree to which they produce the different preceding prosodic cues. Post-hoc pairwise comparisons were followed up with Bonferroni adjustments. See Tables and Figures 5 and 6 for the values of each preceding feature (averaged across all sentence items) from all speakers.

Speech rate. For speech rate (measured as overall duration of the preceding region), analyses revealed a significant a significant interaction between speaker and prosodic context, $F(3, 92) = 6.21$, $p = .001$, partial Eta-squared = .17. Simple effects of speaker revealed that the increase in speech rate in the recorded sentences by S4 Experiment 5 was significantly longer than all of the recorded sentences made by the other speakers in Experiments 1, 3, and 4.

Pre-target interval duration. We excluded the data from S4 from Experiment 5, since this speaker did not produced any significant difference on the parameter. Results of a mixed 2 (Prosodic context: high versus low) X 3 (Speaker: S1 versus S2 versus S3) demonstrate a non-significant interaction, $F(1, 69) = 2.68$, $p = .075$.

F_0. We only compared the prosodic differences produced by the speaker S1 from Experiments 1 and 2 and S2 from Experiment 3, since they were the only speaker to have displayed prosodic differences in F_0 (for S1, mean, maximum, and range; for S2, mean). For mean F_0, there was a significant interaction, $F(1, 46) = 4.82$, $p = .033$, partial Eta-squared = .10, such that S2 produced a greater increase.

Intensity. Comparing across the two speakers who only showed a significant difference (i.e., S1 from Experiments 1 and 2 and S2 from Experiment 3), we revealed no significant interaction in mean intensity, $F(1, 46) = .13$, $p = .718$. There was also no significant interaction for maximum intensity after comparing the production increases across the three speakers who showed a significant difference (S1 from Experiments 1 and 2, S2 from Experiment 3, and S4 from Experiment 5).

Summary. The mixed ANOVAs suggest that the individual speakers can vary in the degree to which they produce the different preceding cues. For overall duration, our analyses suggest that speaker S4 who recorded the stimuli in Experiment 5 produced slower speech rates in the preceding prosody than any of the speakers from the other experiments. For pre-target interval duration, only the speakers from Experiments 1 to 4 (S1, S2, S3) produced a significantly greater increase for high stress context, and these speakers all did it to the same extent. For the F_0 measures, only S1 and S2 (from Experiment 3) produced greater production increases in mean and maximum F_0. The mean F_0 increase was significantly greater in S2. As for F_0 range, only S1 from Experiments 1 and 2 produced a significantly greater increase. Finally, for the intensity measures, both S1 and S2 produced the same degree of increase in mean intensity. There were also no differences in production increases for maximum intensity across the three speakers who showed the increase (S1, S2, and S4).

Table 6.

Preceding prosody F_0 (mean, maximum, and range in Hz) and duration (in ms) three or four syllables before target onset in predicted high versus low stress contexts across the four speakers in Experiments 1 to 5.

Stimuli	Mean Prosodic Variables (SD) [Range]									
	Overall Duration		Pre-target Duration		Mean F_0		Maximum F_0		F_0 Range	
	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress
S1 (Experiment 1)	585.04* (159.22) [385-1000]	553.58 (142.91) [317-940]	74.35*** (10.91) [55-95]	61.71 (13.91) [33-89]	180.84* (15.43) [161-223]	176.11 (14.60) [154-201]	213.97*** (22.57) [175-286]	203.25 (25.99) [165-255]	58.38*** (20.08) [19-100]	44.67 (20.02) [17-90]
S2 (Experiment 3)	551.08 (161.80) [352-1084]	563.83 (161.17) [359-1108]	75.08*** (16.00) [38-112]	62.20 (16.04) [35-107]	187.68** (15.75) [163-221]	172.37 (11.42) [159-203]	217.29 (24.21) [175-272]	203.48 (25.50) [165-203]	49.61 (19.21) [19-83]	47.83 (18.46) [21-92]
S3 (Experiment 4)	589.38 (15.61) [37-98]	595.54 (15.03) [33-101]	84.57*** (15.24) [57-116]	62.29 (13.31) [27-84]	190.62 (19.10) [172-266]	187.05 (9.38) [168-208]	216.44 (19.04) [179-249]	221.91 (21.33) [178-268]	48.11 (17.45) [11-82]	49.99 (20.09) [11-90]
S4 (Experiment 3)	749.71*** (213.47) [414-1403]	707.17 (226.96) [361-1418]	69.91 (12.17) [48-108]	68.27 (13.11) [29-91]	188.08 (9.49) [174-210]	184.26 (10.03) [171-207]	226.07 (20.10) [193-281]	218.20 (16.38) [184-252]	66.23 (19.47) [35-123]	60.12 (18.70) [23-92]

$*p \leq .05$, $**p \leq .01$, $***p \leq .001$ (two-tailed).

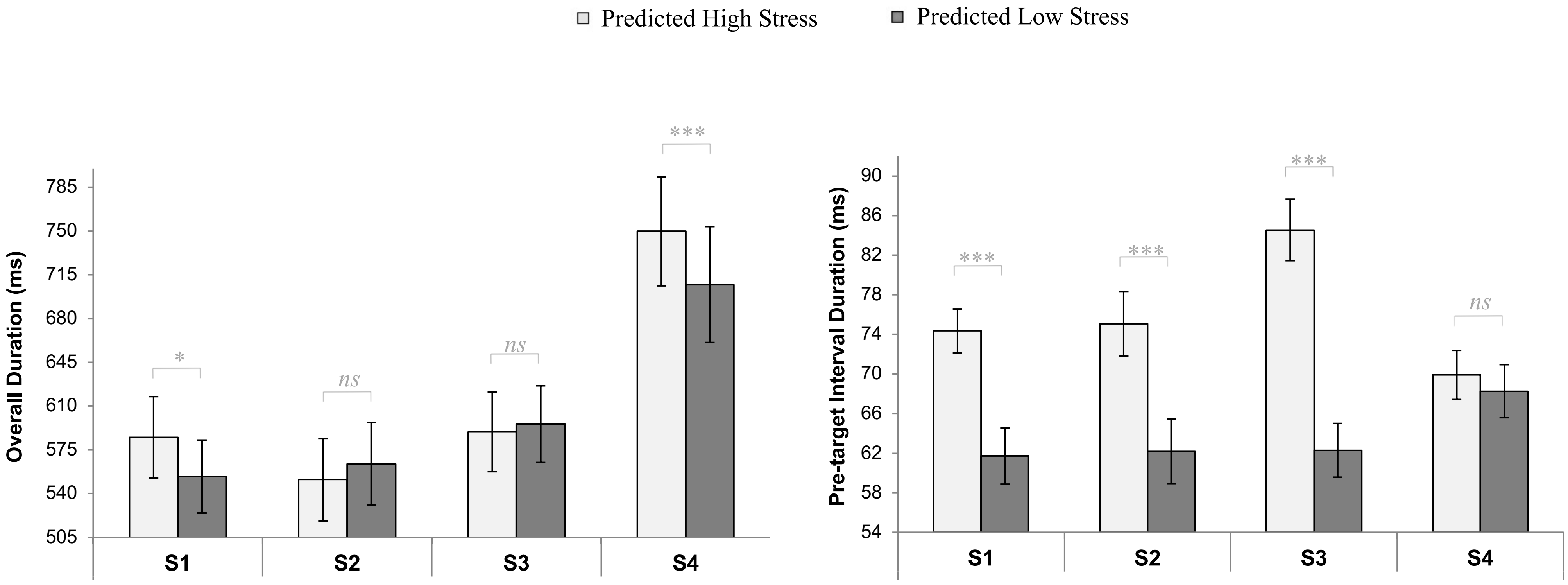


Figure 5. Preceding duration cues (in ms) as a function of predicted high versus low stress contexts in across the four speakers. Error bars indicate standard error of the mean. **p ≤ .05, ***p ≤ .001* (two-tailed).

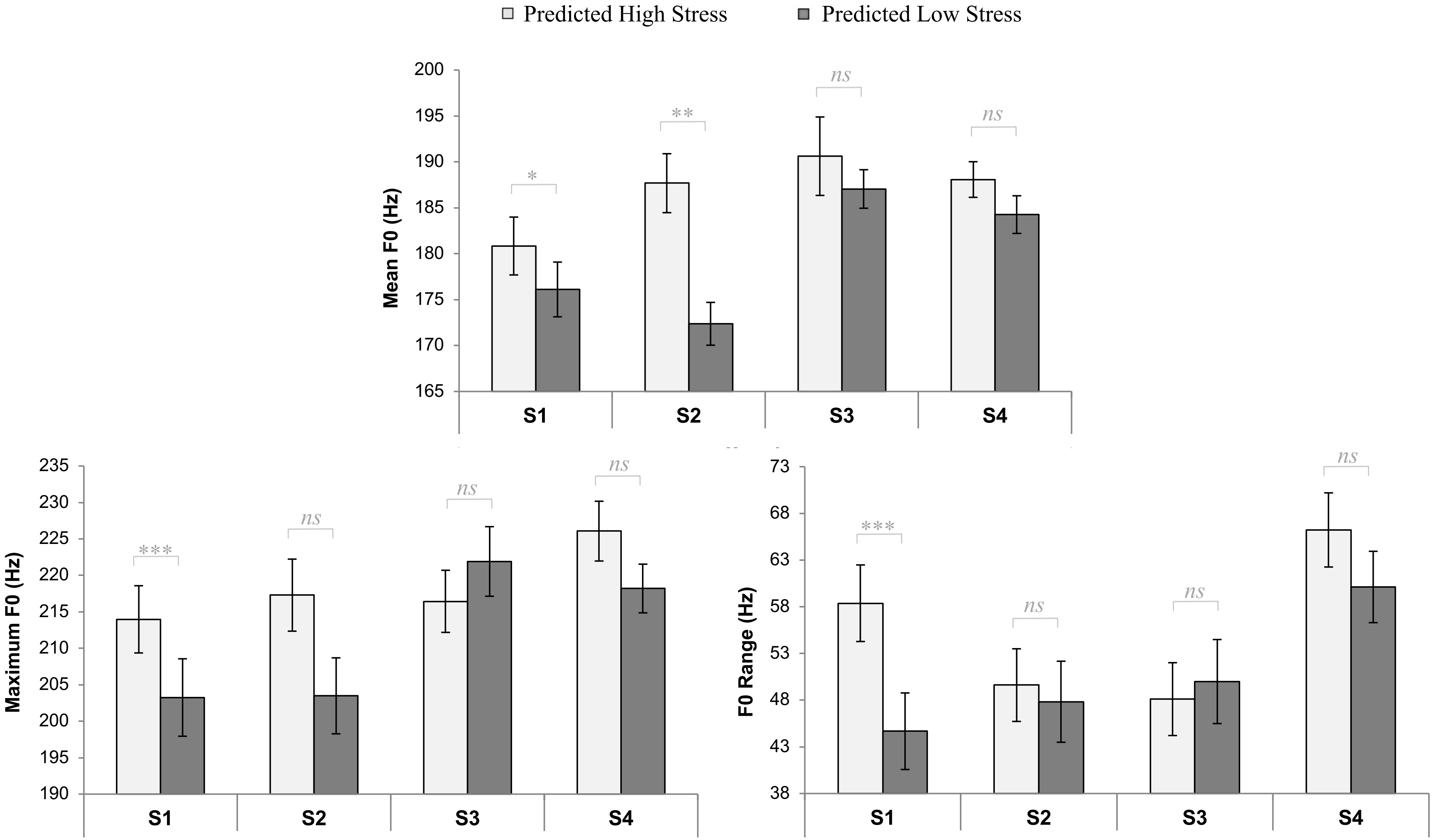


Figure 6. Preceding F_0 cues (in Hertz) as a function of predicted high versus low stress contexts in across the four speakers. Error bars indicate standard error of the mean. **p ≤ .05, **p ≤ .01, ***p ≤ .001* (two-tailed).

Table 7.
Preceding prosody intensity (mean, maximum, and range in RMS) three or four syllables before target onset in predicted high versus low stress contexts.

Stimuli	Mean Prosodic Variables (SD) [Range]					
	Mean Intensity		Maximum Intensity		Intensity Range	
	High Stress	Low Stress	High Stress	Low Stress	High Stress	Low Stress
S1 (Experiment 1)	53.63*** (2.09) [50-58]	52.46 (1.99) [48-56]	59.03*** (1.88) [56-62]	57.32 (1.97) [53-62]	26.94 (7.17) [19-41]	25.63 (6.03) [14-40]
S2 (Experiment 3)	55.90*** (1.37) [54-58]	54.59 (1.31) [53-57]	60.87*** (1.63) [57-63]	59.15 (1.35) [57-62]	28.12 (9.10) [14-51]	28.14 (8.10) [15-45]
S3 (Experiment 4)	48.24 (1.82) [45-52]	48.37 (1.69) [45-52]	53.41 (2.51) [49-58]	53.44 (2.42) [49-58]	29.98 (7.54) [16-44]	29.47 (6.98) [17-41]
S4 (Experiment 5)	50.15 (1.82) [47-55]	49.48 (1.99) [16-54]	55.17** (1.63) [53-58]	53.91 (2.03) [49-58]	24.68 (7.44) [12-39]	24.03 (7.62) [10-41]

***p* ≤ *.01*, ****p* ≤ *.001* (two-tailed).

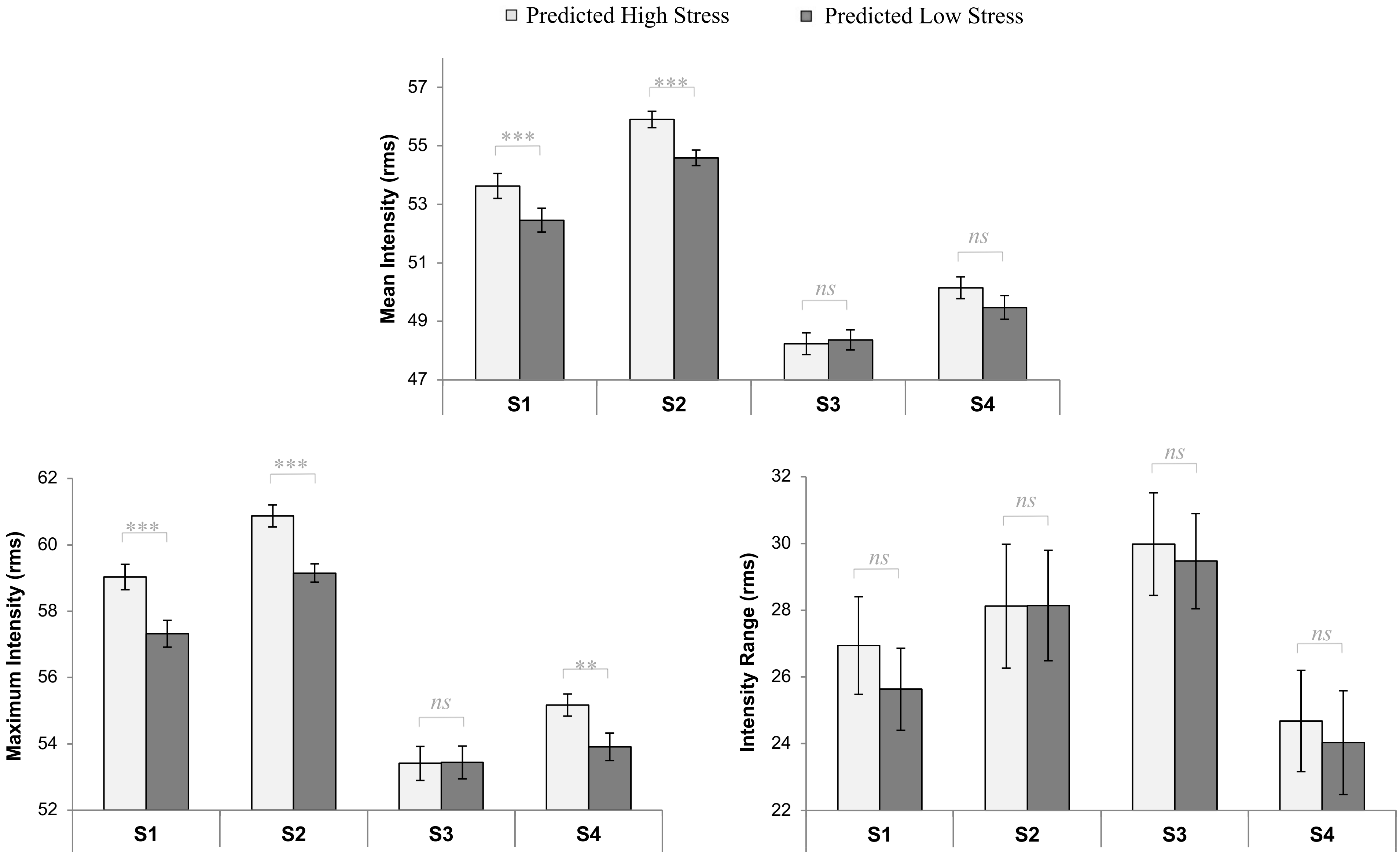


Figure 7. Preceding intensity cues (in root mean square) as a function of predicted high versus low stress contexts in across the four speakers. Error bars indicate standard error of the mean. ***p ≤ .01*, ****p ≤ .001* (two-tailed).

Table 8. *Mean, standard deviation, and range of response time difference (high stress context minus low stress context) across trials (divided into Times 1 to 4) in Experiments 1 to 5.*

	Mean Response Time Difference (SD) [Range]			
Experiment	Time 1	Time 2	Time 3	Time 4
1 (S1)	53.50 (81.94) [-97-210]	27.86 (163.04) [-268-587]	50.41 (107.16) [-77-460]	25.82 (117.61) [-202-302]
2 (S1')	108.11 (147.03) [-134-519]	42.86 (189.50) [-217-749]	48.34 (160.43) [-244-364]	-12.67 (110.68) [-197-199]
3 (S2)	31.45 (108.50) [-142-399]	17.21 (75.32) [-110-147]	2.75 (86.24) [-200-228]	36.55 (89.26) [-154-215]
4 (S3)	12.40 (76.01) [-179-106]	93.97 (165.65) [-113-170]	30.21 (75.23) [-134-170]	46.63 (97.96) [-182-253]
5 (S4)	44.09 (145.56) [-179-486]	21.46 (147.32) [-233-335]	22.65 (96.57) [-147-334]	52.89 (248.05) [-723-682]

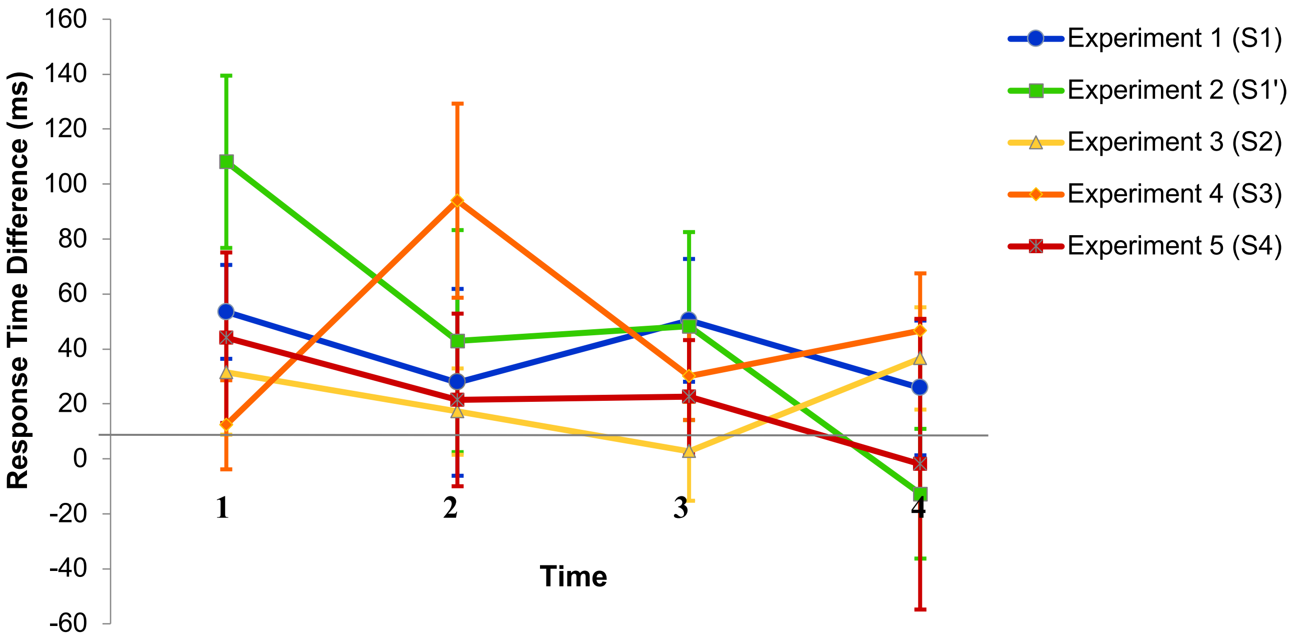


Figure 8. Response time difference (low stress context minus high stress context) across trials (divided into Times 1 to 4) in Experiments 1 to 5. Error bars indicate standard error of the mean.

4.7.3. *Response Time Differences*

We ran a 2-way 2 (Prosodic context: high versus low stress) X 4 (Experiments 1 to 4) mixed-model ANOVA on RT to reveal whether there were any differences in the degree to which participants differ in their RT across the experiments where they have shown to have significantly different RT between high and low stress contexts. In our analyses, we excluded the data from Experiment 5, where there was no significant RT difference between the high and low stress contexts. Results did not show any significant interaction across Experiments 1-4, $F(3, 87) = .90$, $p = .446$, partial Eta-squared = .03.

Response time across trials. Given the similarities in RT across our experiments, we also examined whether there were differences in the pattern of listeners' RT across the 24 sentence trials. RT differences (RT in high stress contexts minus RT in low stress contexts) were divided into and averaged across four separate time sections, with each section containing 6 sentence trials. We conducted a 2 (Experiment) X 4 (Time) mixed ANOVAs on RT difference as a dependent variable to determine whether there was any variation across Experiments 1 to 5 in listeners' response time over the course of the experimental trials. Analyses did not reveal any significant main effect of Time or Experiment. There was also no significant interaction between time and the five experiments (see Table 7 and Figure 8).

4.7.4. *Correlational Analyses*

We conducted a series of Pearson's two-tailed correlational analyses to examine whether there was a link between listeners' RT difference between high and low stress contexts and the degree to which the speaker produced the relevant preceding cues on each sentence item. For each experiment, we calculated, for each sentence item (n = 24), the proportional difference (i.e., percentage change) in RT (averaged across all participants) between predicted high and low stress contexts. We also calculated the proportional difference in the value of each preceding cue produced by the speaker. These were speech rate (i.e., overall duration), pre-target interval duration, mean and maximum F_0, F_0 range, mean and maximum intensity, and intensity range (see Tables 8 to 12 for the correlation results in each experiment).

Across all the five experiments, there was only one significant positive correlation between RT and a preceding cue. This was in Experiment 3 (S2), where there was a positive correlation between the proportional degree of increase in maximum intensity by S2 and the listeners' RT difference ($r = .491$, $p = .015$).

However, across all experiments, there were several significant correlations between various preceding features (see Tables 8 to 12). In Experiment 1 (S1), there were significant correlations between various F_0 cues. The proportional difference in maximum F_0 was positively correlated with mean F_0 ($r = .613$, $p = .001$) and F_0 range ($r = .502$, $p = .012$). Similarly, mean intensity was significantly correlated with maximum intensity ($r = .781$, $p < .001$). However, we observed that some features could also correlate with features from other prosodic dimension. This was seen in intensity range, which was positively correlated with speech rate ($r = .607$, $p = .002$) and negatively correlated with pre-target interval duration ($r = -.472$, $p = .020$).

In Experiment 3 (S2), mean F_0 was positively correlated with maximum F_0 ($r = .897$, $p < .001$), F_0 range ($r = .737$, $p < .001$), as well as mean intensity ($r = .590$, $p = .002$) and maximum intensity ($r = .446$, $p = .029$). Maximum F_0 was positively correlated with F_0 range ($r = .792$, $p < .001$) and mean intensity ($r = .533$, $p = .007$). There was also a positive association between F_0 range and mean intensity ($r = .431$, $p = .035$). Finally, maximum intensity was correlated with mean intensity ($r = .834$, $p < .001$) and intensity range ($r = .550$, $p = .005$).

Table 9. *Pearson's r correlations between proportional differences (% change) in response time (RT) and preceding prosodic cues in Experiment 1 (S1). *p ≤ .05, ***p ≤ .001* (two-tailed).

		RT	Speech Rate	Pretarget	Mean F_0	Max F_0	F_0 Range	Mean Intensity	Max Intensity	Intensity Range
RT	*r*	-	.186	.099	.010	.142	.249	.097	-.068	.261
	p	-	.383	.646	.963	.508	.240	.654	.751	.217
Speech Rate	*r*	-	-	-.074	-.273	-.215	.068	-.403	-.278	**.607****
	p	-	-	.733	.197	.313	.754	.051	.189	**.002**
Pretarget	*r*	-	-	-	-.111	-.327	-.379	-.166	-.295	**-.472***
	p	-	-	-	.607	.119	.068	.438	.161	**.020**
Mean F_0	*r*	-	-	-	-	**.613*****	-.142	.012	-.299	.015
	p	-	-	-	-	**.001**	.509	.956	.156	.944
Max F_0	*r*	-	-	-	-	-	**.502***	.145	.119	.094
	p	-	-	-	-	-	**.012**	.499	.581	.661
F_0 Range	*r*	-	-	-	-	-	-	.106	.370	.293
	p	-	-	-	-	-	-	.623	.075	.164
Mean Intensity	*r*	-	-	-	-	-	-	-	**.781*****	-.026
	p	-	-	-	-	-	-	-	**.000**	.903
Max Intensity	*r*	-	-	-	-	-	-	-	-	.161
	p	-	-	-	-	-	-	-	-	.452
Intensity Range	*r*	-	-	-	-	-	-	-	-	-
	p	-	-	-	-	-	-	-	-	-

Table 10. *Pearson's r correlations between proportional differences (% change) in response time (RT) and preceding prosodic cues in Experiment 2 (S1'). Note the same preceding cues (except pre-target interval duration) from Experiment 1 (S1).* *p ≤ .05, **p ≤ .01, ***p ≤ .001 (two-tailed).

		RT	Speech Rate	Pretarget	Mean F$_0$	Max F$_0$	F$_0$ Range	Mean Intensity	Max Intensity	Intensity Range
RT	*r*	-	.149	-	.081	.247	-.057	-.232	-.126	.145
	p	-	.487	-	.706	.244	.792	.274	.557	.498
Speech Rate	*r*	-	-	-	-.273	-.215	.068	*-.403*	-.278	**.607****
	p	-	-	-	.197	.313	.754	*.051*	.189	**.002**
Pre-target	*r*	-	-	-	-	-	-	-	-	-
	p	-	-	-	-	-	-	-	-	-
Mean F$_0$	*r*	-	-	-	-	**.613*****	-.142	.012	-.299	.015
	p	-	-	-	-	**.001**	.509	.956	.156	.944
Max F$_0$	*r*	-	-	-	-	-	**.502***	.145	.119	.094
	p	-	-	-	-	-	**.012**	.499	.581	.661
F$_0$ Range	*r*	-	-	-	-	-	-	.106	.370	.293
	p	-	-	-	-	-	-	.623	.075	.164
Mean Intensity	*r*	-	-	-	-	-	-	-	**.781*****	-.026
	p	-	-	-	-	-	-	-	**.000**	.903
Max Intensity	*r*	-	-	-	-	-	-	-	-	.161
	p	-	-	-	-	-	-	-	-	.452
Intensity Range	*r*	-	-	-	-	-	-	-	-	-
	p	-	-	-	-	-	-	-	-	-

Table 11. *Pearson's r correlations between proportional differences (% change) in response time (RT) and preceding prosodic cues in Experiment 3 (S2). *p ≤ .05, **p ≤ .01, ***p ≤ .001* (two-tailed).

		RT	**Speech Rate**	**Pretarget**	**Mean F_0**	**Max F_0**	**F_0 Range**	**Mean Intensity**	**Max Intensity**	**Intensity Range**
RT	*r*	-	.049	.298	.040	-.222	-.053	.303	**.491***	.240
	p	-	.819	.158	.852	.297	.807	.150	**.015**	.258
Speech Rate	*r*	-	-	-.355	-.210	-.184	-.003	-.318	-.104	-.071
	p	-	-	.089	.324	.389	.988	.130	.630	.743
Pretarget	*r*	-	-	-	.252	.181	.148	**.404***	.350	.118
	p	-	-	-	.234	.398	.491	**.050**	.094	.584
Mean F_0	*r*	-	-	-	-	**.897*****	**.737*****	**.590****	**.446***	.153
	p	-	-	-	-	**.000**	**.000**	**.002**	**.029**	.475
Max F_0	*r*	-	-	-	-	-	**.792*****	**.533****	.369	.176
	p	-	-	-	-	-	**.000**	**.007**	.076	.410
F_0 Range	*r*	-	-	-	-	-	-	**.431***	.292	.220
	p	-	-	-	-	-	-	**.035**	.166	.301
Mean Intensity	*r*	-	-	-	-	-	-	-	**.834*****	.257
	p	-	-	-	-	-	-	-	**.000**	.225
Max Intensity	*r*	-	-	-	-	-	-	-	-	**.550****
	p	-	-	-	-	-	-	-	-	**.005**
Intensity Range	*r*	-	-	-	-	-	-	-	-	-
	p	-	-	-	-	-	-	-	-	-

Table 12. *Pearson's r correlations between proportional differences (% change) in response time (RT) and preceding prosodic cues in Experiment 4 (S3).* *p ≤ .05, ***p ≤ .001 (two-tailed).

		RT	Speech Rate	Pretarget	Mean F_0	Max F_0	F_0 Range	Mean Intensity	Max Intensity	Intensity Range
RT	*r*	-	-.007	.300	-.006	.291	.073	.061	.311	.111
	p	-	.973	.154	.980	.213	.760	.776	.139	.605
Speech Rate	*r*	-	-	.016	-.061	.065	**.451***	.200	.041	-.358
	p	-	-	.942	.799	.784	**.046**	.350	.849	.086
Pretarget	*r*	-	-	-	-.371	.015	.057	-.037	.033	.025
	p	-	-	-	.107	.949	.812	.863	.877	.908
Mean F_0	*r*	-	-	-	-	.058	-.105	.216	.242	-.128
	p	-	-	-	-	.816	.861	.361	.304	.590
Max F_0	*r*	-	-	-	-	-	**.729*****	.182	.067	-.131
	p	-	-	-	-	-	**.000**	.443	.780	.583
F_0 Range	*r*	-	-	-	-	-	-	-.054	-.096	-.047
	p	-	-	-	-	-	-	.821	.687	.844
Mean Intensity	*r*	-	-	-	-	-	-	-	**.792*****	.044
	p	-	-	-	-	-	-	-	**.000**	.838
Max Intensity	*r*	-	-	-	-	-	-	-	-	*.372*
	p	-	-	-	-	-	-	-	-	*.073*
Intensity Range	*r*	-	-	-	-	-	-	-	-	-
	p	-	-	-	-	-	-	-	-	-

Table 13. *Pearson's r correlations between proportional differences (% change) in response time (RT) and preceding prosodic cues in Experiment 5 (S4). *p ≤ .05, **p ≤ .01, ***p ≤ .001* (two-tailed).

		RT	**Speech Rate**	**Pretarget**	**Mean F_0**	**Max F_0**	**F_0 Range**	**Mean Intensity**	**Max Intensity**	**Intensity Range**
RT	*r*	-	.149	.309	.151	.079	-.017	.221	-.224	.009
	p	-	.486	.142	.482	.714	.936	.300	.292	.968
Speech Rate	*r*	-	-	.104	.132	-.059	-.044	-.268	-.108	**.416***
	p	-	-	.628	.539	.785	.838	.205	.617	**.043**
Pretarget	*r*	-	-	-	-.038	-.083	-.132	.295	-.002	.127
	p	-	-	-	.859	.700	.539	.162	.993	.553
Mean F_0	*r*	-	-	-	-	**.891*****	**.752*****	.191	-.110	.254
	p	-	-	-	-	**.000**	**.000**	.371	.610	.231
Max F_0	*r*	-	-	-	-	-	**.885*****	.219	-.219	.258
	p	-	-	-	-	-	**.000**	.304	.304	.224
F_0 Range	*r*	-	-	-	-	-	-	.051	-.243	.216
	p	-	-	-	-	-	-	.813	.252	.311
Mean Intensity	*r*	-	-	-	-	-	-	-	-.008	.119
	p	-	-	-	-	-	-	-	.971	.579
Max Intensity	*r*	-	-	-	-	-	-	-	-	**-.580****
	p	-	-	-	-	-	-	-	-	**.003**
Intensity Range	*r*	-	-	-	-	-	-	-	-	-
	p	-	-	-	-	-	-	-	-	-

In Experiment 4 (S3), mean and maximum intensity were positively correlated ($r = .792$, $p < .001$). There was also a significant correlation between F_0 range and maximum F_0 ($r = .729$, $p < .001$) as well as speech rate ($r = .451$, $p = .046$).

Finally, for Experiment 5 (S4), mean F_0 was positively correlated with maximum F_0 ($r = .891$, $p < .001$) and F_0 range ($r = .752$, $p < .001$). Maximum F_0 was also significantly positively correlated with F_0 range ($r = .885$, $p < .001$) and intensity range was positively correlated with speech rate ($r = .416$, $p = .043$). However, intensity range was also negatively correlated with maximum intensity ($r = -.580$, $p = .003$).

4.8. General Discussion

This series of experiments is, to our knowledge, the first to examine prosodic focus perception from a cross-speaker perspective. It is also the first to look at the relative roles of different preceding cues using unaltered stimuli recordings. Three important findings have emerged. First, we demonstrate that individual speakers within a given language (i.e., Australian English) can vary in the types of prosodic cues by which they signal later information structure. Second, despite the considerable between-speaker differences, our results across Experiments 1 to 4 indicate that listeners are generally flexible in their prosodic entrainment. Third, when a cue is used, it is used to the same extent.

Our results are interesting in light of previous studies of prosodic entrainment in other languages using the same phoneme detection paradigm. For example, Akker and Cutler (2003) found that both speakers of British English and speakers of Dutch could use the preceding prosody to anticipate a prosodically highlighted word. Similarly, previous works from our laboratory (Ip & Cutler, 2016) have also shown that native speakers of English and Mandarin, two languages with very different intonational systems, can differ in their prosodic production, both in the preceding prosody before focus and in the on-focus cues. In extension of these findings, the current experiments show that production differences in preceding prosody not only differ across speakers of different languages, but also across speakers within a language. Moreover, speakers can have different preferences for one type of cues over the others.

In the segmental literature, a great deal of research has focused on how listeners overcome the "lack of invariance" problem in the mapping of acoustic cues to linguistic categories across speakers (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). However, there has been much less research on the role of talker differences in prosodic production. In the production of on-focus prosodic cues, some of the research to date has found that focus production could differ across individuals in the kinds of prosodic features as well as in the number of strategies (e.g., Ouyang & Kaiser, 2015). Part of the reason for these differences could in principle be due to dialectal differences (e.g., Wang, Wang, & Qadir, 2008) or demographic characteristics such as age (e.g., Fouquet, Pisanski, Mathevon, & Reby, 2016), gender (e.g., Clopper & Smiljanic, 2011; Warren, 2005) or sexual orientations (e.g., Waksler, 2001), all of which are socio-indexical markers that listeners can use to process speech (Kleinschmidt, 2018). However, what is rather surprising about our experiments is that the four speakers who recorded our stimuli were very similar in many respects. They were all female speakers of Australian (Sydney) English from the same age group (i.e., in their late 20s or early 30s) with very similar educational background (i.e., all of them were university postdoctoral researchers or postgraduate students), and above all, they used the same prosodic cues to enhance the actual focused target words (even though there were speaker differences in the degree of production increase). Therefore, beyond the prosodic realisation of focused words, our experiments suggest that there can be cross-speaker variation in the production of preceding prosody even when the production of on-focus cues was the same.

Nevertheless, as also revealed in the crosslanguage studies, the present experiments show listeners can entrain to whatever prosodic features that are immediately available in the speech stream. Listeners could attend to the preceding prosody and respond faster to the phoneme target in predicted high stress contexts regardless of whether the speaker produced the preceding prosody with all the relevant duration, F_0, and intensity cues (Experiment 1), or whether the preceding prosody contains all the preceding cues except pre-target interval duration (Experiment 2), only mean F_0 and intensity cues (Experiment 3), or only pre-target interval duration cues (Experiment 4). Interestingly, the degree to which response times differed between predicted high and low stress was the same across all these talker-specific contexts. This is similar to our previous crosslanguage findings where both English and Mandarin listeners entrain to the preceding prosody to the same extent, even though the predicted high stress sentences in Mandarin only showed significantly greater prefocus maximum F_0 and F_0 range (Ip & Cutler, submitted). On this view, prosodic entrainment may be a common strategy that is used both across different speaker-specific contexts as well as across different languages where prosody is used to express information structure.

Our results indicate functional equivalence of different preceding cues in prosodic entrainment for focus detection. These results are also contrary to previous works showing relative importance of different prosodic cues to prominence (e.g., lexical stress) perception. Using unsynthesised speech recordings from talkers who could freely produce any prosodic cues, our experiments show that listeners are flexible in their prosodic entrainment in that they can use other prosodic cues when one or more cues (e.g., F_0 cues or closure duration of the phoneme target) are uninformative or not consistently produced across the sentence trials. Future research could follow up our findings by addressing further questions on the various factors that may affect the degree to which listeners would engage in prosodic entrainment. First, it is still uncertain whether there are also differences within a language variety in individual listeners' preference for one type of prosodic cue over the other. For example, speakers who tend to produce preceding cues using a particular type of prosodic cue may also prefer to use the same cue in perception. This may happen because processing sensitivity to a particular cue in listening may reflect the individual's attention to that cue in production. Therefore, looking at both between-speaker and between-listener differences may provide a unique insight into the link between prosodic production and perception.

Second, despite the cue equivalence of different preceding features in Experiments 1 to 4, it is still an empirical question as to how each of the different types of preceding cues (e.g., F_0, duration) are processed. For example, it is still uncertain whether the listeners in Experiment 3 and 4, where the speaker only consistently used one or a few preceding cues (e.g., only mean F_0 and intensity cues), were processing the preceding prosody in the same way as the listeners in Experiments 1 or 2, where the speaker provided richer and more robust prosodic cues. Even when listeners may be entraining to prosody to the same extent in different speaker contexts, they may still be processing the different types of suprasegmental properties differently.

Contrary to this view is the possibility that all of the prosodic features may be processed in a similar way. For instance, listeners may be attending to all possible cues based on their relative change. In a previous experiment, Gussenhoven and Rietveld (1999) looked at the role of pitch and found that pitch excursion, rather than pitch height, affects perception of prominence. In their experiment, listeners were asked to judge the prominence of pitch peaks in identical pitch contours superimposed on different artificially manipulated voices. The original utterances were recorded by a woman who had a "deep" voice and were manipulated to have either a male's voice or a more high-pitched female voice by having their original formant values multiplied by a factor of less than or more than 1. Even though the pitch peaks were acoustically identical in both the male and high-pitched female voice contexts, listeners rated the pitch peaks in the artificial male voice as more prominent than the pitch peaks in the artificial high-pitched female voice. This suggests that processing of prosodic cues may be based on their estimate of the span expansion in relation to some baseline register. Applying these findings to other prosodic dimensions, it is possible that listeners are also sensitive to preceding pitch as well as duration and intensity range expansion as the speaker reaches the prosodically highlighted part of the utterance.

Yet another possibility is that listeners were attending to a combination of cues, rather than to a single cue. It is important to note that we measured the strength of different cues in each speaker's preceding prosody based on the proportional difference between the predicted low to high stress contexts, which may indicate the degree to which the speaker increased on each prosodic parameter from the low to high stress contexts. Across all the experiments and prosodic parameters, there was only one significant correlation between a preceding cue and response time difference (i.e., maximum intensity in Experiment 3), which indicates that, overall, the listeners in our experiments were unlikely to have used a single cue. However, there was a tendency for many preceding features to covary, even in the speakers who consistently increased one type of prosodic cue in high stress contexts (e.g., as in the case of S3 in Experiment 4, who only showed a significant increase in pre-target interval duration).

One way in which attention to cue combinations may facilitate focus prediction is by extracting the statistical pattern of the prosodic increases across the different preceding features. In past studies, Cutler (1987) found that listeners no longer engaged in prosodic entrainment when the timing information of a sentence was transposed from another sentence, even when the pitch and intensity information remained intact. In this sense, listeners may not be able to use the preceding prosody to anticipate upcoming focus if the different features are in conflict. For example, it may be harder to entrain to the preceding prosody that has very strong pitch range cues but very low average pitch and no duration cues, compared to a sentence where most of the features in the preceding prosody are patterned in the same direction. On this view, listeners may process all the preceding cues together as one whole, in a gestalt manner.

The literature on auditory perception contains some proposals that listeners are sensitive to statistical covariance of different acoustic features. For example, Stilp, Rogers, and Kluender (2010) found that complex sounds can be processed by collapsing independent but highly correlated acoustic features onto a single perceptual dimension. In their study, they exposed participants to highly-controlled auditory stimuli where they manipulated the spectral shape and the attack/decay of the temporal envelope to be nearly perfectly correlated ($r = \pm 0.97$). During the testing phase, participants did a forced-choice (AXB) discrimination task involving sound pairs where the spectral shape and attack/decay features were correlated in the same way as the exposed sounds (Consistent), or where the features were correlated but in the opposite direction to exposure (Orthogonal), or where only one of the features varied in line with the exposed sounds (Single-cue). Results showed that listeners were better at discriminating the Consistent sound pairs with the covarying properties that were in line with the exposed sounds. However, in a subsequent experiment where there was no passive exposure, Stilp and colleagues found that participants could eventually discriminate both the Consistent and Orthogonal sound pairs over the course of the experimental trials, even though they initially discriminated only the Consistent sound pairs.

These experiments indicate that listeners can extract statistical covariance in different acoustic attributes and perceive them as part of one whole. At the same time, after extended exposure, listeners can also eventually process the remaining variance and start to detect the acoustic features that deviate from the previously experienced variance. Applying these findings to prosodic perception, future research could explore whether processing of prosodic focus operates in the same way. Like the artificial auditory stimuli used by Stilp and colleagues (2010), redundancy is a common feature of prosodic focus. In "stress-accent languages" like English (Beckman, 1986), prominent syllables or words can be highlighted by greater spectral balance (e.g., Sluijter & van Heuven, 1996), longer duration (e.g., Turk & Sawusch, 1996), and higher intensity/loudness (e.g., Kochanski et al., 2005), in addition to higher mean and maximum pitch and pitch range expansion (e.g., Breen, Fedorenko, Wagner, & Gibson, 2010; Gussenhoven & Rietveld, 1999). At the same time, there are also phonological cues (e.g., different pitch contours/pitch accents: Pierrehumbert & Hirschberg, 1990; Selkirk 1984). Whether one type of cue is more informative than the other continues to be a subject of debate.

However, there is much less research on the perception of *prefocus* prosodic cues. Like on-focus production, many of the features in the preceding prosody can be redundant, as indicated in our acoustic analyses and significant correlations between different preceding cues. In such cases, listeners may predict the presence of an upcoming focus by learning to efficiently perceive all the covarying features as one whole. This may explain why the listeners in Experiments 1 to 4 all entrain to prosody to the same extent, even when one set of stimuli had more robust and richer preceding cues than the other. At the same time, entrainment performance over the course of the experimental trials may be temporarily hindered when listeners start to detect prosodic features that are not in line with the other features. Building on this view, future research can examine this possibility by developing a more structured set of sentences to look at prosodic entrainment involving different combinations of preceding cues across the trials.

