UNIVERSITY OF CALIFORNIA

Los Angeles

A Cross-Linguistic Comparison of

Lexical Stress Strength and Macro-Rhythm Strength

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Linguistics

by

Christine Prechtel

2023

© Copyright by

Christine Prechtel

2023

ABSTRACT OF THE DISSERTATION

A Cross-Linguistic Comparison of

Lexical Stress Strength and Macro-Rhythm Strength

by

Christine Prechtel

Doctor of Philosophy in Linguistics
University of California, Los Angeles, 2023
Professor Sun-Ah Jun, Chair

This dissertation investigates the prominence relationship between lexical stress and tonal rhythm across multiple languages and tests whether cross-linguistic differences in tonal rhythm are perceptible to listeners. Specifically, the experiments in this dissertation test Jun's (2014:537) hypothesized inverse correlation between the strength of lexical stress and the strength of tonal rhythm or macro-rhythm (MacR). Lexical stress strength is the difference in duration and intensity between stressed and unstressed syllables, and MacR strength is the presence, regularity, and frequency of word-sized F0 alternations within an utterance. This hypothesis was tested by comparing lexical stress strength with MacR strength in English, Uyghur, and Bengali. The predicted ranking of lexical stress strength is English > Uyghur > Bengali, while the predicted ranking of MacR strength is Bengali > Uyghur > English.

The lexical stress production experiment compared the duration ratios of stress and unstressed vowels of disyllabic nonce words produced by speakers of each language. The results

found that English had the strongest realization of lexical stress, but Uyghur had the weakest stress instead of Bengali (English > Bengali > Uyghur), possibly due to the use of nonce words and Bengali speakers' experience with English. However, the results of the MacR Frequency Index (Jun, 2014:538), which calculates peak-per-Prosodic Word ratio, confirm the predicted strength ranking (Bengali > Uyghur > English). In the MacR perception experiment, participants rated the melodicity of utterances that were phonetically manipulated in two conditions: low-pass filtered (Filtered) and hummed (F0-only) stimuli. The results found that Bengali utterances were rated significantly more melodic than Uyghur in both conditions, and more melodic than English in the Filtered condition, but Uyghur and English were not rated significantly different in either condition.

Overall, the results support the predicted inverse relationship between lexical stress strength and MacR strength, and they also demonstrate that listeners can perceive differences in MacR strength in the predicted direction. This study is the first of its kind to directly test the hypothesized inverse relationship and the perceived MacR strength across languages. The results contribute to our understanding of MacR, its effect on speech rhythm perception, and prosodic typology.

The dissertation of Christine Prechtel is approved.

Claire Moore-Cantwell

Patricia Keating

Sameer ud Dowla Khan

Sun-Ah Jun, Committee Chair

University of California, Los Angeles

2023

For my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I don’t think I can adequately express how grateful I am to the many people in my life who have helped me fulfill a dream. This was a long and grueling journey, and I simply could not have done it without the support and encouragement from my mentors, colleagues, research assistants, friends, and family over the last five and a half years.

First and foremost, I want to express my deep gratitude toward my advisor and committee chair, Sun-Ah Jun, for her years of mentorship, tireless patience, generosity with her time and knowledge, and above all else her faith in me. There were many times throughout my program when it was hard for me to see past my self-doubt and imposter syndrome, but her steadfast encouragement (and occasional gentle admonishment) helped me see my own potential and my capabilities. Thank you for believing in me and helping me make it over the finish line.

I’m grateful to my other committee members for their personal and professional support over the years. Patricia Keating has been a great teacher, mentor, and fount of phonetics knowledge. Thank you for always giving me valuable feedback on my work and for reminding me to not be so hard on myself in my presentations. Claire Moore-Cantwell was a steadfast cheerleader of my research goals long before they had coalesced into my dissertation, and I’m very grateful for her suggestions and insights into my many methodology and statistical analysis questions. Thank you for reaching out and giving me a pep talk when I was really struggling with burnout during the height of the pandemic. Sameer ud Dowla Khan has been an invaluable member of my committee from afar, especially in helping me design and translate my experiment into Bengali. Thank you for your unwavering enthusiasm and confidence in my progress.

Each aspect of the dissertation would not have been possible without the help of research assistants, language consultants, and, of course, the participants themselves. My wonderful undergraduate RAs Leslie Cheng and Rebecca Zhu saved me many weeks of very tedious work with their diligent data processing and annotation. You are the unsung heroes of this project! I'm very grateful to my friend Baisakhi Sengupta for connecting me with Bengali speakers in Kolkata for my production experiments and for introducing me to her father, Subh, who helped with the pilot experiment. She also introduced me to Madhupriya Sengupta, who helped me finalize the production experiments and recruit Bengali speakers. Her feedback was invaluable, and the production experiments are better for it. Special thanks to my colleague Jahnavi Narkar for connecting me with other people who shared my experiment advertisement. Regarding the Uyghur experiments, I'm so grateful for my wonderful Uyghur consultant, Gülnar Eziz, who generously provided multiple recordings, revised translations, and grammatical insights throughout the project. I could not have included Uyghur in my dissertation without her. I also want to thank Mahire Yakup for her insights into Uyghur lexical stress, for helping me connect with other Uyghur speakers, and for sharing my experiment advertisement. I'm grateful for other folks who helped me spread the word about my study, including Mutallip Anwar, Adam McCollum, and Elise Anderson. Most of all, I'd like to thank every participant, regardless of language, who completed my experiments.

I want to acknowledge additional colleagues and mentors whose advice, insights, and general friendliness have contributed to the positive aspects of my graduate experience. In particular, Jessica Rett has been an unwavering pillar of support from the very beginning. I've always appreciated her advice on topics ranging from university bureaucracy to careers outside of academia, and I'm forever grateful that she made semantic and pragmatic theory less

intimidating to me. I also want to thank Jesse Harris for his mentorship and support during my brief foray into pragmatics-intonation research. In an alternate universe, I might have pursued that area for a dissertation topic.

I'm especially grateful to the graduate students who came before me and provided mentorship, camaraderie, and friendship; without you, the first few years of my graduate experience would have been miserable. Special shout-outs to Connor Mayer and Travis Major for encouraging my interest in Uyghur language prosody and for introducing me to Gülnar and Mahire; Eleanor Glewwe and Meng Yang for reminding me to have hobbies outside of work and for keeping the Georgian Choir going; Marju Kaps and Allie Lawn for taking me under their wing when I was dipping my toe into the Psycholinguistics Lab; Brice Roberts for being the best TA I could have asked for when I had to teach a summer phonetics course over Zoom during a global pandemic; Adam Royer for being a delightful human and for mentoring me during times when I felt completely lost in my program; and Jeremy Steffman for always being an excellent work buddy and for keeping me on task with our many pomodoro sessions. I also want to thank Neda Vesselinova, whose friendship and support were invaluable during my first year. I still consider you part of the OG cohort. Speaking of the cohort, I'm so grateful to Hironori Katsuda, Z.L. Zhou, Maddy Booth, Kiki Liu, Andy Xu, and Canaan Breiss for helping me get through this program, both academically and emotionally. We've been through some tough times together, and I'm glad you all stuck around. Also, a special shout-out to John Clayton, one of my dearest friends in L.A., who I consider part of the larger Linguistics/Indo-European graduate cohort.

I could not have finished this dissertation without the support of friends and family. Julie Botnick deserves a special mention here for being an awesome roommate and friend during my first two years of the program and for dragging me along to various hikes and social activities

when I would have otherwise stayed in my room like a recluse. Thanks to the Lane family for being cheerleaders over the past few years, even when I did a poor job of explaining my dissertation over the phone. I'm extremely grateful to my parents and my brother for their steadfast love and support throughout the program. I can't adequately express how much I've cherished your encouragement and prayers along the way. A very special thanks to my oma, whose unwavering faith in my ability to succeed has meant the world to me. I'm glad that you get to celebrate this accomplishment with me. And finally, I want to extend my deep and unending gratitude to my wonderful partner, Emily. The last five and a half years have been long and hard. Thanks for always being there for me.

VITA

2020	M.A., Linguistics
	University of California, Los Angeles
2015	B.A., Linguistics and Spanish
	The Ohio State University

PUBLICATIONS

Prechtel, C. (2022). Lexical Stress Strength vs Macro-Rhythm Strength: An Inverse Relationship Between Prominence Cues. *Proceedings of the 1st International Conference on Tone and Intonation*, 102-106.

Prechtel, C. (2020). Macro-rhythm in English and Spanish: Evidence from Radio Newscaster Speech. *Proceedings of the 10th International Conference on Speech Prosody*, Tokyo, 675-679.

Prechtel, C. (2019). Quantifying Macro-rhythm in English and Spanish. *Proceedings of the 19th International Congress of Phonetic Sciences*, 2896-2900.

Prechtel, C. & Clopper, C. G. (2016). Uptalk in Midwestern American English. *Proceedings of the 8th International Conference on Speech Prosody*, 133-137.

CHAPTER 1

Introduction and Background

1.1. Introduction

Although speech rhythm seems like a relatively straightforward linguistic phenomenon to describe and measure, researchers have found time and again that its exact nature is complex and difficult to capture. Decades of research have attempted to determine the mechanisms and correlates of rhythm in speech, but evidence for definitive acoustic and perceptual correlates has been varied and controversial, with some claiming that speech rhythm can be classified by strictly temporal or isochronous relationships between syllables, feet, or moras (e.g., Abercrombie, 1967; and many others) while others question whether speech is inherently rhythmic at all (Nolan & Jeon, 2014). Nevertheless, listeners seem to hear and have intuitions about differences in rhythmic patterns across languages (e.g., Lloyd James, 1940; Nazzi, Bertoncini, & Mehler, 1998; Ramus & Mehler, 1999; Ramus, Dupoux, & Mehler, 2003; White, Mattys, & Wiget, 2012; Vicenik & Sundara, 2013).

A growing body of research suggests that some of these perceived rhythmic differences are prosodic in nature. More specifically, evidence points to the role of prosodic prominence at both lexical and post-lexical levels in speech rhythm perception (e.g., Niebuhr, 2009; Barry, Andreeva, & Koreman, 2009; Cumming, 2011a, 2011b), and its role in dividing up the continuous speech signal into chunks for word segmentation (e.g., Cutler, Mehler, Norris, & Segui, 1986; Cutler, 1991). Studies have also found that native language affects which prominence cues (duration, intensity, and/or F0) listeners pay attention to in word segmentation (e.g., Bhatara et al., 2013; Ordin & Nespor, 2013; Ordin & Nespor, 2016; Molnar, Carreiras, & Gervain, 2016), reflecting differences in prosodic organization and

prominence realization across languages. Languages that mark lexical stress use duration and intensity as prominence cues (in addition to F0) to find word boundaries. However, in languages that do not mark lexical stress, and thus do not use duration or intensity to mark word boundaries, F0 alternations over a word-sized interval facilitate word segmentation instead (e.g., Kim, 2004; Kim & Cho, 2009; Warner, Otake, & Arai, 2010; Welby, 2007). According to Jun (2014), these patterns of word-sized tonal rhythm, or macro-rhythm, cue word prominence by highlighting word boundaries, similarly to how duration, intensity, and pitch accents highlight word boundaries in languages with lexical stress. Even among lexical stress languages, the strength of duration and intensity and the consistency of tonal alternations vary. Therefore, the language-specific interplay between lexical stress and tonal alternations seems to create differences in speech rhythm perception across languages.