Further research is also needed to examine in more detail why some speakers may produce preceding prosody that is less conducive to prosodic entrainment (e.g., S4 who only consistently produced speech rate cues). This could be due to a number of reasons. First, the lack of entrainment to the preceding prosody could simply be because speech rate is not a strong cue for prosodic entrainment, although this is unlikely, given that speech rate has been found to be highly effective in listeners' anticipation of upcoming word forms (e.g., function words: Dilley & McAuley, 2008; weak syllables: Baese-Berk et al., 2018). A second reason could be, as already mentioned, the lack of covariance in the preceding cues three to four syllables before the target. As revealed in the correlations, there was a significant negative correlation between maximum intensity and intensity range, two parameters that are part of the same prosodic dimension. Finally, it is also noteworthy that the speech rate of S4 tended to be much slower than that of other speakers; the average overall preceding duration in S4's sentences was over 700 milliseconds, compared to an average of around 500 to 600 milliseconds in those of the other speakers. There is probably less need to predict upcoming focus when the speaker's speech rate is relatively slow.

Interestingly, when we also looked at the more local preceding cues (i.e., one or two syllables, or 35 milliseconds, before the target), we found that S4 produced both speech rate cues as well as greater mean and maximum F_0. This may demonstrate that local cues to prosodic entrainment may not be as informative for predicting focus compared to more distal cues. It is also an open question whether this is also the case in Experiment 4 involving the third speaker (S3), who only consistently produced longer pre-target intervals. If local cues to entrainment are unhelpful, as indicated in S4, then the pre-target interval duration in S3 is unlikely to have facilitated the focus anticipation. Future research should further investigate whether a longer pre-target interval before the target phoneme stop may indicate the presence of another cue that was not measured in our analyses (e.g., rhythm or phrasing).

4.8. Conclusion

From the listener's standpoint, holding a conversation can involve two major tasks. First, because utterances tend to be fragmentary and elliptical (Garrod & Pickering, 2004), listeners must predict how the conversation will unfold. Second, all listeners must quickly update their current discourse model whenever the speaker introduces new information or moves to a new topic. Prosodic entrainment can be a useful strategy to overcome these challenges in languages where prosodic focus is related to semantic salience. By entraining to the immediate prosodic contour, listeners can predict where focus will occur in the utterance and get a headstart in navigating the discourse structure. Here, we have demonstrated cue equivalence in this entrainment effect. Even when speakers may differ in the cues they produce, listeners can extract and integrate mostly any type of cue or combinations of cues to search for the most important word in the sentence.

CHAPTER 5

– Prosodic Cues to Juncture –

A version of this chapter will be adapted and submitted to *Laboratory Phonology*: Ip, M. H. K., & Cutler, A. (in prep). Prosodic cues to juncture in production and perception: A crosslanguage perspective.

Portions of the data presented in this chapter were presented at the 9th International Conference on Speech Prosody: Ip, M. H. K., & Cutler, A. (2018). Asymmetric efficiency of juncture perception in L1 and L2. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 289-296). Poznań, Poland: ISCA. doi:10.21437/SpeechProsody.2018-59

5. 0. Abstract

Past research across many languages has identified various prosodic cues that can be used to signal relevant junctures between words and phrases, but very little of this research has been crosslinguistic. Here, we compared how native speakers of languages with different intonation systems (English vs. Mandarin Chinese) use prosody to resolve potential structural ambiguities in both production and perception. Structural ambiguity was manipulated as a function of the location and timing of relevant junctures. Native speakers of English and of Mandarin were asked to resolve sentences that either had an "Early Juncture", where the prosodic juncture occurred earlier in the utterance (e.g., "*He gave her # dog biscuits*"; "*他给她 # 狗饼干*"), or "Late Juncture" sentences, where the juncture occurred later (e.g., "*He gave her dog # biscuits*"; vs. "*他给她狗 # 饼干*". Importantly, the ambiguities used in the study are identical in English and in Mandarin. Our production data show that prosodic disambiguation of this type of ambiguity is realised very similarly in the two languages, but there were crosslanguage differences in the degree to which speakers produced different prosodic juncture cues (e.g., pausing). In our perception experiments, a new disambiguation task was used, requiring speeded responses to select the correct meaning for structurally ambiguous sentences. The perceptual results showed language-specific differences in both disambiguation response time and accuracy. Similar to our production data, there was also crosslanguage variation in perceptual reliance on different prosodic cues to juncture. Finally, listeners' response patterns also differed for native (L1) and non-native (L2) language processing, although there was a significant increase in similarity between the two response patterns with increasing exposure to the L2. Our findings indicate that prosodic cues to juncture may be more language-specific and variable than previously assumed.

– Prosodic Cues to Juncture –

5.1. Introduction

One of the greatest mysteries in the science of language is the human ability to segment continuous streams of speech into meaningful units. While the goal of the speaker is to convey information with as little effort as possible (Zipf, 1949), the listener, on the other hand, is faced with a more challenging task. Language is a complex system where a handful of phonemes and syntactic rules can be recycled to generate a vast repository of words and an infinite range of sentences. Conversational speech is never produced in discrete chunks, and utterances can often convey more than one distinct meaning because words tend to resemble or occur embedded within other words (e.g., "*He gave her son glasses*" vs. "*He gave her sunglasses*"). Yet, virtually all listeners can understand most utterances without much effort. In the face of so much uncertainty, how do listeners access the intended meaning of these ambiguous utterances? To what extent are their processing strategies shared across languages? In what way does language-specific experience affect processing?

The present study will explore these questions by comparing how native speakers of different languages use prosody to resolve structural ambiguity. According to Bolinger (1978), prosody plays a universal role in helping language users signal and detect relevant boundaries in running speech. At the same time, formal language theory suggests that prosody is itself a grammatical structure that can be parsed to the advantage of the listener. It is now widely accepted that prosody is a hierarchical structure where different levels of prosodic constituents, ranging from prosodic words to phonological and intonational phrases, can govern prominence relations and intonational, rhythmic, and pausing patterns across languages (e.g., Beckman, 1996; Ladd, 1986; Liberman & Prince, 1977; Selkirk, 1984; 1986; 2003).

How these constituents are organised in the phonological hierarchy is in large part highly similar across languages (Beckman & Pierrehumbert, 1986). Although prosodic structure is not fully isomorphic with syntactic structure (e.g., Nespor & Vogel, 1986; Price, Shattuck-Hufnagel, & Fong, 1991), all listeners can still reliably identify syntactic boundaries that correspond to prosodic cues from different levels of the hierarchy. This is most evident during early language development when attention to prosodic features coinciding with clause and phrase boundaries can provide a starting tool for infants to learn the syntax of their surrounding language and map auditory word forms onto visual referents (e.g., Gervain & Werker, 2013; Gleitman & Wanner, 1982; Hirsh-Pasek et al., 1987; Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Seidl, 2007; Shukla, White, & Aslin, 2011; Soderstrom, Blossom, Foygel, & Morgan, 2008; Soderstrom, Seidl, Nelson, & Jusczyk, 2003). Over time, these language-general segmentation strategies can also incorporate more language-specific cues for processing smaller prosodic units that coincide with word boundaries and grammatical morphemes (e.g., Demuth & Tremblay, 2007; Demuth, McCullough, & Adamo, 2007; Gout, Christophe, & Morgan, 2004; Johnson & Seidl, 2008; Seidl & Johnson, 2006). In this respect, prosodic cues to juncture can serve a skeletal foundation for integrating different aspects of the speech signal during the early stages of online processing (Frazier, Carlson, & Clifton, 2006).

An extensive body of research across many languages has discovered various ways in which prosody can be used to cue relevant junctures and ultimately disambiguate continuous speech. In the tonal domain, discontinuity across syntactic boundaries can be marked by F_0 change through realisation of specific intonational contours or edge tones, preboundary lowering, and postboundary declination reset (e.g., Danish: Thorsen, 1985; Dutch: Gussenhoven & Rietveld, 1988; Swerts, 1997;

English: Liberman & Pierrehumbert, 1984; Schafer, Speer, Warren, & White, 2000; Streeter, 1978; Watson & Gibson, 2004; French: Vaissière, 1983; Finnish: Hirovenen, 1971; German: Grabe, 1998; Japanese: Beckman & Pierrehumbert, 1986; Kikuyu: Clements & Ford, 1981; Kipare: Herman, 1996; Mandarin: Shih, 2000; Mexican Spanish: Prieto, Shih, & Nibert, 1996; Taiwanese: Peng, 1997; Yorùbá: Laniran, 1992). F_0 changes may also be hierarchically nested across the prosodic structure, for instance, with greater F_0 reset at the utterance level than at phrase- or word-level positions (Fisher & Tokura, 1996; Ladd, 1988; Thorsen, 1985), although this may not be the case in all languages (e.g., French: Michelas & D'Imperio, 2012). In addition, prosodic position in some tone languages (e.g., Taiwanese) can also condition tone sandhi and changes in F_0 range to a greater extent than tonal contexts (Peng, 1994).

In the temporal domain, a great deal of research attention has been accorded to how speakers manipulate pausing and deceleration cues across different prosodic contexts. Apart from the frequent (though optional) prosodic breaks between clauses and phrases (e.g., Cooper & Paccia-Cooper, 1980; Goldman-Eisler, 1972; Grosjean, Grosjean, & Lane, 1979; Krivokapić, 2007), prosodic organisation can also affect the duration of boundary-related segments. For example, vowels in word-initial and word-final syllables tend to be longer (e.g., Dutch: Quené, 1992; English: Beckman & Edwards, 1990; Peterson & Lehiste, 1960), while syllables not adjacent to word boundaries are more likely to be shortened (e.g., Harris & Umeda, 1974; Klatt, 1976; Lehiste, 1972). Likewise, speakers in a variety of languages and regional dialects tend to produce longer segments at phrase-initial and phrase-final positions than at phrase-medial positions (e.g., American English: Cooper & Paccia-Cooper, 1980; Klatt, 1975; Shattuck-Hufnagel & Turk, 1998; Turk & Shattuck-Hufnagel, 2007; British English: Campbell & Isard, 1991; Dutch: Cambier-Langeveld, 1997; Estonian: Krull, 1997; German: Kohler, 1983; Silverman, 1990; Greek: Katsika, 2009; 2016; French: Hirst & Di Cristo, 1984; Tabain, 2003; Hebrew: Berkovits, 1993; Hungarian: Hockey & Zsuzsanna, 1998; Japanese: Takeda, Sagisaka, & Kuwabara, 1989; Mandarin Chinese: Shen, 1993; Swedish: Lindblom & Rapp, 1973; Taiwanese: Peng, 1997). As with F_0 cues, the degree to which these boundary-related segments are lengthened increases cumulatively across the phonological hierarchy, with greater lengthening at higher domains than at lower domains (e.g., Gee & Grosjean, 1984; Michelas & D'Imperio, 2012; Wightmann, Shattuck-Hufnagel, Ostendorf, & Price, 1992).

Finally, prosodic structure can affect the acoustic clarity and articulatory strength of boundary-related segments (for a review see Cho, 2011). Also across a wide range of languages, consonant onsets in boundary-related positions across each level of the prosodic structure are likely to undergo "prosodic strengthening" characterised by spatio-temporal expansion of articulatory gestures and coarticulatory resistance (e.g., Byrd, 2000; Byrd & Saltzman, 2003; Cho, McQueen, & Cox, 2007; Cho, Jun, & Ladefoged, 2002; Cho & Keating, 2001; Fougeron, 2001; 2011; Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2003; Kuzla, Cho & Ernestus, 2007; Onaka, 2003; Redford, Davis, & Miikkulainen, 2004). A similar form of domain-initial strengthening effect can also occur during vowel production, where prosodic position can induce glottalisation, larger lip opening, longer duration, and more enhanced spectral characteristics (Cho, Lee, & Kim, 2011; Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Georgeton & Fougeron, 2014; Georgeton, Antolik, & Fougeron, 2016; Gendrot, Gerdes, & Adda-Decker, 2011; Lehiste, 1960). Prosodic strengthening is therefore a language-universal feature that exists across all types of segments (e.g., affricates: Degenshein & Chitoran, 2001; stops: Kuzla & Ernestus, 2011; Pierrehumbert & Talkin, 1992; fricatives: Kuzla, Cho, Ernestus, 2007; trills:

Spinelli, McQueen, & Cutler, 2003; nasals: Fougeron, 2001; consonant clusters: Cho, Lee, & Kim, 2014; vowel onsets: Georgeton & Fougeron, 2014), but how this strengthening is realised can also depend on language-specific differences in phonetic inventory. For example, 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 (e.g., /d/) undergo VOT shortening to enhance prevoicing (Cho & McQueen, 2005). Similarly, nasals receive greater linguopalatal contact and reduced nasal airflow (i.e., higher velum) in French (Fougeron & Keating, 1996) and slower lip movements and reduced nasal energy in English (Byrd & Saltzman, 1998; Cho & Keating, 2009), but only durational lengthening in Tamil (Byrd, Narayanan, Kaun, & Saltzman, 1997).

These prosodic boundary-related effects can have important implications on the listener's ability to detect relevant junctures in continuous speech. Not only do prosodically conditioned changes facilitate conscious judgements of different boundaries (e.g., Byrd & Saltzman, 1998; Krivokapić & Byrd, 2012), they can also help the listener anticipate whether a particular word begins at the start of a phrase. For example, Cho and colleagues (2007) compared sentences with Prosodic Words or Intonational Phrases (e.g., "*John brought bus tickets for his family*" vs. "*When you get on the bus, tickets must be shown*) and found that listeners can use prosodic position to process sentences. The /ti/ from "*ticket*" in the latter sentence would have undergone consonantal strengthening (i.e., longer VOT) since it was at the start of an Intonational Phrase. Cho and colleagues edited the /ti/ segment out of each sentence and replaced it by the same token coming from another sentence with either the same or different prosodic structure. Using a cross-modal priming paradigm, they found that lexical recognition for "*bus*" was faster if the first syllable of the following word (i.e., /ti/) was strengthened to signal the start of the Intonational Phrase. Likewise, in languages where liaison between words is a common feature (e.g., French), listeners can attend to fine-grained duration of consonants to determine whether a word is at the start or end of a phrase (e.g., /r/ of "*dernier*" in "*dernier oignon*" and "*roignon*" in "*dernier roignon*": Spinelli et al., 2003; see also, Christophe, Peperkamp, Pallier, Block, & Mehler, 2004). These findings demonstrate that domain-initial segmental strengthening serves the same role across languages: to cue prosodic boundaries by enhancing segmental contrasts against competitors. Their effect on perception is therefore the same across languages, even when there is crosslanguage variation in how certain segments are produced.

In addition to these perceptual findings, there is also evidence that listeners can use language-universal prosodic cues even in an unfamiliar language. In an experiment by Carlson, Hirschberg, and Swerts (2005), native speakers of Swedish, American English, and Mandarin Chinese heard single and multi-word fragments of natural Swedish speech extracted from a radio interview and 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 subjects' judgements during both single and multi-word fragments were equally as accurate as those of the Swedish subjects. Likewise, native Mandarin speakers also showed a comparable performance, although it was only in the multiword stimuli. Acoustic analyses of the stimuli revealed that judgement accuracy was correlated with the word's boundary strength in F_0 and glottalisation, and Carlson and colleagues interpreted that participants would have used these cues to segment the unfamiliar speech. A follow-up pilot study using the same Swedish stimuli also found a similar effect in native speakers of Taiwanese, only this time judgement accuracy was correlated with pause duration (Kuo, 2011).

In a similar vein, an investigation by Endress and Hauser (2010) indicates that listeners are even capable of using prosodic cues to process unfamiliar non-native speech with critically different prosodic systems. Endress and Hauser created experimentally manipulated contexts where only prosodic cues were available by using filtered speech where segmental cues were made uninformative. Under these conditions, native adult listeners of American English (a language with mostly word-initial stress) were asked to identify word boundaries in samples of connected speech produced in a foreign language (e.g., Turkish, a language with word-final stress). Listeners could successfully extract words from speech at both the end and middle of intonational phrases even though they had no prior exposure to the test language. As prosody was the only cue that was available, listeners must have employed a universally accessible mechanism that allowed them to use the unfamiliar prosody to organise the speech input at the prelexical level.

However, it is still unclear how universal and language-specific prosodic cues interact in speech segmentation. Language similarities in prosodic juncture effects have been well documented in both the linguistic and psycholinguistic literature, but much less is known about the effect of language differences (Cutler, 2012). Even if there is a common universal substrate that dictates the way we process prosodic junctures (thus, in both a native and an unfamiliar non-native language), this universal substrate may, over the course of development, be gradually shaped by the structure of our mother tongue, leading to prosodic strategies that are particularly optimised for the native language (e.g., Johnson & Seidl, 2008; Seidl & Cristià, 2008; Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle, 2012). On such an account, it is also an empirical question whether listeners can generalise native prosodic strategies to relevant boundaries in a foreign language. Native language strategies may still be used even with some fluency in a second language, because acquisition of second language prosody is a protracted process (Pennington & Ellis, 2000) and learners rarely attain native-like level of prosodic production (Mennen, 2004).

One reason why perception of prosodic junctures may differ could be due to differences in prosodic rhythm. For example, Cutler, Mehler, Norris and Seguí (1983) have revealed that native listeners of English and French can differ in their word segmentation strategies. Across a series of crosslanguage experiments, native listeners of French, but not English, relied on syllabification to locate word boundaries (e.g., response times to French target *bal* was faster when presented with a word *bal-con* that corresponded with the target than a word, *ba-lance*, that did not). On the other hand, native listeners of English would instead segment words based on stress units (e.g., Cutler & Norris, 1988). When processing input from non-native speech, listeners were still sensitive to their native categories, in which case French listeners could still use syllable-timing to their advantage when segmenting the stimulus fragments in English speech. This shows that the specific prosodic structure of the native language can play a crucial role in the listener's detection of critical junctures.

At the same time, however, the different segmentation strategies found in English and French are closely parallel; both stress in English and syllable in French form the foundation for rhythmic structure in each respective language. Listeners may in fact adopt a universally applicable "metrical segmentation strategy" to locate word boundaries by exploiting whatever phonological construct that defines their language (Cutler, 1994; 1996). Other syllable- and stress-timed languages do indeed show the same effect, for instance Korean, Spanish, and Catalan encourage syllabic segmentation (Kim, Davis, & Cutler, 2008; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992) and Dutch encourages stress-based segmentation (Vroomen, Van Zon, & De Gelder, 1996). Most critically, listeners of Tokyo Japanese and Telugu, unrelated languages with rhythmic structures based on neither stress nor syllables, have been found to segment words using the mora (Cutler & Otake, 1999; 2002: Otake, Hatano, Cutler, & Mehler, 1993; Murty, Otake, & Cutler, 2007). Like the domain-initial strengthening data mentioned earlier, these results show that similar segmentation strategies may exist across different listeners, even though the form that it takes depends on the language-specific prosodic system of their native language.

Speakers may also differ in their sensitivity to different prosodic juncture cues. Even if all juncture cues (i.e., pausing, boundary lengthening, F_0 changes, and segmental modification) are universal, languages can still vary in the degree to which these cues are interrelated. For example, major prosodic boundaries (i.e., intonational phrase boundaries) in German are marked by both preboundary lengthening and F_0 reset, and both ERP and behavioural data show that German listeners can only detect these boundaries when pitch and preboundary lengthening cues co-occur (Holzgrefe-Lang et al., 2016). When pause duration was rendered uninformative, German listeners still showed a brain signature associated with phrase boundary detection (i.e., a so-called *Closure Positive Shift*), suggesting that pausing is not an important cue in German (e.g., Steinhauer, Alter, & Friederici, 1999; Männel & Friederici, 2009; Männel, Schipke, & Friederici, 2013). Similar to German speakers, native speakers of English and Russian are also less reliant on pause duration (Aasland & Baum, 2003; Seidl & Cristià, 2008; Volskaya, 2003).

On the other hand, recent studies in Mandarin Chinese revealed better prosodic boundary detection when the stimuli only contained pausing cues compared to when the stimuli only displayed preboundary lengthening and F_0 reset (Yang, Shen, Li, & Yang, 2014). Interestingly, listeners' performance did not differ as a function of whether only pause duration or both pause duration and preboundary lengthening were preserved. Pause duration in Mandarin may therefore be a more powerful cue for boundary detection than preboundary duration or postboundary pitch. Whether the language-specific differences in cue weighting between Mandarin and English and German are due to differences in certain language properties is still an empirical question. These language differences are unlikely to be due to typological distance, since native speakers of Dutch and Swedish also exhibit greater reliance on pause duration (Sanderman & Collier, 1997; Horne, Strangert, & Heldner, 1995; House, Hermes, & Beaugendre, 1998).

5.1.2. *The Present Study: Crosslanguage Production and Perception of Juncture*

The present series of experiment aims to address some shortcomings in previous studies. First, although the literature contains extensive data on prosodic juncture processing involving many languages, the materials used and the prosodic cues in question can differ extensively. This can make it difficult to determine whether the prosodic cues deployed by the language user are language-specific or commonly shared across language. Second, very little research on prosodic juncture processing has been comparative (Cutler, 2012). Even in the handful of recent crosslanguage studies, the structural ambiguity and prosodic cues under investigation are quite different. The languages used for crosslanguage comparison also tend to come from closely related language families with similar prosodic systems (e.g., German and English: O'Brien, Jackson, & Gardner, 2014). Also comparisons of prosodic disambiguation in different languages often involved languages with different prosodic realisation 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).

The experiments we report here, in contrast, compare English and Mandarin in terms of both production and perception. Both languages allow the same kind of structural ambiguity, despite the typological distance and differences in intonation systems. Consider the following examples:

(a) 姥姥 / 給 / 她 # 狗 肉 / 吃
Grandma / gave / her # dog meat / to eat

(b) 姥姥 / 給 / 她 / 狗 # 肉 / 吃
Grandma / gave / her / dog # meat / to eat

The two sentences differ in the direct object, and as a consequence, differ in juncture location. In (a), the juncture (#) is realised 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 "*dog*", so that the personal pronoun becomes a possessive determiner. This ambiguity can occur in English because "*her*" can either be a possessive or an indirect object. It can also occur in Mandarin since speakers ignore alienable versus inalienable distinction in colloquial speech where the possessive particle *-de* can be omitted. Given the identical ambiguous structures, it would be reasonable to expect that speakers in both languages would produce the same prosodic cues to mark the relevant juncture. In perception, both groups of listeners would use prosodic cues to decipher the intended meaning of the ambiguous utterance. However, listeners of different languages may differ in their attention to different boundary cues.

5.2. Production Experiment

5.2.1. *Method*

Participants. We obtained recordings from 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). All of the English speakers reported that they were born and raised in Australia, while the Mandarin speakers were born in Mainland China and had been living in Australia for less than ten years (*M*= 2.84 years; range: 2 months – 9 years). All participants were university students at the time of the experiment. The English speakers were recruited via an undergraduate subject pool and the Mandarin speakers were recruited using advertisements. We excluded additional data from three Mandarin speakers who grew up in Chinese-speaking communities outside of Mainland China (e.g., Taiwan, United States) and from one English-speaking participant who appeared to have disfluency in oral reading. All participants were naïve to the specific purpose of the experiment.

Reading passages. Our materials were three pairs of short reading passages written in English and Simplified Chinese (see Table 1). Each reading 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 contexts provided by the preceding sentences in the passage. In one version, the context provided by the preceding sentences would elicit production of the target sentence with an "Early Juncture", where the boundary occurred earlier in the sentence (e.g., "*He gave her # dog biscuits*"). In another 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"*). Different storylines were used to elicit different timing and location of prosodic juncture (e.g., a vignette

Table 1. *Vignette pairs (Early vs. 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.*

English	Mandarin
"He gave her dog biscuits" /hi: gæɪv hɜ: dɔg bɪskəts/ **Early Juncture: "*He gave her # dog biscuits*"** /hi: gæɪv hɜ: # dɔg bɪskəts/ 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. **Late Juncture: "*He gave her dog # biscuits*"** /hi: gæɪv hɜ: dɔg # bɪskəts/ Adam has just moved to Sydney from Melbourne. His new neighbour is an old lady named Gertrude. Gertrude has have 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.	***"他给她狗饼干"*** /tʰa$_1$ kei$_2$ tʰa$_1$ kou$_3$ pin$_3$kan$_1$/ **Early Juncture: "*他给她 # 狗饼干*"** /tʰa$_1$ kei$_2$ tʰa$_1$ # kou$_3$ pin$_3$kan$_1$/ 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: "*他给她狗 # 饼干*"** /tʰa$_1$ kei$_2$ tʰa$_1$ kou$_3$ # pin$_3$kan$_1$/ 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: so: hɜ: dɐk ɐndɐ ðə tʃeː/ **Early Juncture: "*He saw her # duck under the chair*"** /hi: so: hɜ: # dɐk ɐndɐ ðə tʃeː/ 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. **Late Juncture: "*He saw her duck # under the chair*"** /hi: so: hɜ: dɐk # ɐndɐ ðə tʃeː/ 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.	***"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$_4$/ **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$_4$/ 叶生和玛丽亚是小学同学。他们喜欢玩捉迷藏。叶生喜欢猫在桌子底下。玛丽亚喜欢猫在凳子底下。他们第一次玩捉迷藏是在教室里玩。玛丽亚藏得太慢，叶生很快就发现她藏在哪里。他看见她猫在凳子底下。 **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$_4$/ 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" /hi: gæɪv hɜ: bæɪbɪ mɪlk/ **Early Juncture: "*He gave her # baby milk*"** /hi: gæɪv hɜ: # bæɪbɪ mɪlk/ 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. **Late Juncture: "*He gave her baby # milk*"** /hi: gæɪv hɜ: bæɪbɪ # mɪlk/ 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.	***"他给她婴儿奶粉"*** /tʰa$_1$ kei$_3$ tʰa$_1$ jiŋ$_1$ɚ$_2$ nai$_3$fən$_3$/ **Early Juncture: "*他给她 # 婴儿奶粉*"** /tʰa$_1$ kei$_3$ tʰa$_1$ # jiŋ$_1$ɚ$_2$ nai$_3$fən$_3$/ 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: "*他给她婴儿 # 奶粉*"** /tʰa$_1$ kei$_3$ tʰa$_1$ jiŋ$_1$ɚ$_2$ # nai$_3$fən$_3$/ 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挨饿。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. *Examples of follow-up questions in English and Mandarin Chinese*.

English	Mandarin
“He gave her dog biscuits” /hi: gæɪv hɜ: dɔg bɪskəts/ **Early Juncture: “*He gave her # dog biscuits”*** /hi: gæɪv hɜ: # dɔg bɪskəts/ 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? **Late Juncture: “*He gave her dog # biscuits”*** /hi: gæɪv hɜ: dɔg # bɪskəts/ Questions about Adam and Gertrude’s dog 1.What did Adam give her dog today? 2. Did he give Gertrude any biscuits?	***“他给她狗饼干”*** /t^{h}a$_{1}$ kei$_{2}$ t^{h}a$_{1}$ kou$_{3}$ pin$_{3}$kan$_{1}$/ **Early Juncture: “*他给她 # 狗饼干”*** /t^{h}a$_{1}$ kei$_{2}$ t^{h}a$_{1}$ # kou$_{2}$ pin$_{3}$kan$_{1}$/ 问题关于小周和爱玲 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.　小周有没有给爱玲德狗吃狗饼干？ **Late Juncture: “*他给她狗 # 饼干”*** /t^{h}a$_{1}$ kei$_{2}$ t^{h}a$_{1}$ kou$_{3}$ # pin$_{3}$kan$_{1}$/ 问题关于阿德和老奶奶的狗 1. 阿德今天给她狗什么？ 2. 阿德有没有给老奶奶吃饼干？
“He saw her duck under the chair” /hi: so: hɜ: dɐk undɐ ðə tʃeː/ **Early Juncture: “*He saw her # duck under the chair*”** /hi: so: hɜ: # dɐk undɐ ðə tʃeː/ Questions about Maria 1. Where did Maria hide? 2. Was Maria under the stairs? **Late Juncture: “*He saw her duck # under the chair*”** /hi: so: hɜ: dɐk # undɐ ðə tʃeː/ Questions about Lily’s duck 1. Who does the duck belong to? 2. Where did Mr. Johnson see her duck?	***“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$_{4}$tsɨ$_{5}$ ti$_{3}$ ɕja$_{4}$/ **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$_{4}$tsɨ$_{5}$ ti$_{3}$ ɕja$_{4}$/ 问题关于玛利亚 1.　玛利亚藏在哪里？ 2.　玛利亚有没有藏在桌子底下？ **Late 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$_{4}$tsɨ$_{5}$ ti$_{3}$ ɕja$_{4}$/ 问题关于莉莉的小猫 1.　这只小猫是谁的猫？ 2.　左老师在哪里看见莉莉的猫？
“He gave her baby milk” /hi: gæɪv hɜ: bæɪbɪ mɪlk/ **Early Juncture: “*He gave her # baby milk*”** /hi: gæɪv hɜ: # bæɪbɪ mɪlk/ 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? **Late Juncture: “*He gave her baby # milk”*** /hi: gæɪv hɜ: bæɪbɪ # mɪlk/ 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?	***“他给她婴儿奶粉”*** /t^{h}a$_{1}$ kei$_{3}$ t^{h}a$_{1}$ jiŋ$_{1}$ɚ$_{2}$ nai$_{3}$fən$_{3}$/ **Early Juncture: “*他给她 #　婴儿奶粉”*** /t^{h}a$_{1}$ kei$_{3}$ t^{h}a$_{1}$ # jiŋ$_{1}$ɚ$_{2}$ nai$_{2}$fən$_{3}$/ 问题关于莉莉的小猫 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: “*他给她婴儿 # 奶粉*”** /t^{h}a$_{1}$ kei$_{3}$ t^{h}a$_{1}$ jiŋ$_{1}$ɚ$_{2}$ # nai$_{3}$fən$_{3}$/ 问题关于薄阿姨的孩子 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？

about a man who accidentally gave dog biscuits to a little girl vs. a vignette about a man who gave biscuits to his neighbour's dog).

The English and Chinese reading passages and target sentences were highly comparable in three important ways. First, the English and Chinese target sentences, as well as the storylines, were identical in meaning, except for one minor deviation in translation in the second reading passage where the target ambiguous sentence in English was "*he saw her duck under the chair*", while the target sentence in Chinese was "*他看见她猫在坐凳子低下*" "*he saw her cat/hide under the stool*" (n.b., *猫* can mean either "*cat*" or "*hide*"). Second, both the English and Chinese target sentences involved the same structural ambiguity. In both languages, the "Early Juncture" sentences involved a feminine personal pronoun (i.e., *her/她*) before the juncture, followed by a postboundary compound noun or verb (e.g., *dog biscuit/狗饼干*), while in the "Late Juncture" sentences, the compound noun or compound verb became a simple noun or verb and the personal pronoun became a possessive determiner. Third, we selected target sentences involving pre- or post-boundary consonant onsets that were, in most cases, highly comparable in terms of their manner of articulation (e.g., /dɔg # bɪskəts/ vs. /kou # pinkan/; /bæɪbɪ # mɪlk/ vs. /jiŋɚ # naifən/).

***Recording procedures*.** All participants were tested by the same experimenter, who was fluent in 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 front 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. Participants were encouraged to speak in a way that would "really flesh out the meaning of the entire passage". However, the experimenter did not give any explicit instructions to produce the relevant juncture cues in the target ambiguous sentences. Furthermore, the passages were presented in plain text without any markers (e.g., hashtags) between phrases that would signal the designated boundaries.

After each reading passage, the experimenter asked participants a series of follow-up questions to test their comprehension of the passage (see Table 2 for examples). This was done to confirm that the participants understood the ambiguous sentences. If participants did not know the answers or answered incorrectly, they were encouraged to read the passage by themselves again. When participants finally understood the meaning of the sentences, they were given another chance to produce the passage again. In such cases, only data from the latest recordings were included in our final analyses. Every participant produced all the reading passages. None of the participants had to redo a reading passage more than twice.

Acoustic analyses. Four types of prosodic disambiguation strategies were analysed in Praat (Boersma & Weenink, 2018). These were (1) pause duration, (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 pause insertions, we measured the pause duration of the juncture that would indicate the early juncture in the "Early Juncture" sentences, and the pause duration of the juncture that would indicate the late juncture in the "Late Juncture" sentences. This was done for all sentences, so both the "Early Juncture" and "Late Juncture" sentences had two measures of pause duration, one from the designated early juncture (P1) and one from the designated late juncture (P2). For example, for

Target Sentence: “He gave her dog biscuits”

(a) “*He gave her # dog biscuits*” (Early Juncture)

(b) “*He gave her dog # biscuits*” (Late Juncture)

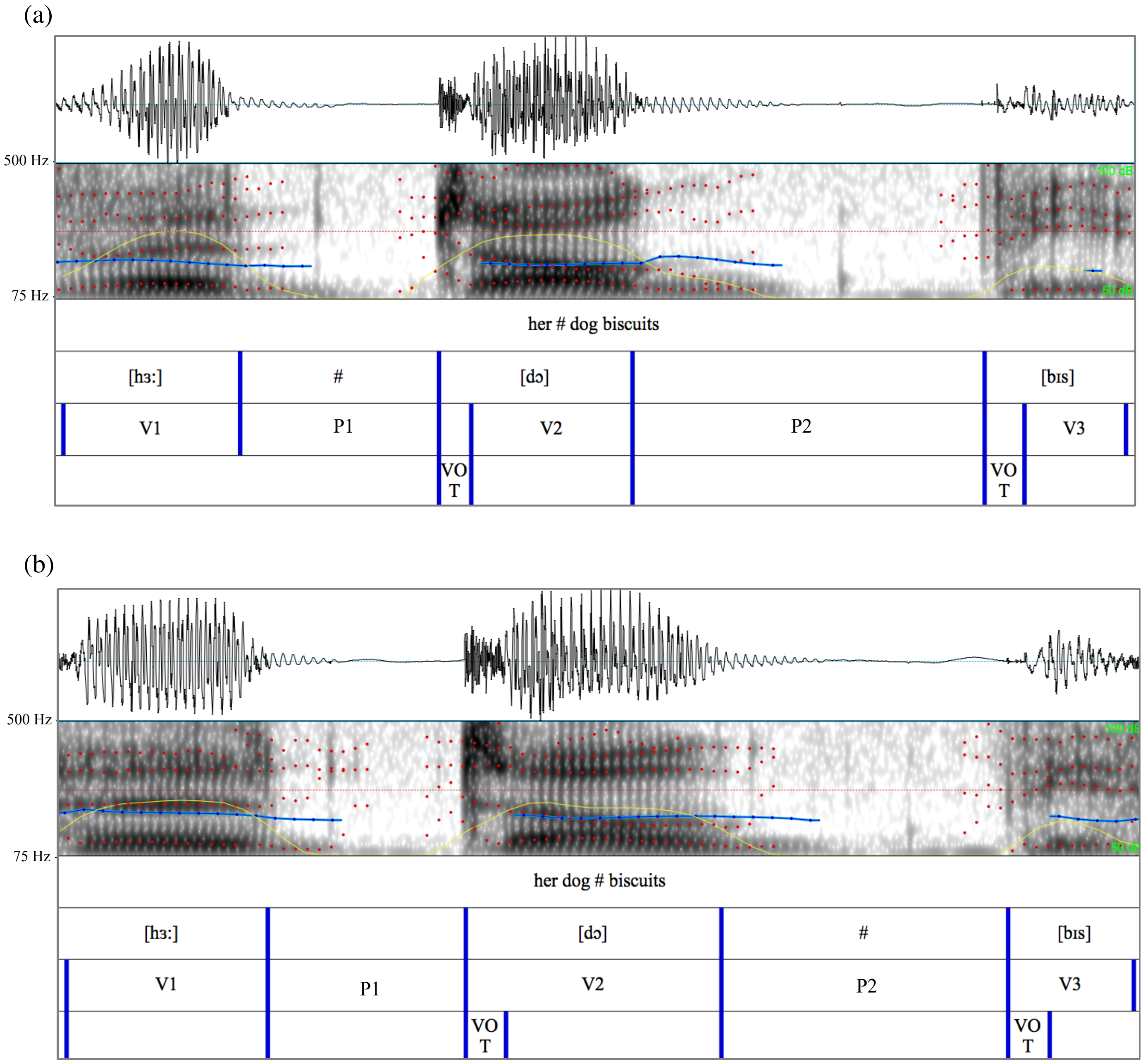


Figure 1.
Waveforms and pitch and amplitude contours of an example sentence pair in (a) “Early Juncture” and (b) “Late Juncture” versions. For both versions, we measured the pause duration of the juncture locations that would indicate the designated early juncture (P1) and the designated late juncture (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.

both juncture versions of the sentence "*He gave her dog biscuits*", we measured the pause duration between "*her*" and "*dog*" as well as the pause duration between "*dog*" and "*biscuits*". If the speaker did not produce any visible pause at one of the designated junctures, as observed on the spectrogram, 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. Each sentence has three measures of vowel duration. These were the preboundary vowel duration of the word before the designated early juncture boundary (V1), the preboundary vowel duration before the designated late juncture boundary (V2), and the postboundary vowel duration after the designated late juncture boundary (V3). For F$_0$ modification, we analysed the mean, minimum, and maximum F$_0$ as well as F$_0$ range of the 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, there was one postboundary nasal duration measure (i.e., /bæɪbɪ # mɪlk/) and 2 measures for VOT duration (i.e., /hɜ: # dɔg # bɪskəts/ and /hɜ: dɔg # bɪskəts/). In Mandarin, there was one measure for affricate duration (i.e., /t^{h}a mau # tsai/), one measure for VOT duration (i.e., /t^{h}a # kou/), and two nasal duration measures (i.e., /t^{h}a # mau tsai/ and /jiŋ$_1$ɚ$_2$ # nai$_3$fən$_3$/).

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)].

5.22. *Results and Discussion*

Prosodic cues to juncture. In each language, acoustic results for each prosodic cue were averaged across all the participants and sentence pairs. For each measurement, a series of pairwise *t*-tests were conducted to examine whether both languages showed similar patterns of production difference between the "Early" and "Late Juncture" versions.

The *t*-test results for all the juncture cues are displayed in Tables 3 to 5. We first measured the pause duration at the designated early and late juncture regions across the two juncture versions. Speakers from both language groups produced a significantly longer pause at the designated early juncture (P1) in "Early Juncture" sentences compared to the same cue in the "Late Juncture" sentences. On the other hand, both the English and Mandarin speakers produced a longer pause at the designated late juncture (P2) in "Late Juncture" sentences compared to the "Early Juncture" sentences.

We next compared the pre- and postboundary vowel durations (V1, V2, and V3) of the "Early" and "Late Juncture" sentences. Preboundary vowel durations were calculated from V1 (the preboundary vowel before the designated early juncture) and V2 (the preboundary vowel before the designated late juncture). In both languages, there was no significant difference in the duration of V1 between the "Early Juncture" and the "Late Juncture" sentences. This shows that the vowel durations of the word *her* or *她* in "Early Juncture" sentences (e.g., "*he gave her # dog* biscuits") were not significantly longer than the same word from the "Late Juncture" sentences (e.g., "*he gave her dog* # biscuits"). On the other hand, both groups of speakers produced a significantly longer preboundary vowel (V2) before the designated late juncture in "Late Juncture" sentences. Further, only Mandarin speakers showed a postboundary lengthening effect for V3, where the vowel duration after the late juncture was longer in the “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. In the analysed data, the English speakers only produced a significantly higher maximum F_0 at the postboundary vowel (V3) after the late juncture. Contrary to our predictions, however, the Mandarin speakers produced a significantly lower mean F_0 in the postboundary region of the designated late juncture in “Late Juncture” sentences. The Mandarin speakers also produced a significantly lower minimum F_0 in the preboundary vowel (V1) before the early juncture.

Finally, for domain-initial segmental strengthening in English, there was no significant difference between “Early” and “Late Juncture” sentences on any of the segmental measures. In Mandarin, there was a significant difference on one of the measures, but in a direction contrary to our predictions; the VOT after the designated late juncture was longer in the “Early Juncture” sentences than the “Late Juncture” sentences. Specifically, the unaspirated /k/ in /kou/ “*dog*” had a longer VOT.

To summarise, both English and Mandarin speakers produced significantly longer pauses at the relevant junctures in both juncture contexts. Speakers from both language groups also produced longer preboundary and postboundary vowel durations in “Late Juncture” sentences. However, neither group produced preboundary lengthening in the “Early Juncture” sentences. The English speakers produced higher maximum F_0 after late juncture in the “Late Juncture” contexts, but the significant effect was in the opposite direction in Mandarin. The Mandarin speakers also showed a lower F_0 before the designated early juncture in “Early Juncture” sentences. For segmental modification, we only found a longer postboundary VOT in Mandarin.

Table 3. *Mean duration of pausing, preboundary lengthening, and postboundary lengthening (in ms) as a function of "Early Juncture" and "Late Juncture" contexts. *p ≤ .05, **p ≤ .01, ***p ≤ .001.*

	English						**Mandarin**					
	Juncture Context			95% CI			Juncture Context			95% CI		
	Early	**Late**	SEM	Lower	Upper	t	**Early**	**Late**	SEM	Lower	Upper	t
Early Juncture Pause (P1)	86.72	75.32	3.774	3.88	18.93	3.02**	57.80	34.20	8.565	6.52	40.69	2.76**
Late Juncture Pause (P2)	51.40	83.26	9.837	51.58	122.46	3.24**	23.92	115.29	17.77	55.94	126.81	5.14***
Early Juncture Preboundary Duration (V1)	85.92	83.99	4.280	-10.46	6.60	.45	113.28	98.85	7.79	-1.11	29.99	1.85
Late Juncture Preboundary Duration (V2)	158.60	182.26	4.972	13.75	33.36	4.76***	193.80	259.60	8.136	49.58	82.02	8.09***
Late Juncture Postboundary Duration (V3)	93.54	91.59	3.539	-9.02	5.10	-.553	111.54	130.77	8.082	3.11	35.35	2.38*

Table 4. *Mean, minimum, maximum F_0 and F_0 range (in Hz) as a function of "Early Juncture" and "Late Juncture" contexts. *p ≤ .05.*

	English						**Mandarin**					
	Juncture Context			95% CI			Juncture Context			95% CI		
	Early	**Late**	SEM	Lower	Upper	t	**Early**	**Late**	SEM	Lower	Upper	t
Early Juncture Preboundary Mean F_0 (V1)	201.82	209.01	3.838	-.46	.14.85	1.88	201.29	206.76	3.250	-1.02	11.95	1.68
Late Juncture Preboundary Mean F_0 (V2)	182.57	175.77	6.645	-6.45	20.06	1.02	210.61	200.94	7.077	-4.45	23.78	1.37
Late Juncture Postboundary Mean F_0 (V3)	148.74	162.27	6.885	-.24	27.29	1.97	198.57	183.07	7.187	-29.85	-1.16	-2.16*
Early Juncture Preboundary Min F_0 (V1)	191.42	197.82	3.611	-.01	13.59	1.77	191.08	198.84	3.861	-.60	-15.46	2.01*
Late Juncture Preboundary Min F_0 (V2)	163.70	158.57	7.035	-8.90	19.16	.73	185.15	172.27	8.184	-3.44	29.20	1.57
Late Juncture Postboundary Min F_0 (V3)	135.06	142.19	8.03	-8.94	23.19	.89	169.75	156.60	6.711	-26.54	.25	-1.96
Early Juncture Preboundary Max F_0 (V1)	214.70	221.22	3.998	-1.45	14.49	1.63	212.42	217.41	4.841	-4.66	14.65	1.03
Late Juncture Preboundary Max F_0 (V2)	202.79	197.71	7.356	-9.59	19.75	.69	235.88	230.75	8.353	-11.52	21.79	.61
Late Juncture Postboundary Max F_0 (V3)	160.49	182.25	9.167	3.42	40.10	2.37*	233.92	218.92	8.825	-32.61	2.61	-1.70
Early Juncture Preboundary F_0 Range (V1)	23.27	23.40	1.605	-3.33	3.07	-.08	21.34	18.57	4.330	-5.87	11.40	.64
Late Juncture Preboundary F_0 Range (V2)	39.09	39.14	6.213	-12.34	12.44	.01	50.74	58.48	6.9688	6.15	21.64	1.11
Late Juncture Postboundary F_0 Range (V3)	25.02	40.76	9.030	-2.31	33.80	1.74	64.18	62.32	6.522	-14.87	11.16	-.28

Table 5. *Domain-initial segmental strengthening, measured in duration (in ms) as a function of "Early Juncture" and "Late Juncture" contexts. *p ≤ .05.*

	English						**Mandarin**					
	Juncture Context			95% CI			Juncture Context			95% CI		
	Early	**Late**	SEM	Lower	Upper	t	**Early**	**Late**	SEM	Lower	Upper	t
Early Juncture Postboundary VOT (V2)	13.18	13.28	.695	-1.48	1.29	-1.40	30.63	20.33	4.609	.76	19.83	2.23*
Late Juncture Postboundary VOT (V3)	8.74	10.74	2.144	-2.45	6.45	.93						
Early Juncture Postboundary Nasal Duration (V2)							93.90	102.20	6.410	-21.50	5.00	-1.28
Late Juncture Postboundary Nasal Duration (V3)	73.04	83.70	6.048	-1.89	23.20	1.76	59.00	67.35	6.657	-5.46	2.22	1.25
Late Juncture Postboundary Affricate Duration (V3)							39.50	47.20	4.590	-1.83	1.72	1.67

Crosslanguage differences. We performed a series of mixed two-way 2 (Language: English vs. Mandarin) × 2 (Juncture Context: “Early Juncture” vs. “Late Juncture”) ANOVAS to investigate whether there were any crosslanguage differences in the degree to which speakers would use the different prosodic cues to mark the designated junctures. Significant crosslanguage differences from the analyses (i.e., significant interactions) are presented in Figures 2 to 4. Bonferroni adjustments were used for follow-up *t*-tests.

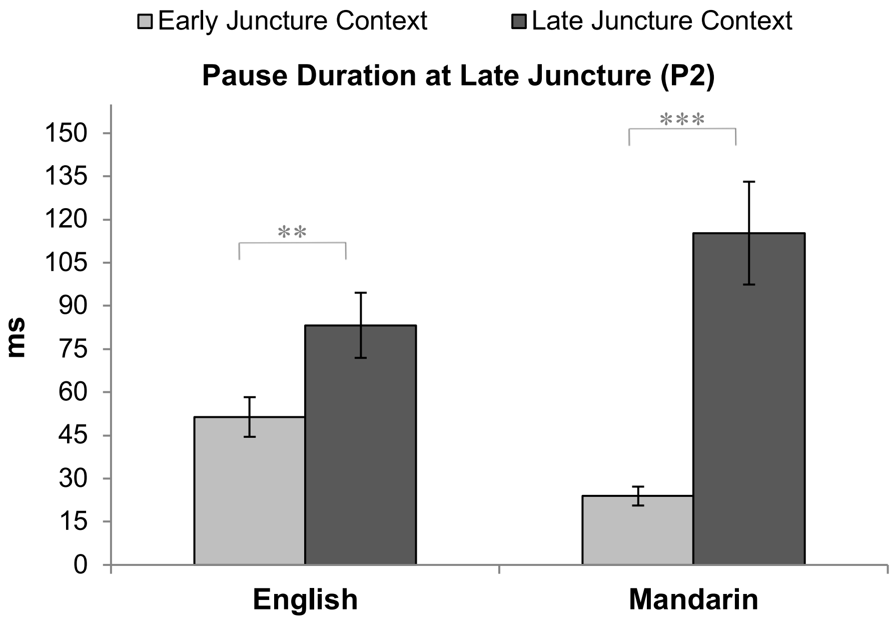


Figure 2. Crosslanguage differences in pause duration (in ms) at the designated Late Juncture marking as a function of “Early Juncture” and “Late Juncture” contexts. Error bars indicate standard error of the mean. **p ≤ . 01, ***p ≤ .001.

For the pause duration of the designated early juncture (P1), there was no significant crosslanguage difference in the degree to which the English and Mandarin speakers increased their pause duration to enhance the early juncture in the “Early Juncture” sentences. However, there was a significant crosslanguage difference in the extent to which the speakers used pausing as a cue to mark the designated late juncture in “Late Juncture” sentences, $F(1, 142) = 8.59, p = .004, \eta_p^2 = .06$. Simple effects tests of juncture context revealed that the effect was greater in Mandarin ($p < .001$) than in English ($p = .002$).

For vowel duration, there was a significant crosslanguage variation in the preboundary vowel duration before the designated late juncture (V2), $F(1, 142) = 19.53, p < .001, \eta_p^2 = .12$, where Mandarin speakers showed a greater increase in preboundary duration than English speakers (all p-values < .001). There was also a crosslanguage difference in the production of postboundary vowel duration after the late juncture (V3), $F(1, 142) = 5.71, p = .018, \eta_p^2 = .04$, but this time, only Mandarin speakers showed a significant increase ($p = .020$). Neither group produced significant increase in preboundary duration before the designated early juncture (V1).

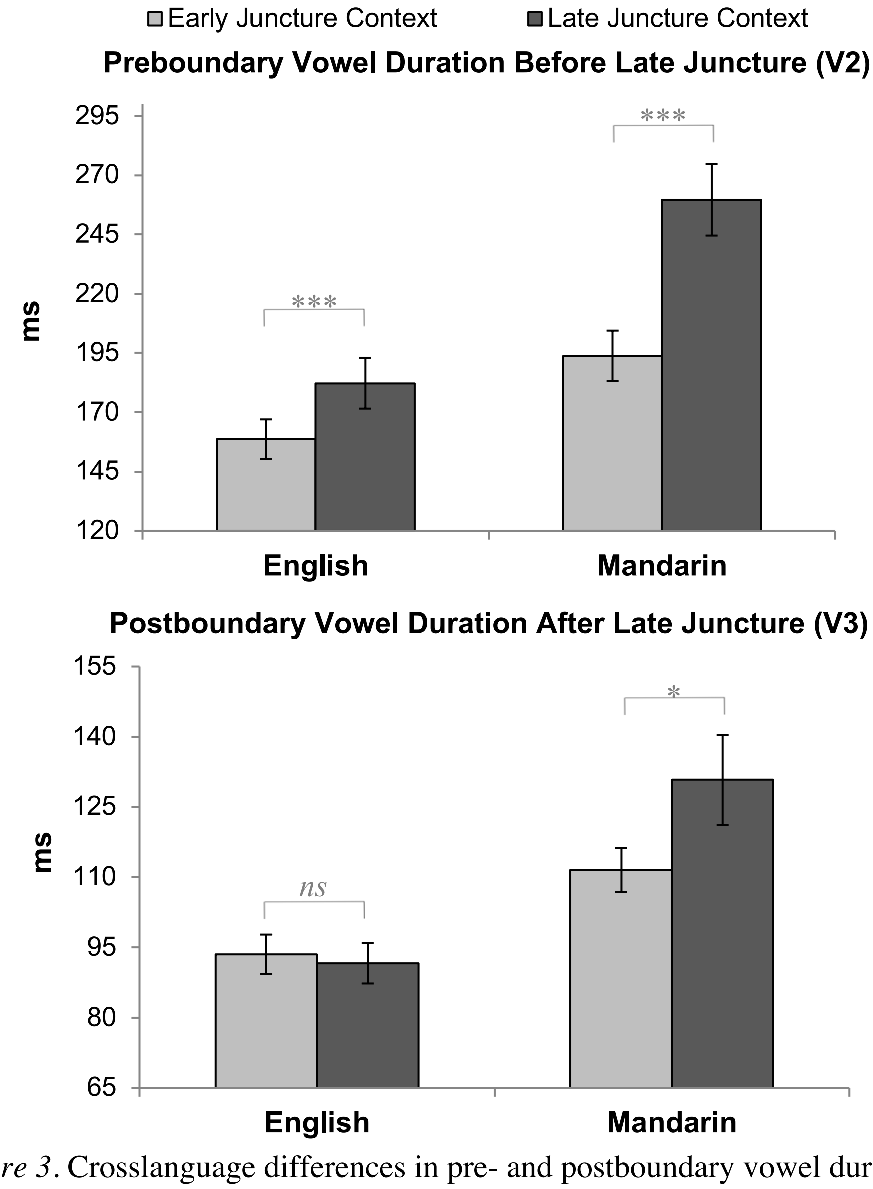


Figure 3. Crosslanguage differences in pre- and postboundary vowel durations (in ms) after the designated late juncture as a function of "Early Juncture" and "Late Juncture" contexts. Error bars indicate standard error of the mean. $*p \leq .05$, $***p \leq .001$.

For mean F_0, there was only a significant language difference in mean F_0 of the postboundary vowels after the designated late juncture, $F(1, 128) = 8.40, p = .004, \eta_p^2$ =.06, where the effect was in the opposite direction to our predictions and was only significant in Mandarin (p = .034), although it was marginally significant in English (p = .054). For maximum F_0, there was significant crosslanguage variation only on the postboundary vowel after the late juncture (V3), $F(1, 128) = 8.32, p = .005, \eta_p^2 = .06$. This time, only English speakers showed a significant increase (p = .021). There were no crosslanguage differences for any other F_0 measures.

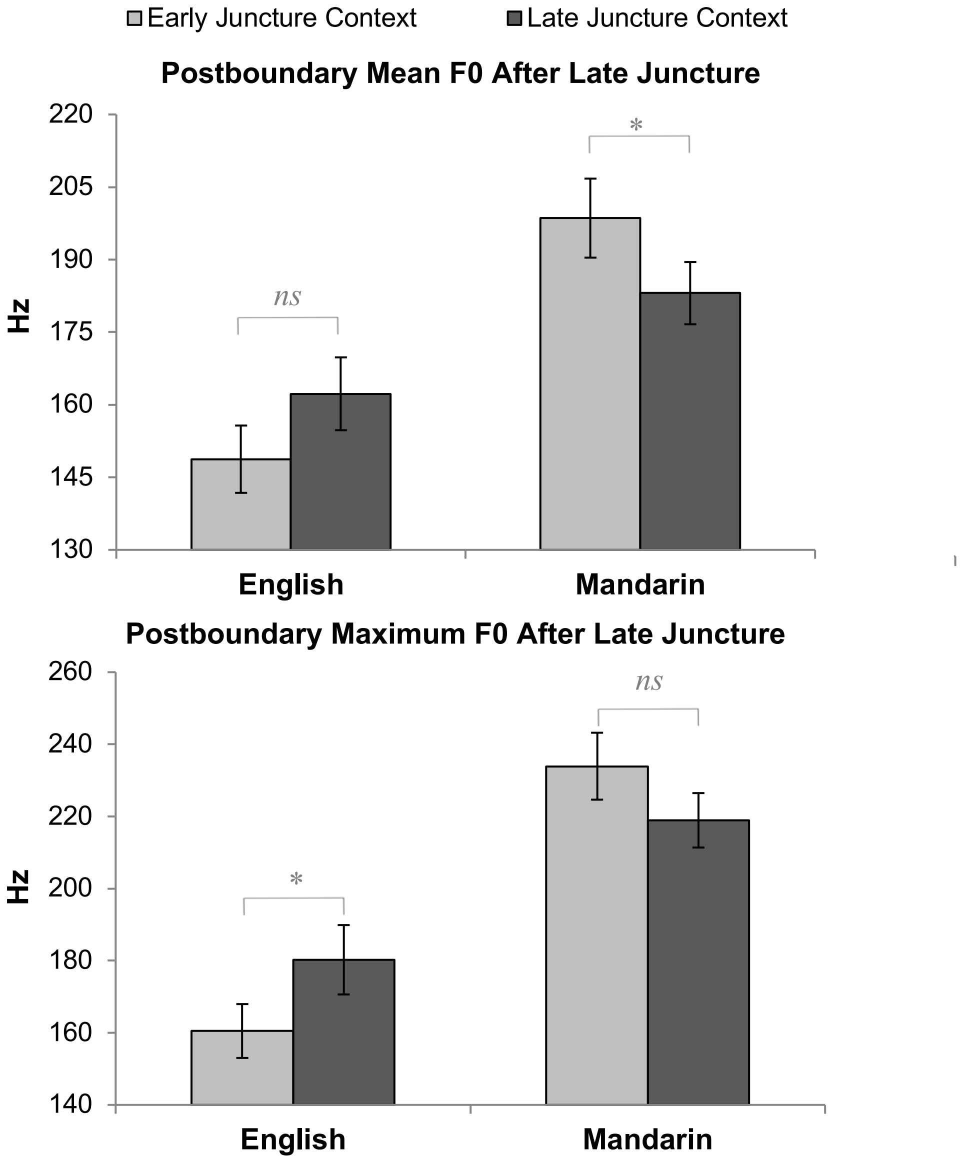


Figure 4. Crosslanguage differences in postboundary mean and maximum F_0 (in Hz) before the designated late juncture as a function of "Early Juncture" and "Late Juncture" contexts. Error bars indicate standard error of the mean. *$p \leq .05$.

We compared the English and Mandarin speakers' domain-initial segmental production only for the cues that were present in both languages, namely postboundary VOT and nasal duration. There was a significant language difference in the postboundary VOT after the designated early juncture in "Early Juncture" sentences, $F(1, 93) = 12.75$, $p = .001$, $\eta_p^2 = .12$, where only the Mandarin speakers produced a significantly longer VOT ($p = .036$). However, neither group of speakers produced a significant increase in postboundary nasal duration.

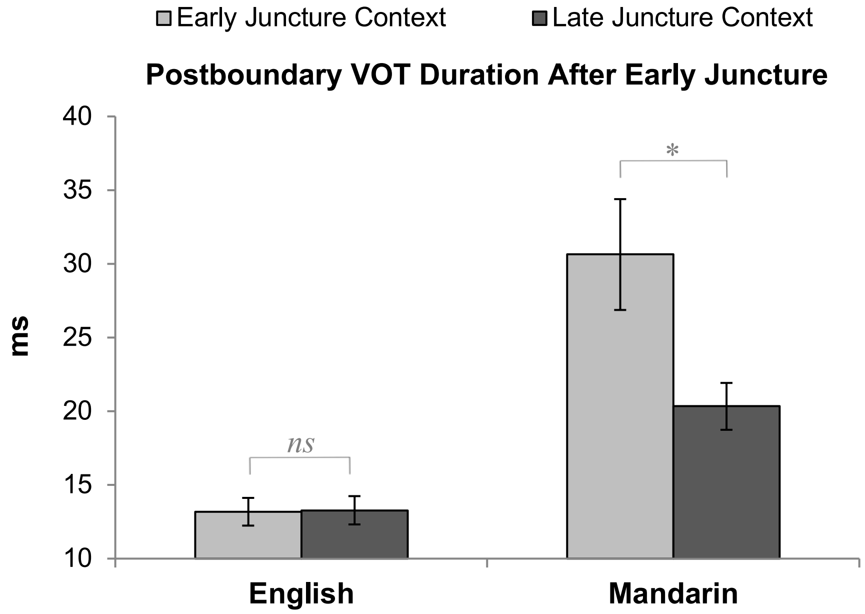


Figure 5. Crosslanguage difference in postboundary VOT duration (in ms) after the designated early juncture as a function of "Early Juncture" and "Late Juncture" contexts. Error bars indicate standard error of the mean. *$p \leq .05$.

Discussion. Our production data suggests English and Mandarin speakers are alike in how they use prosody to mark junctures, but there were language-specific differences in the degree to which different prosodic features were produced. For instance, across both the "Early Juncture" and "Late Juncture" ambiguous sentences, both groups of speakers produced longer pauses at the designated juncture. However, the difference in pause duration at the late juncture position in "Late Juncture" sentences was greater in Mandarin. Similarly, both English and Mandarin speakers produced longer preboundary vowels before the designated late juncture in “Late Juncture” sentences, but the Mandarin speakers produced a greater increase in preboundary duration than the English speakers.

Furthermore, neither language group produced all the boundary-related cues. For example, Mandarin speakers also produced the “Late Juncture” sentences with longer postboundary vowel duration, but this was not the case in English speakers, who produced a higher postboundary maximum F_0. For segmental strengthening, only the Mandarin speakers produced longer domain-initial/postboundary VOT, although this was only in “Early Juncture” sentences. For F_0, contrary to the prosodic context, the Mandarin speakers produced “Late Juncture” sentences with a significantly lower postboundary mean F_0. However, we should note that we have only compared the F_0 cues across the “Early” and “Late Juncture” cues, and the pre- and postboundary values were considered independently. Future research could analyse F_0 reset in more detail by looking at relative F_0 before and after a boundary within the same juncture version.

A reason why speakers did not produce all prosodic juncture cues could be because of the type of stimuli we used. It is important to note that the production experiment involved reading passages where the storyline already provided the referential context necessary for effective disambiguation. Note that it has been proposed that linguistic ambiguity is an advantageous part of communication because listeners can rely on contextual cues (Piantadosi, Tily, & Gibson, 2012). Here, we demonstrate that languages can differ in the degree to which speakers produce certain prosodic cues and omit other cues when the context is informative. By adopting a structured approach involving the identical storylines and ambiguous sentences, we show that speakers can vary in the type of boundary-related prosodic cues they still produce even when the context provided by the storylines made the use of prosody redundant (see also, Speer, Warren, & Schafer, 2011). This suggests that English and Mandarin speakers may differ in the types of prosodic juncture cues they choose to produce. An interesting extension of these production findings is to explore whether a similar pattern of crosslanguage results can also be found in perception. In the following perception experiments, we created a disambiguation task where listeners heard a series of “Early” and “Late Juncture” ambiguous sentences without any contextual cues and were required to press a button to choose the correct interpretation as quickly as possible. Given the identical ambiguous structure, any differences in response time and interpretation accuracy would indicate crosslanguage variation in prosodic juncture perception. On the other hand, any language similarities in perception may indicate that English and Mandarin listeners adopt the same prosodic strategies in disambiguation despite the differences in production.

5.3. Perception Experiment 1

5.3.1. *Method*

Participants. The final sample comprised of 40 native speakers of Australian English (M_{age} = 22.50 years, *SD* = 7.70 years, range: 17.89-53.50 years; 31 females) and of Mandarin (M_{age} = 25.12 years, *SD* = 3.61 years, range: 18.75-38.30 years; 21 females). All of the Mandarin-speaking participants 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). None of the participants reported any hearing or reading impairment.

Materials. Twenty-two syntactically ambiguous experimental sentences in English and Mandarin were chosen (see Appendices J and K), 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, we manipulated the juncture cues based on the timing and location of the boundary (i.e., “Early Juncture” versus “Late Juncture”). In the “Early Juncture” version, the speaker produced a sentence where the 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*”). In the “Late Juncture” version, the same segmentally identical sentence was produced where the 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*” vs. “*Larry gave rat poison to Hannah’s pet rat Rohan*”; “*刘波把老鼠药给珍妮*” vs. “*刘波把老鼠药给珍妮的宠物鼠*”). The English and Mandarin speakers who recorded the stimuli were asked to produce each version of the experimental sentences in a way that would match its corresponding interpretation sentence. In both languages, the “Early” and “Late Juncture” versions for each stimulus sentence pair were segmentally identical. Like the production experiment, the English and Mandarin sentences were highly comparable in terms of their structural ambiguity.

In each language, twelve 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 involving 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 that contained other types of ambiguity.

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 L and M). All participants were given three practice trials and feedback before starting the actual experiment. However, there were no explicit instructions on how to disambiguate the sentences.

At 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 required to choose for each sentence its intended meaning, by pressing the button that matched the correct interpretation sentence. Participants were asked to "pay careful attention to the meaning of each sentence". Participants were told that they were allowed to press the button anytime during the trial while the sentence was being played. A five-second response probe was still available after the sentence was finished, in which participants still have five seconds to press the button before moving to the next trial. Nevertheless, they were told to choose the correct button "as soon as they understood the sentence". Participants were told that they would be tested on both their accuracy and on their speed of comprehension. The interpretation sentences remained on the screen throughout the entire trial. Whether the correct button was the left or right button was counterbalanced across participants.

We recorded participants' response times and number of correct responses. We only included data from participants who correctly disambiguated at least 64% of (i.e., 14 out of 22) the experimental sentences. 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. None of the participants in our final sample failed to respond on more than two occasions during the experimental trials.

At the end, all participants completed a recognition test to judge whether each of the 22 sentences in the list were from the experiment (see Appendices N and O). Half of these sentences were 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%). In addition, the recognition scores made by the English and the Mandarin listeners were not statistically different.

5.3.2. *Results and Discussion*

Response time. More than 90% of participants' correct responses were made by pressing the button after the test sentence was played. Therefore, we measured response time (RT) as the latency duration between the offset of the experimental sentence and participants' button presses. Only data for correct disambiguations were included in our analyses. Control analyses using mixed ANOVAs were performed separately for each language group and revealed no significant effect of the counterbalanced juncture conditions. There was also no significant effect of the counterbalanced button locations.

We conducted a mixed 2-way 2 (Language: English vs. Mandarin) × 2 (Juncture Context: "Early Juncture" vs. "Late Juncture") ANOVA to examine whether there were any crosslanguage differences in RT as a function of the different juncture contexts. Overall, our analyses revealed a significant interaction between language and juncture contexts, $F(1, 78) = 20.00, p < .001, \eta_p^2 = .20$. Follow-up Bonferroni-adjusted pairwise *t*-tests revealed significant differences in RT between "Early Juncture" and "Late Juncture" for both English and Mandarin listeners. However, the response time difference between the juncture contexts was in the opposite direction across the two language groups. Specifically, English listeners were significantly faster at disambiguating the "Late Juncture" sentences ($M = 1109.43$ ms, $SD = 555.64$ ms) compared to "Early Juncture" sentences ($M = 1355.63$ ms, $SD = 704.88$ ms), $t(39) = 3.59, p = .001$, while Mandarin listeners responded more rapidly in “Early Juncture” sentences (*M* = 1072.15 ms, *SD* = 423.11 ms) than “Late Juncture” sentences (*M* = 1219.06 ms, *SD* = 547.61 ms), *t*(39) = -2.67, *p* = .011.

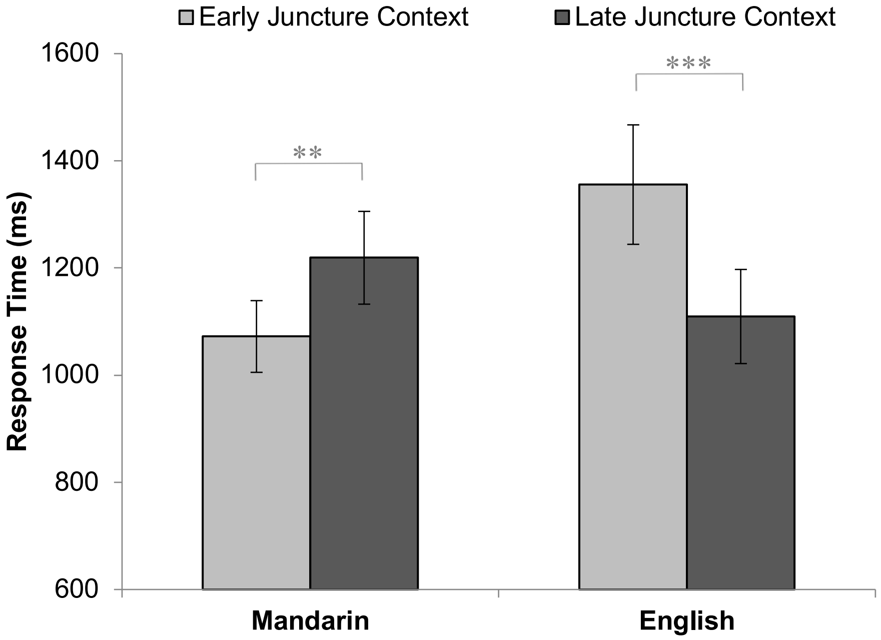


Figure 6. Significant crosslanguage difference in disambiguation response time (in ms) as a function of “Early Juncture” and “Late Juncture” contexts. Error bars indicate standard error of the mean. ***p ≤ .05, ***p ≤ .001.*

Accuracy. On average, the Mandarin-speaking participants in our final sample had 3.3 incorrect disambiguation responses (*SD* = 1.82) throughout the 22 experimental sentences, with an average of 1.63 errors (*SD* = 1.15) in “Early Juncture” sentences, and an average of 1.68 errors (*SD* = 1.59) in the “Late Juncture” sentences. Therefore, both the “Early Juncture” and “Late Juncture” sentences had a similar number of errors.

On the other hand, the English-speaking participants in our final sample had an average of 5.6 incorrect disambiguations (*SD* = 2.1) in the 22 experimental sentences. Across the juncture versions, the English listeners had an average of 3.9 incorrect disambiguations (*SD* = 1.6) for “Early Juncture” sentences, and an average of 1.7 errors (*SD* = 1.07) for “Late Juncture” sentences. Based on our pairwise *t*-test analysis, the English group had significantly more incorrect disambiguations in “Early Juncture” sentences compared to “Late Juncture” sentences, $t(39) = 8.05$, $p < .001$. Overall, the English listeners had significantly more incorrect disambiguation responses than the Mandarin listeners, $t(39) = 6.39$, $p < .001$.

We also examined whether the English and Mandarin samples also differed in the number of 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. On average, the 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.

Table 6. *Number of incorrect responses in English and Mandarin as a function of “Early Juncture” and “Late Juncture” contexts in Perception Experiment 1.*

	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)

Discussion. Our perception experiment revealed significant crosslanguage differences in response time pattern across the different juncture versions. In English, listeners were significantly faster at disambiguating “Late Juncture” sentences than “Early Juncture” sentences. Conversely, Mandarin listeners were faster at disambiguating “Late Juncture” sentences. The English and Mandarin listeners also differed in interpretation accuracy, with more errors made by English listeners.

In our production experiment, we revealed that Mandarin speakers tend to produce “Late Juncture” sentences with longer pauses and pre- and postboundary lengthening compared to the English speakers. In light of the production data, the slower RT in the "Late Juncture" sentences found in Mandarin may indicate that Mandarin listeners were paying attention to the extra increases in boundary-related lengthening and pause duration. At the same time, however, the slower RT might have also been due to the late arrival of the boundary pause. In the second perception experiment, we tested 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 the removal of the pausing cue. If Mandarin listeners do not rely on pausing as a cue to juncture, then the removal of the pausing cue would not affect their accuracy and RT performance. 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.

5.4. Perception Experiment 2

5.4.1. *Method*

Participants. We recruited a new sample of 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). The Mandarin-speaking participants had been living in Australia for an average of 5.22 years (*SD* = 7.32, range: 41 days to 24 years and 9 months). All participants were university students at the time of the experiment and reported no hearing or reading impairment. We excluded additional data from four English listeners and one Mandarin listener who failed to correctly disambiguate at least 64% of the experimental sentences.

Materials and procedures. The procedures were identical to Perception Experiment 1, only this time, the pause durations were rendered uninformative in all experimental sentences. For all experimental juncture sentences, across both the "Early Juncture" and the "Late Juncture" versions, we spliced out both the pause that would indicate the designated early juncture (P1) and the same pause cue that would indicate the late juncture (P2). As a result, there were no interword silences at all in the two positions. In the follow-up recognition test, the final sample of English listeners had an average score of 81.46%, or 17.92 out of 22 (*SD* = 9.59%, range: 68-100%), and the Mandarin listeners scored 92.82%, or an average of 20.42 out of 22 (*SD* = 6.50%, range: 82-100%), which was not statistically different from the recognition scores in Perception Experiment 1.

5.4.2. *Results and Discussion*

Response time. No significant crosslanguage difference appeared between the English and Mandarin listeners' RT, although the comparison was marginally significant, $F(1, 29) = 3.39$, $p = .076$, $\eta_p^2 = .11$. Importantly, the direction of the results was the same as in Experiment 1. The English listeners showed a marginally significantly faster RT in "Late Juncture" sentences (*M* = 1197.23, *SD* = 728.53) compared to "Early Juncture" sentences (*M* = 1452.38, *SD* = 744.84), $t(11) = 2.04$, $p = .066$, while the Mandarin listeners also showed a faster RT in the "Early Juncture" sentences (*M* = 1149.56, *SD* = 649.40) than the "Late Juncture" sentences (*M* = 1235.34, *SD* = 672.95), although this was no significant ($p = .498$). There were no effects of the counterbalanced conditions.

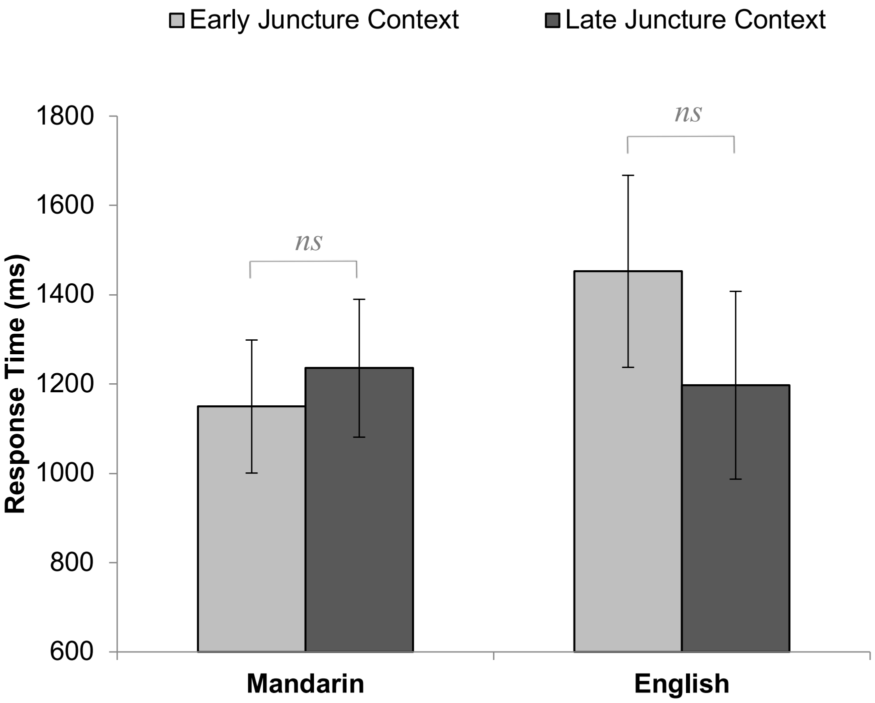


Figure 7. Disambiguation response time (in ms) as a function of "Early Juncture" and "Late Juncture" contexts when pause duration cue was removed. Error bars indicate standard error of the mean.

Accuracy. As in Experiment 1, pairwise *t*-tests were used to compare listeners' accuracy scores. The English listeners in the final sample had an average total of 5.75 (*SD* = 1.49) incorrect responses, with an average of 3.08 errors (*SD* = 1.51) in "Early Juncture" sentences and an average of 2.67 (*SD* = 1.50) errors in "Late Juncture" sentences. In Mandarin, participants in the final sample had an average total of 5.68 errors (*SD* = 1.89), with 2.37 errors (*SD* = 1.17) in "Early Juncture" sentences and 3.32 errors (*SD* = 1.86) in "Late Juncture" sentences.

Table 7. *Number of incorrect responses in English and Mandarin as a function of "Early Juncture" and "Late Juncture" contexts in Perception Experiment 2.*

	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)

We also examined the number of errors made by the excluded four English listeners and the one Mandarin listener. The excluded English listeners had a total average of 9.5 errors (*SD* = .58), with equal average number of errors from each juncture version (*M* = 4.75, *SD* = 1.71). The Mandarin listener had 3 errors in "Early Juncture" sentences and 6 errors in "Late Juncture" sentences.

Discussion. The second perception experiment reveals that Mandarin listeners' disambiguation accuracy was affected when the pausing cue was rendered uninformative. Specifically, their error rate increased from a total average error of 3.30 to 5.68 incorrect responses. The English listeners, however, showed no significant increase in errors. Thus removal of pausing cues affected the Mandarin listeners' performance, but had little effect on the English listeners. It is noteworthy that the pattern of RT difference between the two juncture versions remained unchanged: English listeners' RT in "Late Juncture" sentences was still faster than their RT in "Early Juncture" sentences, although the effect was only marginally significant (possibly due to the small sample size giving low statistical power). More participants would therefore be needed to provide a more definite interpretation of the results. Nevertheless, based on the data so far available, we tentatively conclude that pausing cues may be more likely to be exploited for prosodic disambiguation in Mandarin than in English.

An interesting follow-up question is whether Mandarin speakers may also adopt the same prosodic strategies in a non-native language. Given the identical ambiguous structures and the similar prosodic juncture cues in English and Mandarin, it would be reasonable to expect that Mandarin speakers would transfer their L1 perception strategies to the other language as L2. In the third and final perception experiment, we tested this possibility with Mandarin native speakers listening to the original English sentences from the first perception experiment. If the Mandarin speakers can draw on their L1 experience in processing the L2, then they should show similar response patterns and accuracy rates in English. If L1-optimised prosodic processing requires the presence of L1 speech, however, a different result may ensue.

5.5. Perception Experiment 3

5.5.1. *Method*

Participants. The original sample had a total of 36 native Mandarin speakers. Due to recruitment constraints, most of these participants were those who had already participated in either Experiment 1 or 2. We excluded data from 7 participants who failed to disambiguate at least 64% of the experimental sentences, leaving a remaining total of 29 participants in the final sample. The mean age of the participants in the final sample was 26.28 years (*SD* = 5.42 years, range: 20.73-43.62 years; 14 females). Participants had been living in Australia for any period between 3.65 months to 24.77 years (*M* = 3.21, *SD* = 5.19 years)

Materials and procedures. The procedures were identical to those in the previous experiments, only this time, the stimuli were the original English sentences from Perception Experiment 1. On average, the Mandarin-speaking participants in the final sample scored 19.61 out of 22 (87.82%) in the recognition test (*SD* = 9.46%, range: 64-100%), which was not significantly different from that of the English and Mandarin speakers from the first experiment.

5.5.2. *Results and Discussion*

Response time. Analyses show that Mandarin-speaking participants did not fully transfer their L1 prosodic strategies in a second language. There was no significant difference in RT between the "Early" (*M* = 1377.69 ms, *SD* = 529.85 ms) and "Late Juncture" sentences (*M* = 1343.32 ms, *SD* = 656.59 ms), *t*(28) = 0.36, *p* = .720. As in the previous experiments, there were no significant effects of the counterbalanced conditions.

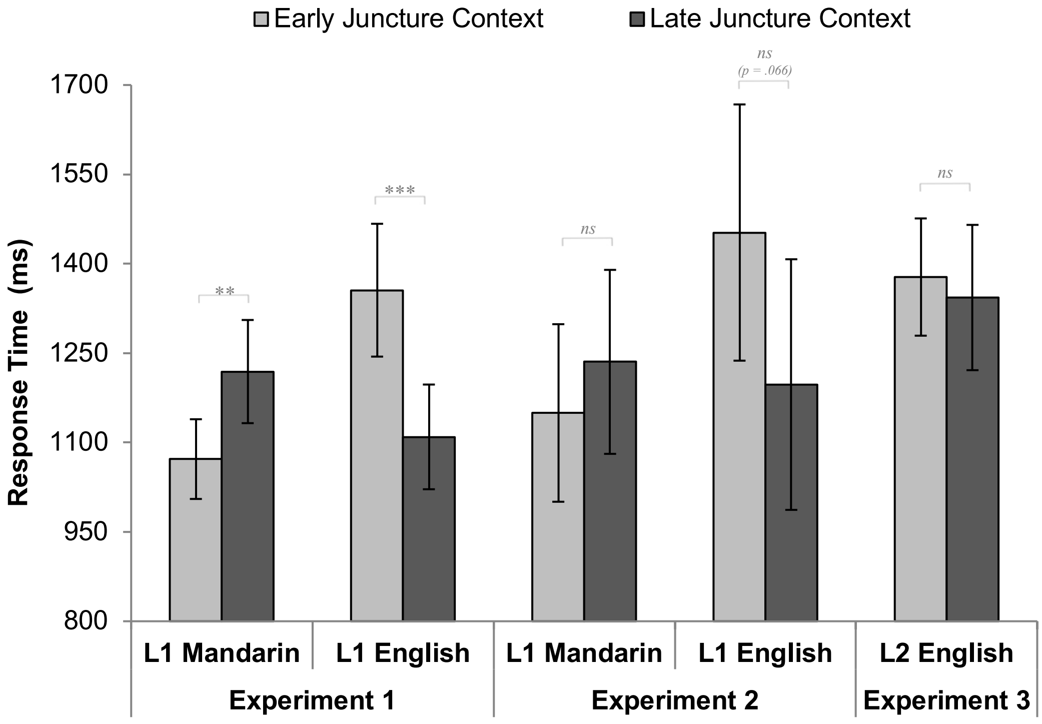


Figure 8. Disambiguation response time (in ms) as a function of "Early Juncture" and "Late Juncture" contexts in Perception Experiment 1 (all prosodic cues present), Experiment 2 (no pausing cue), and Experiment 3 (L2 English). Error bars indicate standard error of the mean. ***p ≤ .01, ***p ≤ .001.*

Accuracy. The Mandarin-speaking participants had, on average, a total of 5.69 incorrect responses (*SD* = 1.95) out of 22 English sentences. Within the juncture sentence versions, there were 3.21 errors (*SD* = 1.42) in the “Early Juncture” sentences and 2.48 errors (*SD* = 1.33) in the “Late Juncture” sentences. In the excluded participants, the total average was 10.29 errors (*SD* = 1.11), with 4.86 incorrect responses (*SD* = 1.07) in the “Early Juncture” sentences and 5.43 incorrect responses (*SD* = 1.40) in the “Late Juncture” sentences. We also compared the total number of incorrect responses made by the Mandarin-speaking participants in this experiment with the number of incorrect responses made by the participants in Perception Experiment 1, when the sentences were presented in their L1. Our analyses revealed that native Mandarin speakers made significantly more disambiguation errors when the sentences were presented in English compared to when the sentences were in their native language, $t(28) = 4.63, p < .001$. However, the native Mandarin speakers in the L2 English context did not make significantly more errors than the native English speakers from Perception Experiment 1.

Table 8. *Number of incorrect responses in English and Mandarin as a function of “Early Juncture” and “Late Juncture” contexts in Perception Experiments 1 to 3.*

		Mean Errors (SD)		
		“Early	“Late Juncture”	Total
Experiment 1	L1 English	3.90 (1.60)	1.70 (1.07)	5.60 (2.10)
	L1 Mandarin	1.63 (1.15)	1.68 (1.59)	3.30 (1.82)
Experiment 2	L1 English	3.08 (1.51)	2.67 (1.50)	5.75 (1.49)
	L1 Mandarin	2.37 (1.17)	3.32 (1.86)	5.68 (1.89)
Experiment 3	L2 English	3.21 (1.42)	2.48 (1.33)	5.69 (1.95)

Length of stay and L2 disambiguation. As our participants were not fully uniform with respect to how long they had spent in non-Mandarin-speaking environments, an additional analysis was conducted to assess whether participants’ RT was related to their exposure to English as a foreign language while living in Australia. Participants’ difference scores in RT were calculated by subtracting their average RT in "Early Juncture" sentences from the RT in "Late Juncture" sentences. A Pearson's correlational analysis was performed to calculate the association between participants' RT difference score and their length of stay in Australia, and the result showed a significant positive correlation, $r = .40$, $p = .032$.

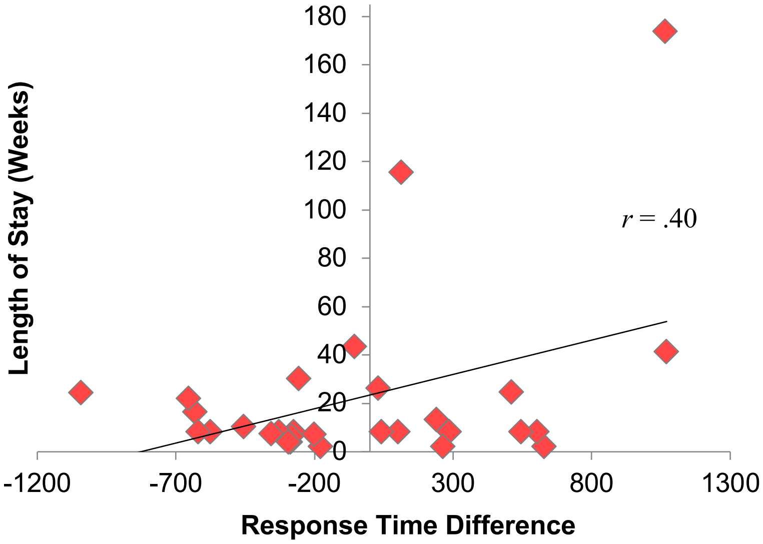


Figure 9. Significant positive correlation between native Mandarin speakers' (*N* = 29) length of stay (i.e., date of testing minus date of arrival) in an English-speaking country (in weeks) and their RT difference scores in English (RT in "Late Juncture" sentences minus RT in "Early Juncture" sentences).

Discussion. The L2 data suggest that native Mandarin speakers do not fully transfer their L1 prosodic strategies to process the same type of structural ambiguity in L2 English. First, there was no significant RT difference between the juncture versions. Second, Mandarin speakers' accuracy rate in L2 English was significantly lower (i.e., had more disambiguation errors) compared to their scores in the first perception experiment when they listened to sentences in their native language. However, their interpretation accuracy in L2 English was comparable to the accuracy scores of the English speakers from the first and second perception experiments.

We also revealed a significant association between Mandarin speakers' length of stay in Australia and the degree to which they showed a response time difference between the two juncture contexts. It is important to note that we measured response time difference as the absolute difference by subtracting their RT in “Early Juncture” sentences form their RT in “Late Juncture” sentences. Given that Mandarin speakers showed a faster RT in “Early Juncture” sentences in their L1, subtracting RT in “Early Juncture” contexts from “Late Juncture” contexts in the L2 experiment would indicate the degree to which listeners showed the same response time pattern as their L1. In other words, the positive correlation we found indicates that Mandarin speakers who had been living in Australia longer were also more likely to disambiguate the L2 English sentences in the same way as their L1 (i.e., faster disambiguation in “Late Juncture” sentences).

5.6. General Discussion

The present experiments provide new findings on how native speakers of two phonologically very different languages may differ in their use of prosody in juncture processing and structural disambiguation. Using English and Mandarin sentences that involved the same structural ambiguity, our production and perception data revealed crosslanguage variation in the degree to which native speakers exploit the different boundary-related prosodic cues. In production, speakers differ in the degree to which they enhance different juncture features. In perception, we discovered crosslanguage variation in listeners’ disambiguation accuracy and response time patterns across ambiguous sentences with different timing and location of prosodic junctures.

According to previous production studies, both English and Mandarin speakers can produce the same prosodic cues to mark relevant junctures. Like many other languages, English and Mandarin speakers can cue prosodic boundaries through a combination of pausing, preboundary durational lengthening, postboundary lengthening, preboundary F_0 lowering, postboundary F_0 reset, and domain-initial segmental strengthening (e.g., Cooper & Paccia-Cooper, 1980; Keating, Cho, Fougeron, & Hsu, 2003; Kuang, 2010; Li

& Yang, 2009; Liberman & Pierrehumbert, 1984; Shen, 1993; Shih, 1988; 2000). This, however, does not guarantee that speakers would always produce these juncture cues to disambiguate speech. Prosodic cues to syntactic disambiguation are unreliable because naïve speakers may not realise such cues under ordinary reading conditions when they were not made aware of the ambiguity or when the referential context is already informative (e.g., Allbritton, McKoon, & Ratcliff, 1996; Snedeker & Trueswell, 2003).

What is interesting about our production findings is that English and Mandarin speakers can still produce at least some of these juncture cues even when the referential context provided by the reading passages made the use of prosody unnecessary. Under such conditions, we were able to discover language-specific differences in the degree to which speakers would optionally mark the different juncture cues. For example, Mandarin speakers were more likely to mark "Late Juncture" ambiguous sentences with greater increases in pause duration and boundary-related vowel lengthening, while English speakers produced the same type of sentences with greater increases in postboundary F_0 reset. Speakers of different languages can vary in their prosodic choices.

In perception, we revealed that native English and Mandarin speakers can differ in how they use prosody to resolve structural ambiguity. First, native English and Mandarin listeners disambiguated the sentences differently as a function of the different juncture locations. Second, listeners may vary in their reliance on different juncture cues (e.g., pausing) during disambiguation. Third, from the accuracy data, English and Mandarin listeners also vary in the degree to which they could use prosody to successfully disambiguate sentences. Finally, our L2 findings provide evidence that listeners do not fully transfer their L1 prosodic strategies even when both the L1 and L2 sentences involve exactly the same type of structural ambiguity.