The goal of this dissertation is to investigate the prominence relationship between lexical stress cues and tonal rhythm cues in languages that mark lexical stress. Specifically, the experiments reported in this dissertation test the hypothesis proposed in Jun (2014:537) that there is an inverse correlation between the strength of lexical stress and the strength of tonal rhythm (macro-rhythm). Lexical stress strength is defined as the size of the duration (and intensity) differences between stressed and unstressed syllables, and tonal rhythm strength is defined as the presence, regularity, and frequency of F0 alternations within an Intonational Phrase (IP). This relationship will be tested in both production and perception experiments.

The rest of this chapter is structured as follows: Section 2 introduces the relevant background literature on speech rhythm (2.1), lexical and phrasal prominence (2.2), cross-linguistic comparisons of prominence strength (2.3), prosodic typology and macro-rhythm

(2.4), and macro-rhythm quantification measures (2.5); while Section 3 provides the primary goals of the dissertation (3.1) and introduces the languages for comparison and their predicted lexical stress and tonal rhythm strength rankings (3.2).

1.2. Background

1.2.1. Speech Rhythm Production and Perception

Much of the early research on speech rhythm has focused on finding acoustic correlates in the speech signal. One of the most influential ideas about speech rhythm is the isochrony hypothesis, or speech rhythm classification (Pike, 1945; Abercrombie, 1967), which posits that speech rhythm is based on the duration intervals of linguistic units such as syllables and feet, and that languages can be classified in terms of how timing is coordinated between these units. Stress-timed languages are said to have patterns of equal duration between stressed or prominent syllables, while syllable-timed languages have equal duration between each syllable. The classification was later updated to include mora-timed languages such as Japanese (e.g., Bloch, 1950; Hockett, 1955; Han, 1962; Ladefoged, 1975).

Although this classification system has been controversial since its conception, and multiple studies have found that rhythm classification is dependent on syllable structure differences both within languages (e.g., Boudreault, 1970; Allen & Hawkins, 1978) and across languages (e.g., Dauer, 1983), decades of speech rhythm research have assumed some version of this hypothesis as the basis for cross-linguistic differences, and numerous measures have been proposed to quantify these differences. For example, Roach (1982) hypothesized that stress-timed languages would have more variability in consonant and vowel intervals than syllable-timed languages because the former allow complex consonant clusters and vowel reduction. This led to a proliferation of rhythm metrics that measured

consonantal and vocalic durations and their variability within an utterance. Ramus, Nespor, and Mehler (1999) introduced measures such as the standard deviation of consonant intervals (ΔC), the standard deviation of vocalic intervals (ΔV), the percentage of consonant intervals (%C), and the percentage of vocalic intervals (%V). They found that syllable-timed languages had a larger percentage of vocalic intervals and smaller variance of consonant intervals than stress-timed languages. Grabe and Low (2002) introduced the Raw and Normalized Pairwise Variability Index measures (rPVI and nPVI respectively), which captured the variability between pairs of consonant or vocalic intervals at both raw values and normalized for speech rate. Other measures that took speech rate into account included the variation coefficient or Varco measures, such as VarcoΔC for consonant intervals (Dellwo, 2006) and VarcoΔV for vowel intervals (White & Mattys, 2007). Perception studies have found that infants can discriminate between languages based on isochrony classification (e.g., Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Jusczyk, & Johnson, 2000; Nazzi & Ramus, 2003), and adult listeners can process speech in moras, syllables, or feet depending on the rhythm type of their first language (L1) (Cutler, Mehler, Norris, & Seguí, 1986, 1992; Otake, Hatano, Cutler, & Mehler, 1993; Cutler & Otake, 1994; Murty, Otake, & Cutler, 2007). However, although the isochrony hypothesis continues to be influential in speech rhythm literature, it also remains controversial. Other studies have failed to support the predictions across languages (e.g., Bolinger, 1968; Lehiste, 1977; Arvaniti, 2009, 2012), finding contradictory results due to differing methodologies (Arvaniti, 2012), properties of the stimuli (Wiget et al., 2010; Arvaniti, 2012; Prieto et al., 2012), and inter-speaker variability (Wiget et al., 2010; Arvaniti, 2012). Indeed, there is evidence that both syllables and stress units are perceived as regular despite being produced in irregular sequences (O'Connor,

1965; Lehiste, 1977; Donovan & Darwin, 1979; Dauer, 1983; Benguerel & D'Arcy, 1986), and stress units themselves are not produced regularly (O'Connor, 1965). This led to the conclusion by Benguerel & D'Arcy (1986), Beckman (1992), and others that speech rhythm is a perceptual phenomenon rather than an acoustic one. Therefore, speech rhythm cannot be captured by isochrony-based rhythm measures alone.

However, even though speech rhythm is now generally understood as perceptual, researchers continue to puzzle over which cues listeners extract from the speech signal to inform their intuitions about rhythm. If speech is not inherently rhythmic as some have claimed (e.g., Nolan & Jeon, 2014), but rather something listeners impose on the continuous speech signal (e.g., Lehiste, 1977; McAuley, 2010; Motz, Erickson, & Hetrick, 2013) to segment it into semi-regular chunks or intervals (Ding et al., 2016; Batterink & Paller, 2017), then what acoustic cues do listeners use for segmentation? Since speech rhythm appears to play a role in both language acquisition (e.g., Cutler et al., 1992) and in word segmentation through patterns of lexical stress cues (e.g., Cutler, Mehler, Norris, & Segui, 1986; Cutler, 1991) and tonal melody over word-sized intervals (e.g., Kim, 2004; Welby, 2007; Kim & Cho, 2009; Warner, Otake, & Arai, 2010; Morrill, Dilley, MacAuley, & Pitt, 2014), there must be patterns within the signal that help listeners chunk word and phrase-sized intervals. Therefore, there should be some acoustic cue or set of cues to measure and quantify.

There is a growing body of evidence that speech rhythm is based on the regularity of acoustic and perceptual prominence in the speech signal. Studies have looked at the syllable "beat" (i.e., stressed syllable) (Allen, 1972a, 1972b, 1975) and its perceptual prominence (Morton, Marcus, & Frankish, 1976; Pompino-Marschall, 1989) and found that speech rhythm is influenced by the onset of the amplitude envelope in the speech signal (Howell,

1988; Goswami et al., 2002). Tilsen and Johnson (2008) found that the periodicity (i.e., rhythmicity) of the amplitude envelope in English was varied, with some utterances exhibiting stress-timed rhythm, some exhibiting syllable-timed rhythm, and some exhibiting phrasal-level rhythm, such as between pitch accents (2008:34). Tilsen and Arvaniti (2013) found that phrasal prominence played an important role in rhythmicity in a cross-linguistic comparison. In their study, different languages exhibited periodicity of the amplitude envelope at different timescales, with English having lower frequency periodicity (corresponding to supra-syllabic or "stress-timed" periods) than other languages traditionally classified as "syllable-timed" (i.e., Greek, Italian, and Spanish) and languages with no lexical stress such as Korean.

To understand perceptual prominence and its role in speech rhythm, we must first understand what the corresponding acoustic correlates of prominence are and how they differ across languages. Indeed, a crucial aspect of studying speech rhythm is examining how lexical stress (word-level prominence) and phrasal accent (post-lexical prominence) interact to inform the perception of rhythm within and across languages.

1.2.2. Lexical and Post-Lexical Prominence

Early studies on lexical stress cues identified multiple acoustic correlates such as duration, intensity, and F0. For example, Fry (1955) tested duration, intensity, and F0 as cues for lexical stress in English and concluded that F0 was the most important cue, followed by duration, and then intensity. Fry (1958) conducted additional stress perception experiments and found that duration and intensity were both salient stress correlates in English, and F0 movement affected the perception of stress and was able to override the duration cue. In general, F0 is frequently treated as an acoustic correlate of lexical stress (see Gordon &

Roettger, 2017, for an overview and references therein), because the stressed syllable often carries a phrase-level prominence, which is marked by F0. However, many of the previous studies on lexical stress that have included F0 as a lexical stress correlate conflate post-lexical (phrasal) prominence with lexical-level prominence. Even in cases where prosodic context, information structure, and location within a phrase are carefully controlled for, the target stimuli may not necessarily be free of the phrasal effects of F0 (Roettger & Gordon, 2017; Vogel, Athanasopoulou, & Pinkus, 2016:134). Studies have demonstrated that F0 can be decoupled from word stress; for example, Huss (1978) found that duration and intensity cues were still present to distinguish noun-verb minimal pairs in English in post-nuclear position, where F0 cues are absent. According to the Autosegmental-Metrical (AM) model of intonational phonology, F0 is not a stress correlate, but rather a phrasal cue that aligns with the head of the word (e.g., the stressed syllable) (Pierrehumbert, 1980; Beckman, 1996; Shattuck-Hufnagel & Turk, 1996; Ladd, 1996/2008). In this model, intonational tunes are composed of prominent pitch targets or movements that mark the head of the word (pitch accents) and pitch targets or movements that mark the edge of a prosodic unit (boundary tones).

Since F0 is a post-lexical prominence cue in non-tonal languages, duration and intensity are two of the most common acoustic cues of word-level stress. Gordon and Roettger (2017) conducted a cross-linguistic survey of acoustic correlates of word stress, which analyzed 110 studies and sub-studies across 72 languages. They found that of all the correlates that were studied (i.e., duration, intensity, VOT, F0, formants, and spectral tilt), duration was both the most frequently measured correlate (included in 100 studies) and the most successful at distinguishing stress in 85 of 100 (sub)studies in 65 out of 72 languages.

They also found that intensity successfully distinguished stress in 49 of the 70 studies, although this included various types of intensity measures (mean, peak, and midpoint). Many of the studies in the database also looked at F0 as a correlate, but as mentioned earlier, this conflates post-lexical prominence with lexical stress. Therefore, when discussing the relative "strength" of lexical stress, this refers to the magnitude or degree of duration and intensity in stressed syllables compared to unstressed syllables. Strong lexical stress is defined as high amplitude and long vowel duration (Jun, 2014:537) compared to unstressed syllables, while weak lexical stress is defined by smaller differences in amplitude and duration between stressed and unstressed syllables.