A possible explanation for the language differences in RT across the different juncture versions could be that the English and Mandarin stimuli might have exhibited different degrees of duration increases. From our production data, “Late Juncture” sentences in the Mandarin stimuli showed a greater increase in pause duration and boundary-related lengthening. At least for pause duration, acoustic analyses found that this was in fact the case in our Mandarin stimuli. The longer pause duration in “Late Juncture” sentences may partly explain why the Mandarin listeners in the first perception experiment have a delayed RT in “Late Juncture” sentences. Mandarin listeners might have been paying more attention to the extra increases in pause duration.

In support of this interpretation, our second perception experiment indicates that disambiguation performance in Mandarin, but not in English, was degraded when pausing duration of the critical juncture was uninformative. Consistent with these findings, recent experiments by Yang and colleagues (2014) showed better Intonational Phrase boundary detection by Mandarin listeners 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. In extension of their findings, we revealed that Mandarin listeners showed greater perceptual reliance on pausing cues under conditions where prosody was the only source of cue that could help them segment ambiguous sentences. This shows that language-specific preference for a given prosodic cue to boundary placement (e.g., durational cues) is far from the whole story; the precise details of a cue’s realisation are also part of the native strategy.

Indeed, there is considerable evidence that even when the same cues (e.g., VOT, domain-initial strengthening) are used across languages, the exact realisation may vary (e.g., Byrd et al., 1997; Cho & McQueen, 2005; Kuzla & Ernestus, 2011; Pierrehumbert

& Talkin, 1992). In the case of juncture pausing, however, it remains an empirical question as to why Mandarin listeners may rely on pause duration to a greater extent. In English, on the other hand, both our perceptual findings and previous ERP data indicate that listeners are less reliant on pausing (e.g., Aasland & Baum, 2003). 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).

What might have induced the language difference in cue weighting? To address this question, it may be important to explore why speakers of certain languages may need to use pause duration. Each language may have a different reason. Mandarin may rely more on pause duration because prosodic cues to relevant junctures may sometimes compete with the use of the same suprasegmental dimension for lexical distinctions. First, Mandarin has only 29 phonemes (7 vowels and 22 consonants) compared to 46 in General Australian English (20 vowels and 26 consonants: Harrington, Cox, & Evans, 1997). At least twelve of the 22 Mandarin consonants involve phonemic distinction based on duration differences (e.g., aspirated vs. unaspirated VOT). Linguistic tones, characterised by differences in intrinsic F_0 shapes, duration, and amplitude, can also provide more opportunities for lexical contrasts. At the same time, ambiguity may also exist when words or syllables with different meaning have the same tone, segments, and even the same written character. For example, the segment /ʂu1/ in high levelled tone can give at least 40 words, many of which can be monosyllabic words (e.g., *書* “*book*”, *叔* “*uncle*”, *梳 “comb”, 殳 “halberd”, 枢 “door hinge”*), not to mention the plethora of meanings that can be conveyed with the same segment in other tones (e.g., /ʂu2/ *熟* “*familiar*”, /ʂu3/ *鼠* “*rat*”, /ʂu4/ *樹* “*tree*”). In written language, there are also cases where the same character can stand for more than one meaning. Ambiguity is therefore may be more prevalent in Mandarin than in English. Providing a simple pause between words or phrases can serve a better alternative for disambiguating ambiguous sentences without altering the F_0 or durational information of the boundary segments. On the other hand, pausing may be redundant in languages where there is not much competition between prosodic and non-prosodic uses of the same suprasegmental dimension.

At the same time, our L2 results indicate that the language differences in juncture perception may reflect more than just the differences in reliance on pause duration increases. Using L2 English sentences that involved the same structural ambiguity as the L1 Mandarin sentences, our L2 experiment show that native Mandarin speakers do not fully transfer their L1 strategies to disambiguate the sentences in English, although they could have done so and achieved an efficient perceptual outcome. This lack of complete L1 to L2 transfer in our experiment cannot be fully explained by duration adjustment differences in English versus Mandarin juncture production.

There was also a significant positive association between the Mandarin speakers' length of stay in Australia and the degree to which their disambiguation RT in L2 reflected the same RT pattern found in L1 (i.e., a slower RT in "Late Juncture" compared to "Early Juncture" sentences). This unexpected finding indicates that longer time spent in a non-native environment increases the chance of L1 to L2 transfer of disambiguation strategies. We also note that age of arrival was also significantly correlated with length of stay, but it did not mediate the link between length of stay and RT in L2 English.

Why did the Mandarin listeners fail to exhibit comparable response patterns and accuracy rate across L1 and L2? There are three possible reasons. One reason could be that disambiguation strategies are indeed specifically tailored to L1 processing. Then it could be the case that it takes time to learn how to assess relative duration as realised on a new (L2) segmental repertoire. Alternatively, listeners may gradually learn to concentrate on those prosodic dimensions that are more reliably related to the boundary occurrence in their native language (Shatzman & McQueen, 2006). Finally, the lack of complete L1 to L2 transfer revealed in our experiments may also suggest that L1 disambiguation is learned as a purely language-specific strategy, and as a result all learners must learn from scratch the prosodic system of their L2. The first and second reasons are related to the listener's episodic experience, while the third reason is related to the issue of phonological abstraction. Further studies are needed to decide between these alternatives.

It is also noteworthy that the pattern of native Mandarin speakers' disambiguation errors in L2 English sentences was comparable with that of the native English speakers from the earlier experiments. Therefore, the native Mandarin speakers can still make use of prosodic cues to disambiguate the English sentences, even though they were not using their L1 strategies to their advantage. Related to these findings, previous studies suggest that speakers do not fully transfer their L1 cues to syntactic structure, although they can exhibit appropriate L2 cues quite early in learning. For example, O'Brien and colleagues (2014) found that prosodic disambiguation in L2 German by native English speakers, and in L2 English by native German speakers, resembled the L2 target cues rather than the cues in the speaker's native language. Similarly, fourth-semester L2 learners of French can correctly produce L2 prosodic cues to resolve relative-clause attachment ambiguity (Dekydtspotter, Donaldson, Edmonds, Fultz, & Petrusch, 2008). All these previous L2 prosodic processing experiments involved different cues across L1 and L2, while our experiments involved a unique case where both English and Mandarin stimuli exhibited the same prosodic cues. Learning to process different cues in the L2 is certainly the whole point of second language learning. If the L2 cues happen to be highly similar to

L1, the transfer of an effective L1 cue to an L2 in which it would be equally effective will nevertheless require an explicit learning process.

Finally, we note that native Mandarin speakers consistently showed higher rates of disambiguation accuracy in their native language compared to the native English speakers. Unlike our production experiment, our perception experiments involved a disambiguation task where only prosody could disambiguate the stimuli sentences. The fact that there were more interpretation errors in English than Mandarin indicates that English listeners may be less likely to rely on prosodic juncture cues for disambiguation.

As already mentioned, how listeners use prosodic cues to disambiguate speech can also be influenced by a multitude of other linguistic factors, including lexical bias, situation-specific contextual information, listeners' knowledge of the speaker, and 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). For instance, Piantadosi, Tily and Gibson (2012) proposed that ambiguity allows efficient communication because listeners can rely on context. Kraljic and Brennan (2005) suggests that speakers use prosody to disambiguate sentences regardless of the listener's needs. Likewise, Straub's Contingent Cueing Hypothesis (1997) states that prosodic cues to boundaries may be less marked if there are alternative sources of disambiguation (e.g., contextual cues). Interestingly, these studies and proposals have been restricted to native English speakers (with the single exception of the Piantadosi et al. computational analyses covering English, German, and Dutch). Certainly, more crosslanguage research is needed to uncover the language-specific effects of prosodic juncture processing.

CHAPTER 6

– General Discussion –

– General Discussion –

6.1. Summary of Research Findings

The overarching aim of the present thesis was to investigate whether prosodic processing is driven by both language-universal/crosslinguistic and language-specific mechanisms. According to Dwight Bolinger (1978), there exist two aspects of prosody that all listeners and speakers would use in the same way to process speech. First, prosody would be used across all languages to enhance prosodic prominence as signals to semantic salience. Second, prosody would play a crosslinguistic role in the organisation of fluent speech into linguistically significant units. If both of these hypotheses were true, we would expect important implications for speech production and perception across languages with different intonation systems. Inspired by Bolinger's proposals, my graduate research examined how prosody is used both in the encoding of information structure and in structural disambiguation. Across a series of production and perception experiments, we adopted either a crosslanguage or a cross-speaker approach to examine prosodic processing in English and Mandarin Chinese. We compared prosodic strategies in speakers of unrelated languages using similar research designs and materials. Through this approach, our experiments can provide a new perspective on both language production and perception.

6.1.1. *Prosodic Focus Production*

We first examined the production of prosodic focus in English and Mandarin Chinese (**Chapter 2**). Structured dialogues were used because they provide a more experimentally rigorous but still ecologically valid way to elicit natural speech. A methodological issue that often arises from previous experiments involving "laboratory speech" (Xu, 2010) is that participants were often explicitly asked to produce speech in a certain way (e.g., explicitly instructed to produce prosodic focus). Further, some previous experiments involved production of focus in sentences that can be rather unnatural (e.g., "*māomī mō māomī*" "*Kitty touches Kitty"*: Xu, 1999). However, research involving spontaneous speech (i.e., speech that is not read or scripted) can also be problematic because it does not provide much experimental control and comparability across individual speakers and language groups. To address these challenges, our structured dialogue scripts can elicit a more naturalistic form of speech while still maintaining experimental control. Further, instead of explicitly asking participants to produce focus, we elicited focus production by manipulating the pragmatic context of the dialogues. Finally, we created an extensive database of focus production (more than 30,000 measurements) from 24 speakers in each language using five dialogues involving different social situations. The fact that there were significant prosodic differences between the focused and unfocused tokens across the five dialogues indicates that our dialogues were successful in eliciting focus production.

The main question that was asked was whether English and Mandarin speakers would differ in the degree to which they would manipulate the different suprasegmental dimensions to enhance prosodic focus. We discovered both crosslanguage similarities and differences across the five focus types (i.e., new-information, wh-question, corrective, confirmatory, and parallel). Overall, consistent with previous data (e.g., Chen & Gussenhoven, 2008; Xu, 1999; Xu & Xu, 2005), English and Mandarin speakers used similar prosodic means to mark focus; both groups of speakers marked focus through lengthened word duration, increased mean and maximum pitch and pitch range, increased mean and maximum intensity and intensity range. There were, however, systematic crosslanguage differences across the five dialogues in the extent to which the different prosodic strategies were used. Specifically, in all of the crosslanguage differences in F_0 production, the Mandarin speakers produced a greater production increase (p. 42), while the English speakers produced a greater increase in intensity in five out of six of the crosslanguage intensity differences (p. 43). Building on the previous findings, our production findings provide evidence that languages can still differ in the exact degree to which speakers can use various aspects of prosody, even when the overall manner of focus production is highly similar.

This was also the first crosslanguage experiment that examined language differences across more than two types of focus. This would be particularly important for our understanding of basic notions of information structure. Gundel and colleagues, for instance, highlighted a logical distinction between two different notions of information structure (Gundel, 1988; Gundel & Fretheim, 2004), namely "relational givenness vs. newness" and "referential givenness vs. newness". "Relational givenness vs. newness" describes the notion of givenness and focus where they are viewed in relation to each other (e.g., in wh-focus, where one person asks "*Who went home*?" and the other person replies "*JOHN went home*", the information that is was John is assessed relative to the predicate "*went home*"). On the other hand, "referential givenness vs. newness" indicates a relation between the referent of a linguistic expression and its discourse status represented in the mind/attention state of the listener (e.g., in new-information focus, where discourse-new referents can be said to be "brand new", "salient", "out of the blue", or "pragmatically non-recoverable from the preceding contexts"). By looking at different forms of focus (e.g., new-information vs. corrective, wh-question, confirmatory, and parallel), we have examined focus as discourse construct that can be viewed from both a relational and a referential point of view. In all, the production experiment is the first crosslanguage study to examine five different types of focus production from an unusually large sample of speakers using structured dialogues that were highly comparable across different languages.

6.1.2. *Prosodic Focus Perception*

In the next experimental chapter (**Chapter 3**), we explored whether the crosslanguage production differences discussed in the preceding chapter are also reflected in focus perception. Using a phoneme detection task (e.g., Akker & Cutler, 2003; Cutler & Darwin, 1981), previous experiments have demonstrated that native speakers of Germanic languages (e.g., English, Dutch) can entrain with the prosodic contour to predict the location of an upcoming prosodically focused word. Importantly, listeners showed faster response time to phoneme targets when the intonation predicted high stress on the target-bearing word, even when the actual word from both the predicted high and low stress contexts was replaced by an acoustically identical neutral version of the same word. Using this paradigm, we explored whether native speakers of English and Mandarin may still show similarities in prosodic focus perception despite the language-specific differences already observed in production.

We hypothesised three possible outcomes. First, there may be language-specific differences in prosodic entrainment. Native listeners of Australian English may a entrain to prosody to forecast upcoming focus in much the same way as native listeners of British or American English (e.g., Cutler, 1976; Cutler & Darwin, 1981), but native listeners of Mandarin may not engage in the same entrainment strategy. This could be because intonation may be less helpful in a tone language, arguably because competition for the same acoustic dimension with lexical tones may reduce their scope for realisation (e.g., Pierrehumbert, 1999). In focus production, it is well known that prosodic effects on focused words are phonetically layered on existing lexical tones, such that F_0 contour shape remains unchanged but F_0 range becomes exaggerated (e.g., Chen & Gussenhoven, 2008). However, previous studies suggest that *prefocus* cues (i.e., intonation contour before focus) may be no different from a neutral sentence with no prosodic focus (Xu, 1999), and some tones (e.g., the low-dipping third tone) may be more prone to F_0 restriction (Lee et al., 2016). Supporting this view is the observation that competing F_0 contour adjustments by lexical tones and intonation can hinder recognition of different intonation categories (e.g., statement vs. questions: Liu & Xu, 2005; Yuan, 2011). All these findings indicate that Mandarin listeners may not engage in a prosodic entrainment strategy to predict upcoming focus.

Alternatively, Mandarin listeners may still engage in entrainment, only they may do so to a lesser extent than English, either because the intonation contour is less informative, or because no effective use is made of the intonation cues, for instance because speakers of tone languages must process pitch information at a lexical level and are therefore less sensitive to sentence intonation (e.g., Gandour et al., 2003; Gussenhoven & Chen, 2000; Himmelmann & Ladd, 2008).

However, what we discovered was that both English and Mandarin listeners engaged in prosodic entrainment. Our response time data show that both English and Mandarin listeners could make effective use of the prosodic cues in the intonation contour to predict the location of upcoming accents (pp. 78-79). Interestingly, all listeners entrained to the same extent, even when the intonation cues in English stimuli were, according to our acoustic analyses, richer and more robust (pp. 83-84). At the same time, our acoustic analyses of the stimulus sentences showed that pitch range was the only preceding prosodic cue that could reliably predict upcoming accent (i.e., the pitch range in the preceding prosody was significantly larger in high stress context than low stress context). However, there was no significant correlation between prefocus pitch range and the Mandarin listeners' response time (p. 86). This may suggest that the listeners anticipated the upcoming focus using whatever cues were available.

To what extent is this entrainment strategy crosslinguistic? In a later experiment, we observed that native Mandarin speakers no longer engaged in prosodic entrainment in a non-native listening context where the English stimuli were used. This lack of native to non-native transference, in spite of more robust cues in the L2, suggests that prosodic entrainment is acquired as a language-specific strategy. In everyday communication, all listeners regardless of language or culture must adopt a strategy where they can efficiently search for the focused word to navigate the utterance information structure. Contrary to previous findings on intonation perception in native listeners of tone languages, our findings have two important implications for our understanding of prosodic perception. First, listeners are flexible enough to attend to separate and subtle cues wherever they are informative, even when they covary with other linguistic functions. Not only is this true for focus production (e.g., Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015), but it is also the case for the perception of prefocus cues that listeners can use to anticipate and predict the location of upcoming focus, even before it is heard. Second, by looking at both the production and perception of prosodic focus, we demonstrate that there is a disconnect between the crosslanguage variation found in production and the underlying processing strategies in perception.

6.1.3. *Prosodic Focus Perception: A Cross-speaker Perspective*

An interesting extension of our crosslanguage findings concerns prosodic entrainment to the speech of different speakers of Australian English (**Chapter 4**). Previous research suggests that listeners' prediction of upcoming speech forms can be influenced by a variety of distal cues from the preceding prosody, including speech rate (e.g., Dilley & Pitt, 2010), pausing (e.g., Gee & Grosjean, 1984), and rhythmic patterns in pitch and timing (e.g., Dilley, & McAuley, 2008; Morrill, Dilley, McAuley, & Pitt, 2014). However, no research to date has examined whether production of these preceding cues may vary across individual speakers. Moreover, no experiments have used unsynthesised speech stimuli to investigate the role of different types of preceding cues in prosodic entrainment.

To address these issues, we conducted a series of experiments using unsynthesised speech materials recorded by different talkers. This can provide a structured but more naturalistic approach to determine whether listeners can entrain with different prosodic features from the preceding prosody. Across a series of experiments, we revealed that listeners could use whatever cue that was available to forecast an upcoming accent. Listeners showed faster response time in predicted high stress contexts (pp. 122-123), regardless of whether the speaker reliably produced all the preceding cues (i.e., speech rate, F_0, intensity, pre-target interval duration), only some preceding cues (e.g., a combination of F_0 and intensity cues), or only one type of prosodic cue (e.g., pre-target interval duration). However, we also found a case where listeners failed to engage in an entrainment strategy in the speech of a particular talker who only consistently produced speech rate and maximum intensity cues.

Future experiments can explore whether listeners attend to the preceding cues as a combination, rather than attending to each of these cues as a single dimension. Attention to cue combination can facilitate prediction of upcoming sound forms where listeners can generalise the statistical pattern of different prosodic cues in the immediate speech stream. For example, outside of speech processing, research in auditory perception suggests that listeners are sensitive to statistical covariance of different acoustic features. This has been revealed in an AXB discrimination task where complex sounds were found to be processed by collapsing independent but highly correlated acoustic features onto a single perceptual dimension (e.g., Stilp, Rogers, & Kluender, 2010). Similarly, in tone perception, native listeners of Northern Vietnamese, where tones are cued by a combination of pitch and voice quality, are more likely to confuse tone with similar pitch excisions compared to native listeners of Southern Vietnamese, where tones are purely pitch-based (Kirby, 2010).

Attending to statistical covariance is a useful crosslinguistic strategy because redundancy is a common feature of all languages. Prominent syllables in English are marked by a number of co-varying cues, including greater intensity, longer duration, and higher pitch and pitch range expansion. However, some suprasegmental features may be less systematically correlated in other languages, such as French, where the last syllables with rising or falling pitch are also longer but not necessarily louder (Vaissière, 1983), or Japanese, where accented morae have little effect on duration or intensity (Beckman, 1982; McCawley, 1968). Future experiments can build on these ideas to examine how listeners engage in focus prediction under conditions where different prosodic dimensions in the sentence are manipulated to covary in different ways. How listeners across different languages exploit the various prosodic features may depend on the degree to which they are interrcorrelated in their native language. On the other hand, like auditory perception, all listeners may also start to attend to the remaining variance from the deviating feature after extended exposure. The latter may be a useful strategy across languages because individual talkers can vary in the kinds of prosodic cues they produce.

6.1.4. *Prosodic Juncture Production*

The last experimental chapter aimed to compare how native speakers of English and Mandarin use prosodic cues to juncture to disambiguate speech in speaking and listening (**Chapter 5**). In the production component of this study, we examined whether both groups of speakers may differ in their production of different prosodic cues to juncture. We tested production of juncture cues using pairs of segmentally identical sentences that can convey different meanings depending on the timing and location of different the critical juncture. In “Early Juncture” sentences, the boundary occurred earlier in the sentence (e.g., *“He gave her # dog biscuits”, “他给她#狗饼干”*), where the feminine singular pronoun before the early juncture “*her*” or “*她*” is an indirect object and the juncture preceded a compound word (i.e., *“dog biscuits*” “*狗饼干*”). In “Late Juncture” sentences, however, the feminine pronoun was a possessive determiner (e.g., *“He gave her dog # biscuits”, “他给她狗# 饼干”*). We tested production of these cues by asking speakers to read aloud different reading passages where the storyline provided the contextual information that could inform the meaning of the ambiguous sentences. The English and Mandarin reading materials were highly comparable because they contained the same storylines and ambiguous sentences with exactly the same structural ambiguity and with the same set of meanings. Further, in many cases, our English and Mandarin sentences contained postboundary word onsets with the same or similar manner or place of articulation (e.g., /hɜ: dɔg # bɪskəts/ vs. /$t^{h}a_{1}$ kou_{3} # $pin_{3}kan_{1}$/).

Our production results show that speakers in both English and Mandarin still use some prosodic cues to disambiguate speech, even when the referential context provided by the storylines made the use of prosody for syntactic disambiguation redundant. Importantly, we extended previous production findings by showing that speakers can differ in the prosodic juncture cues they would choose to realise in conditions where prosodic production of these cues is optional. Like many languages, both English and Mandarin speakers are capable of producing a combination of different prosodic cues to mark juncture: pausing, boundary-related durational lengthening, preboundary F_0 lowering, postboundary F_0 reset, and postboundary segmental strengthening (e.g., Cooper & Paccia-Cooper, 1980; Keating, Cho, Fougeron, & Hsu, 2003; Kuang, 2010 Li & Yang, 2009; Lieberman & Pierrehumbert, 1984; Maeda, 1976; Liao, 1994; Silverman, 1987; Shen, 1985; Shih, 1988; Tseng, 1981). However, it is important to note that previous experiments that looked at these cues involved experimentally manipulated settings and, in many cases, juncture production from trained speakers. Extending previous studies, we here provide evidence that languages can still differ in the types of prosodic juncture cues that speakers persisted in producing despite the available contextual cues (pp. 176-177). In addition, like our crosslanguage focus production experiment, we also found instances where English and Mandarin speakers differed in the exact degree to which they enhance different prosodic cues (pp. 178-181).

6.1.5. *Prosodic Juncture Perception*

Unlike the materials used for production, the perception experiments involved a disambiguation task where only prosody was available as a juncture cue for disambiguating the "Early" and "Late Juncture" sentences. Here, we observed that English and Mandarin listeners disambiguated the sentences differently as a function of the different juncture locations (p. 193). English listeners were significantly faster at disambiguating sentences with late junctures (e.g., "*He gave her dog # biscuits*") compared to sentences with early junctures (e.g., "*He gave her # dog biscuits*"). On the other hand, Mandarin listeners showed the exact opposite results where they were faster at disambiguating sentences with early junctures compared to late junctures. Moreover, there were more incorrect disambiguations in English than Mandarin (p. 194).

Why are "Early Juncture" sentences disambiguated slower than "Late Juncture" sentences in English? One possible reason could be the juncture cues in "Early Juncture" (e.g., "*He gave her # dog biscuit*") sentences are more optional than those in the "Late Juncture" (e.g., "*He gave her dog # biscuit*"). A second reason could be because "Early

Juncture" sentences convey a more outlandish meaning (e.g., giving dog biscuits to another person), but this was unlikely to have been the reason, because we found the converse effect in Mandarin listeners. Another reason could be that the different number of alternative sentences that can be constructed to convey the same meaning of the "Early" and "Late Juncture" sentences. In everyday communication, there are probably more alternative ways of signalling the meaning conveyed by the "Early Juncture" sentences than the "Late Juncture" sentences. "Early Juncture" sentences like "*He gave her # dog biscuits*" can be alternatively expressed as *"He gave some dog biscuits to NAME", "He gave NAME some dog biscuits", or "He gave some dog biscuits to he"* (3 alternative sentence constructions). The "Late Juncture" version of the same sentence can alternatively be expressed as "*He gave NAME's dog some biscuits*" or "*He gave biscuits to NAME's dog*", but "*He gave biscuits to it*" is ungrammatical, so there are at least 2 alternative ways to express the same meaning. Perhaps similar to processing multiple word candidates during word recognition, sentence processing may be slower if there are more alternative ways to express the same meaning.

In Mandarin, the response time difference across the juncture versions was in the opposite direction: listeners disambiguated the "Late Juncture" sentences faster than the "Early Juncture" sentences. Again, this could also be due to the number of alternative constructions that could be used to express the same meaning of the "Late" and "Early Juncture" sentences. The meaning conveyed by the "Late Juncture" sentences (e.g., "*他给她狗 # 饼干*") could alternatively be produced in a way more in line with its citation form where the optional possessive particle "*的*" *-de* is present (n.b., native speakers tend to ignore alienable vs. inalienable distinctions when they omit the possessive particle in colloquial speech). In both "Early" and "Late Juncture" sentences, alternative ways of expressing the same meaning could be done by using the "*把*" *ba-* construction (i.e., "*He* *ba- dog biscuits/biscuits give recipient*"), only this time, there are more alternative ways to express the meaning of the "Late Juncture" sentences because of the optional use of the possessive particle. This may explain why Mandarin listeners needed longer time to disambiguate the "Late Juncture" sentence. Perhaps processing the meaning of an utterance may involve hypothesis-testing between multiple sentence meaning candidates.

There may also be a more language-specific reason for the slower disambiguation of the "Late Juncture" sentences in Mandarin. Processing sentences with no possessive markers may require availability of proper information structure. According to formal analysis by Hsu (2009), sentences with alienable possessum (e.g., "*电影*" "*movie*") can be used without the optional possessive marker *-de* and still be considered acceptable when the object possessor (e.g., "*李安*" "*Ang Lee*") is topicalised (e.g., "*李安, 我看过 [他(的)不少电影]*" "*Speaking of Ang Lee, I've seen [several of (his) movies]*") or when the pragmatic context evokes alternatives. Mandarin listeners may therefore need to integrate other linguistic cues (e.g., information structure) with prosodic juncture cues to disambiguate the "Late Juncture" sentences. Since prosody was the only source of disambiguation cue in our perception experiment, the Mandarin listeners might have needed more time to disambiguate the "Late Juncture" sentences.

Another question that we addressed was whether there were crosslanguage differences in the degree to which listeners attend to pausing cues in juncture perception. Consistent with previous experiments (e.g., Aasland & Baum, 2003; Yang, Shen, Li, & Yang, 2014), English listeners' disambiguation performance and response time pattern remained largely unchanged after the pausing cue was rendered uninformative, but there was a decrease in accuracy in Mandarin. Importantly, the accuracy performance in Mandarin decreased to a level that was comparable with that of the English listeners during both the first experiment (when all juncture cues were intact) and the second experiment where the pausing cue was artificially taken out. This indicates that the better accuracy performance by Mandarin listeners observed in the first experiment would have been due to the availability of the pausing cue.

Why are some listeners more reliant on pausing cues than others? One testable hypothesis is that languages may differ in the degree to which their sound systems produce ambiguity. We speculated that ambiguity may be more prevalent in Mandarin. For instance, many of the consonant contrasts in Mandarin require manipulation of duration cues (e.g., aspirated vs. unaspirated VOT). Lexical tones with different intrinsic pitch contour shapes, duration, and amplitude may also provide less suprasegmental space for the prosodic expressions of juncture cues. Furthermore, identical segments produced in the same tone can convey a multitude of different meanings (e.g., /ʂu1/ can indicate at least 40 different characters/words including singleton words such as “*book*”, *“uncle*”, “*comb*”, *“halberd”, or “door hinge”*). To avoid competition between prosodic and non-prosodic uses of the same cues, Mandarin speakers may devise a strategy where they can rely on pause duration as a useful way to mark boundaries without sacrificing the temporal or pitch cues of the segments.

6.2. Closing Statement

So is prosody really “around the edge of language”? Although prosody is not physically part of the segments or the syntax, it is most certainly a central part of language processing. Here, our production and perception experiments demonstrated how prosody can serve a crucial role in the language user’s ability to process the utterance information structure and organise speech into meaningful units. We have discovered how the use of prosody in languages with different intonation systems can both differ and resemble each other in speech processing. Even when prosody may be produced in the same way across languages, there can still be subtle differences in the degree to which speakers use different prosodic dimensions. Even if languages differ in prosodic production, there may still be a disconnect between the language variation found in production and the underlying processing strategies observed in perception. At the same time, prosodic cues covary with other linguistic functions, and language users are still flexible enough to attend to separate and subtle cues whenever they are informative. Prosodic processing involves a complex interplay between crosslinguistic and language-specific mechanisms.

There are also many unexplored questions in prosodic research that have not been addressed in this thesis. In addition to crosslanguage research, more research is needed to explore how prosody should be taught in language education (e.g., Jackson & O'Brien, 2011; Szczepek Reed, 2015; Yenkimaleki & van Heuven, 2016), how prosody is processed after ingesting alcohol (e.g., Cutler & Henton, 2004), how prosody is processed as a result of sleep deprivation (e.g., Deliens et al., 2015) or psychological stress (e.g., Paulmann, Furnes, Bøkenes, & Cozzolino, 2016), and how prosody is processed by nonhuman animals (e.g., Colbert-White, Tullis, Andresen, Parker, & Patterson, 2018). Answers to these questions will have a great potential, not only for our theoretical understanding of language structure and use, but also for how we can use supralexical aspects of speech to promote everyday communication. Prosody may not be a physical part of the speech segments, but it is central to language processing.

REFERENCES

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: 10.1016/S0093-934X(03)00138-X

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition, 6*, 81-96. doi:10.1017/S1366728903001056

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.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of sentence reference. *Cognition, 73*, 247– 264.

Andreeva, B., Koreman, J., & Barry, W. (2016): Local and global cues in the prosodic realization of broad and narrow focus in Bulgarian. *Phonetica, 73*, 256-278. doi: 10.1159/000448044

Arai, M., van Gompel, R. P., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology, 54*, 218–250. doi: 10.1016/j.cogpsych.2006.07.001

Arnhold, A. (2016). Complex prosodic focus marking in Finnish: expanding the data landscape. *Journal of Phonetics, 56*, 85-109. doi: 10.1016/j.wocn.2016.02.002

Arvaniti, A., & G. Garding . 2007. Dialectal variation in the rising accents of American English. In J. Cole & J. H. Hualde (Eds.), *Papers in Laboratory Phonology 9* (pp. 547-576). Berlin, Germany: De Gruyter Mouton.

Austin, J. L. (1962). *How to do things with words?* Oxford, UK: Oxford University Press.

Avesani, C., & Vayra, M. (2005). Accenting, deaccenting and information structure in Italian dialogues. In L. Dybkjaer & W. Minker (Eds.), *Proceedings of the 6th DIGdial Workshop on Discourse and Dialogue* (pp. 19-24). Lisbon, Portugal.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baese-Berk, M. M., Dilley, L. C., Henry, M. J., Vinke, L., & Banzina, E. (2018). Not just a function of function words: Distal speech rate influences perception of prosodically weak syllables. *Attention, Perception, and Psychophysics*. doi: 10.3758/s13414-018-1626-4

Baese-Berk, M., Heffner, C., Dilley, L., Pitt, M., Morrill, T., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25*, 1546-155. doi: 10.1177/0956797614533705

Baltaxe, C. A. M. (1983). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research, 27*, 97-105. doi: 10.1044/jshr.2701.97

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48. doi:10.18637/jss.v067.i01

Beckman, M. (1982). Segment duration and the 'mora' in Japanese. *Phonetica, 39*, 113-135. doi: 10.1159/000261655

Beckman, M. (1986). *Stress and non-stress accent*. Dordrecht, The Netherlands: Foris.

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

Beckman, M. E. (1997) A typology of spontaneous speech. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 7-26). New York, NY: Springer.

Beckman, M. E., & Edwards, J. (1990) Lengthening and shortening and the nature of prosodic constituency. In J. Kingston and 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: 10.1017/S095267570000066X

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*, 289-300.

Berkovits R. (1993). Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics*. *21*, 476-489.

Best, C. T. (1994). Learning to perceive the sound pattern of English. In C. Rovee-Collier, & L. Lipsitt (Eds.), *Advances in infancy research (Volume 8)* (pp. 217-304). Hillsdale, NJ: Ablex Publishers.

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech, 46*, 183-216. doi: 10.1177/00238309030460020701

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38*, 365-391. doi: 10.1177/002383099503800403

Birch, S., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language, 34*, 232-267. doi: 10.1006/jmla.1995.1011.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*, 37-49.

Blutner, R., & Sommer, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language, 27*, 359-367. doi: 10.1016/0749-596X(88)90061-7

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: 10.1006/jmla.1995.1034

Bolinger, D. L. (1958). A theory of pitch accent. *Word, 14*, 109-149.

Bolinger, D. L. (1964). Around the edge of language: Intonation. Harvard Educational Review, 34, 282-296. doi: 10.17763/haer.34.2.4474051q78442216

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.

Bolinger, D. L. (1986). *Intonation and its parts: Melody in spoken English*. Palo Alto, CA: Stanford University Press.

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes, 25*, 1024-1043. doi: 10.1080/01690960903036836

Breen, M., Dilley, L. C., McAuley, J. D., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech. *Language, Cognition and Neuroscience, 29*, 1131-1146. doi: 10.1080/23273798.2014.894642

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*, 1044-1098. doi: 10.1080/01690965.2010.504378

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sounds*. MIT Press: Cambridge, MA.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93-125. doi: 10.1016/S0010-0277(96)00719-6

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychometric Bulletin and Review, 18*, 1189-1196. doi: 10.3758/s13423-011-0167-9

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 306-323. doi: 10.1080/23273798.2014.894642

Brunellière, A., Auran, C., & Delrue, L. (in press). Does the prosodic emphasis of sentential context cause deeper lexical-semantic processing? *Language, Cognition and Neuroscience*. doi: 10.1080/23273798.2018.1499945

Burdin, R. S., Phillips-Bourass, S. P., Turnbull, R., Yasavul, M., Clopper, C. G. & Tonhauser, J. (2015). Variation in the prosody of focus in head- and head/edge-prominence languages. *Lingua, 165, Part B*, 254-276. doi: 10.1016/j.lingua.2014.10.001

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica, 57*, 3-16. doi: 10.1159/000028456

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: 10.1017/s0952675700001019

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modelling the dynamics of boundary- adjacent lengthening. *Journal of Phonetics, 31*, 149-180. doi: 10.1016/S0095-4470(02)00085-2

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.

Campbell, W. N., & Isard, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics, 19*, 37-47.

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Individual differences in speech production and perception* (pp. 123-145). Frankfurt, Germany: Peter Lang International Academic Publishers. doi: 10.3726/978-3-653-05777-5

Cao, J. (1992). On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics, 20.*

Cao J. (2012). Pitch prominence and tonal typology for low register tone in Mandarin. In *Proceedings of the 3rd International Symposium on Tonal Aspects of Languages*. Nanjing, China.

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: 10.1016/j.specom.2005.02.013

Caspers, J., Bosma, E., Kramm, F., & Reya, P. (2012). Deaccentuation in Dutch as a second language: Where does the accent go to? *Linguistics in the Netherlands, 29*, 27-40.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 27-55). New York: Academic Press.

Chafe, W. (1987) Cognitive constraints and information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse: Outcome of a Symposium* (pp. 21-51). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chávez-Peón, M. (2010). *The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec*. (Ph.D. dissertation). University of British Columbia, Vancouver, Canada.

Chen, A. (2012). The prosodic investigation of information structure. In M. Krifka & R. Musan. (Eds.), *The expression of information structure* (pp. 249-286). Berlin, Germany: De Gruyter Mouton.

Chen, G. T. (1972). *A comparative study of pitch range of native speakers of Midwestern English and Mandarin Chinese: An acoustic study* (Doctoral dissertation, University of Wisconsin-Madison, Madison, USA).

Chen, S. (2005). The effects of tones on speaking fundamental frequency and intensity ranges in Mandarin and Min dialects. *Journal of the Acoustical Society of America, 117*, 3225–3230.

Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics, 34*, 176-201. doi: 10.1016/ j.wocn.2005.05.002

Chen, Y., & Braun, B. (2006). Prosodic realization in information structure categories in standard Chinese. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of Speech Prosody*. Dresden: TUD Press.

Chen, Y., & Gussenhoven, C. 2008. Emphasis and tonal implementation in Standard Chinese. Journal of Phonetics 36:724-746. doi: 10.1016/j.wocn.2008.06.003

Chen, Y., Xu, Y., & Guion-Anderson, S, (2014). Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica 71*, 249-270. doi: 10.1159/000371891

Cho, T. (2011). Laboratory phonology. In N. C. Kula, B. Botma, & K. Nasukawa (Eds.), *Bloomsbury companion to phonology* (pp. 343–368). London/New York: The Bloomsbury Continuum.

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

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*,193-228. doi: 10.1006/jpho.2001.0153

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155-190. doi: 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: 10.1016/j.wocn.2009.08.001

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics, 39*, 344-361. doi: 10.1016/j.wocn.2011.02.005

Cho, T., Lee, Y., & Kim, S. (2014). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics, 46*. 128-146. doi: 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: 10.1016/j.wocn.2005.01.001

Cho, T., McQueen, J., & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: the case of domain-initial strengthening in English. *Journal of Phonetics, 35*, 210-243. doi:10.1016/j.wocn.2006.03.003

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper and Row.

Choudhury, A., & Kaiser, E. (2016). Interaction between prosody and focus types: Evidence from Bangla and Hind. In R. Balusu and S. Sundaresan (Eds.), *Proceedings of formal approaches to South Asian languages* (pp. 176-195).

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of*

Speech, Language, and Hearing Research, 57, 1468-1479. doi: 10.1044/2014_JSLHR-L-13-0279

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*, 221–268. doi: 10.1080/016909698386528

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: 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.

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 39*, 237-245. doi: 10.1016/j.wocn.2011.02.006

Colbert-White, E.N., Tullis, A., Andresen, D.R., Parker, K.M., & Patterson, K.E. (2018). Can dogs use vocal intonation as a social referencing cue in an object choice task? *Animal Cognition, 21*, 253-265. doi: 10.1007/s10071-018-1163-5

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*, 207-228. doi: 10.1177/00238309020450030101

Cooper, W. E., Eady, S. J. & Mueller, P. R. (1985) Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America, 77*, 2142-2156. doi: 10.1121/1.392372

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Connell, B. (2017). Tone and intonation in Mambila. In L. J. Downing, & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 132-166). Berlin: De Gruyter.

Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (2018). English learners' use of segmental and suprasegmental cues to stress in lexical Access: An eye-tracking study. *Language Learning,* 1-34. Doi: 10.1111/lang.12288

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.

Cruttenden, A. (1993). The de-accenting and re-accenting of repeated lexical items. In D. House & P. Touati (Eds.), *Proceedings of the ESCA workshop on Prosody* (pp. 16–19). Lund, Sweden: Reprocentralen Lund University.

Cruttenden, A. (2006). The de-accenting of old information: A cognitive universal? In G. Bernini & M. L. Schwartz (Eds.), *The pragmatic organization of discourse in the languages of Europe* (pp. 311-356). Berlin, Germany: De Gruyter Mouton.

Cutler, A., (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics, 20*, 55-60.

Cutler, A. (1982). Prosody and sentence perception in English. In J. Mehler, E. C. Walker, & M. Garrett (Eds.), *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities* (pp. 201-216). Hillsdale, N.J: Erlbaum.

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

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua, 92*, 81-104. doi: 10.1016/0024-3841(94)90338-7

Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 87-99). Mahwah, NJ: Erlbaum.

Cutler, A. (1997). Prosody and the structure of the message. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 63-66). Heidelberg, Germany: Springer.

Cutler, A. (2012). *Native listening: Language experience and recognition of spoken words*. MIT Press: Cambridge, MA.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics, 59*, 165-79. doi: 10.3758/BF03211886

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception and Psychophysics, 29*, 217-224.

Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science, 13*, 258 -262. doi: 10.1111/1467-9280.00447

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7*, 49–59. doi: 10.1016/0010-0277(79)90010-6

Cutler, A., & Foss, D.J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20*, 1-10. doi: 10.1177/002383097702000101

Cutler, A., & Henton, C. G. (2004). There's many a slip 'twixt the cup and the lip. In H. Quené & V. van Heuven (Eds.) *On speech and language: Studies for Sieb G. Nooteboom* (pp. 37-45). Utrecht, The Netherlands: Landelijk Onderzoekschool Taalwetenscha.

Cutler, A., Klein, W., & Levinson, S. C. (2005). The cornerstone of twenty-first century psycholinguistics. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cutler, A., & Ladd, D.R. (Eds.) (1983). *Prosody: Models and measurements*. Heidelberg: Springer.

Cutler, A., Mehler, J., Norris, D., Seguí, J. (1983). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385-400. doi: 10.1016/0749-596X(86)90033-1

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113–121. doi: 10.1037/0096-1523.14.1.113

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America, 105*, 1877-1888. doi: 10.1121/1.427141

Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language, 46*, 296-322. doi: 10.1006/jmla.2001.2814

Cutler, A., Otake, T., & Bruggeman, L. (2012). Phonologically determined asymmetries in vocabulary structure across languages. *Journal of the Acoustical Society of America, 132*, EL155-EL160. doi:10.1121/1.4737596.

Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language, 14*, 145-167. doi: 10.1017/S0305000900012782

Cutler, A., & van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Speech and Language, 44*, 171-195. doi: 10.1177/00238309010440020301

Dahan, D., & Bernard, J. -M. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech, 39*, 341-374. doi: 10.1177/002383099603900402

Dalton, P., & Hardcastle, W. J. (1977). *Disorders of fluency and their effects on communication*. London, UK: Elsevier.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language, 47*, 292-314. doi: 10.1016/S0749-596X(02)00001-3

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: constraints on formant perception. *Journal of the Acoustical Society of America, 76,* 1636-47. doi: 10.1121/1.391610

Darwin, C. J. (2007). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B Biological Sciences, 363,* 1011-1021. doi: 10.1098/rstb.2007.2156

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218 –244. doi: 10.1037/0096-1523.28.1.218

Degenshein, R., & Chitoran, I. (2001). Dholuo interdentals: Fricatives or affricates? Evidence from domain-initial strengthening. *Journal of the Acoustical Society of America, 115*, 2542. Retrieved from https://www.researchgate.net/publication/252296463_Dholuo_interdentals_Fricatives_or_affricates_Evidence_from_domain-initial_strengthening

Deliens, G., Stercq, F., Mary, A., Slama, H., Cleeremans, A., Peigneux, P., Kissine, M., 2015. Impact of acute sleep deprivation on sarcasm detection. PLOS ONE 10, e0140527. doi: 10.1371/journal.pone.0140527

Demuth, K. & Tremblay, A. 2007. Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language, 34*, 1-29.

Demuth, K., McCullough, E., & Adamo, M. (2007). The Prosodic (re)organization of determiners. In H. Caunt-Nulton, S. Kulatilake, & I. Woo (Eds.), *Proceedings of the 31st annual Boston University Conference on Language Development (BUCLD 31) (Vol. 1, pp. 196-205)*. Somerville, MA: Cascadilla Press.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics, 32*, 493-516. doi: 10.1016/j.wocn.2004.05.002

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.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. S. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America, 119*, 719-722. doi: 10.1121/1.2151799

DiCanio, C., & Hatcher, R. (2018). On the non-universality of intonation: Evidence from Triqui. *Journal of the Acoustical Society of America, 144*, 1941-1941. doi: 10.1121/1.5068494

DiCanio, C., Benn, J., & García, R. C. (2018). The phonetics of information structure in Yoloxóchitl Mixtec. *Journal of Phonetics, 68*, 50-68. doi: 10.1016/j.wocn.2018.03.001

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*, 294-311. doi: 10.1016/j.jml.2008.06.006

Dilley, L. C., & Pitt, M. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*, 1664-1670. doi: 10.1177/0956797610384743

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*, 274–294. doi: 10.1016/j.jml.2010.06.003

Dilley, L.C., Morrill, T., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalizability. *Frontiers in Language Sciences, 4*, 1-13. doi: 10.3389/fpsyg.2013.01002

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: 10.1006/jpho.1996.0023

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

Downing, L. J., Mtenje, A., & Pompino-Marschall, B. (2004). Prosody and information structure in Chichewa. In S. Fuchs & S. Hamann (Eds.), *ZASPiL 37, Papers in Phonetics and Phonology* (pp. 167–187). Berlin, Germany: ZAS.

Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford, UK: Oxford University Press.

Duanmu, S. (2004). Tone and non-tone languages: An alternative to language typology and parameters. *Language and Linguistics, 5*, 891-924.

Duběda, T., & Mády, K. (2010). Nucleus position within the intonation phrase: a typological study of English, Czech and Hungarian. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association* (pp. 126-129). Makuhari, Japan.

Dwyer, M. 2004. More is better: The impact of study abroad program duration. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*, 151-163.

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

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36*, 680-703. doi: 10.1016/j.wocn.2008.05.001

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development, 81*, 472-479. doi: 10.1111/j.1467-8624.2009.01408.x

Fiedler, I., & Jannedy, S. (2013). Prosody of focus marking in Ewe. *Journal of African Linguistics, 34*, 1-46. doi: 10.1515/jall-2013-0001

Fear, B. D., Cutler, A., & Butterfield, S. (1994). The strong/week syllable distinction in English. *Journal of the Acoustical Society of America, 97*, 1893-1904. doi: 10.1121/1.412063

Félix-Brasdefer, J. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language Learning, 54*, 587-653.

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development, 67*, 3192–3218. doi: 10.2307/1131774

Flemming, E. (2008). The role of pitch range in focus marking. Slides from a talk given at the Workshop on Information Structure and Prosody, Studiecentrum Soeterbeeck, Ravenstein, Netherlands.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics, 29*, 109-135. doi.10.006/jpho.2000.0114

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: 10.1121/1.418332

Fouquet, M., Pisanski, K., Mathevon, N., & Reby, D. (2016). Seven and up: Individual differences in male voice fundamental frequency emerge before puberty and remain stable throughout adulthood. *Royal Society Open Science, 3*, 160395. doi: 10.1098/rsos.160395

Fowler, C. A., & Housum, J. (1987). Talker's signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504. doi: 10.1016/0749-596X(87)90136-7

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

Fraundorf, S., Watson, D., & Benjamin, A. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory & Language, 63*, 367–386. doi: 10.1016/j.jml.2010.06.004.

Frota, S. (2000). *Prosody and focus in European Portuguese*. New York, NY: Garland.

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America, 104*, 505-510. doi: 10.1121/1.413004

Fung, H. S. H., & Mok, P. (2018). Temporal coordination between focus prosody and pointing gestures in Cantonese. *Journal of Phonetics, 71*, 113-125. doi: 10.1016/j.wocn.2018.07.006

Fry, D. B. (1955), Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America 27*, 765-768. doi: 10.1121/1.1908022

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*, 126-152. doi: 10.1177/002383095800100207

Gandour, J., Dzemidzic M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., & Lurito J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language, 84*, 318-336. doi: 10.1016/S0093-934X(02)00505-9

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*, 468-492. doi: 10.1006/jmla.2000.2784

Garrod S., & Pickering M. J. (2004) Why is conversation so easy? *Trends in Cognitive Sciences, 8*, 8-11. doi: 10.1016/j.tics.2003.10.016

Gaskell, M. G. & Marslen-Wilson, W. D. (1997) Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes 12*, 613–56. doi: 10.1080/016909697386646

Gee, J. P., & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science, 8*, 59–84. doi: 10.1016/S0364-0213(84)80025-7

Gendrot, C., Gerdes, K., & Adda-Decker, M. (2011). Impact of prosodic position on vocalic space in German and French. In *Proceedings of the 17th international congress of phonetic sciences* (pp. 731–734). Hong-Kong, China.

Genzel, S., Ishihara, S., & Surányi, B. (2015). The prosodic expression of focus, contrast and givenness: a production study of Hungarian. *Lingua, 165*, 183-204. doi: 10.1016/j.lingua.2014.07.010

Genzel, S., Renans, A., & Kügler, F. (2018). Focus and its prosody in Akan and Ga. . In *Proceedings of the 9th International Conference on Speech Prosody* (pp. 724-728). Poznan, Poland.

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: 10.1016/j.wocn.2014.02.006

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: 10.1044/2016_JSLHR-S-15-0044

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications, 4*, Article 1490. doi: 10.1038/ncomms2430

Gleitman, L., & Wanner, E. (1982). Language acquisition: The state of the art. In E. Wanner, & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3-48). Cambridge, UK: Cambridge University Press.

Goedemans, R.W.N., & E. van Zanten. (2007). Stress and accent in Indonesian In V. J. van Heuven & E. van Zanten (Eds.), *Prosody in Indonesian languages* (pp. 35-62). Utrecht, The Netherlands: LOT (Netherlands Graduate School of Linguistics).

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory Language, 28*, 508–518.

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

Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. H. A., Cattarossi, L., Nespor, M., & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences, 111*, 5837-5841. doi: 10.1073/pnas.1318261111

Gordon, M. (2007). The intonational realization of contrastive focus in Chickasaw. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 69-82). Dordrecht, The Netherlands: Springer.

Gordon, M. K., & Roettger, T. B. (2017). Acoustic correlates of word stress: a cross-linguistic survey. *Linguistic Vanguard, 3*, 1-11. doi: 10.1515/lingvan-2017-0007

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: effects of age and experimental procedure. *Frontiers in Psychology, 9*, 477. doi: 10.3389/fpsyg.2018.00477

Gow, D, W., and McMurray, B. (2007) Word recognition and phonology: The case of English coronal place assimilation. J.S. Cole & J. Hualdo (Eds.) *Papers in Laboratory Phonology 9*. (pp. 173-200). New York: Mouton de Gruyter.

Gout, A., Christophe, A., & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access. II: Infant data. *Journal of Memory and Language, 51*, 548-567. doi: 10.1016/j.jml.2004.07.002

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

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech, 46*, 375–401. doi: 10.1177/00238309030460040201

Greif, M. (2010). Contrastive focus in Mandarin Chinese. In *Proceedings of Speech Prosody 2010* (pp. 2-5). Chicago, USA.

Grosjean. F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology, 11*, 58-8 1.

Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. Moravscik, & J. Wirth (Eds.), *Studies in syntactic typology* (pp. 209-239). Amsterdam, The Netherlands: John Benjamins.

Gundel, J. K., & Thorstein. F. (2004). Topic and focus. In Horn, R. Laurence, & G. Ward (Eds.). *The handbook of pragmatics* (pp. 175-196). Oxford, UK: Blackwell.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274-307. doi: 10.2307/416535

Gu, W., & Lee, T. (2007). Effects of focus on prosody of Cantonese speech: A comparison of surface feature analysis and model-based analysis. In *Proceedings of the International Workshop Paralinguistic Speech 2007*. Saarbrücken, Germany.

Gu, Z., Mori, H., & Kasuya, H. (2003). Prosodic variations in disyllabic meaningful words with different stress patterns in Mandarin Chinese. *Acoustical Science and Technology, 24*, 111-119.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274-307. doi: 10.2307/416535

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics, 19*, 377-419. doi: 10.1017/S0022226700007799

Gussenhoven, C. (2000). On the origin and development of the Central Franconian tone contrast. In A. Lahiri (Ed.). *Markedness and change*. Berlin, Germany: Mouton de Gruyter.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In *Proceedings of the 1st International Conference on Speech Prosody* (pp. 47-57). Aix-en-Provence, France: ISCA.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In B. Yuan, T. Huang & X. Tang (Eds.), *Proceedings of The Sixth International Conference on Spoken Language Processing, Vol. I* (pp. 91-94). Beijing, China: China Military Friendship Publish.

Gussenhoven, C., & Rietveld, A.C.M. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics, 16*, 355-369. Retrieved from https://www.researchgate.net/publication/246272347_Fundamental_frequency_declination_in_Dutch_Testing_three_hypotheses/download

Gussenhoven, C., & Rietveld, T. (1999). On the speaker dependence of the perceived prominence of F0. *Journal of Phonetics, 26*, 371-380. doi: 10.1006/jpho.1998.0080

Gussenhoven, C., & Rietveld, T. (2000). The behavior of H* and L* undervariations in pitch range in Dutch rising contours. *Language and Speech, 43*, 183-203. doi: 10.1177/00238309000430020301

Gussenhoven, C., & Teeuw, T. (2007). A moraic and a syllabic H-tone in Yucatec Maya. In E. Herrea Z., & P. M. Butrageño (Eds.), *Fonología instrumental: Patrones fónicos y variación lingüística* (pp. 49–71). Mexico City: Colegio de México.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English, Part 2. *Journal of* Linguistics*, 3*, 199-244. doi: 10.1017/S0022226700016613

Hamlaoui, F., Zygis, M., Engelmann, J., & Wagner, M. (2018). Acoustic correlates of focus marking in Czech and Polish. *Language and Speech*. doi: 10.1177/0023830918773536

Harrington, J., Cox, F., & Evans, Z. (1997) An acoustic study of broad, general and cultivated Australian English vowels. *Australian Journal of Linguistics, 17*, 155-184. doi: 10.1080/07268609708599550

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.

Hart, J. T., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge, UK: Cambridge University Press

Hartmann, K., Zimmermann, M., 2007a. In place – out of place? Focus in Hausa. In K. Schwabe & S. Winkler (Eds.), *On information structure, meaning and form:*

Generalizing across languages (pp. 365-403). Amsterdam, The Netherlands: Benjamins.

Hawkins, S., & P. Warren. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics, 22*, 493-511.

Hay, J.F., Sato, M., Coren, A.E., Moran, C.L., & Diehl, R.L. (2006) Enhanced contrast for vowels in utterance focus: A cross-language study. *Journal of the Acoustical Society of America, 119*, 3022-3033. doi: 10.1121/1.2184226

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago, IL: Chicago University Press.

Hayes, B., & Lahiri, A. (1991). Bengali Intonational Phonology. *Natural Language and Linguistic Theory, 9*, 47-96. doi: 10.1007/BF00133326

Hedberg, N., & Sosa, J. M. (2008). The prosody of topic and focus in spontaneous English dialogue. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Crosslinguistic perspectives on meaning and intonation* (pp. 101-120). Dordrecht, The Netherlands: Springer.

Heffner, C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes, 28*, 1275–1302. doi: 10.1080/01690965.2012.672229

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics, 31*, 39-62. doi: 10.1016/S0095-4470(02)00071-2

Hellmuth, S. (2005). No deaccenting in (or of) phrases: Evidence from Arabic for cross-linguistic and cross-dialectal prosodic variation. In S. Frota, M. Vigaro, & M.

Freitas (Eds.), *Prosodies: With special reference to Iberian languages* (pp. 99-121). Berlin, Germany: De Gruyter Mouton.

Herman, R. (1996). Final lowering in Kipare. *Phonology, 13*, 171-196.

Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with "articulatory" pauses. *Language and Speech, 26*, 203-214. doi: 10.1177/002383098302600302

Himmelmann, N. P., & Ladd, D. R. (2008). Prosodic description: An introduction for fieldworkers. *Language Documentation and Conservation, 2*, 244-274.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 269–286. doi: 10.1016/S0010-0277(87)80002-1

Hirst, D., & Di Cristo, A. (1998). *Intonation system: A survey of twenty languages*. Cambridge, UK: Cambridge University Press.

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

Hoeschele, M., & Fitch, W. T. (2016). Phonological perception by birds: budgerigars can perceive lexical stress. *Animal Cognition, 19*, 643. doi:10.1007/s10071-016-0968-3

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: 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. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.5554&rep=rep1&type=pdf

House, D., Hermes, D., & Beaugendre, F. (1998). Perception of tonal rises and falls for accentuation and phrasing in Swedish. In R.H. Mannell & J. Robert-Ribes (Eds.), Proceedings of the International Congress of Phonetic Sciences (pp. 2799–2802). Sydney.

House, D., Karlsson, A., Svantesson, J. -O., & Tayanin, D. (2009). The phrase-final accent in Kammu: Effects of tone, focus and engagement. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (pp. 2439-2442). Brighton, UK.

Huang, B., & Liao, X. (2002). *Modern Chinese*. Beijing, China: Higher Education Press.

Hsu, Y.-Y. (2009). Possessor extraction in Mandarin Chinese. *University of Pennsylvania Working Papers in Linguistics, 15*, 94-104. Retrieved from https://repository.upenn.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1078&context=pwpl

Hsu, C.-H., Evans, J.P., & Lee, C.-Y. (2015). Brain responses to spoken F0 changes: Is H special? *Journal of Phonetics, 51*, 82-92. doi: dx.doi.org/10.1016/j.wocn.2015.02.003.

Hyman, L. (1990). Boundary tonology and the prosodic hierarchy. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 109-126). Chicago, IL: The University of Chicago Press.

Hyman, L. (2001) Tone systems., In M. Haspelmath, E. Koenig, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook Vol 2* (pp. 1367-1380). Berlin, Germany: Mouton de Gruyter.

Ife, A., Vives, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics, 4,* 55-84.

Ip, M. H. K., & Cutler, A. (2016). Cross-language data on five types of prosodic focus. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of Speech Prosody 2016* (pp. 330-334). Boston, USA. doi: 10.21437/SpeechProsody.2016-68

Ip, M. H. K., Shaw, J. A., & Cutler, A. (submitted). Prosodic strategies of focus expression across languages.

Ito, K., & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In P. Augurzky & D. Lenertova (Eds.), *Methods in empirical prosody research* (pp. 231-259). Berlin, Germany: De Gruyter Mouton.

Ito, K., & Speer, S.R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*, 541-573. doi: 10.1016/j.jml.2007.06.013

Ito, K., Speer, S. R., & Beckman, M. E. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. In *Proceedings of Speech Prosody* (pp. 279-282).

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jackson, C. N., & O'Brien, M. G. (2011). The interaction between prosody and meaning in second language speech production. *Unterrichtspraxis, 44*, 1-11. doi: 10.1111/j.1756-1221.2011.00087.x

Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese* (Doctoral dissertation, The Ohio State University, OH, USA). Retrieved from https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu1239977570

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.). *Casual Speech: Data and Analysis* (pp. 29-54). Tokyo, Japan: The National Institute for Japanese Language.

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy, 13*, 440-455. doi: 10.1080/15250000802329321

Jongman, A., Wang, Y., & Sereno, J. (2006). Perception and production of Mandarin tone. P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics (Vol. 1: Chinese)*. Cambridge, UK: Cambridge University Press.

Jun, S. -A. (2011). Prosodic marking of complex NP focus, syntax, and the pre-/post-focus string. In M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press.

Jun, S. -A. (Ed) (2014). *Prosodic typology II: The phonology of intonation and phrasing*. Oxford, UK: Oxford University Press.

Jun, S. -A., & Lee, H. -J. (1998) Phonetic and phonological markers of contrastive focus in Korean. In *Proceedings of the 5th International Conference on Spoken Language Processing* (pp. 1295-1298). Sydney, Australia.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y. & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402-420. doi: 10.1006/jmla.1993.1022

Karlsson, A., House , D., Svantesson , J., & Tayanin, D. (2010). Influence of lexical tones on intonation in Kammu. In W. Hess (Ed.), *Proceedings of the 12th Annual*

Conference of the International Speech Communication Association (pp. 1740-1743). Makuhari, Japan

Kazanina, N. (2017). Predicting complex syntactic structure in real time: Processing of negative sentences in Russian. *Quarterly Journal of Experimental Psychology, 70*, 2200-2218. doi: 10.1080/17470218.2016.1228684

Kaiser E. (2006). Negation and the left periphery in Finnish. *Lingua, 116*, 314-350. doi: 10.1016/j.lingua.2004.08.008

Kang, K.-H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America, 124*, 3909–3917. doi: 10.1121/1.2988292

Katsika A. (2009). Boundary- and prominence-related lengthening and their interaction. *Journal of the Acoustical Society of America, 125*, 2572-2572.

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

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language, 87*, 771-816. doi: 10.1353/lan.2011.0076

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.

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *Journal of the Acoustical Society of America, 132*, 1050-1060. doi: 10.1121/1.4730893

Kember, H., Choi, J., Yu, J., & Cutler, A. (revised). The processing of linguistic prominence.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech, 51*, 343-359. doi: 10.1177/0023830908099069

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: 10.1016/j.jml.2011.12.007

Kiss, K. E. (1998). Identificational focus versus information focus. *Language, 74*, 245-273. doi: 10.2307/417867

Kirby, J. P. (2011) Dialect experience in Vietnamese tone perception. *Journal of the Acoustical Society of America, 127*, 3749-3757. doi: 10.1121/1.3327793

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16*, 29-59. doi: 10.1080/0269993943000121

Klatt D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3,* 129–140.

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: 10.1121/1.380986

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7*, 279-312.

Kochanski, G., Grabe, E., Coleman, J. & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustic Society of America, 118*, 1038-1054. doi: 10.1121/1.1923349

Kohler, K. (1983). Prosodic boundary signals in German. *Phonetica, 40*, 89-134.

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

Kratochvil, P. (1998). Intonation in Beijing Chinese. In: D. Hirst, & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 417-431). Cambridge, UK: Cambridge University Press.

Krifka, M. (2006). Basic notions of information structure. In M. Krifka & R. Musan (Eds.), *Interdisciplinary studies on information structure*. Berlin, Germany: De Gruyter Mouton.

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex, 23*, 1836-1848. doi: 10.1093/cercor/bhs164

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

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics, 40*, 430-442. doi: 10.1016/j.wocn.2012.02.011

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 *Proceedings of Interspeech 2010* (pp. 1748-1751). Makuhari, Japan: ISCA.

Kuo, G. (2011). Prosodic boundaries and the Taiwanese tone sandhi group. *UCLA Working Papers in Phonetics, 109*, 40-59. Retrieved from https://escholarship.org/content/qt1dz69593/qt1dz69593.pdf

Kügler, F. (2017). Tone and intonation in Akan. In L. J. Downing, & A. Rialland (eds.), *Intonation in African tone languages* (pp. 89-129). Berlin: Mouton de Gruyter.

Kügler, F., & Skopeteas, S. (2007). On the universality of prosodic reflexes of contrast: The case of Yucatec Maya. In J. Trouvin & W. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1025–1028). Saarbrücken: University of Saarland.

Kuijpers, C., & van Donselaar, W. (1998). The influence of rhythmic context on schwa epenthesis and schwa deletion in Dutch. *Language and Speech, 41*, 87-108. doi: 10.1177/002383099804100105

Kuzla, C., & Ernestus, M. (2011). Prosodic conditioning of phonetic detail in German plosives. *Journal of Phonetics*, *39*, 143-155.

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

Ladd, D. R. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook, 3*, 311-340. doi: 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.

Ladd, D. R. (1990a). Intonation: emotion vs. grammar. [Review of book *Intonation and its uses*, by D. L. Bolinger]. *Language, 66*, 806-816.

Ladd, D. R. (1990b). Metrical representation of pitch register. In J. Kingston and M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 35-57). Cambridge, UK: Cambridge University Press.

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

Lai, W., & Dilley, L. C. (2016). Cross-linguistic generalization of the distal rate effect: Speech rate in context affects whether listeners hear a function word in Chinese Mandarin. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of Speech Prosody 2016* (pp. 1124-1128). Boston, USA. doi: 10.21437/SpeechProsody.2016-231

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge, UK: Cambridge University Press.

Laniran. Y. O. (1992). Intonation in tone languages: The phonetic implementation of tones in Yoruba. (Doctoral dissertation, Cornell University, USA).

Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics,* 31, 203-250. doi: 10.1016/S0095-4470(02)00098-0

Leben, W. R., & Ahoua, F. (2006). Phonological reflexes of emphasis in Kwa languages of Cote d'Ivoire. In P. Newman & L. M. Hyman (Eds.), *Studies in African linguistics* (pp. 145–158). Columbus, OH: The Department of Linguistics and the Center for African Studies, Ohio State University.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics, 53*, 157-165. doi: 10.1080/01690960701728261

Lee, A., Chiu, F., & Xu, Y. (2017). Focus perception in Japanese: Effects of focus location and accent condition. *Proceedings of Meetings on Acoustics, 29*, 60007. doi: doi.org/10.1121/2.0000441.

Lee, Y. -C., Wang, T., & Liberman, M. (2016). Production and perception of Tone 3 focus in Mandarin Chinese. *Frontiers in Psychology 7*, 1-13. doi: 10.3389/fpsyg.2016.01058.

Lee, Y. -C., Wang, B., Chen, S., Adda-Decke, M., Amelot, A., Nambu, S., & Liberman, M. (2015). A crosslinguistic study of prosodic focus. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4754-4758).

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica, 5*, 1-54.

Lehiste, I. (1970). *Suprasegmental*. Cambridge, MA: MIT Press.

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

Lehiste, I. & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America, 31*, 428-435. doi: 10.1121/1.1907729

Lengeris, A. (2012). Prosody and second language teaching: Lessons from L2 speech perception and production research. In J. Romero-Trillo (Ed.), *Pragmatics and prosody in English language teaching* (pp. 25-40). New York, NY: Springer. doi: 10.1007/978-94-007-3883-6_3

Li, X. -Q., & Ren, G. -Q. (2012). How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: An ERP study. *Neuropsychologia, 50*, 1882-1894. doi: 10.1016/j.neuropsychologia.2012.04.013

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

Liang, J., & van Heuven, V., J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes, Volume 2: Experimental studies in word and sentence prosody* (pp. 27-61). Berlin: Mouton de Gruyter.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461. doi: 10.1037/h0020279

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.

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

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America, 32*, 451-454.(doi: 10.1121/1.1908095

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172-179. doi: 10.1177/002383096300600306

Liberman, M., & Pierrehumbert, J. (1984) Intonational invariance under Changes in pitch range and length. In M. Aronoff & R. Oerhle (Eds.), *Language sound structure* (pp. 157-233). Cambridge, MA: MIT Press.

Lin, H. (2006). Mandarin neutral tone is a phonologically low tone. *Journal of Chinese Language and Computing, 16*, 121-134.

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

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech, 104*, 109-138. doi: 10.1177/00238309040470020101

Liu, F., & Xu, X. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*, 70-87. doi: 10.1159/000090090

Ma, W., Zhou, P., Singh, L., & Gao, L. (2017). Spoken word recognition in young tone language learners: age-dependent effects of segmental and suprasegmental variation. *Cognition, 159*, 139–155. doi: 10.1016/j.cognition.2016.11.011

Mády, K. (2015). Prosodic (non-)realisation of broad, narrow and contrastive focus in Hungarian: A production and a perception study. In *Proceedings of the 16th Annual Conference of the International Speech Communication Association* (pp. 948-952). Dresden, Germany.

Maekawa, K. (1997). Effects of focus on duration and vowel formant frequency in Japanese. In Y. Sagisaka, W. N. Campbell, & N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 129–153). New York, NY: Springer.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*, 1994-1997. doi: 10.1016/j.cub.2009.09.064

Mang, E. (2001). A cross-language comparison of preschool children's vocal fundamental frequency in speech and song production *Research Studies in Music Education, 16*, 4-14. doi: 10.1177/1321103X010160010201

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: 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: 10.1016/jdcn.2013.01.003

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102. doi: 10.1016/0010-0277(87)90005-9

Maskikit-Essed, R., & Gussenhoven, C. (2016). No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay. *Phonology, 33*, 353-389. doi: 10.1017/S0952675716000154

Mattys, S. L. & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language, 36*, 87–116. doi: 10.1006/jmla.1996.2472

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 960–977.

McAllister, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech, 34*, 1-26. doi: 10.1177/002383099103400101

McCawley, J. D. (1968) *The phonological component of a grammar of Japanese*. The Hague, the Netherlands: Mouton de Gruyter.

McClelland, J. L. & Elman, J. L. (1986) The TRACE model of speech perception. *Cognitive Psychology 18*,1–86. doi: 10.1016/0010-0285(86)90015-0

McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21-46. doi: 10.1006/jmla.1998.2568

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1363-1389. doi: 10.1037/0096-1523.25.5.1363

Mehler, J., Jusczyk, E W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

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: 10.1016/j.wocn.2012.08.004

Mirman, D., Magnuson, J., Graf Estes, K. & Dixon, J. A. (2008) The link between statistical segmentation and word learning in adults. *Cognition 108*:271–80. doi: 10.1016/j.cognition.2008.02.003

Mennon, I. (2004). Bi-directional interference in the intonation of. Dutch speakers of Greek. *Journal of Phonetics, 32*, 54–563. doi: 10.1016/j.wocn.2004.02.002

Morrill, T. H., Dilley, L. C., McAuley, J, & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition, 131*, 69–74. doi: 10.1016/j.cognition.2013.12.006.

Mullennix, J. W. & Pisoni, D. B. (1990) Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics 47*, 379– 90.

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual test of rhythmic similarity: I. Mora rhythm. *Language and Speech, 50*, 77-99. doi: 10.1177/00238309070500010401

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six month olds! detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy, 1*, 123-147. doi: 10.1207/S15327078IN0101_11

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Berlin, Germany: Mouton de Gruyter.

Nolan, F. (2006). Intonation. In B. Aarts, & A. McMahon (Eds.), *Handbook of English linguistics* (pp. 433-457). Oxford, UK: Blackwell Publishing Ltd.

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology, 53*, 146–193. doi: 10.1016/j.cogpsych.2006.03.001

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299-325. doi: 10.1017/S0140525X00003241.

Norris, D. G., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*,191–243. doi: 10.1006/cogp.1997.0671

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology, 53*, 146-193. doi:10.1016/j.cogpsych.2006.03.001.

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

Onaka, A. (2003). Domain-initial strengthening in Japanese: An acoustic and articulatory study. In *Proceedings of the 15th International Congress of the Phonetic Sciences* (pp. 2091-2094). Barcelona, Spain.

Otake, T., & Higuchi, M. (2008). The role of Japanese pitch accent in spoken-word recognition: Evidence from middle-aged accentless dialect listeners. In *Proceedings of INTERSPECH* (pp. 1097-1100). Brisbane, Australia: ISCA.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 358-378. doi: 10.1006/jmla.1993.1014

Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, *1*, 57-72. doi: 10.1080/01690965.2013. 805795.

Paulmann, S., Furnes, D., Bøkenes, A. M., & Cozzolino, P. J. (2016). How psychological stress affects emotional prosody. *PLOS ONE. 11*:e0165022. doi: 10.1371/journal.pone.0165022

Peng, S.-H. (1994). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics, 25*, 371-400.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84*, 372–389. doi: 10.1111/0026-7902.00075

Peterson, G. E., & Lehiste, I. (1960) Duration of syllable nuclei in English, *Journal of the Acoustical Society of America, 32*, 693-703.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*, 280-91. doi: 10.1016/j.cognition.2011.10.004

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434-446. doi: 10.1044/jshr.2904.434

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*, 427-459. doi: 10.1037/0033-2909.134.3.427

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., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack, M. E. (Eds.), *Intentions in communication* (pp. 271–311). 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.

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.

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

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20*, 331-350.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510-532. doi: 10.1037/0033-2909.114.3.510

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition, 2*, 85-115. doi: 10.1075/arla.2.05ram

Remijsen, B. (2002). Lexically contrastive accent and lexical tone in Ma'ya. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 585–614). Berlin: Mouton de Gruyter.

Redford, M. A. (2013). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics, 34*, 569-589. doi: 10.1017/S0142716411000877

Redford, M. A., Davis, B. L., & Miikkulainen, R. (2004). Phonetic variability and prosodic structure in mothers. *Infant Behavior & Development, 27*. 477-498. doi: 10.1016/j.infbeh.2004.05.001.

Redford, M. A., Stine, J. A., & Vatikiotis-Bateson, E. (2014). A question of scope? Direct comparison of clear and in-focus speech productions. In *Proceedings of the International Seminar on Speech Production* (pp. 352-355).

Repp, B. H. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics, 18*, 481-495. doi: 10.1121/1.2028239

Rialland, A., & Robert, S. (2001). The intonational system of Wolof. *Linguistics 39*, 893-939. doi: 10.1515/ling.2001.038

Rochemont, M. S. (1986). *Focus in generative grammar*. Philadelphia, PA: Benjamins.

Romøren, A. S. H. and Chen, A. (2015). Quiet is the new loud: Pausing and focus in child and adult Dutch. *Language and Speech, 58*, 8-23. doi: 10.1177/0023830914563589

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, 75-116.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-olds. *Science, 274*, 1926-1928. doi: 10.1126/science.274.5294.1926

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulatory information in spoken-word recognition. *Journal of Memory and Language, 71*, 145–163. doi: 10.1016/j.jml.2013.11.002

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*, 348–351. doi: 10.1111/1467-9280.00364

Sanderman, A. A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech, 40*, 391–409.

Schafer, A. .J., Speer, S. R., Warren, P., & White, S. D. (2000) Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistics Research, 29,* 169-182. doi: 10.1023/A:1005192911512

Schneider, W., Eschman, A., & Zuccolotto, A. (2002) E-Prime User's Guide. Pittsburgh: Psychology Software Tools Inc.

Sebastián-Gallés, M., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31*, 18-32. doi: 10.1016/0749-596X(92)90003-G

Seddoh, S. A. (2002). How discrete or independent are "affective prosody" and "linguistic prosody"? *Aphasiology, 16,* 683–692. doi: 10.1080/02687030143000861

Seidenberg, M.S., Tanenhaus, M.K., Leiman, J.M., & Bienkowsky, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489-537. doi: 10.1016/0010-0285(82)90017-2

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57*, 24–48. doi: 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: 10.1111/j.1467-7687.2008.00704.x

Seidl, A., & Johnson, E. K. (2006) Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9*, 565–573. doi: 10.1371/journal.pone.0083546

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3*, 371-405.

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.

Selkirk, E. (2004). Bengali intonation revisited. In Lee, Gordon, & D. Buring (Eds.), *Topic and focus: A cross-linguistic perspective* (pp. 217–246). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Selkirk, E., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 313–337). Chicago, IL: The University of Chicago Press.

Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of phrase-final lengthening in English. *Journal of the Acoustical Society of America, 103*, 2889–2889.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics, 68,* 1-16.

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

Shen, W., Vaissière, J., & Isel, F. (2013). Acoustic correlates of contrastive stress in compound words versus verbal phrase in Mandarin Chinese. *Computational Linguistics and Chinese Language Processing, 18*, 45-58.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory, 3*, 83-109.

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.

Sholicar, J. R., & Fallside, I. F. A prosodically and lexically constrained approach to continuous speech recognition. In *Proceedings of the 2nd Biennial Conference on Speech Science and Technology* (pp. 106-111). The Australasian Speech Science and Technology Association.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences, 108*, 6038-6043. doi: 10.1073/pnas.1017617108

Silipo, R., & Greenberg, S. (2000). Prosodic stress revisited: Reassessing the role of fundamental frequency. In *Proceedings of the NIST Speech Transcription Workshop*.

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.

Singh, L. & Fu, C. S. L. (2016). A new view of language development: the acquisition of lexical tone. *Child Development, 87*, 834–854. doi: 10.1111/cdev.12512

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011b). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia, 49*, 3831–3846. doi: 10.1016/j.neuropsychologia.2011.09.044

Sluijter, A. M. C., & van Heuven, V. J. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. *Phonetica, 52*, 71–89. doi: 10.1159/000262061

Sluijter, A. M., & Heuven, V. J. van (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*, 2471-2485. doi: 10.1121/1.417955

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts in conversational and clear speech. *Journal of Phonetics, 36*, 91–113. doi: 10.1016/j.wocn.2007.02.002

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: 10.1016/S0749-596X(02)00519-3

Soderstrom, M., Blossom, M., Foygel, I., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language, 35*, 869-902. doi: 10.1017/S0305000908008763

Soderstrom, M., Seidl, A., Kemler Nelson, D., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*, 249–267. doi: 10.1016/S0749-596X(03)00024-X

Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology, 2*(1), 39–98.

Spierings, M. J., & ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proceedings of Royal Society B: Biological Sciences, 281*, 20140480. doi:10.1098/rspb. 2014.0480

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

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).

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

Stevens, C. J., Keller, P. E., & Tyler, M. D. (2013). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music, 41*, 59-74. doi: 10.1177/0305735611415749

Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Sciences, 107*, 21914-21919. doi: 10.1073/pnas.1009020107

Storkel, H. L., Armbruster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research 49*. 1175-1192. doi: 10.1044/1092-4388(2006/085)

Streeter, L. (1978). Acoustic determinants of phrase boundary perception, *Journal of the Acoustical Society of America, 64*, 1582-1592.

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

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics, 30*, 629-654. doi:10.1006/jpho.2002.0178

Szczepek Reed, B. (2012). Prosody in conversation: Implications for teaching English pronunciation. In J. Romero-Trillo (Ed.), *Pragmatics and prosody in English language teaching* (pp. 147-168). Dordrecht, The Netherlands: Springer.

Tabain M. (2003). Effects of prosodic boundary on /aC/ sequences: acoustic results. *Journal of the Acoustical Society of America,* 113,516-531.

Taft, M., & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen, & O. J.-L. Tzeng (Eds.), *Language Processing in Chinese* (pp. 151-172). Amsterdam, The Netherlands: Elsevier.

Takeda, K., Sagisaka, Y., & Kuwabara, H. (1989). On sentence-level factors governing segmental duration in Japanese. *Journal of the Acoustical Society of America, 86*, 2081–2087.

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: 10.1126/science.7777863

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica, 158*, 407-424. doi: 10.1515/SEM.2006.017

Thorsen, N.G. (1985). Intonation and text in Standard Danish. *Journal of the Acoustical Society of America, 77*, 1205-1216.

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes, 23*, 689-708. doi: 10.1080/01690960701728261

Toro, J. M., and Hoeschele, M. (2017). Generalizing prosodic patterns by a non-vocal learning mammal. *Animal Cognition, 20*, 179. doi: 10.1007/s10071-016-1036-8

Tremblay, A., Broersma, M., & Coughlin, C. E. (2017). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition, 21*, 1–13. doi: 10.1017/S136672891700030X

Tsang, K. K., & Hoosain, R. (1979). Segmental phonemes and tonal phonemes in comprehension of Cantonese. *Psychologia, 22*, 222- 224.

Turnbull, R., Burdin, R. S., Clopper, C. G., & Tonhauser, J. (2015). Contextual predictability and the prosodic realisation of focus: A cross-linguistic comparison. *Language, Cognition and Neuroscience, 30*, 1061–1076. doi:10.1080/23273798. 2015.1071856

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America, 99*, 3782-3790. doi:10.1121/1.414995.

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

Ulbrich, C., & Mennen, I. (2015). When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism, 16*, doi: 10.1177/1367006915572383.

Vallduví, E. (1991). The role of plasticity in the association of focus and prominence. *Eastern States Conference in Linguistics, 7*, 295-306.

Vallduví, E. (1992). *The informational component*. New York, NY: Garland Publishers.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53-66). Heidelberg: Springer.

Vaissière, J. (2005). Perception of intonation. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 1-28). Oxford, UK: Blackwell.

van Berkum, J. J., Zwitserlood, P., Hagoort, P. & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*, 701–718. doi: 10.1016/S0926-6410(03)00196-4

van Katwijk, A.F.V. (1974) Accentuation in Dutch: An experimental linguistic study. Amsterdam, The Netherlands: Van Gorcum.

van Kuijk, D., & Boves, L. (1999). Acoustic characteristics of lexical stress in continuous telephone speech. *Speech Communication, 27*, 95-111. doi: 10.1016/S0167-6393(98)00069-7 van Zanten, E. A., & van Heuven, V. J. (1998). Word stress in Indonesian: Its communicative relevance. *Journal of the Humanities and Social Sciences of Southeast Asia and Oceania, 154*, 129-147.

Vallduví, E. 1991. The role of plasticity in the association of focus and prominence. *Proceedings of the Eastern States Conference on Linguistics (ESCOL), 7*, 295-306.

Vallduví, E. (1992). *The Informational Component*. New York, NY: Garland.

Vanlancker-Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics, 24*, 45–57. doi: 10.1017/S0142716403000031

Venditti, J. J., Jun, S. -A., & Beckman, M. E. (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean and English. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 287-311). Mahwah, NJ: Lawrence Erlbaum Publishers.

Vitevitch, M. S., & Luce, P. L. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory of Language, 40*, 374-408. doi: 10.1006/brln.1999.2116

Vogel, I., Athanasopoulou, A., & Pincus, N. (2016). Prominence, contrast and the Functional Load Hypothesis: An acoustic investigation. In J. Heinz, R. Goedemans, & H. van der Hulst (Eds.), *Dimensions of Phonological Stress* (pp. 123-167). Cambridge, UK: Cambridge University Press.

Volskaya, N. (2003). Virtual and real pauses at clause and sentence boundaries. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 499-502). Barcelona: Causal Productions.

Vroomen, J., Van Zon, M., & De Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misconceptions and word spotting. *Memory and Cognition, 24*, 744-755.

Waksler, S. (2001). Pitch range and women's sexual orientation. *Word, 52*, 69–77.

Wang, L., Bastiaansen, M. C. M., Yang, Y., and Hagoort, P. (2012). Information structure influences depth of syntactic processing: event-related potential evidence for the Chomsky illusion. *PLoS ONE, 7*, e47917. doi: 10.1371/journal.pone.0047917

Wang, B., Wang, L. & Qadir, T. (2011). Prosodic realization of focus in six languages/dialects in China. In Wai-Sum Lee & Eric Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong 2011* (pp. 144-147).

Wang, B., Wang, L. & Qadir, T. (2011). Prosodic realization of focus in six languages/dialects in China. In Wai-Sum Lee & Eric Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong 2011* (pp. 144-147).

Warren. P. (2005). Patterns of late rising in New Zealand English: Intonational variation or intonational change? *Language Variation and Change, 17*, 209-230

Watson, D. G, & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19,* 713-755. doi: 10.1111/j.1467-9582.2005.00130.x

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science, 32*, 1232-1244. doi: 10.1080/03640210802138755

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology, 3*, 580. doi: 10.3389/fpsyg.2012.005580

Werker, J. F., & Tee, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63. doi: 10.1016/S0163-6383(84)80022-3

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*, 25-47. doi: 10.1159/000261901

Wichmann, A., House, J., & Rietveld, T. (1997). Peak displacement and topic structure. In A. Botinis, G. Kouroupetroglou, & G. Carayannis (Eds.), *Intonation: Theory, models and applications* (pp. 329-332). Athens, Greece: ESCA and University of Athens, Department of Informatics.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of he Acoustical Society of America, 92*, 1707-1717.

Winters, S., & O'Brian, M. G. (2012). Perceived accentedness and intelligibility: the relative contributions of F0 and duration. *Speech Communication, 55*, 486–507. doi: 10.1016/j.specom.2012.12.006

Wong, P., & Strange, W. (2017). Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PLoS One, 12*, e0182337. doi: 10.1371/journal.pone.0182337

Wong, P., Fu, W. M., & Cheung, E. Y. L. (2017). Cantonese-speaking children do not acquire tone perception before tone production: A perceptual and acoustic study of three-year-olds' monosyllabic tones. *Frontiers in Psychology, 8*, 1450. doi: 10.3389/fpsyg.2017.01450

Wu, M. H. (2019). Effect of F0 contour on perception of Mandarin Chinese speech against masking. *PLoS One, 14*, e0209976. doi: 10.1371/journal.pone.0209976

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics, 27*, 55–105. doi: 10.1006/jpho.1999.0086

Xu, Y. (2008). In defense of lab speech. *Journal of Phonetics, 38*, 329-336. doi: 10.1016/j.wocn.2010.04.003

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*, 159-197. doi: 10.1016/j. wocn.2004.11.001

Xu, Y., Chen, S. -W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *Linguistic Review, 29*, 131-147. doi: 10.1515/tlr-2012-0006

Xue, A., Hagstrom, F., & Hao, G. (2002). Speaking F0 characteristics of bilingual Chinese-English speakers: A functional system approach. *Asian Pacific Journal of Speech, Language, and Hearing 7*, 55–62. doi: 10.1179/136132802805576544

Yang, X., Shen, X., Li, W., & Yang, Y. (2014). How listeners weight acoustic cues to intonational phrase boundaries. *PLoS ONE, 9*, e102166.

Yenkimaleki, M., & van Heuven, V. J. (2016). The effect of teaching prosody awareness on interpreting performance: An experimental study of consecutive interpreting from English into Farsi. *Perspectives: Studies in Translation Theory and Practice, 26*, 84-99. doi: 10.1080/0907676X.2017.1315824

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. Journal of Memory and Language, 68(2), 123-139. doi: 10.1016/j.jml.2012.09.004

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *Journal of the Acoustical Society of America, 130*, 4063-4069. doi: 10.1121/1.3651

Zerbian, S. (2006). *Expression of information structure in the Bantu language Northern Sotho* (Doctoral dissertation, Humboldt University, Berlin, Germany). Retrieved from http://www.zas.gwz-berlin.de/fileadmin/material/ZASPiL_Volltexte/zp45/zaspil45.pdf

Zerbian, S. (2017). Sentence intonation in Tswana (Sotho-Tswana group). In L. J. Downing, & A. Rialland (eds.), *Intonation in African tone languages* (pp. 89-129). Berlin: Mouton de Gruyter.

Zipf, G. (1949). *Human behavior and the principle of least effort*. New York, USA: Addison-Wesley.

Zubizarreta, M. L. (1994). On Some Prosodically Governed Syntactic Operations. In G. Cinque, J. Koster, J. -Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths toward universal grammar: Studies in honor of Richard S. Kayne* (pp. 473-485). Washington, DC: Georgetown University Press.

Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. Cambridge, MA: MIT Press.

APPENDICES

Appendix A

(Ethical Approval Letter)

Locked Bag 1797
Penrith NSW 2751 Australia Office of Research Services

ORS Reference: H11216

8 July 2015

Professor Anne Cutler The MARCS Institute

Dear Anne,

I wish to formally advise you that the Human Research Ethics Committee has approved your research proposal H11216 “Universal and Language-specific in Speech Processing: The Case of Prosody“, until 1 March 2018 with the provision of a progress report annually if over 12 months and a final report on completion.

Conditions of Approval

1. A progress report will be due annually on the anniversary of the approval date. 2. A final report will be due at the expiration of the approval period.

3. Any amendments to the project must be approved by the Human Research Ethics Committee prior to being implemented. Amendments must be requested using the HREC Amendment Request Form: http://www.uws.edu.au/__data/assets/pdf_file/0018/491130/HREC_Amendment_Request_Form.pdf

4. Any serious or unexpected adverse events on participants must be reported to the Human Ethics Committee via the Human Ethics Officer as a matter of priority.

5. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the Committee as a matter of priority

6. Consent forms are to be retained within the archives of the School or Research Institute and made available to the Committee upon request.

Please quote the registration number and title as indicated above in the subject line on all future correspondence related to this project. All correspondence should be sent to the email address humanethics@uws.edu.au.