As previous lexical stress studies have demonstrated, F0 is a perceptually salient prominence cue, even if its domain is post-lexical rather than lexical. Studies on prominence cue perception such as Kohler (2008) found that F0 (i.e., pitch accent) was a stronger prominence cue than syllable duration and overall acoustic energy in German. In addition to marking prominence on the stressed syllable, F0 also marks boundaries of prosodic units. The size and type of prosodic units vary by language, as well as the combinations of pitch accents and boundary tones that demarcate these units. The repetition of language-specific tonal patterns (i.e., F0 movement) can facilitate the perception of speech rhythm. Niebuhr (2009) found that patterns of F0 movement in German cued the perception of speech rhythm. Barry, Andreeva, and Koreman (2009) investigated the perceptual weight of duration, F0, intensity, and vowel quality measures in speech rhythm perception in Bulgarian, German, and English, and they found an interaction of the measures in perceived rhythmicity; F0 changes within the foot were a strong secondary cue to duration for English and German listeners and an equally important cue for Bulgarian listeners. Cumming (2011a, 2011b)

investigated the interdependence of duration and F0 cues in rhythm perception in Swiss German, Swiss French, and Metropolitan French and found that the relative weighting of each cue was influenced by the listener's native language; in Swiss German, duration was more highly weighted than F0, while the cues were weighted more equally in both French varieties.

Studies have also found that the alternation between tonal targets (from low to high or high to low) facilitates prosodic grouping when the tonal patterns align with language-specific prosodic patterns (e.g., Dilley & Shattuck-Hufnagel, 1999; Dilley & McAuley, 2008). These tonal alternations have also been found to facilitate word segmentation in languages such as Korean (Kim, 2004; Kim & Cho, 2009), French (Welby, 2007), German (Niebuhr, 2009), and Japanese (Warner, Otake, & Arai, 2010). In Korean, for example, the H to L alternation is a strong cue for prosodic word-sized boundaries because the prosodic unit of Accentual Phrase, which is slightly larger than a word, ends with a H target and begins with a L target, so only the H-to-L alternation across two syllables facilitated word segmentation (Kim, 2004; Kim & Cho, 2009). In addition to real languages, studies have found that tonal alternations also facilitate word segmentation in artificial languages (e.g., Shukla, Nespor & Mehler, 2007; Tyler & Cutler, 2009), and the listeners' native language affects which prominence cues (duration, intensity, and/or F0) listeners attune to in word segmentation (e.g., Bhatara et al., 2013; Ordin & Nespor, 2016; Molnar, Carreiras, & Gervain, 2016).

1.2.3. Comparing Prominence Strength Across Languages

It has long been observed that the strength of lexical stress correlates differs across languages (see Gordon & Roettger, 2017, for an overview). However, despite the proliferation of

studies examining acoustic correlates of prominence within languages, there are surprisingly few studies comparing correlates across languages. Delattre (1966) studied the effect of syllable type (open vs closed syllable), syllable position (final vs non-final) and stress on syllable duration and intensity differences in English, German, Spanish, and French. Comparing the three languages with variable stress (English, German, Spanish), the results found that English had both the largest average duration difference and intensity difference between stressed and unstressed syllables regardless of syllable type and position, followed by German, followed by Spanish. While French had larger duration differences than Spanish when comparing word-final stressed syllables with non-final unstressed syllables because French has fixed word-final prominence, Delattre found that unstressed syllables had slightly higher intensity than stressed syllables (0.5 dB), thus concluding that intensity is not a stress correlate in French.

Botinis et al. (2002) investigated the effects of syllable position, stress, focus, and tempo (speech rate) on segmental (consonant and vowel) durations in American English, British English, Greek, and Swedish using nonsense disyllabic CVCV words contained within a carrier sentence under different focus conditions and speech rates. However, while they found significant within-language differences in vowel durations for stressed vs unstressed syllables, they did not do a cross-linguistic comparison, in part because they only measured 6 productions of one speaker per language.

Andreeva, Barry, & Koreman (2014) compared the phonetic realization of prominence cues (F0, duration, intensity, and vowel spectrum) in five languages (Bulgarian, Russian, French, German and Norwegian) in both pitch accented and de-accented conditions. They found systematic cross-linguistic differences in the degree to which the acoustic

measures changed between the two phrasal accent conditions. However, the authors explicitly stated that they were not investigating word stress or accent.

Mairano, Santiago, and Romano (2015) appears to be the only study that directly and explicitly compares lexical stress realization across languages. They specifically compared vowel durations between accented (perceptually prominent) and unaccented syllables in five languages (English, German, Spanish, French, Italian) to test whether durational differences were greater in Germanic languages (classified as stress-timed) than Romance languages (classified as syllable-timed). The results confirmed their predictions, with English having the most extreme duration differences and Spanish and French having the smallest differences. It should be noted that the data were collected from participants reading the same short text translated into their native language, and thus did not control for variables such as vowel identity, speech rate, and syllable structure.

Given the findings of previous literature, prosodic prominence and phrasing clearly play an important role in speech rhythm perception, both at the lexical and phrasal levels. The current study will focus on the interplay between lexical stress and the regularity of F0 movement at the phrasal level.

1.2.4. Prosodic Typology and Macro-rhythm

Based on the AM model of intonational phonology of typologically diverse languages, Jun (2005, 2014) proposed a model of prosodic typology that captures cross-linguistic similarities and differences in prosodic structure. According to the prosodic typology, languages can be classified in terms of prominence and phrasing at the lexical and post-lexical levels (Jun, 2005, 2014). At the lexical or word level, prosodic units include moras, syllables and feet, and prominence can be marked by one or a combination of stress, lexical pitch accent, and

tone, or not marked at all. At the post-lexical/phrasal level, prosodic units include the Accentual Phrase (AP), Intermediate Phrase (ip), and Intonational Phrase (IP), and prominence can be marked by the head of the phrase such as a pitch accent (Head), by a boundary tone at the phrase edge (Edge), or by both (Head/Edge). Languages can be Head-prominent like English, Edge-prominent like Seoul Korean, or both like Bengali (2005, 2014). The combination of prominence and phrasing at multiple levels of the prosodic hierarchy determines the organization of F0 movement within an utterance, as well as how prominence is marked.

Even when languages have similar prominence and phrasing characteristics, they can still prosodically differ in other ways. For example, English and Greek are both Head-prominent languages that mark lexical stress, but Greek has more regular F0 alternations within an utterance than English. Since prominence and phrasing alone could not capture these differences in tonal rhythm, an additional parameter was added to the prosodic typology proposed in Jun (2014). This parameter, macro-rhythm (MacR), is defined as phrase-medial tonal rhythm (i.e., the regularity of high/low F0 alternations). The unit of tonal rhythm is equal to or slightly greater than a Prosodic Word (PWord) or Accentual Phrase (AP), which is a content word plus surrounding unaccented function words and/or clitics (2014:522). MacR strength can differ across languages under the following three rules: the presence of alternating L and H tones (Figure 1), the uniformity of the rise-fall slope shape (Figure 2), and the frequency of the L/H intervals (Figure 3).

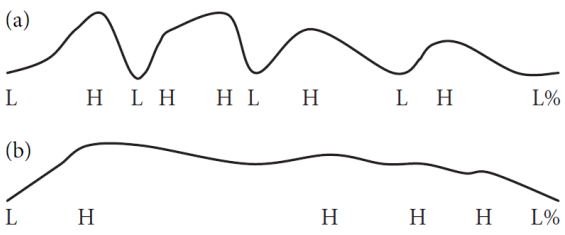

Figure 1: Schematic pitch contours that differ in the presence of L/H alternations (taken from (2) in Jun 2014:525). The number of H and L alternations in contour (a) is greater than contour (b), thus showing stronger macro-rhythm.

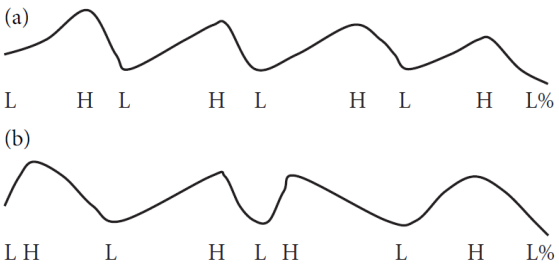

Figure 2: Schematic pitch contours that differ in the similarity or uniformity of the rise-fall slope shape (taken from (3) in Jun, 2014:525). The rise-fall units in contour (a) are more regularly shaped than contour (b), thus showing stronger macro-rhythm.

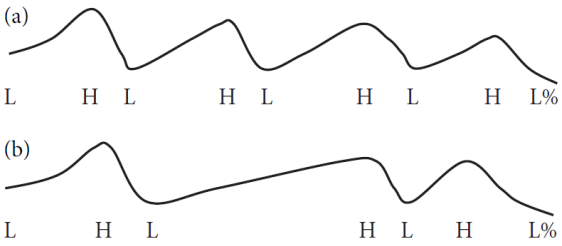

Figure 3: Schematic pitch contours that differ in the regularity of L/H alternation intervals (taken from (4) in Jun, 2014:525). The distances between peaks and valleys in contour (a) is more regular than contour (b), thus showing stronger macro-rhythm.

These three rules correspond to phonological criteria: the most common type of phrase-medial tone in a language's tonal inventory (Figure 1), the number of phrase-medial tones in the tonal inventory (Figure 2), and the frequency of f0 rise per word in a phrase (Figure 3) (Jun, 2014:526). Languages in which the most common phrase-medial tone is

rising (e.g., L+H*) or falling (e.g., H*+L) will have stronger MacR than languages whose most common tone is level (e.g., H*, L*), which corresponds to the first rule. Languages with fewer types of phrase-medial tones will have less variable f0 slope shapes and therefore stronger MacR than languages with more types of tones, corresponding to the second rule. Languages in which every word is marked by a tone will have stronger MacR than languages with less or more frequent tone marking per word, corresponding to the third rule. The model can therefore predict the relative strength of MacR in any language based on its prosodic structure and tonal inventory.

There is a small but growing body of research comparing MacR strength across languages. Burdin et al. (2014) examined the realization of prosodic focus in four typologically unrelated languages that mark lexical stress: American English, Paraguayan Guaraní, Moroccan Arabic, and K'iche'. While they did not quantify MacR strength, they argued that differences in MacR strength could account for the differences in each language's phonetic realization of prominence in focus-marking. American English and Paraguayan Guaraní are both Head-prominence languages, but they differ in their phonological organization. Since Paraguayan Guaraní has a smaller inventory of pitch accents than American English and the most common pitch accent is rising (compared to H* in English (Dainora, 2001, 2006)), the former language was predicted to have stronger MacR than the latter (Paraguayan Guaraní > English). In contrast, Moroccan Arabic and K'iche' are both classified as Head/Edge-prominence languages. K'iche' has a small tonal inventory and marks nearly every content word (i.e., AP) with a rising phrasal accent. In contrast, Moroccan Arabic has a larger tonal inventory and does not have an accentual phrase (AP), although it does tonally mark intermediate phrase (ip) boundaries instead. These ips vary in

length, and the boundary edges are often marked with plateau contours rather than rising or falling contours. Therefore, K'iche' was predicted to have stronger MacR than Moroccan Arabic (K'iche' > Moroccan Arabic). The study measured the following prosodic cues of focus: deaccenting, pitch accent type, phrasing (i.e., phrase breaks), and duration (i.e., word length). It should be noted that the first two prosodic cues were not applicable to the Head/Edge-prominence languages. They found that English used all four cues in focus-marking, Paraguayan Guaraní used deaccenting and duration, Moroccan Arabic used phrasing and duration, and K'iche' used none of these cues. The authors concluded that the variability in focus-marking is partly due to prominence type and partly due to MacR strength. That is, languages with weaker MacR (English and Moroccan Arabic) were more likely to use focus-marking strategies that disrupted regular F0 alternations than languages with stronger MacR (Paraguayan Guaraní and K'iche') (2014:275).