This protocol covers the following researchers:

Anne Cutler, Paola Escudero, Jason Shaw, Martin Ip

Yours sincerely

Professor Elizabeth Deane
Presiding Member,
Human Researcher Ethics Committee

Appendix B

(Dialogue Scripts in English)

Dialogue 1: The Street Vendor

Vendor = Experimenter
Buyer = Participant

(Buyer is browsing around)

Vendor: Hello. Haven't seen you for a while! What are you after?
Buyer: I'm after a [**SWEATER**]*wh-foc*. A present for a friend. His birthday is next week.
Vendor: Good timing! We've just got new arrivals in.
Buyer: (*pointing to an item*) What about that one over there?
Vendor: That's the women's section over there, and that's actually a jacket. It's a jacket you want to buy?
Buyer: No, no, I want to buy a [**SWEATER**]*correct-foc*.
Vendor: OK, what kind of sweater are you looking for?
Buyer: I was thinking of a [**BLUE**]*wh-foc* sweater...
Vendor: Ok, let me see what brown sweaters we have…
Buyer: No, no, I said I was thinking of a [**BLUE**]*correct-foc* sweater.
Vendor: A blue sweater you said?
Buyer: That's right, a [**BLUE**]*confirm-foc* sweater.
Vendor: Ok, let me see....hmmm…
Buyer: Oh wait, I know I told you I was looking for a blue sweater. Actually I just remembered his favourite colour. That's [**GREEN**]*new-foc*. Let's make it a green sweater.
Vendor: Ok, let me see...
Buyer: Oh wait a minute. Maybe I should buy [**TWO**]*new-foc* new sweaters. My sister's coming, and it would be nice to buy her a sweater too. Say a [**GREEN**]*parallel-foc* sweater for my friend and a [**RED**]*parallel-foc* sweater for my sister.
Vendor: (*hands over two new sweaters*) Here you go – that will be 200 dollars each.
Buyer: 200 dollars each! That's way too much for me. I've never had a 200 dollar sweater1. Can't you make it less given that I'm buying more than one sweater? I am buying [**TWO**]*confirm-foc* sweaters.
Vendor: How about 100 dollars each?
Buyer: I was more thinking of about [**FIFTY**]*new-foc* dollars for each sweater. I'd be happy to pay fifty dollars for the green sweater and another fifty dollars for the red sweater. So 100 dollars for two sweaters.
Vendor: hmm… what about 80 dollars for each sweater?
Buyer: 80 dollars is still too much…(*looking at the green sweater*)… Oh look! There's a [**STAIN**]*new-foc* on the green sweater. Maybe you can reduce your price a bit since there is a stain on one of your sweaters.

Dialogue 2: A Criminal Investigation

Inspector = Experimenter
Student = Participant

(*Inspector is questioning a high school student at a crime scene*)

Inspector: I first need to know where you were at lunch when your fellow students died. Were you in the courtyard?
Student: No, I was reading in the [**LIBRARY**]*correct-foc*.
Inspector: (*frowning*) What? Were you eating in the library?
Student: No, I was [**READING**]*correct-foc* in the library.
Inspector: Tell me what happened.
Student: Well, I was reading a book and my friends were browsing through magazines when suddenly we heard a huge noise. It was as if all the books have fallen off the shelves.
Inspector: Did you go to see where the noise came from?
Student: Yes, but when I was on my way, I suddenly heard two [**GUNSHOTS**]*new-foc*, so I ran away…
Inspector: So on your way, you heard two books dropped.
Student: No, I heard two [**GUNSHOTS**]*correct-foc*.
Inspector: Ah, and there was more than one gunshot?
Student: Yes that's right, I heard [**TWO**]*confirm-foc* gunshots.
Inspector: Hmm…..two gunshots….hmm..
Student: Oh wait, I remember something. Before I heard the gunshots, I heard two people [**WHISPERING**]*new-foc* having an [**ARGUMENT**]*new-foc*.
Inspector: Do you know what they were arguing about? Were they arguing over boyfriends or girlfriends?
Student: No, I think they were having an argument over a [**BOOK**]*correct-foc* they had read. I could make out what they were saying because they were whispering very loudly.
Inspector: Really?
Student: Yes, I think they were always arguing over the books they read. Three days ago, they were also having an argument. But that argument was different. Strangely enough, the argument they were having that time was about a book they hadn't even [**READ**]*new-foc*!

Dialogue 3: Where Is My Ring?

Police = Experimenter
John = Participant

(*John is a rich man, and he suspects that one of his rings is stolen at a jewelry store*)

Police: Okay, I would like to ask you some questions about what happened at the Harrison's jewelry store. Was it your first time at Harrison's?
John: No, it was my [**SECOND**]*correct-foc* time.
Police: And what were you planning to buy in the jewelry store?
John: I was not planning to buy anything. I came to pick up my [**ENGAGEMENT RING**]*new-foc*. I also came here to pick up a few other rings that I brought here to be repaired.
Police: Your engagement ring, was it a sapphire ring?
John: No, it was a [**RUBY**]*correct-foc* ring.
Police: The ruby ring is for your fiancée?
John: Yes, the ruby ring is for my [**FIANCÉE**]*confirm-foc*.
Police: Who did you give the ring to?
John: I gave it to [**MARY**]*wh-foc*. And that was the last time I saw my <u>engagement ring</u>.
Police: Did you show the ring to anybody else? Who else did you show the ruby ring?
John: I only showed [**MARY**]*correct-foc* the ruby ring.
Police: Did you show Mary any other rings?
John: No, I only showed <u>Mary</u> the [**RUBY**]*correct-foc* ring.
Police: And where did you see Mary put the ruby ring?
John: I think she put the ring on the [**COUNTER**]*wh-foc*, next to other jewels on display. I looked at the jewels on the <u>counter</u> twice. The first time, it was there. But the second time I looked for it again in the display <u>counter</u>, my ring was [**MISSING**]*new-foc*. The ring was not there when I looked for it the <u>second</u> time.
Police: Mary said she already returned the ring to you. Did you look for it in your own bag?
John: I am checking my bag now… (*looking into his bag*)… Nope, I cannot find my <u>ruby</u> ring. I know you won't believe me, but <u>Mary</u> did not <u>return</u> the <u>ruby1</u> ring. So it is not <u>in my bag</u>. The ring must be missing. I now have nothing to give to my <u>fiancée</u>.
Police: Where are all your other rings?
John: My other rings are here. But as for my [**ENGAGEMENT RING**]*new-foc*, I still don't have it!
Police: That's strange indeed. I wonder why she would take the ruby from you….
John: (*looking for the ring in his bag again – this time he found it!*)… Wait! I found my ruby ring! I remember now! Oh I am so sorry. I was wrong when I told you Mary did not return my ring. Mary [**RETURN**]*correct-foc* the ruby ring. The ring is not <u>missing</u>. It's [**IN MY BAG**]*correct-foc*.

Dialogue 4: Teacher and Student

Student = Experimenter
Teacher = Participant

Teacher: Your mother said you are struggling to get good grades in [**GEOLOGY**]*new-foc*.
Student: My grades in biology?
Teacher: No, your grades in [**GEOLOGY**]*correct-foc*. So let's do some revision.
Student: Good idea. This is my first question. Is the Earth's mantle above the crust?
Teacher: No, the mantle is [**BELOW**]*correct-foc* the crust.
Student: But this doesn't make sense. If the mantle is below the crust, how does the lava in the mantle get past the Earth's crust? Does lava get through the crust when we dig a huge hole in the ground?
Teacher: No, lava gets through when [**VOLCANO**]*new-foc* es are formed. The lava flows out to the surface every time a volcano erupts. So we can still see lava from the mantle, even though it is below the crust.
Student: Ah!
Teacher: And here's the interesting part! Lava cools over time and it forms the most beautiful mountain ranges. That's how [**MT WILSON**]*new-foc* was formed. Many parts of Mt Wilson have layers of lava up to one [**HUNDRED AND FIFTY METRES THICK**]*new-foc*. And these layers are [**FOURTEEN MILLION YEARS**]*new-foc* old.
Student: One hundred and fifty metres thick of lava?
Teacher: Yes, one [**HUNDRED AND FIFTY METRES THICK**]*confirm-foc* from fourteen million years ago.
Student: In Mt Wilis you said?
Teacher: No, in [**MT WILSON**]*correct-foc*.
Student: Where is Mt Wilson? Is that is in the US?
Teacher: No, Mt Wilson is in [**SYDNEY**]*correct-foc*, in the [**BLUE MOUNTAINS**]*new-foc*.
Student: Oh, in Sydney? I never knew that Sydney has one hundred and fifty metres of lava that is fourteen million years old.
Teacher: Well actually, [**WEST**]*correct-foc* of Sydney. If you travel west from Sydney, you'll reach the Blue Mountains. And you'll know that the lava formation there was one hundred and fifty metres thick, fourteen million years ago. That's what you'll see in Mt Wilson. (*Looking at the time*)….Oh, my time's up. I need to leave. Goodbye, I hope you've learned something useful today for your revision in geology.

Dialogue 5: The Job Interview

Interviewer = Experimenter
Applicant = Participant

Employer: So you are applying to be a cameraman here?
Applicant: No, I am applying to be a [**REPORTER**]*correct-foc*.
Employer: Oh I see, there must be an admin mistake in our feed…So you just graduated from the Australian Catholic University in Sydney?
Applicant: No, I graduated from the Australian [**NATIONAL**]*correct-foc* University in Canberra.
Employer: Excuse me. So what made you want to work for us?
Applicant: I am interested in the kinds of news you report. Other news companies are focused on [**LOCAL**]*parallel-foc* news, at a local level, but you work at a [**NATIONAL**]*parallel-foc* level. When I was a student at the Australian National University in Canberra, I developed an interest in national news. Compared to local news, I find national news more interesting. So I want to become a national news reporter.
Employer: That is good to know. Well, since you are interested in news at a national level, I may ask you some scenario questions about life as a reporter.
Applicant: Sure.
Employer: Say if someone tells you there is a fire at the opera house. What would you do? At first, would you immediately deliver a report about it before others?
Applicant: No, I would not deliver the report. But what I'll do is, I would first call the emergency service.
Employer: But the emergency service would already be at the opera house. So you would call them to bring more firemen to the opera house?
Applicant: No, I would call the emergency service not to bring more firemen, but to [**CONFIRM WHAT HAS HAPPENED**]*correct-foc*. Because if there were a fire, the emergency service would already be there and know the situation before anybody else.
Employer: I see.
Applicant: That way, you will confirm what has happened at the opera house and get the true version of events. But there is one thing that I won't do that other reporters do.
Employer: Oh? And what is that?
Applicant: I will give some disclosure about the fire, but I will not give [**FULL**]*correct-foc* disclosure to my readers all in one go.
Employer: Why? Because you don't want to give unverifiable information?
Applicant: No, that's not the reason. If I give full disclosure all in one go, I will ruin the suspense for my readers. And adding this suspense in news report is a good way to attract interests in readers.
Employer: What do you mean by suspense? Are you going to use it as a marketing strategy? Writing a news report is not like a commercial advertisement, you know.
Applicant: No, a news report is not like a commercial ad, but it's like a [**DETECTIVE**]*correct-foc* story. Writing a report is like writing a detective story.

Appendix C

(Dialogue Scripts in Mandarin)

Dialogue 1: 摆摊的小贩

小贩 = 实验者
顧客 = 受试者

(*顾客到处看*)

小贩：你好! 很久不见了! 你今天想买什么呀?

顾客：我想买件[**毛衣**]*wh-foc*。是送给我朋友的。她下星期过生日。

小贩：你来的正是时候! 我们刚来了新货。

顾客：(*用手指着那边那件*) 你看那边的那一件怎么样?

小贩：那是男装。而且那不是毛衣是夹克。你想买夹克吗?

顾客：不不，我想买件[**毛衣**]*correct-foc*。

小贩：好，没问题。你想买哪一款式的毛衣?

顾客：我想买件[**红色**]*wh-foc* 毛衣…

小贩：好，让我看看有什么款式的褐色毛衣…

顾客：不对不对，我说我想买件[**红色**]*correct-foc* 毛衣。

小贩：啊? 你说你想要件红色毛衣?

顾客：没错，[**红色**]*confirm-foc* 毛衣。

小贩：好，你等一下，让我找一找…

顾客：啊! 等一下，我刚才告诉你我想要件红色毛衣。我刚想起来我朋友最喜欢的颜色。是[**绿色**]*new-foc*。还是给我件绿毛衣吧。

小贩：好，让我找一找….

顾客：啊! 在等一下，我想买[**两**]*new-foc* 件毛衣。我弟弟要来了。我想我也应该给他买件毛衣。我要买件[**绿色**]*parallel-foc* 的毛衣给我朋友，[**蓝色**]*parallel-foc* 的毛衣给我弟弟。

小贩：（*拿给顾客两件毛衣*）给你啊，每件毛衣两百块。

顾客：两百块! 太贵了! 我从来没有买过一件两百块的毛衣。既然我都要买超过一件了，你能不能便宜点? 我现在要跟你买[**两**]*confirm-foc* 件毛衣。

小贩：那么，一百块一件，行不行?

顾客：我想花[**五十**]*new-foc* 块买一件毛衣。我愿意花五十块买一件绿色毛衣，再加五十块买一件蓝色的毛衣。所以一百块两件毛衣。

小贩：那么…每件八十块，行不行?

顾客：唉呀! 八十块还是太贵了….(*正在看着绿毛衣*) … 哎，你看! 你看这绿毛衣这块儿[**脏了**]*new-foc* … 能不能在便宜点，既然绿毛衣都脏了。

Dialogue 2: 刑事调查

警察 = 实验者
学生 = 受试者

(*警察在询问学生*)

警察：首先，我想知道，在吃午餐的时候，你的同学被杀了 – 当时你在哪里?你在院子里吗?

学生：我不在，我在[**图书馆**]*correct-foc* 里读书。

警察：(*皱著眉头*) 你说什麼? 你在图书馆里吃东西?

学生：我没有，我在图书馆里[**读书**]*correct-foc*。

警察：告诉我怎麽回事。

学生：嗯…我当时在图书馆读书。我的朋友正在读杂誌。突然我们听到很大的声音。就好像图书馆全部的书都从书架上掉下来。

警察：你有过去那里看吗?

学生：我有啊，但是当我正在走过去的时候， 我突然间听到两声[**枪响**]*new-foc*。所以我吓跑了。

警察：那麽说….在你走过去的时候，你突然间听到两声砲响?

学生： 不是，我听到[**枪响**]*correct-foc*。

警察：啊，而且还不只一次抢响，对吗?

学生：是啊没错，我听到[**两声**]*confirm-foc* 枪响。

警察：嗯...两声枪响…嗯...

学生：哦! 等等! 我想起来了。在我听到枪响之前，我听到两个人在[**叽叽喳喳**]new-*foc* [**争论**]*new-foc*。

警察：那麽，你知道他们在争论著什麽? 他们是不是在为了男朋友女朋友争论？

学生：应该不是，我想他们在争论他们所读过的一本[**书**]*correct-foc*。我听到他们在争论什麽，因为他们叽叽喳喳很大声。

警察：真的吗?

学生：是啊，这两人，他们经常在争论他们所读过的书。三天前，他们就有过争论。但是，那一次的争论不一样。非常怪，他们那一次所争论的是他们从来没有[**读过**]*new-foc* 的一本书 !

Dialogue 3: 我的戒指在哪里?

警察 = 实验者
温西山 = 受试者

警察：我想问你一些问题关于在福发珠宝店发生的事。这是你第一次在福发珠宝店吗?

温西山：不是，是我[**第二次**]*correct-foc* 了。

警察：那么，你在福发珠宝店想买什么?

温西山：我没打算买什么。我是来取我的[订婚戒指]*new-foc* 的。并且，我也是来取我另外的几个戒指。

警察：你的订婚戒指，是蓝宝石的吗?

温西山：不是，是[**红宝石**]*correct-foc* 的。

警察：这红宝石戒指，是给你未婚妻的吗?

温西山：是啊，这红宝石戒指是给我[未婚妻]*confirm-foc* 的。

警察：你把这戒指交给谁了?

温西山：我把它交给了[马小姐]*wh-foc* 从那之后，我在也没有见到我的订婚戒指。

警察：你有没有给其他人看过这个戒指? 你还给谁看过这个红宝石戒指?

温西山：我只给[**马小姐**]*correct-foc* 看过我的红宝石戒指。

警察：你有没有给马小姐看过你的其它戒指?

温西山：没有，我只给马小姐看过我的[**红宝石**]*correct-foc* 戒指。

警察：那么，你有看到马小姐把红宝石戒指放到哪里了?

温西山：我看到她把戒指放到[**柜台**]*wh-foc* 上，在其它的首饰旁边。我看了柜台两次。第一次，我戒指在那里。可是过后我看了又看在柜台上，戒指[**不见**]*new-foc* 了。戒指不在那了第二次。

警察：马小姐说她已经把戒指还给你了。你有没有找一下你的手提包啊?

温西山：我看一下… (*在查看他的手提包*)…. 没有，我找不到我的红宝石戒指。你可能不相信我，但是马小姐没有把红宝石戒指还给我。红宝石戒指不在我手提包里。所以这戒指肯定不见了。我没有订婚戒指给我的未婚妻了。

警察：那么， 你的其它戒指呢?

温西山：我的其它戒指都在这。但是我的[**订婚戒指**]*new-foc*，我还没有呢!

警察：这真是太奇怪。她为什么拿你的红宝石戒指呢….

温西山：(*在一次查看他的手提包 – 这一次, 他找到了!*)…. 哎，等等！我找到我红宝石戒指了! 我记得了! 哦, 我真是对不起。是我搞错了告诉你马小姐没有把戒指还给我。马小姐[**还给我**]*correct-foc* 了。戒指没有不见。它在我的[**手提包里**]*correct-foc*。

Dialogue 4: 老师和学生

老师 = 实验者
小学生 = 受试者

老师：从你的成绩看，我认为你有些困难在[**地理学**]new-foc。
小学生：你说我有些困难在生理学?
老师：我说你有些困难在[**地理学**]*correct-foc*。我们还是複习一下吧。
小学生：好吧。我第一个问题是，地球的地幔是不是在地壳的上面?
老师：不是，地幔在地壳的[**下面**]*correct-foc*。
小学生：这怎麽有可能呢？ 如果地幔在地壳的下面，那麽地幔的岩浆怎麽能通过地壳啊？ 难道说我们一定要在地上挖个大洞才能让地幔的岩浆通过地壳？
老师：不一定，因为[**火山**]*new-foc* 的爆发，才使岩浆冒出来。每次火山爆发，我们都能看见岩浆流出来。我们每次都能看到地幔的岩浆，虽然地幔在地壳的下面。
小学生：啊!
老师：还有你知不知道，岩浆凝固後就会变成一座山。世界上很多美丽的风景都是从地幔的岩浆造成的。[**无功山**]*new-foc* 就是其中之一从岩浆造成的。无功山 有一[**百五十米厚**]*new-foc* 的岩浆。而且，这些岩浆有[**十四亿年**]*new-foc* 的进化。
小学生：一百五十米厚的岩浆?
老师：没错，一[**百五十米厚**]*confirm-foc* 的岩浆，十四亿年的进化。
小学生：你说在武当山?
老师：不是，在[**无功山**]*correct-foc*。
小学生：无功山在哪里? 是不是在济南?
老师：不是，无功山在[**沈阳**]*correct-foc*，在[**玉林县**]*new-foc*。
小学生：啊，在沈阳啊？我从来没听说过辽宁有一百五十米厚，十四亿年的岩浆。
老师：啊… 实际上，在沈阳…在沈阳的[**西**]*correct-foc* 边。如果你在沈阳往西走，你会来到玉林县。等你到玉林县你就会知道那里的山有一百五十米厚，十四亿年的岩浆… (*看时间*)…哦，到时间了，我要走了。再见，我希望你今天学到了新的知识关於地理学。

Dialogue 5: 工作面试

雇主 = 实验者
求职 =受试者

雇主：… 你来申请做摄影师吗?

求职：不是，我来申请做[**记者**]*correct-foc*。

雇主：哦，我明白了，一定是人事部在什麼地方搞错了….那麼，你刚刚从澳洲天主大学在悉尼毕业的吗?

求职：不是，我从澳洲[**国立**]*correct-foc* 大学在堪培拉毕业的。

雇主：不好意思。那麼，为什麼你想在我们这里工作?

求职：我对你们的新闻报告很有兴趣。其它公司专注[**当地**]parallel 新闻，当地水準，而你们专注[**国际**]*parallel-foc* 新闻。当我在澳洲国立大学学习的时候，我展开了兴趣对国际新闻。相比跟当地新闻，我认为国际新闻还更加有兴趣。所以我想做国际新闻的记者。

雇主：哪太好了。既然你对国际新闻有兴趣，我想问你一些问题关於做记者的职业。

求职：好。

雇主：假如有人告诉你，悉尼歌剧院着火了，你会怎麼做? 你会不会立即抢先报导这条新闻?

求职：我不会立即报导这条新闻。但是，我将会做的是打紧急电话 。

雇主：啊? 救火车应该早已经到达悉尼歌剧院了。难道说你打紧急电话因为你想叫更多的消防员?

求职：我打紧急电话不是叫更多的消防员，而是我想确定[**到底发生了什麼事**]*correct-foc*。因为如果歌剧院着火了，消防员早就会在那里了。他们会比谁都知道歌剧院的情形。

雇主：啊..

求职：这样做，我们能确定到底发生了什麼事， 然後得到真实的消息。但是有一点我不会做像其他记者。

雇主：哦? 那是什麼呀?

求职：我会在电视报导一点消息， 可是我不会马上报导[**全部**]*correct-foc* 消息。

雇主：为什麼呀? 是不是因为你不想给任何不肯定的消息呀?

求职：不是，我不想马上报导全部消息，因为我不想破坏观众的好奇心。好奇的观众才能继续看我们的新闻报导。

雇主：好奇心? 你这什麼意思啊? 你是不是把新闻报导当成一种类的商业广告?

求职：不是，新闻报导不是商业广告，但是它很像一种类的[**侦探**]*correct-foc* 节目。新闻报告其实就像侦探故事。

Appendix D

(Prosodic Entrainment Stimuli Sentences in English)

Note: Target-bearing words are italicised. The capitalised words are words with the predicted accent in the (a) predicted high and (b) predicted low stress sentences.

Experimental Sentences

1.
(a) I wish he weren't going to a *PARTY* on Monday
(b) I wish he weren't going to a *party* on MONDAY

2.
(a) The old lady thought she saw three *PIXIES* in her garden
(b) The old lady thought she saw three *pixies* in her GARDEN

3.
(a) All the contestants were in a state of PANIC when their names were called out
(b) All the contestants were in a state of panic when their NAMES were called out

4.
(a) Getting an Academy Award was the very *PEAK* of his extremely long career
(b) Getting an Academy Award was the very *peak* of his EXTREMELY long career

5.
(a) Her servants finally found a *PERFECT* way to disguise the stain
(b) Her servants finally found a *perfect* way to DISGUISE the stain

6.
(a) A crowd of activists threw *POWDER* at the mayor's face
(b) A crowd of activists threw *powder* at the mayor's FACE

7.
(a) None of the students could solve the *PUZZLES* the Russians had made
(b) None of the students could solve the *puzzles* the RUSSIANS had made

8.
(a) That summer four years ago I ate roast *PEANUTS* for every meal
(b) That summer four years ago I ate roast *peanuts* for EVERY meal

9.
(a) My friends and I used to meet in the *PARK* every day
(b) My friends and I used to meet in the *park* every DAY

10.
(a) They want to inform my *PARTNER* that I was sent home from work
(b) They want to inform my *partner* that I was sent HOME from work

11.
(a) Most of the jurors find it odd that the millionaire was *PARDONED* after the verdict
(b) Most of the jurors find it odd that the millionaire was *pardoned* AFTER the verdict

12.
(a) The hotel wants to hire more *PORTERS* to deal with the increase in guests
(b) The hotel wants to hire more *porters* to deal with the increase in GUESTS

13.
(a) Our clock no longer works ever since the *PENDULUM* went missing
(b) Our clock no longer works ever since the *pendulum* went MISSING

14.
(a) The surgeons must quickly remove her *PANCREAS* to delay the cancer from advancing
(b) The surgeons must quickly remove her *pancreas* to delay the CANCER from advancing

15.
(a) The Greeks once lived in a society where citizens had the *POWER* to demand their leaders' dismissal
(b) The Greeks once lived in a society where citizens had the *power* to demand their leaders' DISMISSAL

16.
(a) In some convents nuns still use *PADLOCKS* to seal their gates from the outside world
(b) In some convents nuns still use *padlocks* to seal their GATES from the outside world

17.
(a) Down on the farm we were amused to see a *PARROT* who could sing in French
(b) Down on the farm we were amused to see a *parrot* who could sing in FRENCH

18.
(a) Unfortunately the geologist didn't have enough time to *POLISH* all his minerals for the show
(b) Unfortunately the geologist didn't have enough time to *polish* ALL his minerals for the show

19.
(a) The naval officer shook hands with a *PIRATE* who rescued him from the fire
(b) The naval officer shook hands with a *pirate* who RESCUED him from the fire

20.
(a) A child who witnessed the crime said the gunman used his *PENCIL* to scare her away

(b) A child who witnessed the crime said the gunman used his *pencil* to SCARE her away

21.

(a) I was quite shocked to see the Archduke's *POODLES* eating truffles for lunch

(b) I was quite shocked to see the Archduke's *poodles* eating TRUFFLES for lunch

22.

(a) It is sad that the chief commander will *PUNISH* his men for saving the foreigners

(b) It is sad that the chief commander will *punish* his men for SAVING the foreigners

23.

(a) Marine scientists were angry when they discovered *PETROL* inside the whale's eyes

(b) Marine scientists were angry when they discovered *petrol* inside the whale's EYES

24.

(a) These tourists said they would like to *PICNIC* in the desert

(b) These tourists said they would like to *picnic* in the DESERT

Filler Sentences

4 filler sentences with early occurrence of the phoneme target

1. *PARSLEY* is the only thing you should add to the salad

2. In *POLAND* watching movies like "Home Alone" is now a Christmas tradition

3. Kim is *PAINTING* her own face with green and yellow ink for the soccer finale

4. You should not *PONDER* over what colour dress you will wear

4 filler sentences with late occurrence of the phoneme target

5. The examiner failed us on our driver's license after we told her she was too *PICKY*

6. According to researchers, children under eleven don't understand what a *PARTICLE* is

7. If something goes wrong during the flight the lead stewardess must tell the *PILOTS*

8. Many seafood lovers are unaware that some of the fish they eat may have *POISON* in their scales

16 filler sentences with no phoneme target

9. Shareholders sometimes take TOO much risk to make themselves rich

10. At the meeting the climatologists told the winery owners that they will NEVER survive if there’s no rain

11. His new house is of EXACTLY the same height as the surrounding high rises

12. Anna’s colleagues NEARLY fell down the stairs when they were getting off the train

13. After the earthquake our family had to SCAVENGE for food

14. Their new show was not good enough to AMAZE the audience

15. The giant ran towards the garden and DEVOURED all the flowers

16. Several folks from the village were DANCING in the streets

17. Magicians can use their cunning skills to CONTROL the audience’s emotions

18. In Congolese culture newlyweds are NOT allowed to smile on their wedding day

19. To get rid of such a massive amount of snow an ELECTRIC shovel is more convenient

20. Construction workers often work in all KINDS of weather conditions

21. The dressmakers at the fashion firm used METAL as material for their couture gowns

22. Quite a few travellers were arrested after COCAINE was found in their luggage

23. Everyone is talking about the HUNTER who lost his way in the woods

24. More than a THOUSAND cars were sold last year even though the economy wasn’t so good

Appendix E

(Prosodic Entrainment Stimuli Sentences in Mandarin)

Experimental Sentences in Mandarin (with rough IPA transcriptions)

1.
(a) 他們上星期去*爬山*踩了很多野花
(b) 他們上星期去*爬山*踩了**很多**野花

tʰa1	mən2	ʂaŋ4	ɕin1	tɕʰi1	tɕʰy4	**pʰa2**	**ʂan1**	tsʰai3	lə5	xən3	two1	jɛ3	xwa1
他	們	上	星	期	去	*爬*	*山*	踩	了	**很**	**多**	野	花

"他們"	"上星期"	"去"	"*爬山*"	"踩-了"	"**很多**"	"野花"
3.PL.M	last week	go	***HIKING***	stamp-PFV	**MANY**	wild flowers

"*They stamped on a lot of wild flowers while out hiking last week"*

2.
(a) 他想馬上回家因為他的*朋友*想偷他的錢
(b) 他想馬上回家因為他的*朋友*想偷他的**錢**

tʰa1	ɕjaŋ2	ma3	ʂaŋ4	xwei2	tɕja1	jin1	wei2	tʰa1	tɤ5	**pʰəŋ2**	**joʊ4**	ɕjaŋ3	tʰou1	tʰa1	tɤ5	tɕʰjɛn2
他	想	馬	上	回	家	因	為	他	的	*朋*	*友*	想	偷	他	的	錢

"他"	"想"	"馬上-回-家"	"因為"	"他-的"	"*朋友*"	"想"
3.s.M	want	quickly-return-home	because	3.s.M-GEN	***FRIENDS***	want

"偷"	"他-的"	"**錢**"
steal	3.s.M-GEN	**MONEY**

"He wants to quickly return home because he suspects that his friends want to steal his money"

3.
(a) 笑死人了, 這幾位遊客想穿***皮衣***在沙灘上溜達
(b) 笑死人了, 這幾位遊客想穿*皮衣*在**沙灘**上溜達

ɕjau4 sɨ3 ɻən2 lɤ5 ʈʂɤ4 tɕi3 wei4 jou2 kʰɤ4 ɕjɛŋ3 ʈʂwaŋ1 **pʰi2 ji1** ʈʂai4 ʂa1 tʰan1 ʂaŋ1 ljou1 dɤ5
笑 死 人 了, 這 幾 位 遊 客 想 穿 ***皮 衣*** 在 **沙灘** 上 溜 達

"笑死人了" "這" "幾-位" "遊客" "想" "穿" "***皮衣***"
laugh-die-people-PRF ART few-CLF tourist want wear ***JACKET***

"在" "**沙灘**" "上" "溜達"
LOC **BEACH** on-PREP stroll

"How funny! These tourists want to wear their leather jackets while strolling down the beach"

4.
(a) 昨天我看見倆個愛人在***蘋果樹***下偷偷地親嘴
(b) 昨天我看見倆個愛人在*蘋果樹*下偷偷地**親嘴**

tswo2 tʰjɛn1 wo3 kʰan4 tɕʰjɛn4 ljaŋ3 kʰɤ4 ai4 ɻən2 ʈʂai4 **pʰiŋ2 kwo3 ʂu4** ɕja4 tʰou1 tʰou1
昨 天 我 看 見 倆 個 愛 人 在 ***蘋 果 樹*** 下 偷 偷

dɤ5 tɕʰin1 tswei3
地 **親 嘴**

"昨天" "我" "看-見" "倆-個" "愛人" "在" "***蘋果樹***"
yesterday 1s see-R.COMP two-CLF lover LOC ***APPLE TREE***

"下" "偷偷-的" "**親嘴**"
under-PREP secretly-ADV **KISS**

"*Yesterday I saw two lovers kissing in secret under the apple tree"*

5.
(a) 沒有人在中國能相信***葡萄***能製造香水
(b) 沒有人在中國能相信 *葡萄*能製造**香水**

mei2 jou3 ɻən2 tsai4 ʈʂʊŋ1kwo3 nəŋ2 ɕjɑŋ1 ɕin4 **pʰu2 tʌ5** nəŋ2 ʈʂɨ 4 tsau4 ɕjaŋ1 ʂwei3
沒 有 人 在 中 國 能 相 信 ***葡 萄*** 能 製 造 **香 水**

"沒有" "人" "在-中國" "能" "相信" "***葡萄***" "能" "製造" "**香水**"
NEG people LOC-China can believe ***GRAPES*** can create **PERFUMES**

"*No one in China believes that grapes can be used to make perfumes*"

6.

(a) 我將家裡的一套***盤子***送給我的偶像

(b) 我將家裡的一套*盤子*送給我的**偶像**

wo3 tɕjaŋ1 tɕja1 li3 tɤ5 ji2 tʰau4 **pʰaŋ2 tsɨ5** sʊŋ4 kei3 wo3 tɤ5 ou3 ɕjaŋ4
我 將 家 裡 的 一 套 ***盤 子*** 送 給 我 的 **偶 像**

“我” “將” “家-裡-的-一-套” “***盤子***” “送-給”
1s FUT home-PREP.LOC-GEN-one-CLF ***PLATE*** give-R.COMP

“我-的” “**偶像**”
1.s-GEN **IDOL**

“*I shall give away my dinnerware as a present for my idol*”

7.

(a) 很多演員認為這***牌子***的鞋已經過時了

(b) 很多演員認為這*牌子*的鞋已經**過時**了

xəŋ3 twɔ1 jan3 ɥɛn2 ɻən4 wei2 ʈʂɤ4 **pʰai2 tsɨ5** tɤ5 ɕje2 ji2 tɕiŋ1 kwo4 ʂɨ2 lɤ5
很 多 演 員 認 為 這 ***牌 子*** 的 鞋 已 經 **過 時** 了

“很多” “演員” “認為” “這” “***牌子***-的” “鞋” 已經 “**過時**” “了”
Many actors think ART ***BRAND***-GEN shoe already **OUTDATED** PRF.COS

“*A lot of actors think that the shoes made by this brand are no longer in fashion*”

8.

(a) 聽說村裡那個長得像***螃蟹***的男孩要結婚

(b) 聽說村裡那個長得像*螃蟹*的男孩要**結婚**

tʰin1 ʂwo1 tsʰun1 li3 na4 kɤ5 ʈʂaŋ3 dɤ5 ɕjaŋ4 **pʰaŋ2 ɕɛ4** tɤ5 nan2 xai2 jau4 tɕjɛ2 xwən1
聽 說 村 裡 那 個 長 得 像 ***螃 蟹*** 的 男 孩 要 **結 婚**

“聽-說” “明天” “村-裡” “那-個” “長-得” “像”
heard-say tomorrow village-LOC.PREP ART-CLF look-A.COMP like

“***螃蟹***-的” “男孩” “要” “**結婚**”
CRAB-GEN boy AUX.FUT **MARRY**

“*It has been rumoured that that boy from the village who looks like a crab will get married tomorrow*”

9.
(a) 你可以看見他肚子***膨脹***得越來越大
(b) 你可以看見他肚子*膨脹*得越來越大

ni3 kɤ2 yi3 kʰan4 tɕʰjɛn4 tʰa1 tu4 tsɨ5 **pʰəŋ2 ʈʂaŋ4** tɤ5 ɥe4 lai2 ɥe4 da4
你 可 以 看 見 他 肚 子 ***膨 脹*** 得 越 來 越 大

"你" "可以" "看-見" "他" "肚子" "***膨脹***-得"
2.s can see-R.COMP 3.s.M stomach ***SWOLLEN***-D.COMP

"越來越" "大"
more and more **BIG**

"*You can see that his stomach is getting bigger and bigger*"

10.
(a) 我挺驚訝他會申請那套***便宜***的房子給自己住
(b) 我挺驚訝他會申請那套*便宜*的房子給**自己**住

wo3 tʰiŋ3 tɕiŋ1 ja4 tʰa1 xwei4 ʂəŋ1 tɕʰiŋ3 na4 tʰau4 **pʰjɛn2 ji5** tɤ5
我 挺 驚 訝 他 會 申 請 那 套 ***便 宜*** 的

faŋ2 tsɨ5 kei3 tsɨ4 tɕi3 ʈʂu4
房 子 給 **自 己** 住

"我" "挺" "驚訝" "他" "會" "申請" "那-套" "***便宜***-的"
1.s quite surprised 3.s.M FUT apply ART-CLF ***CHEAP***-D.COMP

"房子" "給" "**自己**" "住"
house give **SELF** live

"*I am quite surprised that he will apply to live in that cheap house by himself*"

11.
(a) 沒想到她乾女兒的***脾氣***能讓她得癌症
(b) 沒想到她乾女兒的*脾氣*能讓她得**癌症**

mei2 ɕjaŋ3 dau4 tʰa1 kan1 ny3 ɚ2 tɤ5 **pʰi2 tɕʰi4** nəŋ2 ɻan4 tʰa1 tɤ3 ai2 ʈʂəŋ4
沒 想 到 她 乾 女 兒 的 ***脾 氣*** 能 讓 她 得 **癌 症**

"沒" "想-到" "她" "乾-女兒-的" "***脾氣***"
NEG think-R.COMP 3.s.F adopted-daughter-GEN ***TEMPER***

"能" "讓" "她" "得" "**癌症**"
can CAUS 3.s.F acquire **CANCER**

"*Nobody would have thought that her adopted daughter's temper led her to have cancer*"

12.
(a) 身體虛弱的年輕人需要吃***排骨***來增加營養
(b) 身體虛弱的年輕人需要吃*排骨*來增加**營養**

ʂən1 tʰi3 ɕy1 ɻwo4 tɤ5 njɛn2 tɕʰiŋ1 ɻən2 ɕu1 jau4 ʈʂʰɨ1 **pʰai 2 ku3** lai2 tsəŋ1 tɕja1 jiŋ2 jaŋ3
身 體 虛 弱 的 年 輕 人 需 要 吃 ***排 骨*** 來 增 加 **營養**

"身體-虛弱-得" "年輕人" "需要" "吃" "***排骨***" "來" "增加"
body-weak-A.COMP young people need eat ***RIBS*** in order to add

"**營養**"
NUTRIENTS.

"*Young people who are physically weak need to eat some ribs to gain more nutrients*"

13.
(a) 這些狗仔隊能***破壞***總統的名聲
(b) 這些狗仔隊能*破壞***總統**的名聲

ʈʂɤ4 ɕjɛ1 gou2 tsai3 twei4 nəŋ2 **pʰwo4 xwai4** tsʊŋ2 tʰʊŋ3 tɤ5 miŋ2 ʂəŋ1
這 些 狗 仔 隊 能 ***破 壞*** **總 統** 的 名 聲

"這-些" "狗仔隊" "能" "***破壞***" "**總統**-的" "名聲"
ART-CLF paparazzi can ***RUIN*** **PRESIDENT**-GEN reputation

"*These paparazzis can ruin the president's reputation*"

14.
(a) 紅樓夢裡的姑娘長得***漂亮***因爲她們吃過仙丹
(b) 紅樓夢裡的姑娘長得*漂亮*因爲她們吃過**仙丹**

xʊŋ2 lou2 məŋ4 li3 tɤ5 ku1 njaŋ5 ʈʂaŋ3 tɤ5 **pʰau 4 ljɛn5** jin1 wei2 tʰa1 mən2
紅 樓 夢 裡 的 姑 娘 長 得 ***漂 亮*** 因 爲 她 們

ʈʂʰɨ2 kwo4 ɕjɛn1 tan1
吃 過 **仙 丹**

"紅樓夢-裡-的" "姑娘" "長-得" "*漂亮*"
"Dream Red Mansion-LOC-GEN" maiden look-D.COML ***BEAUTIFUL***

"因爲" "她們" "吃-過" "**仙丹**"
because 3.s.F.PLR eat-ASP **MAGIC POTION**

"*That maiden from "Dream of the Red Chamber" was beautiful because she once swallowed a magic potion*"

15.

(a) 我家大姐的行爲像***叛徒***因爲她取笑我們家的秘方
(b) 我家大姐的行爲像*叛徒*因爲她**取笑**我們家的秘方

wo3	tɕja1	da4	tɕjɛ3	tɤ5	ɕiŋ2	wei2	ɕjaŋ4	**p^{h}an4**	**t^{h}u2**	jin1	wei2	t^{h}a1	tɕ^{h}u3	ɕjau4
我	家	大	姐	的	行	爲	像	*叛*	*徒*	因	爲	她	**取**	**笑**

wo3	mən4	tɕja1	dɤ5	mi4	faŋ1
我	們	家	的	秘	方

"我"	"家"	"大姐-的"	"行爲"	"像"	"***叛徒***'	"因爲"	"她"	"**取笑**"
1.s	family	sister-GEN	behaviour	like	***TRAITOR***	because	2.s.F	**MOCK**

"我們-家-的"	"秘方"
1.PL-family-GEN	secret recipe

"*Our oldest sister acted like a traitor when she made a mockery of our family recipe*"

16.