Polyanskaya, Busà, and Ordin (2019) quantified and compared MacR strength between Italian and English, which are both Head-prominence languages that mark lexical stress. Although they have similar tonal inventory sizes, one of the most common pitch accents in Italian is a rising accent (L+H*) (Jun, 2014:528), whereas the most common pitch accent in English is H*. In addition, English tends to deaccent certain types of content words (Schmerling, 1976; Ladd, 1996/2008) and given information (Katz & Selkirk, 2011), whereas Italian accents content words whether new or given (D'Imperio, 2001; Avesani & Vayra, 2005). Therefore, Italian was predicted to have stronger MacR than English (Jun, 2014), and the results of the study confirmed the predicted strength ranking. Similarly, Prechtel (2020) quantified and compared the MacR strength of Spanish and English. Like Italian, the most common pitch accent in Spanish is rising (L+ <H*) (Aguilar, de-la-Mota, &

Prieto, 2009; de-la-Mota, Butragueño, & Prieto, 2010; Estebas-Vilaplana & Prieto, 2010). In addition, every content word in Spanish (with some exceptions) is pitch accented (Hualde & Prieto, 2015). Therefore, Spanish was also predicted to have stronger MacR than English, and the results of the study confirmed the predicted strength ranking.

Nagao and Ortega-Llebaria (2021) investigated the interaction between micro-rhythm and macro-rhythm in the speech of L1 Japanese speakers learning English. Micro-rhythm refers to traditional speech rhythm classifications (i.e., isochrony hypothesis), where the domain of rhythm is smaller than the word (see Jun, 2014:524). According to the prosodic typology, (Tokyo) Japanese is predicted to have stronger MacR than English (Jun, 2014:535) because it is a Head/Edge-prominence pitch accent language that marks AP boundaries with a rising tone. In this study, two L1 Japanese speakers imitated a 2-minute English speech sample two times, with the second attempt produced over a month after the first one. Their productions were compared against the original sample spoken by an L1 English speaker. The authors found that the acquisition of English MacR by the two native Japanese speakers was dependent on the acquisition of lexical stress at the micro-rhythmic level. They concluded that a precise description of L2 rhythm requires the inclusion of both micro-rhythm and MacR measures.

Most recently, Kaland (2022) compared MacR strength between Greek, German, and European Portuguese, three Head-prominence languages that mark lexical stress. While the three languages have similarly sized tonal inventories, the most common pitch accent is L*+H in Greek (Jun, 2014, based Arvaniti & Baltazani, 2005), H* in German (Grice, Baumann, & Benzmüller, 2005), and H* in European Portuguese (Frotà, 2014). In addition, European Portuguese often does not mark the stressed syllable with a pitch accent, regardless

of whether the information is new or given, resulting in very little alternation of low-high tones within an utterance (Frotà, 2014). Therefore, the predicted MacR strength ranking was Greek > German > European Portuguese. The results of the study confirmed the strength ranking, with Greek having the strongest MacR, followed by German, followed by European Portuguese.

One aspect of the prosodic typology that has been given little attention in the literature so far is the relationship between the strength of lexical stress and MacR within a language. Jun hypothesized an inverse correlation between MacR and stress strength; that is, languages with strong phonetic realization of stress (i.e., higher amplitude and longer duration) are expected to have weaker MacR than languages with weak stress (2014:537). This has implications for the differences in stress realization across prominence types. For example, unlike Head-prominence languages, in which stress is generally realized with longer duration and increased intensity, the phonetic realization of stress in Head/Edge-prominence languages is typically weak. Jun suggests that this weak realization of stress is probably due to the presence of edge tones. It is either the case that a language with weak stress needs an edge tone to boost word prominence, or that a language does not need strong stress because the edge tone consistently marks word prominence (2014:537). In the case of Edge-prominence languages, lexical stress is not marked at all, so they are predicted to have stronger MacR than languages of other prominence types. Therefore, this inverse relationship in prominence-marking provides a clear, testable hypothesis for comparing languages of different prominence types.

1.2.5. MacR Quantification

Various measures have been proposed to phonetically quantify MacR. Jun (2014:538) introduced two such measures: MacR Variability Index (MacR_Var) and MacR Frequency Index (MacR_Freq). MacR_Var is calculated by taking the sum of the standard deviations of rising slope (rSD), falling slope (fSD), peak-to-peak distance (pSD), and valley-to-valley distance (vSD), as summarized in (1).

(1) MacR_Var = rSD + fSD + pSD + vSD

This measure is intended to capture the regularity or uniformity of the rise-fall slope. Languages with strong MacR are predicted to have less variability than languages with weak MacR. Prechtel (2020) found only a marginal difference in the MacR_Var Index between Spanish and English in both read speech and radio newscaster speech, and Spanish appeared to have a higher index, or *more* variability than English in newscaster speech, despite other evidence that Spanish had stronger MacR than English. This result could reflect one of the issues with the MacR_Var measure, which is that it does not specify the source(s) of variability; that is, it collapses the distinction between variability in F0 height and variability in the slope rise/fall of F0 movement. Polyanskaya et al. (2019) illustrated this potential issue by comparing a hypothetical Language A with low variability in the temporal domain (i.e., has regular L/H alternations) and high variability in the frequency domain (i.e., variable height of F0 peaks) to hypothetical Language B with high variability in the temporal domain and low variability in the frequency domain. In this scenario, the MacR_Var Index of Language A may be higher than Language B, despite Language A having more regular L/H alternations. To mitigate this issue, they calculated the MacR_Var Index by substituting the standard deviation with the coefficient of variation or Varco measure (SD divided by the mean) because it is robust to idiosyncratic differences in mean F0. However, they found no

significant difference between Italian and English, suggesting that the variability index measure should be revised to better capture differences in F0 magnitude and distance intervals, perhaps as separate quantification measures.

The MacR_Freq Index is calculated by dividing the number of F0 peaks per sentence by the number of PWords in the sentence, as summarized in (2). A language with stronger MacR will have a MacR_Freq ratio value close to 1, meaning that each PWord has one F0 peak.

(2) $\text{MacR\_Freq} = \frac{\text{Number of f0 peaks per sentence}}{\text{Number of PWords per sentence}}$

This measure is intended to capture the frequency of the F0 rise, corresponding to rule 3 (Figure 3). Prechtel (2019, 2020) found that Spanish had a MacR_Freq Index closer to 1 than English in both speech styles, meaning that Spanish had more consistent peak-to-PWord ratios than English. Although Polyanskaya et al. (2019) did not calculate MacR_Freq directly, they measured the number of F0 turning points, and found that Italian had significantly more turning points than English. This appears to be a more reliable measure for capturing MacR strength differences, but with a few caveats. First, this measure is dependent on phonological structure (e.g., PWord boundaries) in a way that other quantification measures are not. Other proposed measures are strictly based on the phonetic realization of F0 and do not explicitly reference linguistic structure. Second, while MacR_Freq Index works for languages like Italian and English, where each word generally has a maximum of one peak (or two turning points), it is not reliable for a language where each word can have multiple peaks, such as a contour tone language. While multiple turning points per word would raise the MacR_Freq Index of the utterances in a contour language, it would still have

weaker MacR than a language with one peak per PWord because the peaks occur at less regular intervals.

More recent work has adapted various speech rhythm metrics to capture MacR strength. Polyanskaya et al. (2019) used the Normalized Pairwise Variability Index (nPVI) to calculate variability in the duration distance intervals between F0 peaks and valleys. This measure is intended to correspond to both the presence of L/H alternations (Figure 1) and to the regularity of the peaks (Figure 3). As summarized in (3), nPVI calculates the difference in duration between each pair of successive F0 intervals, takes the absolute value of the difference, and divides it by the mean duration of the pair to normalize for speech rate.

(3) $$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

Lower variability in distance intervals between F0 targets corresponds to stronger MacR because less variation suggests the presence of L/H alternations occurring at regular intervals. The results of Polyanskaya et al. (2019) found that Italian had significantly lower nPVI values between L targets and between L and H targets compared to English, but there was no significant difference between H targets. They suggest that this finding is the result of Italian having more frequent phonological L* and L + H* targets than English, and so Italian speakers must purposefully plan the L targets to align with the stressed syllable. In contrast, Prechtel (2020) found that Spanish only had significantly lower nPVI values between H points compared to English, but no significant difference between L targets and between L and H targets, despite Spanish also having a default bitonal pitch accent with a phonological L target like Italian. It is possible that these cross-linguistic differences in nPVI between Italian and Spanish are due to differences in the default pre-nuclear pitch accent. That is, since the most common pre-nuclear pitch accent in Spanish has a delayed peak (L + >H*), it

is possible that the L targets are more variable than H targets compared to the L + H* pitch accent in Italian.

Nagao and Ortega-Llebaria (2021) adapted Vacro measures between F0 targets. To compare MacR strength between the L1 Japanese speakers' two English productions and the L1 English speaker's productions, the authors measured the distance intervals between tonal events (i.e., L and H targets), between peaks, and between valleys and then calculated normalized Varco scores. They found that although both speakers produced peak-to-peak intervals closer to the L1 English production, only one student produced native-like productions for all three measures, and this student also had more native-like productions of lexical stress.

The most recent study has taken a different approach to quantifying MacR than the previous literature. Kaland (2022) proposed the Contour Length Increase (CLI), which calculates the overall length of the F0 contour in an utterance compared to a flat line. For each phrase, Pythagoras' theorem ($a^2 + b^2 = c^2$) is used to calculate contour length, where *a* represents the duration difference between consecutive F0 points, *b* represents the F0 difference between the points, and *c* represents the length of the contour, as shown in Figure 4. The difference between the interval length (L_i = sum of all *a*) and contour length (L_c = sum of all *c*) is expressed as a percentage of length increase, as summarized in (4).

(4) Contour length increase (CLI) = $\frac{L_C}{L_i/100} - 100$

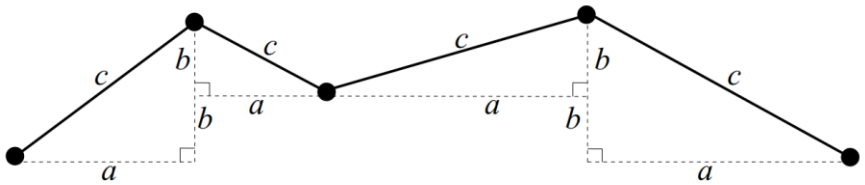

Figure 4: Example of a stylized f0 contour (solid line), where each dot represents an F0 point, and each line between points is the hypotenuse (c) of a right triangle (shown with dotted lines). Contour length is calculated as the sum of all *c*'s using Pythagoras' theorem. Adapted from Kaland (2022:5235).

This measure assumes that languages with stronger MacR will have more F0 alternations than languages with weaker MacR, and thus have a greater increase in overall F0 length compared to an F0 track with little to no movement. In Kaland's study, the results of the CLI quantification found that Greek had the largest percent length increase, followed by German, followed by European Portuguese, supporting the predicted MacR strength ranking.

Finally, Polyanskaya et al. (2019) looked at the magnitude of successive F0 turning points to capture variability in the frequency domain; that is, the magnitude of L targets (F0 minima) following H targets (F0 maxima). They found that Italian had larger displacement in L/H alternations than English, which they argue is the surface realization of the phonological L and H tones in the Italian pitch accent. The magnitude of F0 excursions between L and H targets has also been used to capture the degree of perceived "singsongy" intonation reported in speakers diagnosed with Autism Spectrum Disorder (ASD) (Wehrle, Cangemi, Krüger, & Grice, 2018; Wehrle et al., 2020). While not directly related to MacR strength, this measure, which Wehrle and colleagues call Spaciousness, was able to capture differences in F0 excursions between pitch points in ASD and neurotypical speakers of German. More generally, measuring differences in F0 magnitude is useful for determining variability in the frequency domain. That is, differences in the degree of displacement in L/H alternations between languages could affect the perception of MacR because the presence of F0

alternations (and their regularity) may be more noticeable in some languages and therefore contribute to stronger perceived MacR than languages with smaller magnitude of L/H alternations.

1.3. Current Study

1.3.1. Primary Goals

This dissertation is motivated by two primary goals. The first goal is to expand upon Jun's (2014) hypothesis that languages that mark AP prosodic units with edge tones (thus having strong MacR) tend to have "weaker" or no lexical stress. This leads to the prediction that there is an inverse relationship between lexical stress strength and tonal rhythm strength (2014:537), which has not yet been directly tested.

The second goal is to test whether differences in MacR strength are readily perceptible to listeners. While MacR quantification captures phonetic differences in F0 movement, which are motivated by the intonational phonology of a given language, it remains unclear how perceptible these differences in tonal rhythm are to listeners. If the function of MacR is to mark word prominence and facilitate word segmentation, then listeners are expected to have some sensitivity to tonal alternations in an utterance. Given the hypothesized inverse relationship between lexical stress and MacR strength, listeners are also expected to show a higher degree of sensitivity to tonal rhythm when the acoustic cues of lexical stress are weaker than when the cues are stronger.

So far, only one study has compared the perception of MacR strength, although rather indirectly and from a sociolinguistic perspective. In a study investigating the intonational characteristics associated with Jewish English, Burdin (2020) found that sentences with more frequent and salient rising pitch accents (i.e., having stronger MacR) were perceived as

sounding more Jewish. However, this result depended on the specific sociolinguistic characteristics of the listeners. Jewish listeners only identified the most extreme rises as sounding Jewish, specifically evoking the type of Jewish speaker who is older and speaks Yiddish. Of the non-Jewish listeners, only the ones with some familiarity with Yiddish associated less extreme rise/rise-fall contours as sounding Jewish while all non-Jewish listeners associated the frequent rises/strong MacR with older speakers. Therefore, these results suggest that listeners can perceive tonal regularity in an utterance, and that these differences can be associated with different language varieties.

1.3.2. Languages Chosen for Comparison

Three languages were chosen to test the predicted inverse correlation between the strength of lexical stress correlates and the strength of MacR. These languages are expected to have an inverse strength ranking between lexical stress strength and MacR strength. In other words, if the lexical stress strength ranking is Language A > Language B > Language C, then the MacR strength ranking is predicted to be Language C > Language B > Language A. For this comparison, at least one language should have strong realization of duration and intensity, and at least one language should have strong predicted MacR according to the prosodic typology model. Therefore, the following three languages were chosen for analysis: American English, Kolkata Bengali, and Uyghur.

English is a West Germanic language with SVO word order. It was chosen for analysis because of its well-documented strong acoustic realization of stress (i.e., duration and intensity) (e.g., Mairano et al., 2015). This strong realization could be a consequence of variable (not fixed) stress. In monomorphemic words, primary stress falls on the last three syllables (e.g., Chomsky & Halle, 1968; Domahs, Plag, & Carroll, 2014), and disyllabic

words tend to have stress on the first syllable (Cutler & Carter, 1987), while longer words have other patterns, depending on factors like syllable weight, affixation, and word-final vowel (e.g., Chomsky & Halle, 1968; Moore-Cantwell & Sanders, 2017; Moore-Cantwell, 2020). Regarding phrasal prominence, the most common pitch accent is H* (Dainora 2001, 2006), as opposed to a bitonal pitch accent. In addition, English frequently deaccents certain word categories (Schmerling, 1976; Ladd, 2008) and given/old information (Katz & Selkirk, 2011), whereas languages with stronger MacR like Spanish tend to mark every content word, given or new (e.g., Hualde & Prieto, 2015). Jun's model therefore predicts that English has medium MacR strength. Previous studies phonetically quantifying MacR strength found that English has weaker MacR than Italian (Polyanskaya et al., 2019) and Spanish (Prechtel, 2019, 2020). Therefore, English is expected to have stronger realization of lexical stress cues compared to the other languages, but not stronger MacR.

Bengali is an Eastern Indo-Aryan language spoken primarily in Bangladesh and India, including the states of West Bengal, Assam, Bihar, Jharkhand, Mizoram, Tripura (Khan, 2008:11). There are sizable speaker populations in the United States, the United Kingdom, Nepal, Singapore, and several other countries (Gordon, 2005). Like many of the languages spoken in the region, Bengali has SOV word order (Khan, 2016). The prosodic typology model predicts that it will have strong MacR because it has an AP unit (Khan, 2008, 2014), and the most common tonal pattern of the AP is a rise (i.e., L* Ha), with the stressed syllable on the left edge having a L tone and the AP-final syllable on the right edge having a H tone (Khan, 2008, 2014). Although this prediction is based on the intonational model of Bangladeshi Bengali specifically (Khan, 2008, 2014), the AP-sized default L/H pattern is also consistent with Hayes and Lahiri's (1991) intonational model of Kolkata Standard

Bengali. If the MacR strength prediction is correct, then stress realization is expected to be weaker than English. Indeed, most studies have found that while Bengali has fixed initial stress, the acoustic correlates of stress (duration and intensity) are relatively weak (e.g., Chatterji, 1921; Ferguson & Chowdhury, 1960; Kawasaki & Shattuck-Hufnagel, 1988; Hayes & Lahiri, 1991). Therefore, Bengali is expected to have stronger MacR and weaker lexical stress cues than English.

Uyghur is a southeastern Turkic language primarily spoken in the Xinjiang Uyghur Autonomous Region (XUAR) in the People's Republic of China, although there are diasporic communities in Kazakhstan, Kyrgyzstan, Uzbekistan, Australia, the United States, and elsewhere (Nazarova & Niyaz, 2013:xix). Like other Turkic languages, Uyghur is highly agglutinating with SOV word order and a rich case marking and agreement system (Engesæth, Yakup, & Dwyer, 2010; Nazarova & Niyaz, 2013). Although Uyghur is not included in the MacR typology data in Jun (2014), Major and Mayer's (2018) model of intonational phonology indicates that it has an AP unit that is typically marked with a rising tonal pattern (L H). Therefore, Uyghur is also predicted to have strong MacR, but it differs from Bengali in a few ways. First, stress is acoustically realized as duration (Yakup, 2013; Yakup & Sereno, 2016; Major & Mayor, 2018); intensity has not been found to be a reliable cue (Yakup & Sereno, 2016). However, lexical stress appears to be marked independently of F0. In other words, the F0 targets that contribute to tonal rhythm do not necessarily align with the lexically stressed syllable, which makes Uyghur an example of a lexical stress language with strictly edge-marking intonation. This is typologically unusual because the AM model assumes that phrasal prominence (i.e., F0) aligns with the lexically prominent syllable. However, while rare, this pattern is also attested in a few other languages such as

Wolof (Rialland & Robert, 2001), Kuot (Lindström & Remijsen, 2005), and Farasani Arabic (although in Farasani Arabic this pattern occurs only in neutral focus contexts (Abbas & Jun, 2021)). This contrasts with Bengali, where the stressed syllable bears the AP-initial tonal target. Second, unlike Bengali, the location of stress in Uyghur is variable (Yakup, 2013; Yakup & Sereno, 2016; Major & Mayer, 2018) and weight-sensitive (Hahn, 1991; Engesæth et al., 2010; McCollum, 2020), although there is a strong preference for stress on the last syllable of the word (Nadzhip, 1971; Hahn, 1998). In general, stress falls on the penultimate (Hahn, 1991) or leftmost (Engesæth et al., 2010) syllable if heavy, otherwise it falls on the last syllable. However, certain CVC suffixes tend to avoid stress (Engesæth et al., 2010), and stress tends to avoid high vowels, which undergo devoicing and reduction processes (e.g., Hahn, 1991). Despite the independence of lexical stress realization and phrasal prominence, the preference for word-final stress may be a way to boost prominence at the AP/word edge that is already marked by F0. Since Uyghur stress is consistently marked by duration, it is expected to have stronger lexical stress realization than Bengali but weaker realization than English. This also suggests that Uyghur will have weaker MacR than Bengali but stronger MacR than English. Therefore, the predicted language rankings can be summarized as the following in (5):

(5) Predicted language ranking

a. Degree of lexical stress strength: English > Uyghur > Bengali

b. Degree of MacR strength: Bengali > Uyghur > English

The rest of this dissertation is organized as follows: Chapter 2 reports the results of a lexical stress production experiment, in which the duration ratios of stressed and unstressed syllables were compared across languages; Chapter 3 reports the results of a MacR

production experiment, which quantified MacR of prosodically annotated recordings of *The North Wind and the Sun* short story; Chapter 4 reports the results of a MacR perception experiment in which participants rated tonal rhythm or melody of utterances in each language; and Chapter 5 discusses the results of the experiments.

CHAPTER 2

Lexical Stress Production Experiment

This chapter presents the details of the lexical stress production experiment. The purpose of this experiment was to test the prediction that English had the strongest realization of lexical stress, followed by Uyghur, followed by Bengali. To test this prediction, ratios of the duration of stressed and unstressed vowels were compared across the three languages. Larger durational differences indicate stronger realization of lexical stress, so the predicted language ranking is English > Uyghur > Bengali.

2.1. Methods

2.1.1. Stimuli

The stimuli consisted of 8 nonce disyllabic CVCV words with trochaic stress. Both vowels were low unrounded /a/ and the consonants were oral or nasal stops. The purpose of using nonce words in the experiment was to be able to directly compare vowel duration ratios and control segmental contexts across all three languages.

The target words were embedded in two separate carrier phrases, which were in turn embedded into a short 4-sentence paragraph. The two carrier sentences were treated as two prosodic conditions: first repetition and second repetition. The target word in the first repetition condition was predicted to have the nuclear pitch accent in English, thus the most prominent in the sentence, while the second repetition condition was predicted to be less prosodically prominent (e.g., Fowler & Housum, 1987; Aylett & Turk, 2004). Each paragraph followed a similar structure to the English example in (6):

(6) I baked a pastry called <u>bada</u> today. It's a type of pastry filled with cheese and fruit. I heard that <u>bada</u> should be very tasty. I'm excited to eat it!