(a) 時代雜誌的分析家預測音樂會的***票***將要跌下來
(b) 時代雜誌的分析家預測音樂會的*票*將要**跌**下來

ʂɨ2	tai4	tsa2	ʈʂɨ4	tɤ5	fən3	ɕi1	tɕja1	jy4	ts^{h}ɤ4	jin1	ɥe4	xwei4	tɤ5	**p^{h}jau4**	tɕjaŋ1	jau4	tjɛ2
時	代	雜	誌	的	分	析	家	預	測	音	樂	會	的	***票***	將	要	**跌**

ɕja4	lai2
下	來

"時代雜誌-的"	"分析家"	"預測"	"音樂會-的"	"***票***"	"將要"
Times-Magazine-GEN	analysts	predict	concert-GEN	***TICKET***	shall.FUT

"**跌**"	"下來"
DOWN	C.D.COMP

"*Analysts from the Times Magazine predict that the price of the concert will go down*"

17.
(a) 餐廳經理聽到***砲響***都嚇呆了
(b) 餐廳經理聽到*砲響*都嚇**呆**了

tsʰan1 tʰin1 tɕin1 li3 tʰiŋ1 tau4 **pʰau4 ɕjaŋ3** tou1 ɕja4 tai1 lɤ5
餐 廳 經 理 聽 到 ***砲 響*** 都 嚇 **獃** 了

“餐廳-經理” “聽-到” “***砲響***” “都”
restaurant-manager hear-R.COMP ***BOMB EXPLOSION*** dou.ADV

“**嚇獃**-了”
SCARED STIFF-COS.PRF

“*The restaurant manager was scared stiff after he heard an explosion*”

18.
(a) 有些護士喜歡向嬰兒的***屁股***打針
(b) 有些護士喜歡向嬰兒的*屁股*打**針**

jou3 ɕjɛ1 xu4 ʂɨ5 ɕi3 xwaŋ1 ɕjaŋ4 jin1 ɚ2 tɤ5 **pʰi4 ku5** ta3 tʂəŋ1
有 些 護 士 喜 歡 向 嬰 兒 的 ***屁股*** 打 **針**

“有-些” “護士” “喜歡” “向” “嬰兒-的” “***屁股***”
Some-CLF nurse like to.PREP infant-GEN ***GLUTE***

“打-**針**”
apply-**INJECTION**

“*Some nurses prefer performing glute injections on toddlers*”

19.
(a) 李先生逛超市時看見一位***胖子***買紅豆
(b) 李先生逛超市時看見一位*胖子*買**紅豆**

li3 ɕjɛn1 ʂən1 kwaŋ4 tʂʰau1 ʂɨ4 ʂɨ2 kʰan4 tɕjɛn4 ji2 wei4 **pʰaŋ4 tsɨ5** mai3 xʊŋ2 tou4
李 先 生 逛 超 市 時 看 見 一 位 ***胖 子*** 買 **紅 豆**

“李先生” “逛-超市-時” “看-見” “一-位” “***胖子***”
Mr. Lee stroll-market-TEMP see-R.COMP one-CLF ***OBESE PERSON***

“買” “**紅豆**”
buy **RED BEANS**

“*While doing grocery shopping, Mr Lee saw an obese guy buying red beans*”

20.
(a) 在羅馬有三個***騙子***往我們的方向走
(b) 在羅馬有三個*騙子*往**我們**的方向走

tsai4 lwo2 ma3 jou3 san1 kɤ4 **pʰjɛn4 tsɨ5** waŋ3 wo3 mən2 tɤ5 faŋ1 ɕjaŋ4 tsou3
在 羅 馬 有 三 個 ***騙 子*** 往 **我 們** 的 方 向 走

"在-羅馬" "有" "三-個" "***騙子***" "往" "**我們**-的-方向" "走"
LOC-Rome have three-CLF ***SWINDLER*** PREP **1.PL-GEN**-direction walk

"*When we were in Rome, three swindlers were walking in our direction*"

21.
(a) 住在山裡的那位小伙子買了一個***鋪頭***在蘭州
(b) 住在山裡的那位小伙子買了一個*鋪頭*在**蘭州**

ʈʂu4 tsai4 ʂan1 li3 tɤ5 na4 wei4 ɕjau2 xwo3 tsɨ5 mai3 lɤ5 ji2 kɤ4 **pʰu4 tʰou2**
住 在 山 裡 的 那 位 小 伙 子 買 了 一 個 ***鋪 頭***

tsai4 lan2 ʈʂou1
在 **蘭 州**

"住-在" "山-裡-的" "那-位" "小伙子" "買-了" "一-個"
live-LOC mountain-PREP-GEN ART-CLF young man buy-PRF one-CLF

"***鋪頭***" "在-**蘭州**"
SHOP LOC-**LANZHOU**

"*The young man who lived in the mountains bought a retail shop in Lanzhou*"

22.
(a) 小學生聽到喊叫后***怕***得貓在桌子底下
(b) 小學生聽到喊叫后 *怕*得貓在**桌子**底下

ɕjau3 ɕye2 ʂəŋ1 tʰiŋ1 tau4 xan3 tɕjau4 xou4 **pʰa4** tɤ5 mau1 tsai4 tswo1 tsɨ5 ti3 ɕja4
小 學 生 聽 到 喊 叫 后 ***怕*** 得 貓 在 **桌 子** 底下

"小學生" "聽到" "喊叫" "后" "***怕***-得" "貓"
Primary-school students hear-R.COMP scream after ***SCARED***-A.COMP hide

"在" "**桌子**" "底-下"
LOC **TABLE** under-D.COMP

"*After hearing someone screaming, the primary-school students were so scared that they hid under the table*"

23.
(a) 老奶奶每天一個人站在門前***盼望***她兒子從戰爭回家
(b) 老奶奶每天一個人站在門前*盼望*她兒子從**戰爭**回家

lau2 nai3 nai5 mei3 tʰjɛn1 ji2 kɤ4 ɹən2 ʈʂan4 tsai4 mən2 tɕʰjɛn2 **pʰan4 wɑŋ4** tʰa1 ɚ2 tsɨ5 tsʰʊŋ2
老 奶 奶 每 天 一 個 人 站 在 門 前 ***盼望*** 她 兒 子 從

ʈʂan4 ʈʂəŋ1 xwei2 tɕja1
戰 爭 回 家

"老奶奶" "每天" "一個人" "站-在" "門-前" "***盼望***' "她-兒子"
Old lady every day alone stand-LOC door-front ***YEARN*** 2.s.F-son

"從-**戰爭**" "回-家"
from.PREP-**WAR** return-home

"*Every day, the old lady stood in front of her doorstep and yearned for her son's return from war*"

24.
(a) 我很驚愕她開車***碰撞***了一支大象
(b) 我很驚愕她開車*碰撞*了一支**大象**

wo2 xən3 tɕiŋ1 ɤ4 tʰa1 kʰai1 ʈʂʰɤ1 **pʰəŋ4 ʈʂʰwaŋ4** lɤ5 ji4 ʈʂɨ1 ta4 ɕjaŋ4
我 很 驚 愕 她 開 車 ***碰 撞*** 了 一 支 **大 象**

"我" "很-驚愕" "她" "開車" "***碰撞***-了" "一-支" "**大象**"
1.s. very-horrified 3.s.F drive ***COLLIDE***-PRF.COS one-CLF **ELEPHANT**

"*I am completely horrified by the fact that she collided with an elephant while she was driving*"

Filler Sentences

4 filler sentences with early occurrence of the phoneme target and their translations in English

1.
她想***陪着***她的母親去澳大利亞參加婚禮
"*She wants to accompany her mother when they go to Australia for the wedding ceremony*"

2.
啤酒能有時候讓嗓子難受
"*Beer can sometimes make your throat feel uncomfortable*"

3.
在***派出所***我遇到了很多人
"*At the police station I encountered a lot of people"*

4.
我很***佩服***很多非常勇敢的哲學家
"*I really admire many philosophers who are very brave*"

4 filler sentences with late occurrence of the phoneme target

5.
這倆位園丁花三天三夜設計一個很華麗的***盆景***
"*The two gardeners spent three days and nights designing a beautiful bonsai tree"*

6.
研究地理的工程師喜歡在松樹的***旁邊***休息
"*The engineers who do research in geology prefer to take a rest next to the pine tree*"

7.
我們的手上粘滿了肥皂***泡沫***
"*Our hands are filled with soap bubbles*"

8.
地爭災民現在的食物要求很***頻道***
"The earthquake victims currently have a very *urgent* need for food"

16 filler sentences with no phoneme target

9.
真的沒看出來他的藝術眼光有**那麼**差
"*I have never really noticed that his taste for art can be that bad*"

10.
大公司的會計師**老是**埋怨他們公司的經濟困難
"*Accountants from big companies are always complaining about their company's financial problems*"

11.
我在俄羅斯下火車**差點兒**跌倒了 (note: focused word is disyllabic)
"*I almost fell down when I was getting off the train in Russia*"

12.
這塔**正好**更附近的樓一樣高
"*This tower is of exactly the same height as the surrounding buildings*"

13.
藥劑師知道怎麼**混和**中藥和其他的香料來提高藥的味道
"*Pharmacists know how to mix Chinese herbal medicine with other ingredients to enhance the medicine's flavour*"

14.
安娜卡列尼娜曾經和渥倫斯基**說過**她的生命很痛苦
"*Anna Karenina did tell Vronsky that she has suffered a lot in her life*"

15.
調查人員發現房間的溫度很**熱**
"*The investigator discovered that the room's temperature was very hot*"

16.
機場海關**沒有**沒受了走私的偽造手袋
"*At the airport the customs officers did not confiscate the smugglers' counterfeit handbags*"

17.
工會說建築工人總是在**危險**的環境工作
"*The unions said the construction workers' are working under very dangerous conditions*"

18.
快樂的夫婦從來沒有在公共場合吵過架
"*Couples who are happily married would never quarrel in public*"

19.
生存在城市的麻雀經常喜歡從**垃圾桶**掏食物吃
"*Finches living in big cities often like to scavenge for food from trash cans*"

20.
我知道有些人喜歡在**酒店**開會
"*I know there are people who prefer setting up conferences in hotels"*

21.
所有的律師同意馬路的衛生是**清潔工**的責任
"*All the lawyers unanimously agree that the hygiene in our streets is the cleaners' responsibility*"

22.
歷害的魔術師能用他的**手法**來影響其他人的心情
"*Skilled magicians can use his legerdemain to influence other people's mood*"

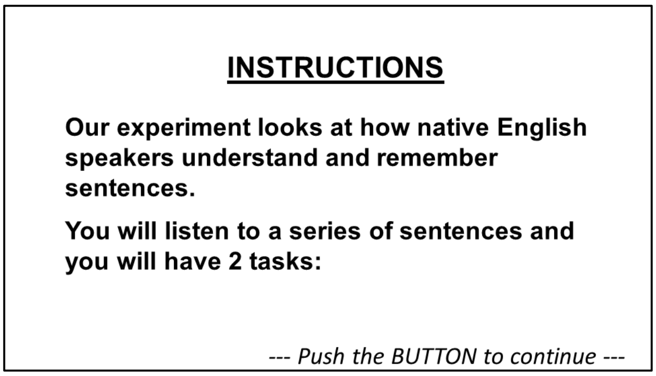
23.
沒見過一個**模特**有那麼多的學問
"*I have never met a model who is that knowledgeable*"

24.
藝術館失踪了一**千**幅畫因為晚上值勤的員工光顧看電視
"*A thousand paintings are missing at the art gallery because the night staff were watching TV*"

Appendix F

(Prosodic Entrainment Experiment: Instructions in English)

Slide 1



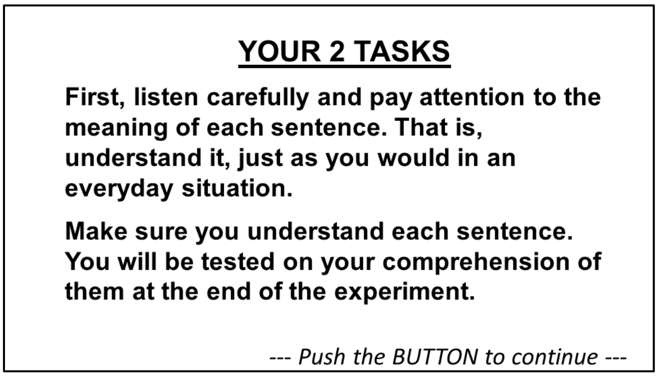
INSTRUCTIONS

Our experiment looks at how native English speakers understand and remember sentences.

You will listen to a series of sentences and you will have 2 tasks:

--- Push the BUTTON to continue ---

Slide 2



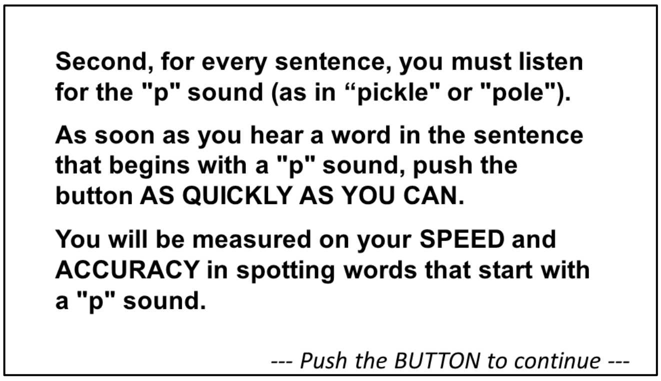
YOUR 2 TASKS

First, listen carefully and pay attention to the meaning of each sentence. That is, understand it, just as you would in an everyday situation.

Make sure you understand each sentence. You will be tested on your comprehension of them at the end of the experiment.

--- Push the BUTTON to continue ---

Slide 3



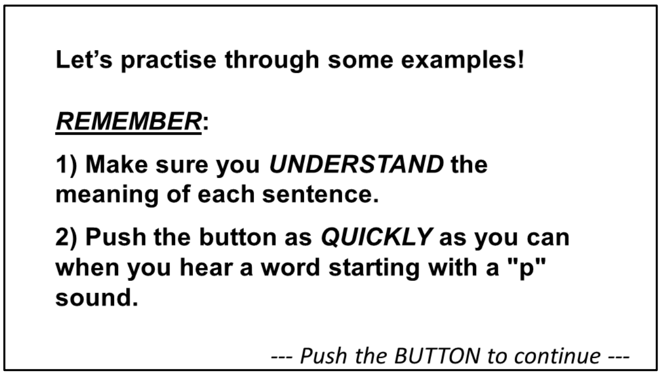
Second, for every sentence, you must listen for the "p" sound (as in “pickle" or "pole").

As soon as you hear a word in the sentence that begins with a "p" sound, push the button AS QUICKLY AS YOU CAN.

You will be measured on your SPEED and ACCURACY in spotting words that start with a "p" sound.

--- Push the BUTTON to continue ---

Slide 4



Let’s practise through some examples!

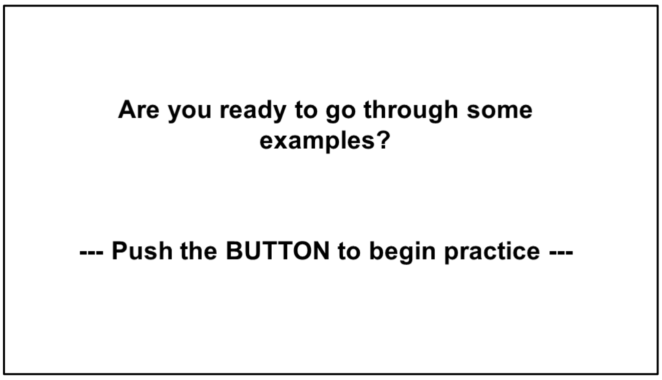
***REMEMBER*:**

1) Make sure you *UNDERSTAND* the meaning of each sentence.

2) Push the button as *QUICKLY* as you can when you hear a word starting with a "p" sound.

--- Push the BUTTON to continue ---

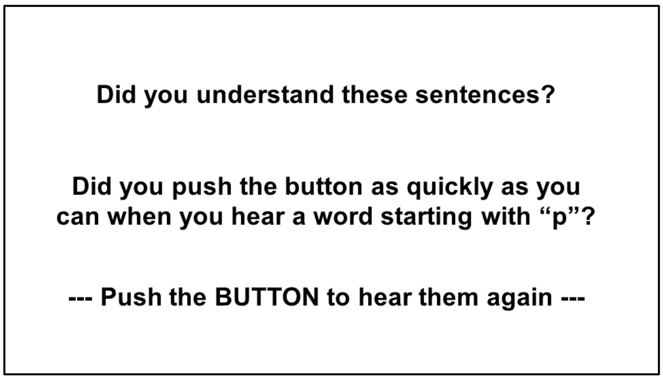
Slide 5



Are you ready to go through some examples?

--- Push the BUTTON to begin practice ---

Slide 6

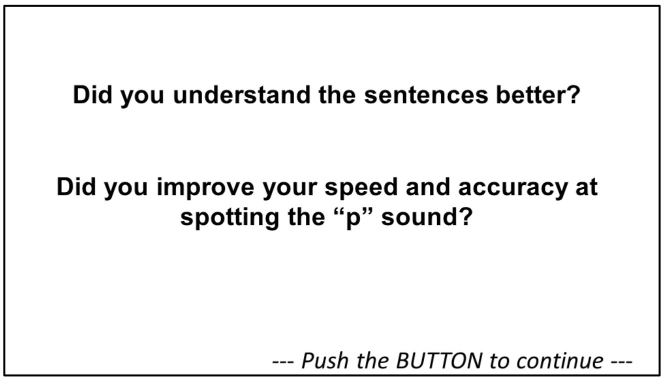


Did you understand these sentences?

Did you push the button as quickly as you can when you hear a word starting with “p”?

--- Push the BUTTON to hear them again ---

Slide 7

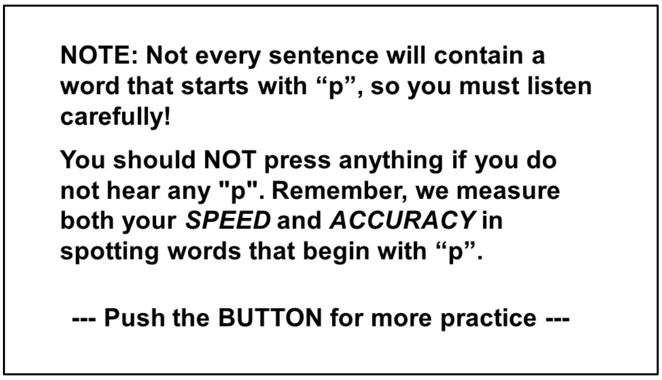


Did you understand the sentences better?

Did you improve your speed and accuracy at spotting the “p” sound?

--- Push the BUTTON to continue ---

Slide 8

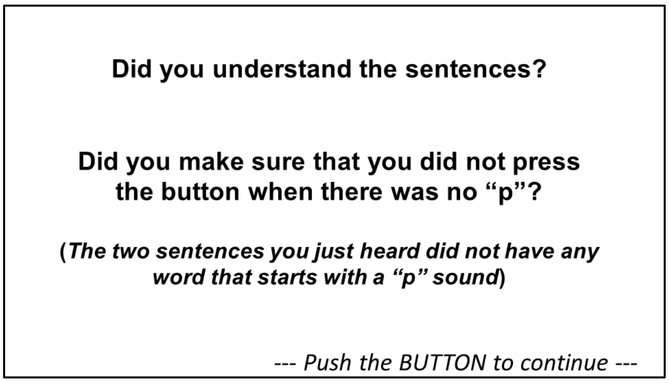


NOTE: Not every sentence will contain a word that starts with “p”, so you must listen carefully!

You should NOT press anything if you do not hear any "p". Remember, we measure both your *SPEED* and *ACCURACY* in spotting words that begin with “p”.

--- Push the BUTTON for more practice ---

Slide 9



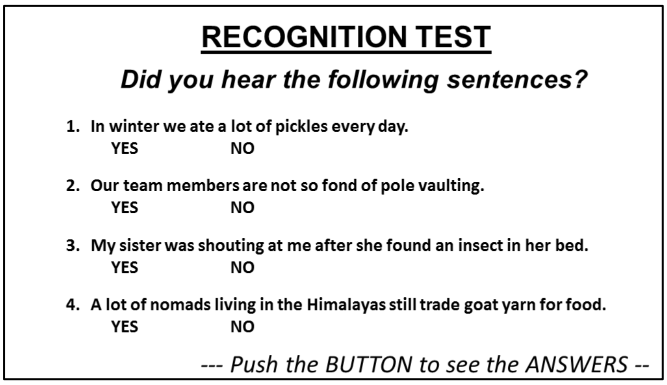
Did you understand the sentences?

Did you make sure that you did not press the button when there was no "p"?

(*The two sentences you just heard did not have any word that starts with a "p" sound*)

--- Push the BUTTON to continue ---

Slide 10



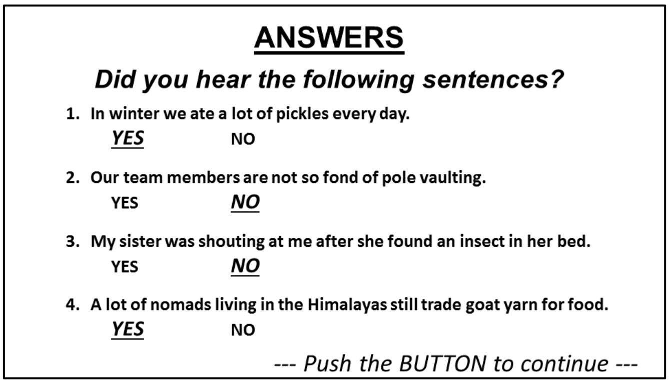
RECOGNITION TEST

Did you hear the following sentences?

1. **In winter we ate a lot of pickles every day.**
   YES NO
2. **Our team members are not so fond of pole vaulting.**
   YES NO
3. **My sister was shouting at me after she found an insect in her bed.**
   YES NO
4. **A lot of nomads living in the Himalayas still trade goat yarn for food.**
   YES NO

--- Push the BUTTON to see the ANSWERS --

Slide 11



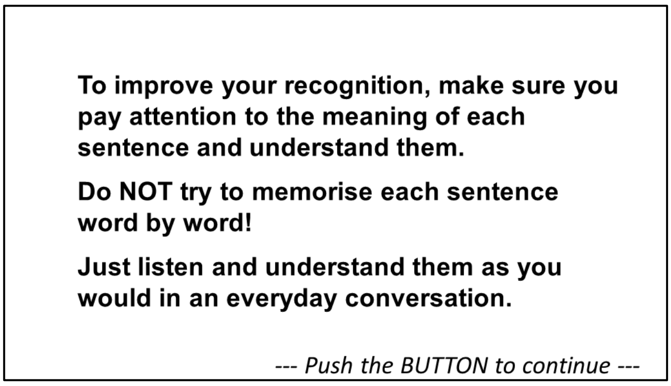
ANSWERS

Did you hear the following sentences?

1. **In winter we ate a lot of pickles every day.**
   YES **NO**
2. **Our team members are not so fond of pole vaulting.**
   YES ***NO***
3. **My sister was shouting at me after she found an insect in her bed.**
   YES ***NO***
4. **A lot of nomads living in the Himalayas still trade goat yarn for food.**
   YES **NO**

--- Push the BUTTON to continue ---

Slide 12



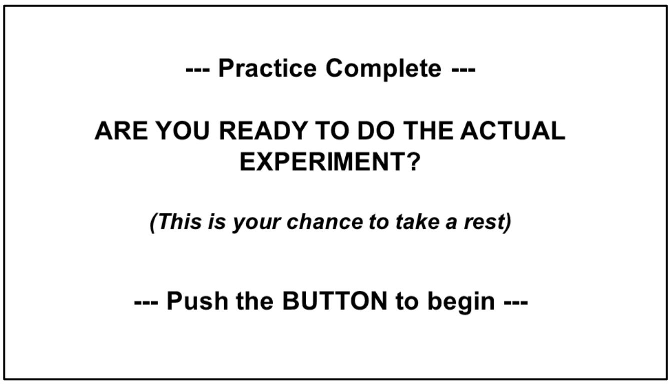
To improve your recognition, make sure you pay attention to the meaning of each sentence and understand them.

Do NOT try to memorise each sentence word by word!

Just listen and understand them as you would in an everyday conversation.

--- Push the BUTTON to continue ---

Slide 13



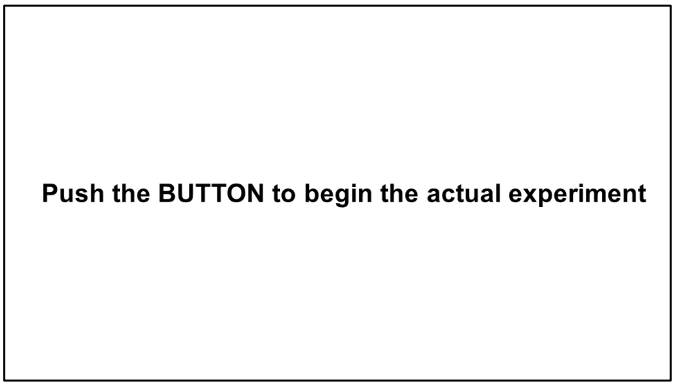
--- Practice Complete ---

ARE YOU READY TO DO THE ACTUAL EXPERIMENT?

(This is your chance to take a rest)

--- Push the BUTTON to begin ---

Slide 14

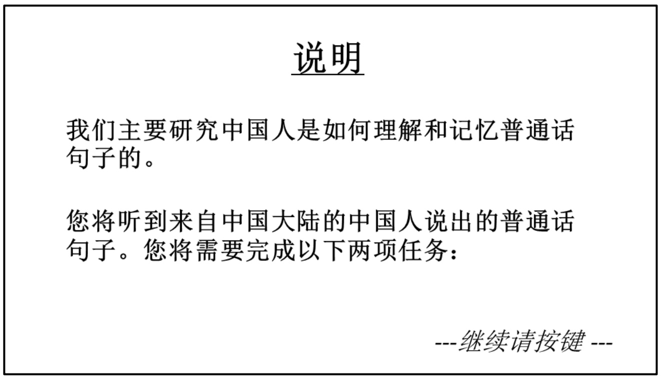


Push the BUTTON to begin the actual experiment

Appendix G

(Prosodic Entrainment Experiment: Instructions in Simplified Chinese)

Slide 1



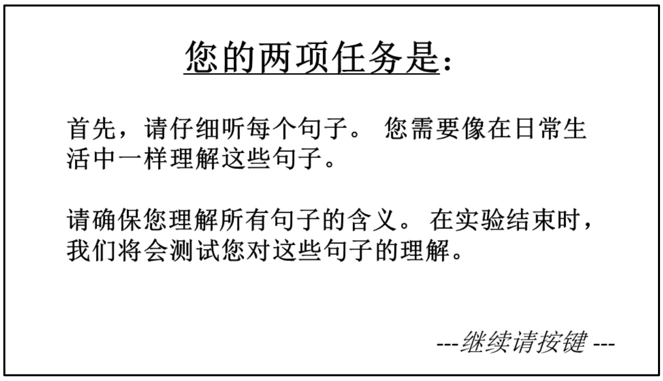
说明

我们主要研究中国人是如何理解和记忆普通话句子的。

您将听到来自中国大陆的中国人说出的普通话句子。您将需要完成以下两项任务：

---*继续请按键* ---

Slide 2



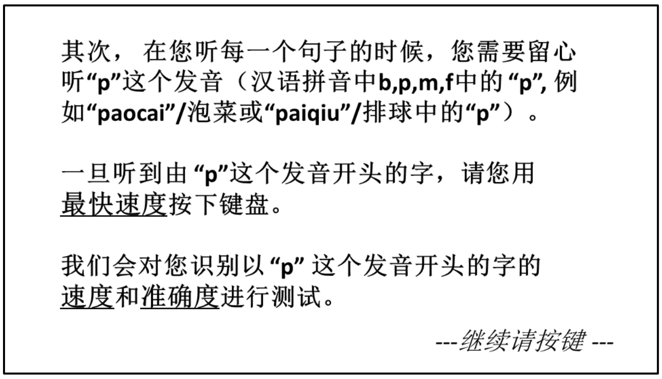
您的两项任务是：

首先，请仔细听每个句子。 您需要像在日常生活中一样理解这些句子。

请确保您理解所有句子的含义。 在实验结束时，我们将会测试您对这些句子的理解。

---*继续请按键* ---

Slide 3



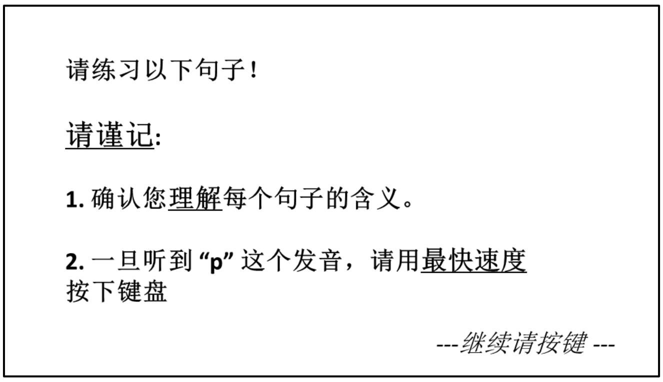
其次， 在您听每一个句子的时候，您需要留心听“**p**”这个发音（汉语拼音中**b,p,m,f**中的 “**p**”, 例如“**paocai**”/泡菜或“**paiqiu**”/排球中的“**p**”）。

一旦听到由 “**p**”这个发音开头的字，请您用最快速度按下键盘。

我们会对您识别以 “**p**” 这个发音开头的字的速度和准确度进行测试。

---*继续请按键* ---

Slide 4



请练习以下句子！

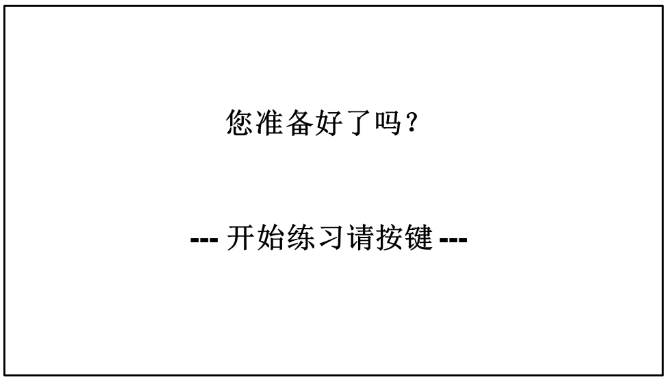
请谨记:

1. 确认您理解每个句子的含义。

2. 一旦听到 “**p**” 这个发音，请用最快速度按下键盘

---*继续请按键* ---

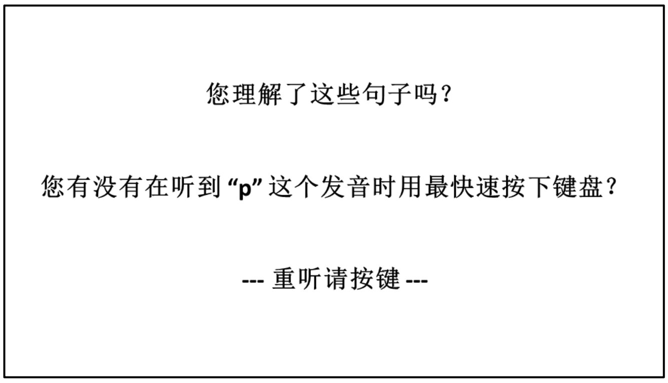
Slide 5



您准备好了吗？

--- 开始练习请按键 **---**

Slide 6

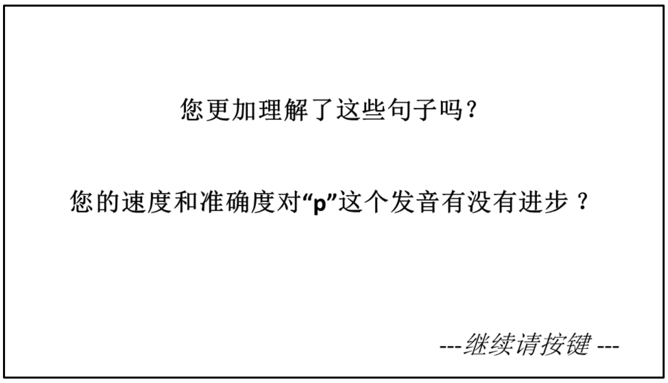


您理解了这些句子吗？

您有没有在听到 “**p**” 这个发音时用最快速按下键盘？

--- 重听请按键 ---

Slide 7

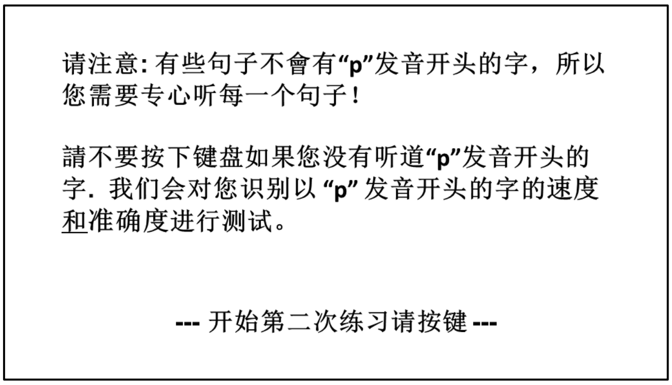


您更加理解了这些句子吗？

您的速度和准确度对“**p**”这个发音有没有进步 ？

---*继续请按键* ---

Slide 8

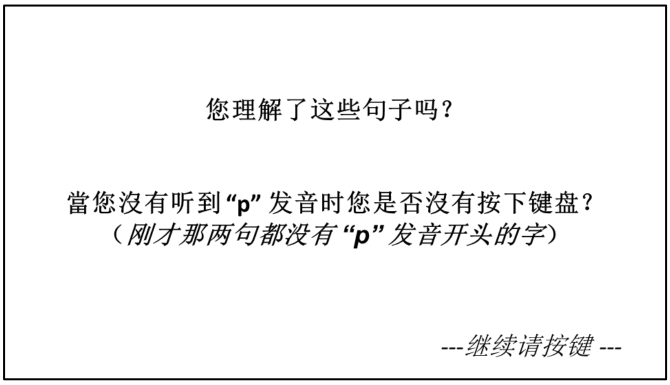


请注意: 有些句子不會有“**p**”发音开头的字，所以您需要专心听每一个句子！

請不要按下键盘如果您没有听道“**p**”发音开头的字. 我们会对您识别以 “**p**” 发音开头的字的速度和准确度进行测试。

--- 开始第二次练习请按键 **---**

Slide 9

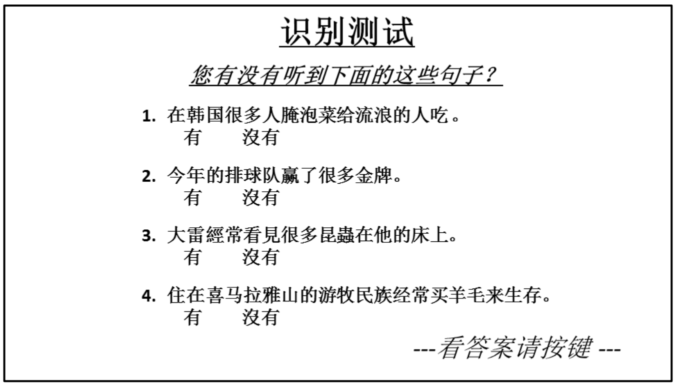


您理解了这些句子吗？

當您沒有听到“p” 发音时您是否沒有按下键盘？
（*刚才那两句都没有 “p” 发音开头的字*）

---继续请按键 ---

Slide 10



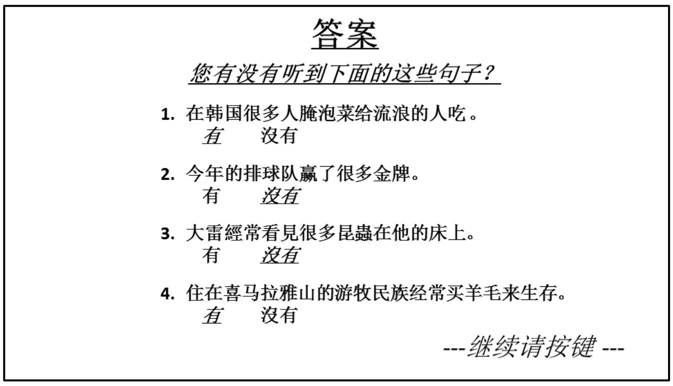
<u>识别测试</u>

<u>您有没有听到下面的这些句子？</u>

1. 在韩国很多人腌泡菜给流浪的人吃。
   有　　沒有
2. 今年的排球队赢了很多金牌。
   有　　沒有
3. 大雷經常看見很多昆蟲在他的床上。
   有　　沒有
4. 住在喜马拉雅山的游牧民族经常买羊毛来生存。
   有　　沒有

---看答案请按键 ---

Slide 11



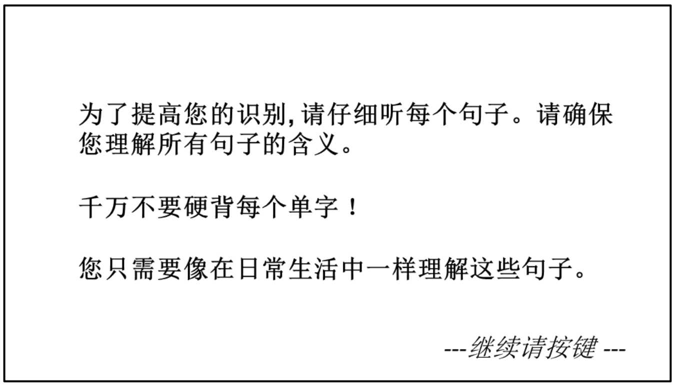
<u>答案</u>

<u>您有没有听到下面的这些句子？</u>

1. 在韩国很多人腌泡菜给流浪的人吃。
   <u>有</u>　　沒有
2. 今年的排球队赢了很多金牌。
   有　　*<u>沒有</u>*
3. 大雷經常看見很多昆蟲在他的床上。
   有　　*<u>沒有</u>*
4. 住在喜马拉雅山的游牧民族经常买羊毛来生存。
   <u>有</u>　　沒有

---继续请按键 ---

Slide 12



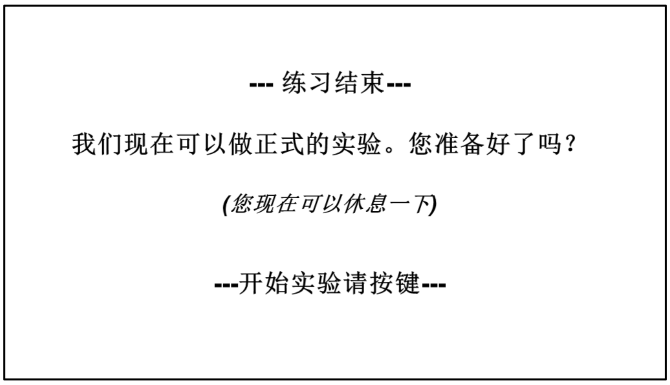
为了提高您的识别，请仔细听每个句子。请确保您理解所有句子的含义。

千万不要硬背每个单字！

您只需要像在日常生活中一样理解这些句子。

---继续请按键 ---

Slide 13



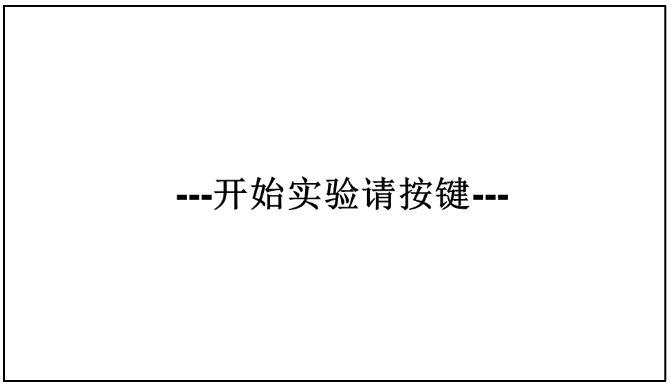
--- 练习结束**---**

我们现在可以做正式的实验。您准备好了吗？

(**您现在可以休息一下**)

---开始实验请按键**---**

Slide 14



---开始实验请按键**---**

Appendix H

(Recognition Test in English)

RECOGNITION TEST

Did you hear the following sentences? Please circle your response.

1) The very peak of his acting career was not when he received the Golden Globe's award.

YES NO

2) After the earthquake, our family had to scavenge for food.

YES NO

3) That summer four years ago, I ate roast peanuts for every meal.

YES NO

4) Most of the jurors find it odd that the millionaire was pardoned after the verdict

YES NO

5) No one in the farm was surprised to see the parrot when it sang in German.

YES NO

6) Down on the farm we were amused to see a parrot who could sing in French.

YES NO

7) The porter stole a tourist's suitcase while he was working in the lobby.

YES NO

8) Three fairies appeared in my grandmother's backyard yesterday.

YES NO

9) Magicians can use their cunning skills to control the audience's emotions.

YES NO

10) Everyone is talking about the hunter who lost his way in the woods.

YES NO

11) The teacher called her partner and told him that their daughter was sent home from school.

YES NO

12) The giant ran towards the gate and devoured all the flowers.

YES NO

13) The countess's dogs are very spoiled because they eat caviar every morning.

YES NO

14) Most of the farmers in the village say they like to dance when they hear music.

YES NO

15) Unfortunately the geologist didn't have enough time to polish all his minerals for the show.

YES NO

16) Several of my friends from Wall Street are now in danger of losing their wealth.

YES NO

17) Some students always party, even when they should be revising for the exams.

YES NO

18) The soldiers couldn't break the code the foreigners had used.

YES NO

19) All the contestants were in a state of panic when their names were called out.

YES NO

20) The dressmakers at the fashion firm used metal as material for their couture gowns.

YES NO

Appendix I

(Recognition Test in Mandarin)

识别测试

您有没有听到下面的这些句子？请在答案上画圈

1 *我认为这牌子的衣服还是太土了*
有 沒有

2 我在俄罗斯下火车差点儿跌倒了
有 沒有

3 没有人在中国能相信葡萄能制造香水
有 沒有

4 我挺惊讶他会申请那套便宜的房子给自己住
有 沒有

5 大家都很高兴因为那个长得像螃蟹的女孩要结婚
有 沒有

6 听说村里那个长得像螃蟹的男孩要结婚
有 沒有

7 我对我的朋友很失望因为他们现在都很贪财
有 沒有

8 这些游客在市场买了很多西瓜
有 沒有

9 厉害的魔术师能用他的手法来影响其他人的心情
有 沒有

10 机场海關沒有沒收走私的伪造手袋
有 沒有

11 很多人喜欢用大盘子吃意粉

有　　　　　　沒有

12 所有的律师同意马路的卫生是清洁工的责任

有　　　　　　沒有

13 我的大哥在香港岛买了一套很小的公寓

有　　　　　　沒有

14 我的同事经常说我应该讲话大点声

有　　　　　　沒有

15 昨天我看见俩位爱人在苹果树下偷偷地亲嘴

有　　　　　　沒有

16 有些老人依靠卖奶粉来生存

有　　　　　　沒有

17 我的家人喜欢在加拿大爬山多过游泳

有　　　　　　沒有

18 今天我把我的两千块杯子送给了我的最崇拜的歌星

有　　　　　　沒有

19 我家大姐的行为像叛徒因为她取笑我们家的秘方

有　　　　　　沒有

20 艺术馆失踪了一千幅画因为晚上值勤的员工光顾看电视

有　　　　　　沒有

Appendix J

(Ambiguous Sentences in English)

Participants heard ambiguous sentences produced with an "Early Juncture" or "Late Juncture" (italicised) and were asked to choose the correct interpretation sentence by pressing either the left or right button (note that some designated early junctures may be optional). Button locations and juncture versions were counterbalanced.

EXPERIMENTAL SENTENCES

1.

LEFT BUTTON
Simon gave dog biscuits to Mary

RIGHT BUTTON
Simon gave biscuits to Mary's dog Max

"Simon gave her # dog biscuits" OR "Simon gave her dog # biscuits"

2.

LEFT BUTTON
The boy gave cat food to Mrs. Hubbard

RIGHT BUTTON
The boy gave food to Mrs. Hubbard's cat Tommy

"Last night, the boy gave her # cat food" OR "Last night, the boy gave her cat # food"

3.

LEFT BUTTON
Larry gave horse radishes to Anne everyday

RIGHT BUTTON
Larry gave radishes to Anne's horse Albert everyday

"Everyday, Simon gave her # horse radishes" OR "Everyday, Simon gave her horse # radishes"

4.
LEFT BUTTON
Tonight, John the butler will serve catfish to Madame Aubert

RIGHT BUTTON
Tonight, John the butler will serve fish to Madame Aubert's cat Felix

"Tonight, John the butler will serve her # catfish for dinner" OR "Tonight, John the butler will serve her cat # fish for dinner"

5.

LEFT BUTTON
Last night, Nora got into trouble because she served ladybirds to Mme Aubert

RIGHT BUTTON
Last night, Nora got into trouble because she served birds to Mme Aubert

"Last night, Nora got into trouble for serving her # ladybirds" OR "Last night, Nora got into trouble for serving her lady # birds"

6.

LEFT BUTTON
David accidentally gave rat poison to Hannah

RIGHT BUTTON
David accidentally gave poison to Hannah's pet rat Rohan

"David accidentally gave her # rat poison" OR "David accidentally gave her rat # poison"

7.

LEFT BUTTON
In the novel "Peter Pan", Wendy made pancakes for Peter Pan

RIGHT BUTTON
In the novel "Peter Pan", Wendy made cakes for Peter Pan

"In the movie, Wendy made Peter Pan cookies. But in the actual novel, Wendy made Peter # pancakes" OR "In the movie, Wendy made Peter Pan cookies. But in the actual novel, Wendy made Peter Pan # cakes"

8.

LEFT BUTTON
Mrs. Fields fed goat's milk to her baby

RIGHT BUTTON
Mrs. Fields fed milk to her baby goats

"This morning, Mrs. Fields fed her baby # goat's milk" OR "This morning, Mrs. Fields fed her baby goats # milk"

EXPERIMENTAL SENTENCES

9.

LEFT BUTTON
Sam gave the baby milk to Sophie

RIGHT BUTTON
Sam gave milk to Sophie's baby

"An hour ago, Sam gave her # baby milk" OR "An hour ago, Sam gave her baby # milk"

10.

LEFT BUTTON
We fed some fishcake crumbs to Brigitte

RIGHT BUTTON
We fed some cakecrumbs to Brigitte's fish Harry

"We fed her # fishcake crumbs" OR "We fed her fish # cakecrumbs"

11.

LEFT BUTTON
The children accidentally gave duckweeds to Janet Farmer

RIGHT BUTTON
The children accidentally gave weeds to Janet Farmer's pet duck

"The children accidentally gave her # duckweeds to eat" OR "The children accidentally gave her duck # weeds to eat"

12.

LEFT BUTTON
The old sorcerer fed dragonfruits to her pet

RIGHT BUTTON
The old sorcerer fed fruits to her pet dragon

"For breakfast, the old sorcerer fed her pet # dragonfruits" OR "For breakfast, the old sorcerer fed her pet dragon # fruits"

13.

LEFT BUTTON
The tour guide made the ginseng tea from Korea

RIGHT BUTTON
The tour guide made ginseng tea for the Korean tourist

"The tour guide made # the Korean ginseng tea" OR "The tour guide made the Korean # ginseng tea"

14.
LEFT BUTTON
This year, the host family gave away fans that are from Japan

RIGHT
This year, the host family gave away fans to the Japanese

"This year, the host family gave # the Japanese fans" OR "This year, the host family gave the Japanese # fans"

15.

LEFT BUTTON
The waiter served salads to the Greeks

RIGHT BUTTON
The waiter served the Greek salads to all

"The waiter served # the Greek salad" OR "The waiter served the Greeks # salad"

16.

LEFT BUTTON
The chef cooked spaghetti for the Sicilians

RIGHT BUTTON
The chef cooked Sicilian-styled spaghetti

"Our chef cooked # the Sicilian spaghetti" OR "Our chef cooked the Sicilians # spaghetti"

EXPERIMENTAL SENTENCES

17.

LEFT BUTTON
In the morning, the hotel chef serves toasts to the French tourists

RIGHT BUTTON
In the morning, the hotel chef serves French-style toasts

"Every morning the hotel chef serves # the French toast" OR "Every morning the hotel chef serves the French # toast"

18.

LEFT BUTTON
Gertrude narrated her travel stories to the English people

RIGHT BUTTON
Gertrude narrated her stories about her travels in England

"Today, Gertrude told # the English travel stories" OR "Today, Gertrude told the English # travel stories"

19.

LEFT BUTTON
The ambassadors will present the Chinese with porcelain vases

RIGHT BUTTON
The ambassadors will give vases that are made of Chinese porcelain

"Next year, the German ambassadors will give # the Chinese porcelain vases" OR "Next year, the German ambassadors will give the Chinese # porcelain vases"

20.