The first sentence introduced the target word (underlined in the above example), the second sentence defined the target word as a type of food or drink, the third sentence embedded the target word into a subordinate clause, and the final sentence declared a desire to try the food or drink. The paragraphs were designed to introduce the target words as plausibly real words. See Appendix A for the list of utterances used in each language.

2.1.2. Participants

A total of 28 speakers (10 English, 10 Bengali, and 8 Uyghur) were included in this experiment. All participants, regardless of language, had to meet the following eligibility requirements: they must be at least 18 years old, a native speaker of the target language (and dialect, if specified), and able to record themselves in a very quiet room using some type of head-mounted microphone. Before starting the experiment, participants also filled out a background questionnaire, which asked basic demographic information such as age, gender, birthplace (city or general region), places lived for more than 12 months, language background, and formal language education.

English participants were recruited through the UCLA psychology subject pool and compensated with course credit. In addition to the eligibility requirements stated in the previous paragraph, participants must have grown up in the U.S. and natively speak American English. A total of 29 speakers participated, although 11 participants were excluded because they did not meet the eligibility criteria or had poor audio quality. From 18 speakers, 10 speakers (5 females and 5 males) were randomly selected for analysis in this experiment. Table 1 summarizes their background information. One of the speakers was born in China but moved to the U.S. by age 4. The age range was 19-23 years old, and all participants reported acquiring English before age 5.

Participant	Age	Gender	Birthplace	Other languages acquired + approx. age of acquisition
355010_en_27	20	Female	Los Angeles, CA	Vietnamese, birth
355342_en_29	19	Female	Overland Park, KS	Malayalam, birth Tamil, age 2 Spanish, age 4
358843_en_37	19	Female	Staten Island, NY	N/A
359757_en_39	23	Male	Northridge, CA	Hebrew, 12
359790_en_40	20	Male	Riverside, CA	N/A
362322_en_41	21	Female	La Palma, CA	N/A
362351_en_43	21	Male	Hefei, China	Mandarin, birth
381996_en_46	19	Female	Burbank, CA	Armenian, birth
382428_en_48	21	Male	Torrance, CA	N/A
382515_en_52	20	Female	Los Angeles, CA	Farsi, birth Hebrew, 5

Table 1: Language and demographic information of the English participants.

Bengali participants were recruited through word-of-mouth and online advertising, and they received $10 for their participation, paid through an online cash transfer app. To be eligible to participate, speakers must have grown up in or around Kolkata, India. A total of 16 participants completed the experiment, but 6 were excluded because of poor audio quality, excessive background noise, or disfluent reading, so 10 participants (5 females and 5 males) were included for analysis. Table 2 summarizes their background information. The age range was 21-39 years old, and all participants reported acquiring English before age 5.

Participant	Age	Gender	Birthplace	Other languages acquired + approx. age of acquisition
412389_be_02	30	Female	Kolkata	English, 3 Hindi, 10
424843_be_07	26	Male	Kolkata	English, 1 Hindi, 1
446946_be_08	23	Female	Near Kolkata	English, 3 Hindi, 5 Sanskrit, 12
449053_be_10	22	Female	Kolkata	English, 5 Hindi, 6
450012_be_11	34	Male	Kolkata (south)	English, 1 Hindi, 3
451069_be_15	39	Female	Kolkata	English, 5 Hindi, 12
476993_be_18	21	Male	Kolkata	English, 7 Hindi, 8
487510_be_19	30	Female	Kolkata	English, 6 Hindi, 6
489487_be_20	29	Male	Kolkata	English, 1 Hindi, 1 French, 22
491610_be_22	29	Female	Kolkata	English, 2 Hindi, 11 German, 27 French, 29

Table 2: Language and demographic information of the Bengali participants.

Uyghur participants were recruited through word-of-mouth and online advertising, and they could choose the form of compensation, if any, such as monetary compensation ($10) or request that the money be donated to a Uyghur organization of their choice. Some participants opted to forgo payment altogether for privacy reasons. To be eligible to participate, speakers must have been born in Xinjiang Uyghur Autonomous Region (XUAR), China, and currently live within the U.S. or Canada. Only Uyghurs originally from XUAR were included because

there is evidence that some Uyghur dialects spoken in Kazakhstan do not mark stress with duration (Major & Mayer, 2019). A total of 12 speakers participated, but 4 were excluded due to poor audio quality or disfluent reading, so 8 speakers (5 females and 3 males) were included for analysis. The experiment was initially presented in Uyghur Latin script but was later changed to Perso-Arabic script after feedback from participants. About half of the participants completed the Latin script version. Both versions of the stimuli are included in Appendix A. Table 3 summarizes the participants' demographic information. The age range was 28-73 years old, and all participants reported acquiring Uyghur from birth.

Participant	**Age**	**Gender**	**Birthplace (city or region)**	**Other languages acquired + approx. age of acquisition**
395319_uy_02	47	Female	Korla	Mandarin, 7 English, 26 Turkish, 28
433461_uy_08	28	Female	Qeshqer	Mandarin, age unknown English, age unknown
436078_uy_09	31	Female	Qeshqer	Mandarin, age unknown English, age unknown
472570_uy_10	49	Male	Ghulja	Mandarin, 8 English, 30
520256_uy_11	38	Male	Hotan	Mandarin, ~11 (6th grade) English, ~19 (college sophomore)
520300_uy_12	48	Female	Ürümqi	Mandarin, 10 English, 18
551530_uy_14	73	Male	Ili Kazakh Autonomous Province	English, age unknown
556214_uy_15	50	Female	Ghulja	Mandarin, 9 English, 19 Turkish, 35

Table 3: Language and demographic information of the Uyghur participants.

While a few English participants were monolingual (see Table 1), most participants in this study were multilingual, having acquired at least one other language from a young age. Most Bengali speakers began acquiring English by age 5, and all reported receiving some formal English language education in school. The average age for Uyghur speakers was 46 years old, which is older than the average English (20.3) and Bengali (28.3 years) speakers in this study, but they were also more likely to have received some formal Uyghur language education. Most Uyghur speakers indicated that they learned Mandarin Chinese starting in primary school and received formal education through university level.

2.1.3. Procedure

The experiment was conducted online through *LabVanced* (Finger et al., 2017) and consisted of two parts. In Part 1, participants were presented with the stimuli described in section 2.1.1. of this chapter, while in Part 2, participants were presented with the short story *The North Wind and the Sun*, which was used to analyze MacR production (Chapter 3). The two parts were divided by an optional break. At the start of the experiment, participants were instructed to record in a very quiet room and use some type of head-mounted microphone. Examples of acceptable microphone recording setups included gaming headsets, wired earbuds with attached microphone, and Apple AirPods. This was to ensure that the microphone was always the same distance away from the speaker's mouth throughout the recording, as well as to ensure better audio quality for analysis.

After completing the consent form and background questionnaire, participants were provided with detailed instructions on how to do the experiment. All participants listened to two audio examples of the task in Part 1 in the target language, which accompanied the displayed paragraph text. The first audio example was a real word in each language with the same

characteristics as the target words, e.g., a CaCa word with trochaic stress, while the second audio example was the nonce word ‘waga’, which was used to reinforce the target pronunciation using an unfamiliar word. For Uyghur speakers only, the instructions also went into greater detail about stress placement. Because stress in Uyghur tends to fall on the final syllable (Nadzhip, 1971; Hahn, 1998) and is also affected by syllable weight (Hahn, 1991; Engesæth et al., 2010; McCollum, 2020), there are only a few lexical stress minimal pairs, so the instructions additionally explained that the nonce words should be pronounced the same way as the real words with stress on the first syllable. After listening to the audio examples, participants recorded two practice trials to familiarize them with the reading task and the recording setup. Both practice trials used nonce words; the first trial contained nonce word ‘waga’ like the audio example, and the second trial contained nonce word ‘mawa.’ Once participants finished the practice trials, they could begin the experiment trials at their own pace.

2.1.4. Analysis

All words were acoustically analyzed in *Praat* (Boersma & Weenink, 2022). For each sound file, a TextGrid was created with two interval tiers; the first labeled the target words and their repetition number, while the second tier divided the word into individual segments. Only the vowels of the target word were labelled and extracted for analysis (Figure 5). It should be noted that the second repetition of the target word in the Uyghur data contained an additional accusative suffix *–ning* (e.g., *baganing*), which was excluded from analysis despite being part of the morphological word. After the vowel duration for each token were extracted using a Praat script, the duration of the first and second vowels were converted into log-transformed ratios to normalize for speech rate (Beckman, 1986).

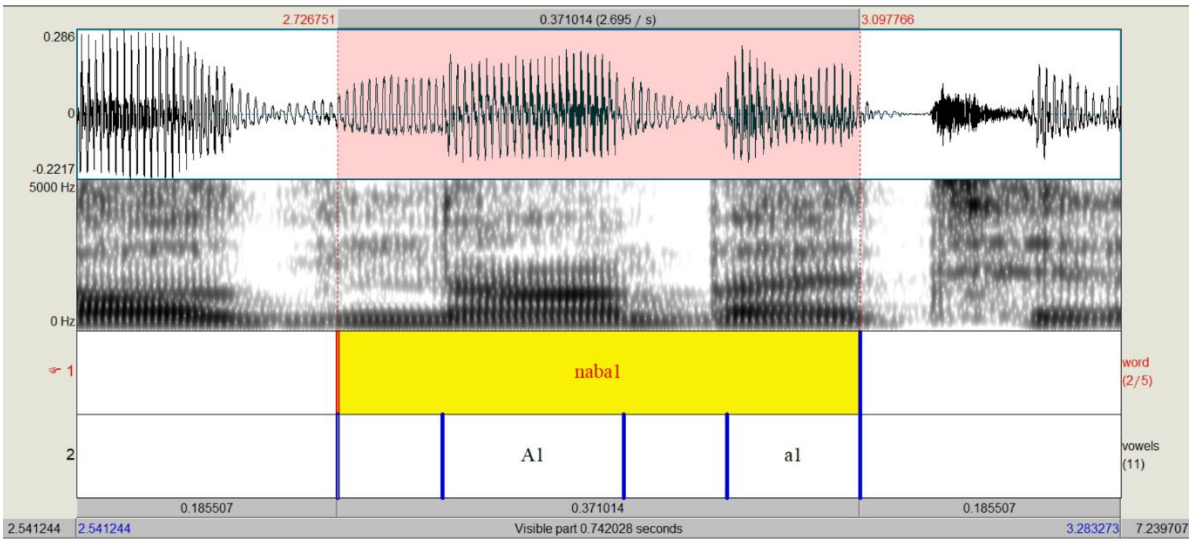

Figure 5: Example of a target word produced by an English speaker. Tier 1 shows the word boundaries for 'naba' labelled by repetition number ('1'), and Tier 2 shows the boundaries of each segment. Only the stressed and unstressed vowels are labelled, represented as capitalized 'A' and lowercase 'a' respectively.