LEFT BUTTON
Bette only bought sunglasses for her son

RIGHT BUTTON
Bette only bought glasses for her son

"Bette only bought her # sunglasses" OR "Bette only bought her son # glasses"

EXPERIMENTAL SENTENCES

21.

LEFT BUTTON
Mr. Johnson saw Jessica kneel under the table

RIGHT BUTTON Mr. Johnson saw Jessica's pet duck, Donald, under the table

"Yesterday, Mr. Johnson saw her # duck under her table" OR "Yesterday, Miss. Johnson saw her duck # under her table"

22.

LEFT BUTTON
Holly was fooling around in the churchyard

RIGHT BUTTON
Holly's pet got into the churchyard

"Holly wondered whether the vicar had seen her monkey around the churchyard" OR "Holly wondered whether the vicar had seen her monkey # around the churchyard"

FILLER SENTENCES

1. - Syntactic ambiguity

LEFT BUTTON
Adam has never tried meat before – he has no idea that meat can be really tasty

RIGHT BUTTON
Adam has never tried good-quality meat before – he has no idea that good-quality meat can be really tasty.

"Adam has no idea how good meat tastes"

---

2. - Syntactic ambiguity

LEFT BUTTON
Going away to visit relatives can be really tiring

RIGHT BUTTON
Our relatives who are currently visiting can be really tiring

"*Visiting relatives can be really tiring"*

FILLER SENTENCES

3. Syntactic ambiguity (attachment ambiguity)

LEFT BUTTON
The nun used a telescope to see the woman standing on the hill

RIGHT BUTTON
The nun saw the woman who standing on the hill carrying her a telescope

"The nun saw the woman on the hill with a telescope"

4. -– Syntactic ambiguity (attachment ambiguity)

LEFT BUTTON
Old Grandma Fensby used could only look at the injured dog with one of her eyes.

RIGHT BUTTON
Old Grandma Fensby looked at the injured dog who only had one eye

"*Old Grandma Fensby was staring at the injured dog with only one eye"*

5. - Semantic ambiguity

LEFT BUTTON
Joan also loves her own mother

RIGHT BUTTON
Joan also loves Natasha's mother

"*Natasha loves her mother and Joan does too"*

6. - Semantic ambiguity

LEFT BUTTON
At the train station, both Richard and Marcel said goodbye to their own wives

RIGHT BUTTON
At the train station, Richard said goodbye to his wife, who was also farewelled by Marcel

"*At the train station, Richard said goodbye to his wife and Marcel did too"*

FILLER SENTENCES

7. – Anaphoric ambiguity

LEFT BUTTON
The parrot got really oily

RIGHT BUTTON
The kitchen got really oily

“*The parrot stood on the kitchen table and it tried to get some oil out of the oil jar. It spilled everywhere. It soon got really oily.*

---

8. – Anaphoric ambiguity

LEFT BUTTON
The horse got really muddy

RIGHT BUTTON
The hill got really muddy

“*The horse tried to go up a hill while it was raining. It was very steep. It soon got really muddy from all the thumping”*

---

9. – Lexical ambiguity involving homonyms

LEFT BUTTON
Sasha wants to eat some vegetables

RIGHT BUTTON
Sasha wants to go to space

“*Sasha wants to ride on a rocket”*

---

10. – Lexical ambiguity involving homonyms

LEFT BUTTON
Lola went to the saving bank

RIGHT BUTTON
Lola went to the bank of the river

“Lola went to the bank for a swim”

FILLER SENTENCES

11. – Lexical ambiguity involving homonyms

LEFT BUTTON
Joe wants to find another red sock

RIGHT BUTTON
Joe wants to light his cigarette

"*With a cigarette in his hand, Joe looked in the drawer for a match*"

---

12. Lexical ambiguity involving homonyms

LEFT BUTTON
I saw military tanks

RIGHT BUTTON
I saw water tanks

"After the riots, I saw tanks in the town square"

Appendix K

(Ambiguous Sentences in Mandarin)

Participants heard ambiguous sentences produced with an "Early Juncture" or "Late Juncture" (italicised) and were asked to choose the correct interpretation sentence by pressing either the left or right button. (note that some designated early junctures may be optional). Button locations and juncture versions were counterbalanced.

EXPERIMENTAL SENTENCES

1.

左键：

马大叔把狗骨头喂给小红

右键：

马大叔把骨头喂给小红的狗

"马大叔喂她 # 狗骨头" OR "马大叔喂她狗 # 骨头"

2.

左键：

大卫不小心把狗饼干喂给老奶奶

右键：

大卫不小心把饼干喂给老奶奶的狗

"大卫不小心喂她 # 狗饼干吃" OR "大卫不小心喂她狗 # 饼干吃"

3.

左键：

王力强每天煮好洋葱给他媳妇吃

右键：

王力强每天准备好葱给他媳妇的羊吃

"王力强每天给她 # 洋葱吃" OR "王力强每天给她羊 # 葱吃"

4.

左键：

刘老师不小心把猪排骨汤给这位吃素的女孩

右键：

刘老师不小心把排骨汤给这位吃素女孩的宠物猪

“刘老师不小心给她 # 猪排骨汤喝” OR “刘老师不小心给她猪 # 排骨汤喝”

5.

左键：

王元把鸡蛋糕给马大姐吃

右键：

王元把蛋糕给马大姐的鸡吃

“王元给她 # 鸡蛋糕” OR “王元给她鸡 # 蛋糕”

6.

左键：

林姐把狗肉给李燕吃

右键：

林姐把肉给李燕的狗吃

“林姐给她 # 狗肉吃” OR “林姐给她狗 # 肉吃”

7.

左键：

凌天生不小心给梦瑶吃象牙

右键：

凌天生不小心给梦瑶的大象吃牙

“凌天生不小心给她 # 象牙吃” OR “凌天生不小心给她象 # 牙吃”

8.

左键：为了补充营养, 医院的护士也冲婴儿奶粉给这位孕妇

右键：为了补充营养, 医院的护士也冲奶粉给这位孕妇的婴儿

“为了补充营养, 医院的护士也给她 # 婴儿奶粉” OR“为了补充营养, 医院的护士也给她婴儿 # 奶粉”

9.

左键：

这位老仙人煮小龙虾给他宠物

右键：

这位老仙人煮虾给他宠物小龙

“这位老仙人喂他宠物 # 小龙虾吃” OR “这位老仙人喂他宠物小龙 # 虾吃”

10.

左键：

今天中国大使在首尔把很贵的参茶送给很多韩国人

右键：

今天中国大使在首尔把很贵的人参茶送给韩国

“中国大使今天给韩国人 # 参茶” OR “中国大使今天给韩国 # 人参茶”

11.

左键：

大姨把蛇泡酒给小芳喝

右键：

大姨把泡酒给小芳的宠物蛇喝

“大姨给她 # 蛇泡酒喝” OR “大姨给她蛇 # 泡酒喝”

12.

左键：

小平把奶油递给刘成

右键：

小平把油递给刘成的奶

“小平递给他 # 奶油” OR “小平递给他奶 # 油”

13.

左键：

李碧月把奶粉递给浩然

右键：

李碧月把粉递给浩然的奶

“李碧月递给他 # 奶粉” OR “李碧月递给他奶 # 粉”

14.

左键：

昨天早上小毛不小心炖牛羊肉给他姥姥吃

右键：

昨天早上小毛不小心炖羊肉给他姥姥的老牛吃

“昨天早上小毛不小心喂他姥 # 牛羊肉吃” OR “昨天早上小毛不小心喂她老牛 # 羊肉吃”

15.

左键：

刘波不小心把老鼠药给珍妮

右键：

刘波不小心把药给珍妮的宠物鼠

“刘波不小心给她 # 老鼠药吃” OR “刘波不小心给她老鼠 # 药吃”

16.

左键：杨七郎把花椒水给了谷姨

右键：杨七郎给谷姨的花浇水

“杨七郎给她 # 花椒水” OR “杨七郎给她花 # 浇水”

17.

左键：

这淘气的小男孩向周姥姥吐牛奶

右键：

这淘气的小男孩给周姥姥的宠物兔喝牛奶

“这淘气的小男孩给她吐牛奶” OR “这淘气的小男孩给她兔 # 牛奶”

18.

左键：

吴老师看见小学生如玲藏在桌子底下

右键：

吴老师看见小学生如玲的宠物在桌子底下

“吴老师看到她 # 猫在桌子低下” OR “吴老师看到她猫 # 在桌子低下”

19 ·

左键：

总经理让金长江在这个舞台上霸占位置

右键：

总经理让金长江的爸在这个舞台上站着

“总经理让他霸占 # 在这个舞台上” OR “总经理让他爸 # 站在这个舞台上”

20.

左键：

曹小姐给小李一块抹布 – 小李用这块抹布作清洁

右键：

曹小姐给小李的妈一块布 – 小李的妈用这块布作清洁

“曹小姐给她 # 抹布用作清洁” OR “曹小姐给她妈 # 布用作清洁”

21.

左键：王厨师只煮素的米汤给这些出家和尚

右键：王厨师只煮粟米汤给这些出家和尚

“王厨师只给她们煮素 # 米汤” OR “王厨师只给她们煮粟米汤”

EXPERIMENTAL SENTENCES

22.

左键：

方方的妈生了一对双胞胎女儿 – 她们很可爱

右键：

方方生了一对孖生女儿 – 她们很可爱

"方方的妈 # 生一对女儿很可爱" OR "方方的孖生一对女儿很可爱"

FILLER SENTENCES

1. – Garden path sentence (relative vs. complement clauses)

左键：我的邻居只想种苹果

右键：我的邻居只欣赏会种苹果的农民

"我家邻居只喜欢种苹果的农民"

2. – Garden path sentence (relative vs. complement clauses)

左键：张少爷只喜欢去法国读书
右键：张少爷只喜欢法国的大学学生

"张少爷只喜欢跟去法国上过大学的学生聊天"

FILLER SENTENCES

3. – Semantic ambiguity

左鍵：瑪利亞的母親叫她盡量吃多點

右鍵：瑪利亞的母親叫她盡量吃少點

“瑪利亞已經三天沒吃飯了。今天她媽煮飯時跟她說，“你能吃多少就吃多少！”

---

4. – Semantic ambiguity

左鍵：赵老爷开始埋怨因为钱花得太多了

右鍵：赵老爷开始埋怨因为他有太多鲜花了

“趙老爺又開始跟他的秘書埋怨因為他花太多了”

---

5. – Anaphoric ambiguity

左键：爷爷只给小海三块糖

右键：爷爷只给莉莉三块糖

爷爷给小海和莉莉买东西吃。他给买了很多糖，也买了很多火腿。小海很高兴。莉莉也很高兴。可是爷爷只给她三块糖。”

FILLER SENTENCES

6. — Pragmatic ambiguity

左键：王娜是个正中农村人

右键：王娜是个冒牌农村人

“王娜是个正儿八经的乡村大妞”

7. – Tonal ambiguity

左键：我们给他行李

右键：我们给他行礼

“徐校长还没有进屋时我们就已经给他行礼了”

8. – Tonal ambiguity

左键：我看见很多金鱼

右键：我看见很多雨水

“我在我朋友家的鱼缸里看见很多鱼”

FILLER SENTENCES

9. – Lexical ambiguity involving homonyms

左键：今天，这几位学生学了普通话的元音和辅音

右键：今天，这几位学生学了语言变化的原音

“薛老师今天在中文课教外国学生怎么发普通话的元音和辅音”

10. – Lexical ambiguity involving homonyms

左键：小丽荣看见很多星星

右键：小丽荣看见很多猩猩

“小丽蓉在动物园看见很多大猩猩”

11. – Lexical ambiguity involving homonyms

左键：秀英想吃枇杷

右键：秀英想学弹琵琶

“秀英报名了琵琶课程”

12. – Lexical ambiguity involving homonyms

左键： 潘经理不知道他老婆的目地

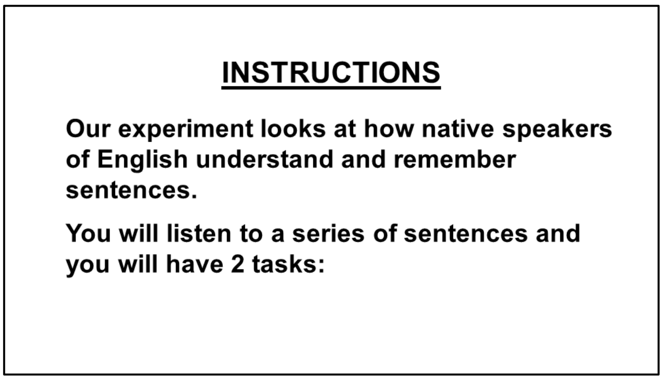
右键： 潘经理不知道他老婆的墓地在哪里

“经理在陵园找不到他老婆的墓地”

Appendix L

(Juncture Perception Experiment Instructions in English)

Slide 1

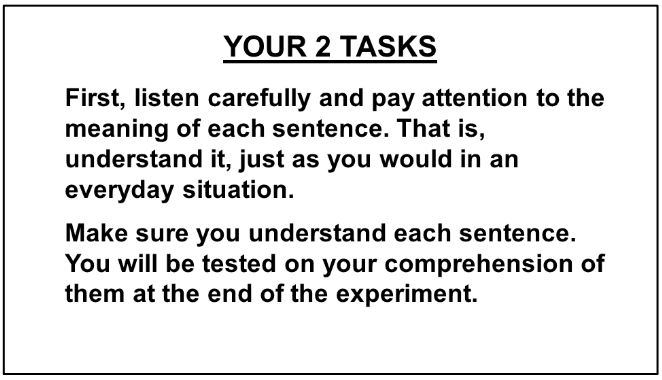


INSTRUCTIONS

Our experiment looks at how native speakers of English understand and remember sentences.

You will listen to a series of sentences and you will have 2 tasks:

Slide 2

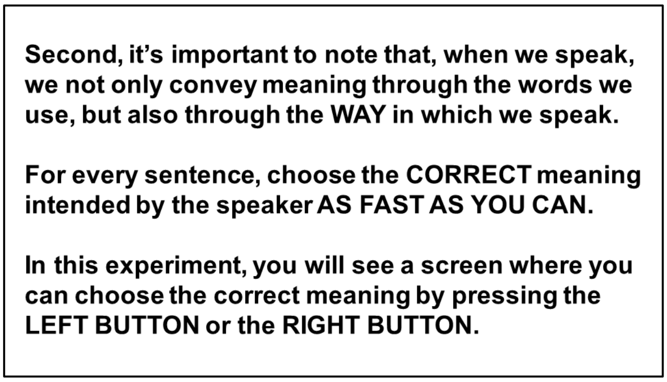


YOUR 2 TASKS

First, listen carefully and pay attention to the meaning of each sentence. That is, understand it, just as you would in an everyday situation.

Make sure you understand each sentence. You will be tested on your comprehension of them at the end of the experiment.

Slide 3

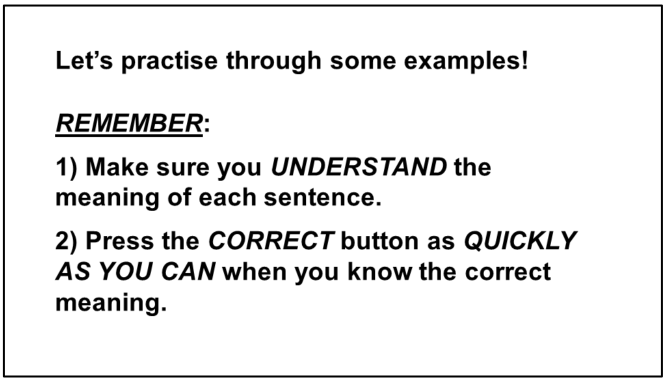


Second, it's important to note that, when we speak, we not only convey meaning through the words we use, but also through the WAY in which we speak.

For every sentence, choose the CORRECT meaning intended by the speaker AS FAST AS YOU CAN.

In this experiment, you will see a screen where you can choose the correct meaning by pressing the LEFT BUTTON or the RIGHT BUTTON.

Slide 4



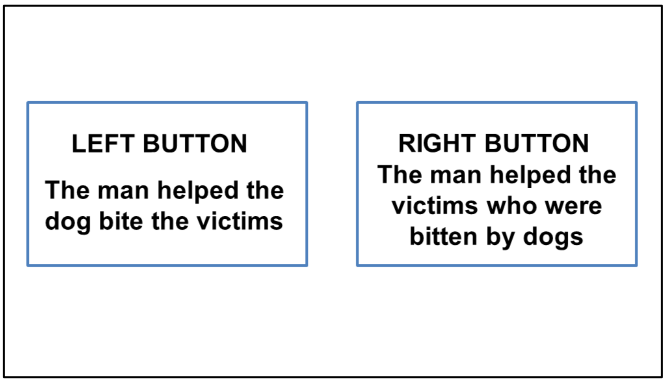
Let's practise through some examples!

***REMEMBER*:**

1) Make sure you *UNDERSTAND* the meaning of each sentence.

2) Press the *CORRECT* button as *QUICKLY AS YOU CAN* when you know the correct meaning.

Slide 5



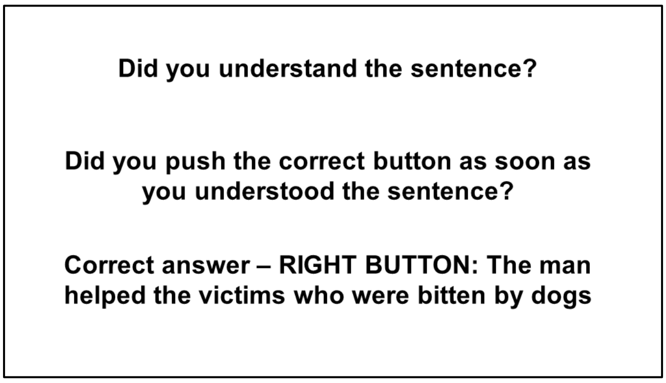
LEFT BUTTON

The man helped the dog bite the victims

RIGHT BUTTON

The man helped the victims who were bitten by dogs

Slide 6

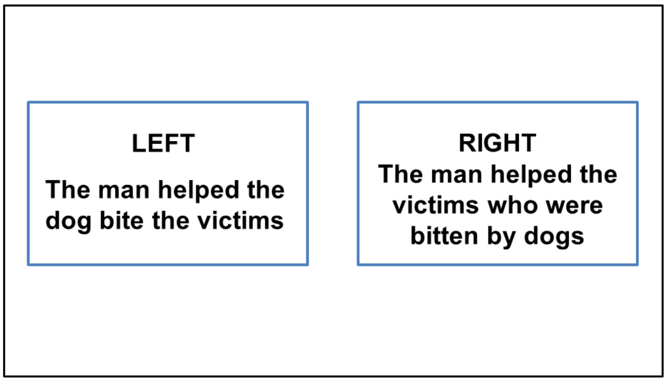


Did you understand the sentence?

Did you push the correct button as soon as you understood the sentence?

Correct answer – RIGHT BUTTON: The man helped the victims who were bitten by dogs

Slide 7



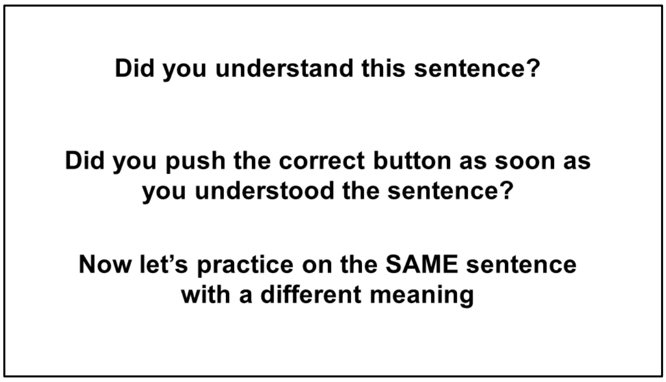
LEFT

The man helped the dog bite the victims

RIGHT

The man helped the victims who were bitten by dogs

Slide 8

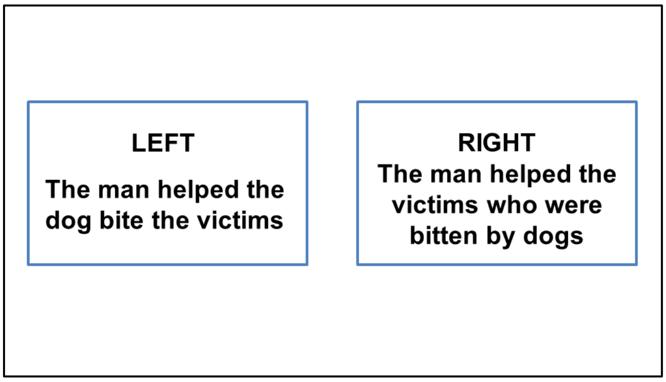


Did you understand this sentence?

Did you push the correct button as soon as you understood the sentence?

Now let's practice on the SAME sentence with a different meaning

Slide 9



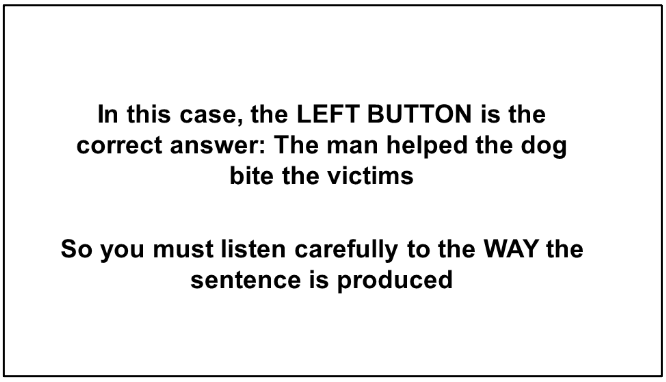
LEFT

The man helped the dog bite the victims

RIGHT

The man helped the victims who were bitten by dogs

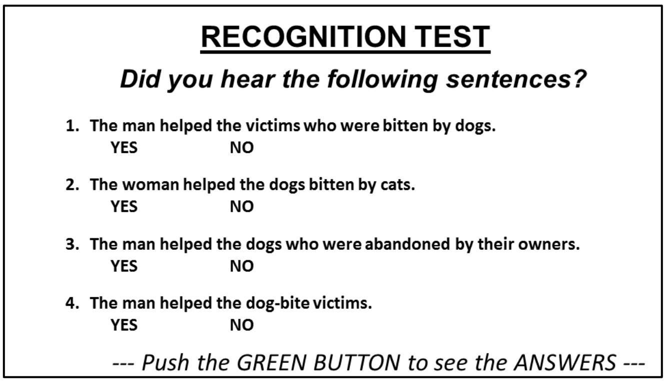
Slide 10



In this case, the LEFT BUTTON is the correct answer: The man helped the dog bite the victims

So you must listen carefully to the WAY the sentence is produced

Slide 11



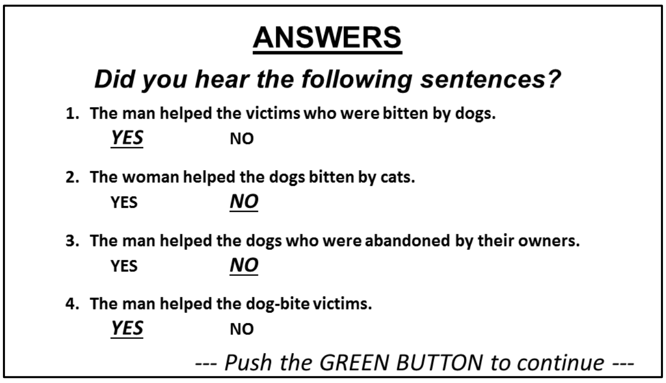
RECOGNITION TEST

Did you hear the following sentences?

1. The man helped the victims who were bitten by dogs.
   YES NO
2. The woman helped the dogs bitten by cats.
   YES NO
3. The man helped the dogs who were abandoned by their owners.
   YES NO
4. The man helped the dog-bite victims.
   YES NO

--- Push the GREEN BUTTON to see the ANSWERS ---

Slide 12



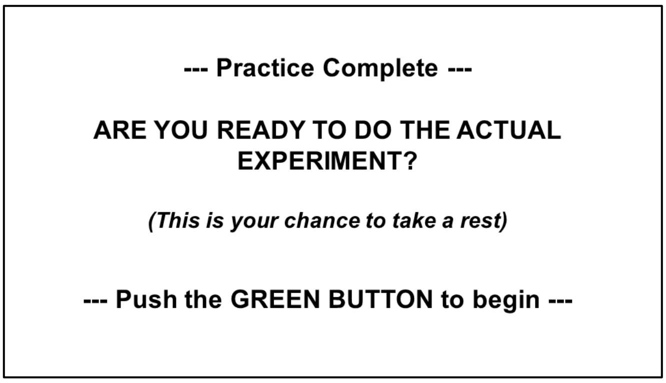
ANSWERS

Did you hear the following sentences?

1. The man helped the victims who were bitten by dogs.
   YES NO
2. The woman helped the dogs bitten by cats.
   YES ***NO***
3. The man helped the dogs who were abandoned by their owners.
   YES ***NO***
4. The man helped the dog-bite victims.
   YES NO

--- Push the GREEN BUTTON to continue ---

Slide 13



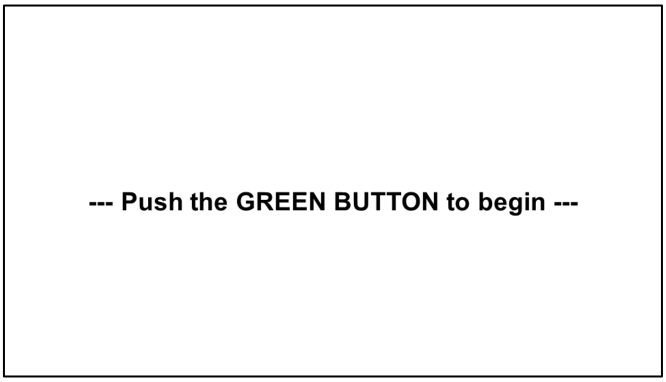
--- Practice Complete ---

ARE YOU READY TO DO THE ACTUAL EXPERIMENT?

(This is your chance to take a rest)

--- Push the GREEN BUTTON to begin ---

Slide 14

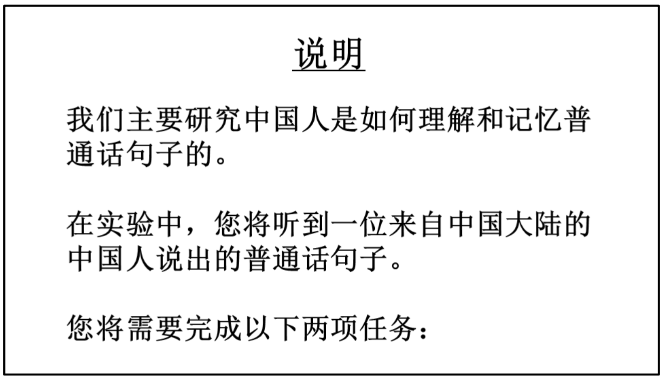


--- Push the GREEN BUTTON to begin ---

Appendix M

(Juncture Perception Experiment Instructions in Simplified Chinese)

Slide 1



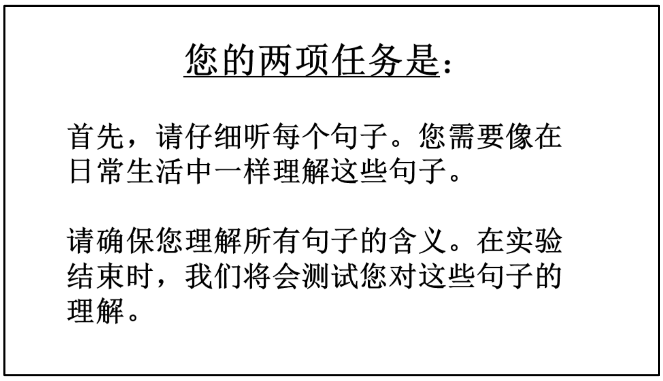
说明

我们主要研究中国人是如何理解和记忆普通话句子的。

在实验中，您将听到一位来自中国大陆的中国人说出的普通话句子。

您将需要完成以下两项任务：

Slide 2

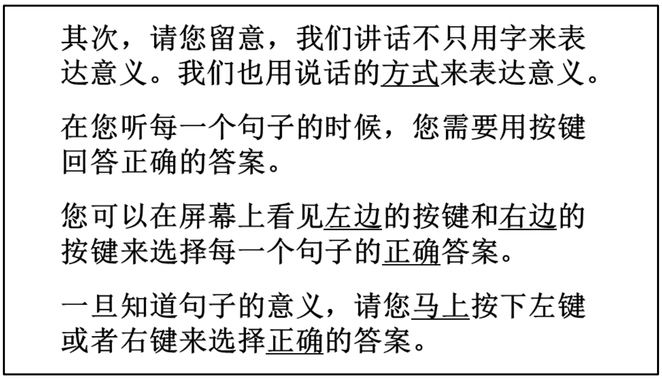


您的两项任务是：

首先，请仔细听每个句子。您需要像在日常生活中一样理解这些句子。

请确保您理解所有句子的含义。在实验结束时，我们将会测试您对这些句子的理解。

Slide 3



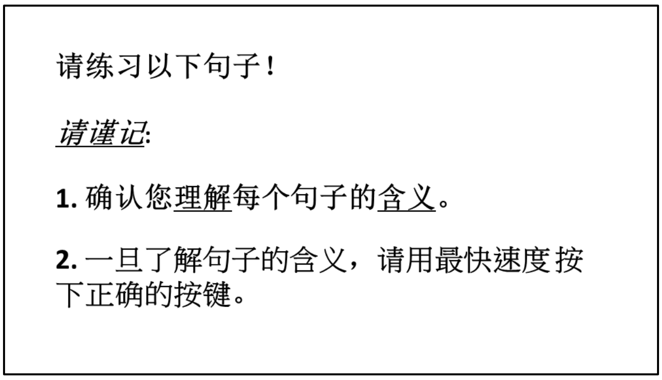
其次，请您留意，我们讲话不只用字来表达意义。我们也用说话的方式来表达意义。

在您听每一个句子的时候，您需要用按键回答正确的答案。

您可以在屏幕上看见左边的按键和右边的按键来选择每一个句子的正确答案。

一旦知道句子的意义，请您马上按下左键或者右键来选择正确的答案。

Slide 4



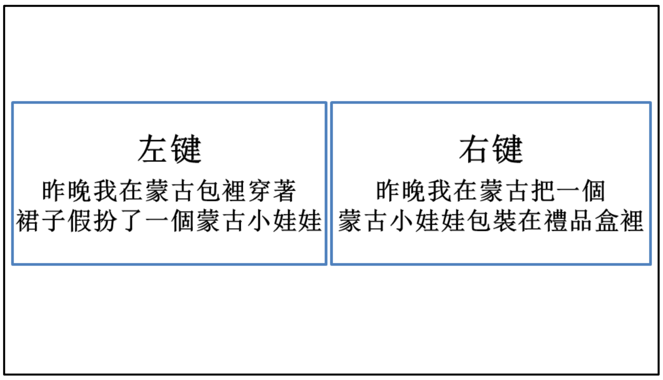
请练习以下句子！

请谨记:

1. 确认您理解每个句子的含义。

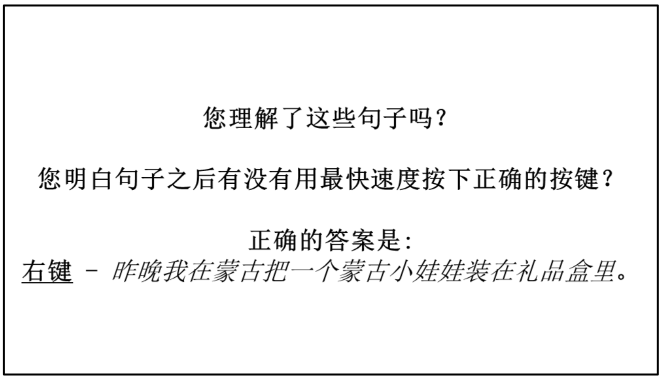
2. 一旦了解句子的含义，请用最快速度按下正确的按键。

Slide 5



左键 昨晚我在蒙古包裡穿著 裙子假扮了一個蒙古小娃娃	右键 昨晚我在蒙古把一個 蒙古小娃娃包裝在禮品盒裡

Slide 6



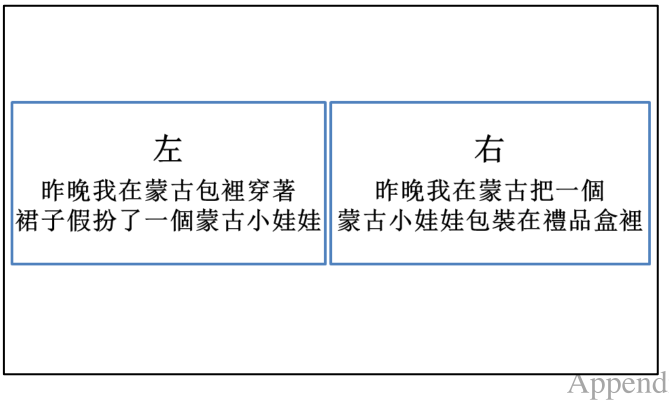
您理解了这些句子吗？

您明白句子之后有没有用最快速度按下正确的按键？

正确的答案是：

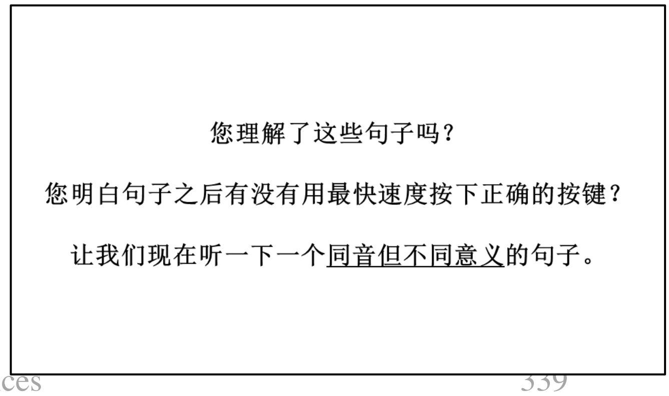
右键 – *昨晚我在蒙古把一个蒙古小娃娃装在礼品盒里*。

Slide 7



左 昨晚我在蒙古包裡穿著 裙子假扮了一個蒙古小娃娃	右 昨晚我在蒙古把一個 蒙古小娃娃包裝在禮品盒裡

Slide 8

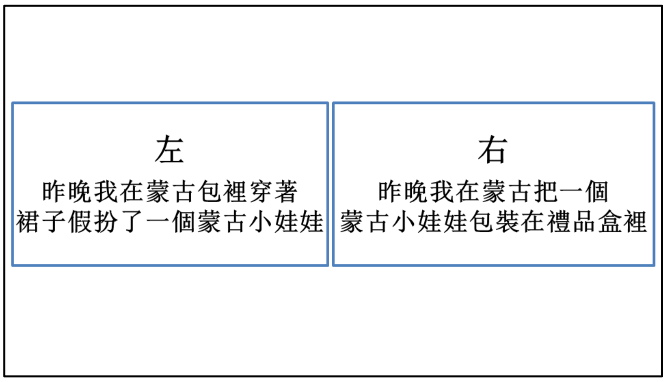


您理解了这些句子吗？

您明白句子之后有没有用最快速度按下正确的按键？

让我们现在听一下一个同音但不同意义的句子。

Slide 9



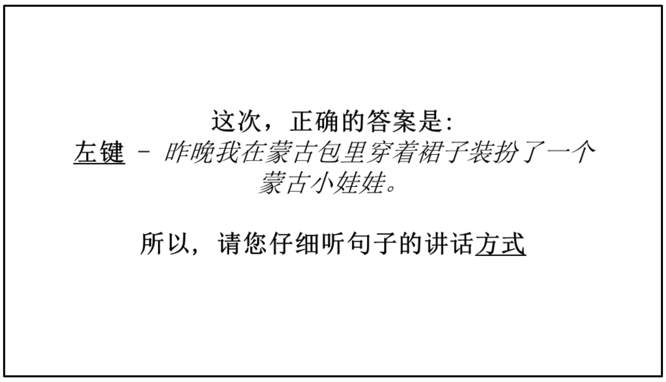
左

昨晚我在蒙古包裡穿著裙子假扮了一個蒙古小娃娃

右

昨晚我在蒙古把一個蒙古小娃娃包裝在禮品盒裡

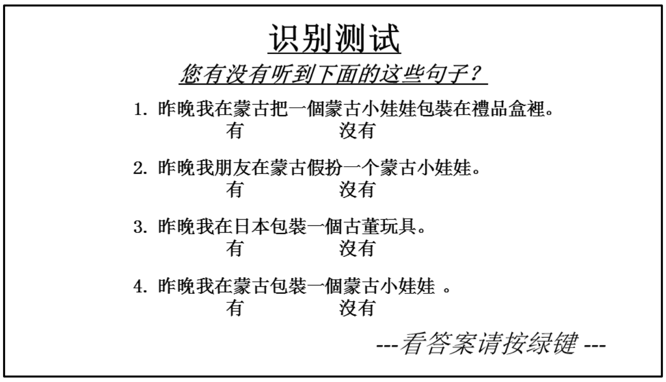
Slide 10



这次，正确的答案是：
左键 – *昨晚我在蒙古包里穿着裙子装扮了一个蒙古小娃娃。*

所以，请您仔细听句子的讲话方式

Slide 11



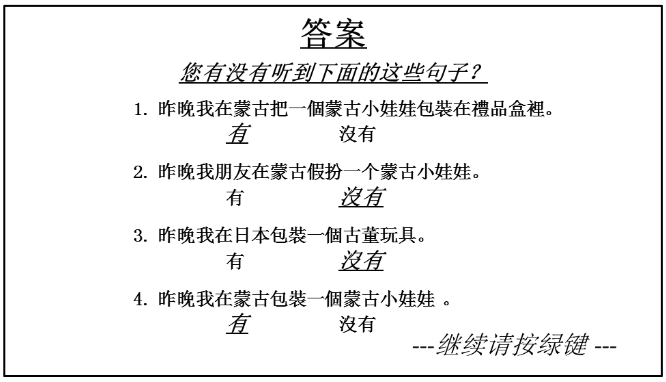
识别测试

您有没有听到下面的这些句子？

1. 昨晚我在蒙古把一個蒙古小娃娃包裝在禮品盒裡。
   有　　　沒有
2. 昨晚我朋友在蒙古假扮一个蒙古小娃娃。
   有　　　沒有
3. 昨晚我在日本包裝一個古董玩具。
   有　　　沒有
4. 昨晚我在蒙古包裝一個蒙古小娃娃 。
   有　　　沒有

---*看答案请按绿键* ---

Slide 12



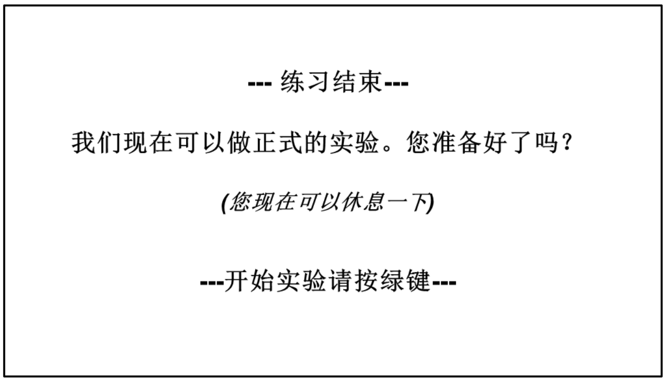
答案

您有没有听到下面的这些句子？

1. 昨晚我在蒙古把一個蒙古小娃娃包裝在禮品盒裡。
   有　　　沒有
2. 昨晚我朋友在蒙古假扮一个蒙古小娃娃。
   有　　　*沒有*
3. 昨晚我在日本包裝一個古董玩具。
   有　　　*沒有*
4. 昨晚我在蒙古包裝一個蒙古小娃娃 。
   有　　　沒有

---*继续请按绿键* ---

Slide 13



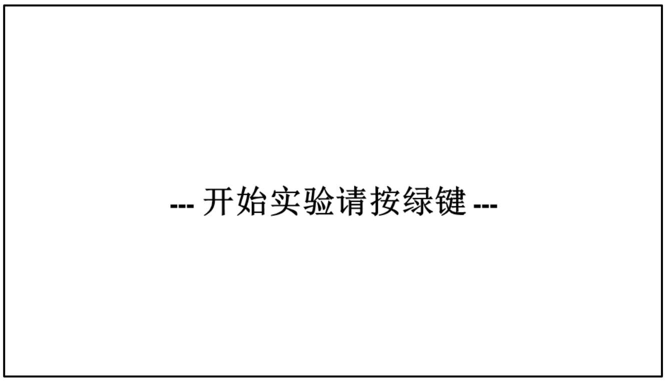
--- 练习结束---

我们现在可以做正式的实验。您准备好了吗？

(您现在可以休息一下)

---开始实验请按绿键---

Slide 14



--- 开始实验请按绿键 ---

Appendix N

(Recognition Test in English)

RECOGNITION TEST

Did you hear the following sentences? Please circle your response.

1. The oil spilled everywhere in the chair and child got really oily after he sat on it
   YES NO

2. Simon gave her dog biscuits.
   YES NO

3. David accidentally gave rat poison to Hanna
   YES NO

4. The boy gave her cat food
   YES NO

5. Gertrude told the Estonian travel stories
   YES NO

6. Gertrude told the English travel stories
   YES NO

7. The sorcerer cooked some vegetables for his guests for dinner
   YES NO

8. Most of the restaurant regular guests at this Greek restaurant are Sicilians
   YES NO

9. Lola went to the bank for a swim
   YES NO

10. Sasha wants to go to space
    YES NO

11. We saw lot of soldiers with machine guns in front of the city hall
    YES NO

12. Holly wondered whether the vicar had seen her monkey around the churchyard
    YES NO

13. The waiter had no idea that he served the salad to the French tourists
    YES NO

14. Joe said he cannot find a matching red sock

YES NO

15. Mr Jonsen saw Jessica's pet duck under the table

YES NO

16. Several of my friends say they prefer to spend their money on travelling

YES NO

17. Betty didn't buy anything at the shop

YES NO

18. None of the children want to feed her hungry fish.

YES NO

19. The host family gave away the Japanese-styled fans

YES NO

20. Sam gave her baby formula

YES NO

21. The maid did not get into trouble last night, even though she accidentally served insects to her lady.

YES NO

22. Every morning, the hotel chef serves the French toast

YES NO

Appendix O

(Recognition Test in Mandarin)

识别测试

您有没有听到下面的这些句子？请在答案上画圈

1. 我家大姨的邻居不喜欢种田
有　　　沒有

2. 大卫不小心喂她狗饼干吃
有　　　沒有

3. 刘波不小心把老鼠葯给珍妮
有　　　沒有

4. 王力强每天给她羊葱吃
有　　　沒有

5. 刘老师不小心给她羊肉汤喝
有　　　沒有

6. 刘老师不小心给她猪排骨汤喝
有　　　沒有

7. 李小姐只給小菊的爷爷冲奶粉
有　　　沒有

8. 很多韩国人在日本喜欢喝奶茶
有　　　沒有

9. 薛老师今天在中文课教外国学生怎么发普通话的元音和辅音
有　　　沒有

10. 王娜是个正宗农村人
有　　　沒有

11. 我们看见很多律师给法官行礼
有　　　沒有

12. 凌天生不小心给她象牙吃

有　　　沒有

13. 小丽荣说她在动物园从来没有看过猩猩

有　　　沒有

14. 大姨不知道她现在应该给不给她的宠物老鼠煮粟米汤

有　　　沒有

15. 吴老师看到她猫在桌子底下

有　　　沒有

16. 我朋友都说他们喜欢出去旅游，不喜欢去走街

有　　　沒有

17. 我很惊讶张少爷不想去外国读书

有　　　沒有

18. 很多人在中国和韩国喜欢吃泡菜

有　　　沒有

19. 这淘气的小男孩向周姥姥吐牛奶
有　　　沒有

20. 中国大使今天给韩国人参茶

有　　　沒有

21. 在医院的医生很累因为他们刚刚帮一位孕妇生了一对双胞胎女儿
有　　　沒有

22. 林姐给她狗肉吃

有　　　沒有