2.2. Results

A total of 398 tokens were included for analysis. Within each language group, English had 155 tokens (4 excluded), Uyghur had 94 tokens (34 excluded), and Bengali had 149 tokens (11 excluded). Words were excluded if there was a disfluency in the pronunciation, the stress was realized on the second syllable, or there was an issue with the audio quality on that word (e.g., background noise or electronic interference). Most of the excluded tokens in Uyghur were instances where stress was realized on the second syllable.

Figures 6-8 show the mean duration ratio distribution (in natural log) for each speaker in each language, capturing the individual speaker variability within each language group. English speakers (Figure 6) tended to have more within-speaker variability in vowel duration ratios compared to Uyghur (Figure 7) and Bengali (Figure 8) speakers. In general, the vowel duration ratios in English were larger than the other two languages, indicated by the higher values.

Uyghur, in contrast, tended to have the smallest duration ratios, indicated by the lower values. Bengali duration ratio values appeared to be somewhere in between English and Uyghur.

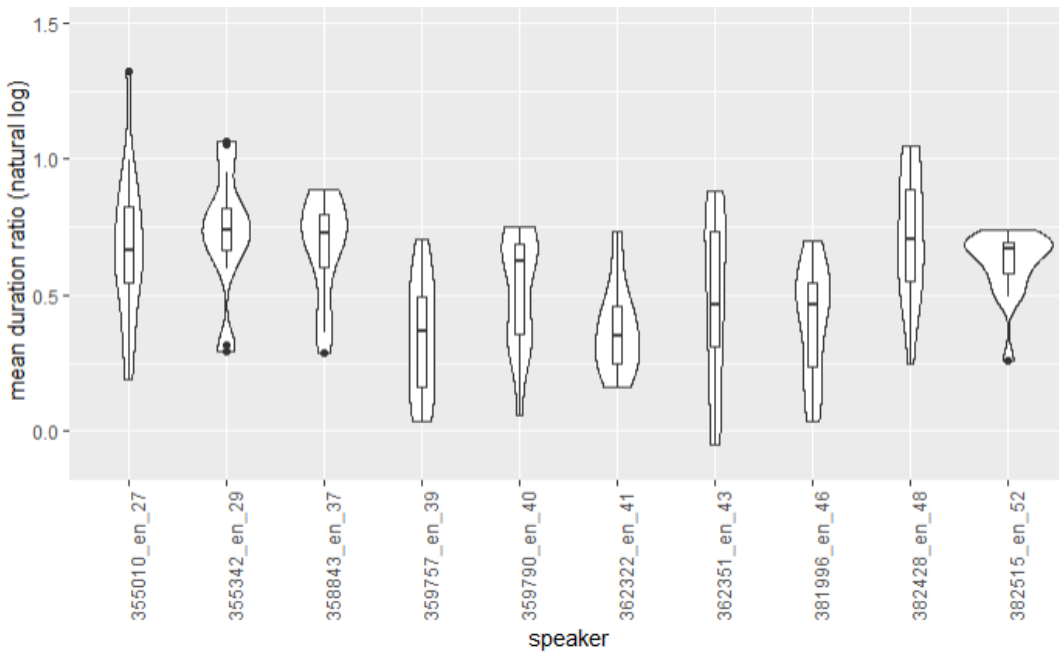

Figure 6: Mean duration ratio distribution of each English speaker (n=10).

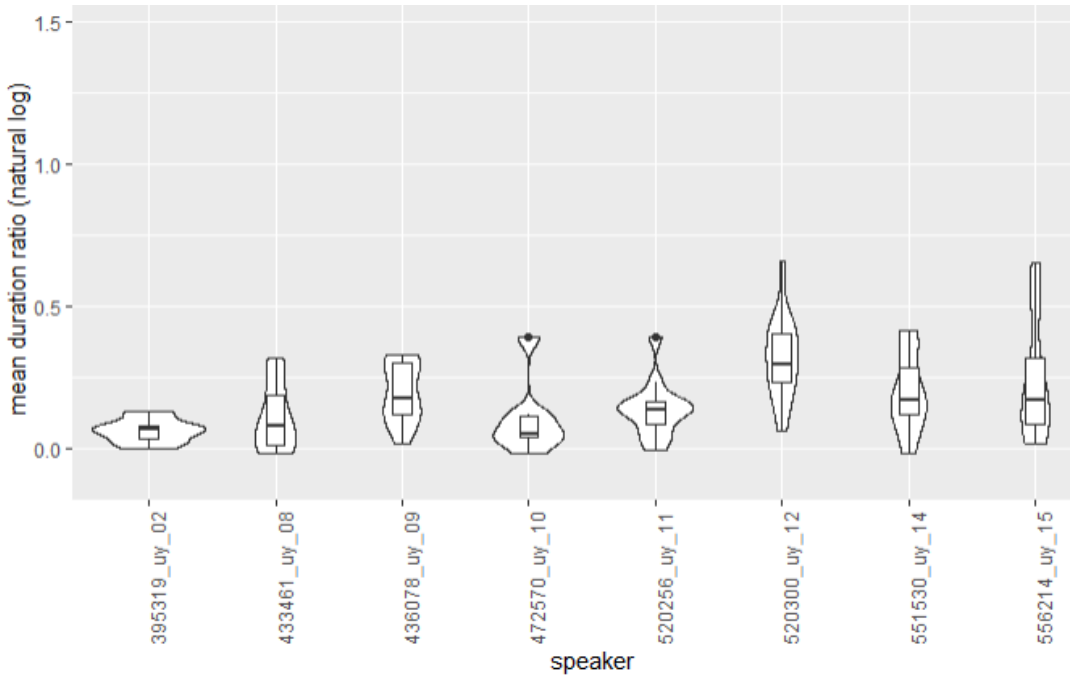

Figure 7: Mean duration ratio distribution of each Uyghur speaker (n=8).

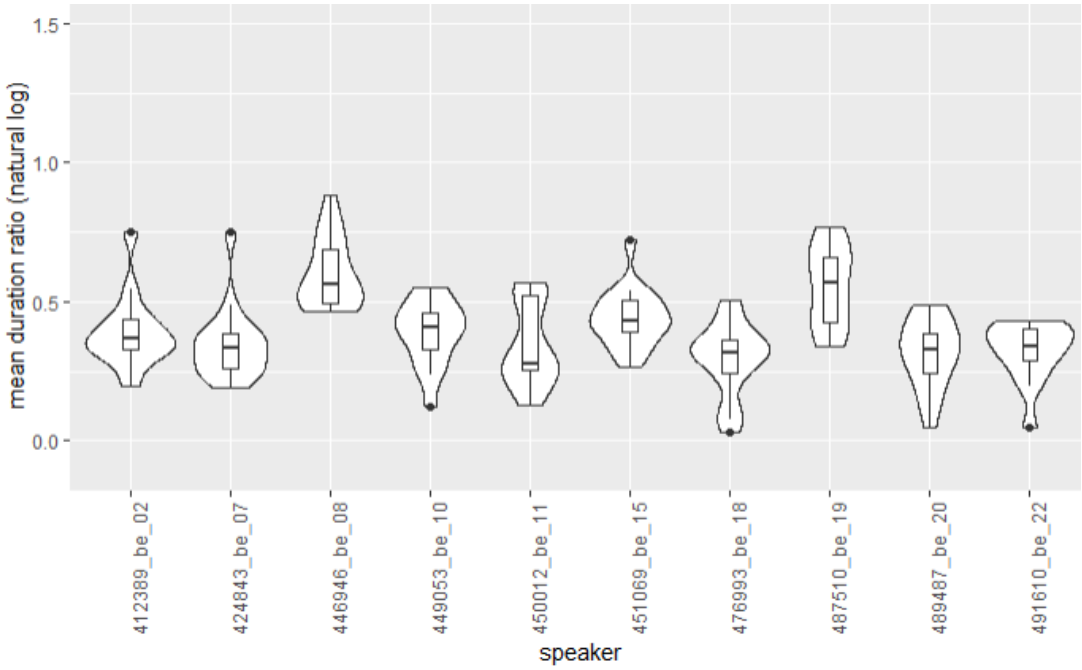

Figure 8: Mean duration ratio distribution of each Bengali speaker (n=10).

In addition to comparing individual speakers, the mean duration ratios were compared for each word within each language. Figures 9-11 show the mean distribution of vowel duration ratios for each word within each language. In Figure 9, English speakers tended to have larger variability in duration ratios than Uyghur (Figure 10) and Bengali (Figure 11). In contrast, the Bengali speakers appear to have the most consistent ratios across word type, while the Uyghur speakers produced words with variability between English and Bengali.

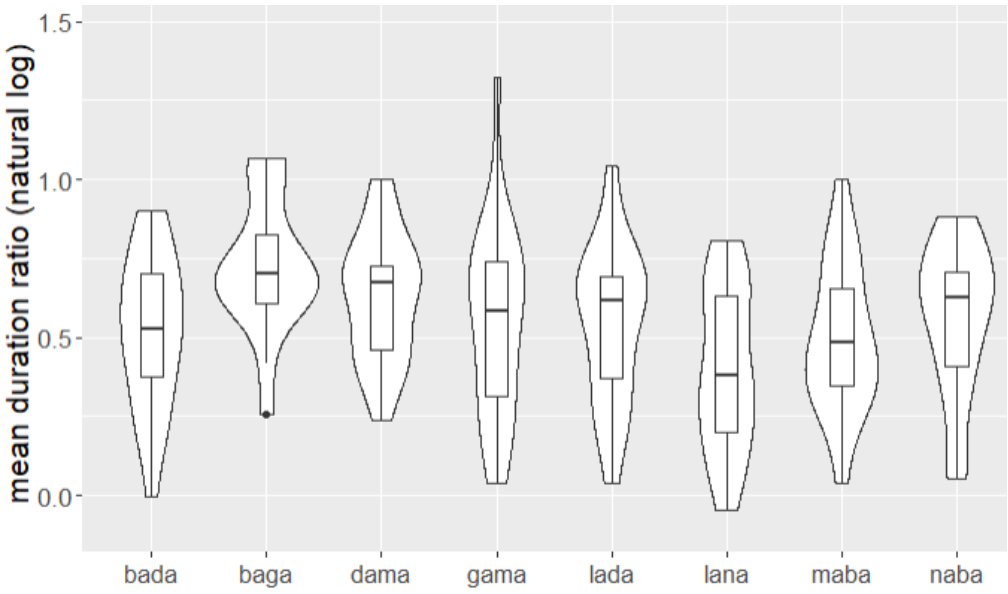

Figure 9: Distribution of the mean vowel duration ratios of each target word produced by English speakers.

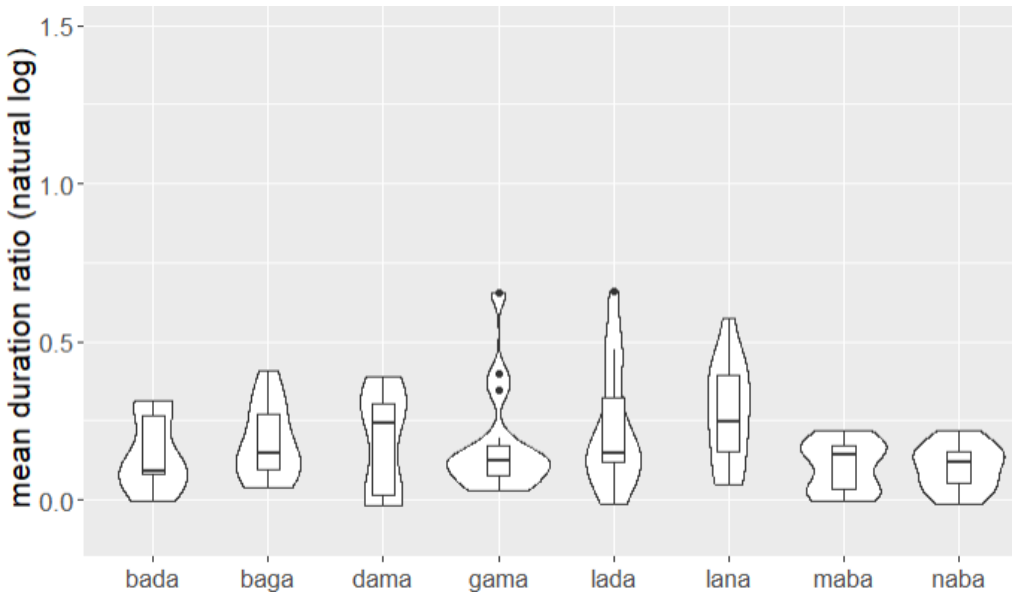

Figure 10: Distribution of the mean vowel duration ratios of each target word produced by Uyghur speakers.

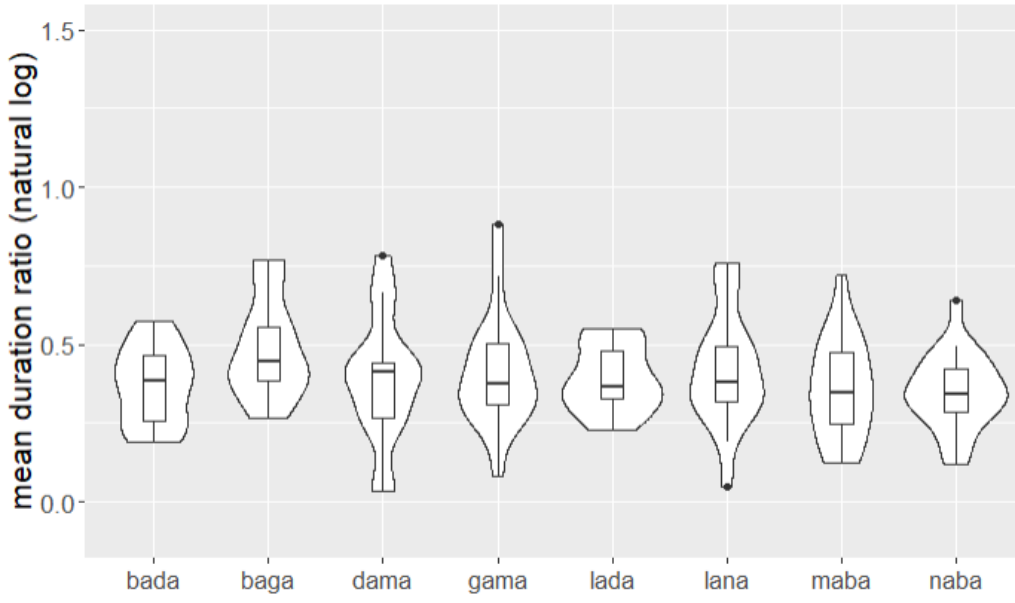

Figure 11: Distribution of the mean vowel duration ratios of each target word produced by Bengali speakers.

Linear mixed effects models were run in RStudio (RStudio Team, 2020) using the *lme4* package (v1.1-26; Bates et al., 2015) and the *lmerTest* package (v3.1-3; Kuznetsova, Brockhoff, & Christensen, 2017) for p-values, with duration ratio as the dependent variable, language and repetition as predictors, and participant and word as random intercepts. The results found that there was no significant difference between repetitions, and English ratios were significantly larger than both Uyghur ratios ($\beta = 0.38$, $t = 6.79$, $p < 0.0001$) and Bengali ratios ($\beta = 0.15$, $t = 2.92$, $p = 0.007$). A post-hoc pairwise comparison using the *emmeans* package (v1.7.2; Lenth, 2022) found that Bengali ratios were significantly larger than Uyghur ratios ($\beta = 0.23$, $t = 4.06$, $p = 0.001$). These results are reflected in Figure 12, which shows the distribution of vowel duration ratios in each language. Both the distributions of English and Bengali ratios are shown to have higher values than Uyghur, meaning that the duration of the first vowel was much longer compared to the second vowel in both English and Bengali compared to Uyghur.

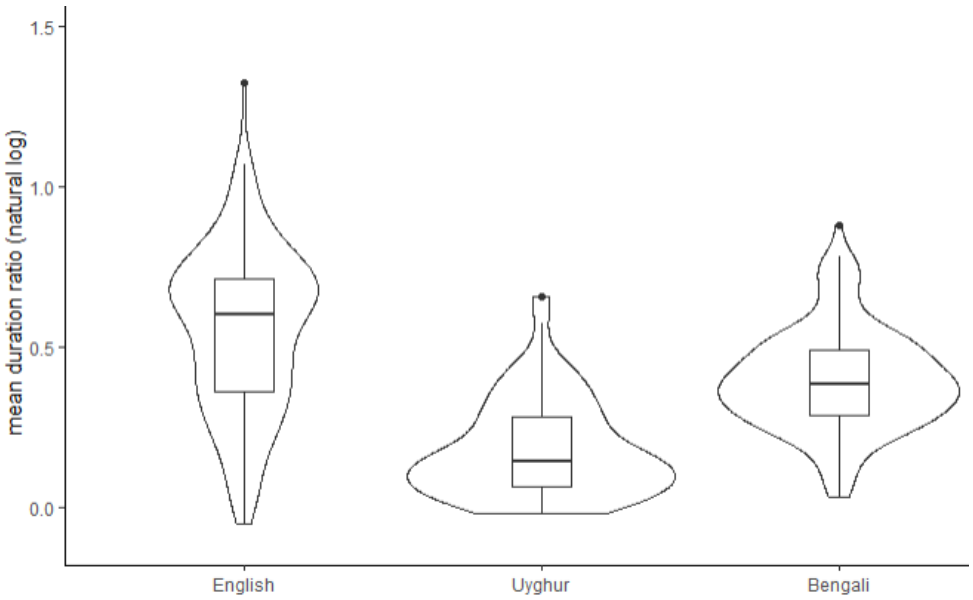

Figure 12: Distribution of vowel duration ratios in English, Uyghur, and Bengali.

2.3. Discussion

While the results of the lexical stress production experiment support the predication that English has larger vowel duration ratios than both Uyghur and Bengali, the Bengali ratios were unexpectedly large. Regardless of prosodic context (i.e., repetition), English duration ratios were significantly larger than both Uyghur and Bengali, but Bengali ratios were significantly larger than Uyghur ratios, contrary to the predicted language ranking. This is surprising, given that previous literature on Bengali word stress has not found duration to be a reliable cue. The results suggest that English and Bengali speakers in this experiment realized stress more similarly to each other than Bengali and Uyghur speakers, despite the expectation that duration as a cue would be the least reliable for Bengali. The unexpected result for Bengali stress realization may be due, in part, to the speaker demographic that participated in this experiment. As discussed in section 2.1.2., most Bengali participants began acquiring English at a young age and all of them also received formal English education. In addition, the participants were mostly university students and young professionals who used English more frequently in their daily lives. Therefore, as English-Bengali bilinguals, their production of unfamiliar nonce words may have been influenced by English, although their tonal rhythm patterns seemed largely unaffected.

The potential interference of English-like lexical stress was likely due to the use of nonce words as stimuli instead of real ones. Since the carrier sentences were designed to introduce the nonce word as a plausible loanword, it is possible that the Bengali speakers treated them as English borrowings. An earlier pilot experiment included both real and nonce target words to test whether speakers would produce them differently, and the results found no significant duration differences (Prechtel, 2021). Based on the results of the pilot, the current experiment did not include real words from the target language as a control. However, the pilot experiment only

included one speaker per language, and the Bengali participant was older than the participants in this experiment (> 45 years old). Therefore, it is possible that younger Bengali speakers who used English more frequently produced the nonce words more like an English word.

The results also found no significant difference between prosodic contexts within languages. This is a bit surprising given that English stress realization is affected by prosodic prominence (e.g., Anderson, Pierrehumbert, & Liberman, 1984; Li & Post, 2014), and second mention contexts elicit segment reduction in both spontaneous speech (e.g., Fowler & Housum, 1987) and read speech (e.g., Baker & Bradlow, 2009; Fowler, 1988; see Clopper & Turnbull, 2018 for an overview of phonetic reduction). However, while duration ratios were numerically smaller in the second repetition in all three languages, prosodic context did not significantly affect stress realization, probably because the stimuli were short nonce words.

To summarize the findings of this chapter, the predicted language ranking for lexical stress strength was English > Uyghur > Bengali, but the results of this experiment found that the actual ranking was English > Bengali > Uyghur. Contrary to expectation, Bengali speakers produced the nonce words more similarly to English speakers, while Uyghur speakers tended to have small differences in duration between the first and second syllable. Therefore, while English had the strongest lexical stress realization of the three languages as predicted, Uyghur had the weakest realization instead of Bengali. The next chapter will report the results of the MacR production (i.e., quantification) experiment, which tests whether the predicted strength ranking holds for tonal rhythm.

CHAPTER 3

Macro-Rhythm Production Experiment

This chapter presents the details of the MacR production experiment. The purpose of this experiment was to test the prediction that Bengali has the strongest MacR, followed by Uyghur, followed by English (Bengali > Uyghur > English). To test this prediction, participants read a short story, which was prosodically annotated and compared across the three languages.

3.1. Methods

3.1.1. Stimuli

Participants read a version of *The North Wind and the Sun* story, which was taken from an online collection of translations into many languages and dialects (Aesop Language Bank Team, 2010). As mentioned in Chapter 2, participants read this story after completing the lexical stress production task. See Appendix B for each story translation.

3.1.2. Participants

The data of 24 participants (8 per language) was included for analysis, and each language had 5 female and 3 male speakers. Most speakers included here were also analyzed in the lexical stress production experiment in Chapter 2 (6 English, 8 Bengali, 6 Uyghur). Tables 4-6 summarize the demographic and language background information of the participants for each language.

Participant	Age	Gender	Birthplace	Other languages acquired + approx. age of acquisition
355010_en_27	20	Female	Los Angeles, CA	Vietnamese, birth
358843_en_31	19	Male	Denver, CO	Spanish, 12 Hebrew, 18
359757_en_39	23	Male	Northridge, CA	Hebrew, 12
359790_en_40	20	Male	Riverside, CA	N/A
362322_en_41	21	Female	La Palma, CA	N/A
381996_en_46	19	Female	Burbank, CA	Armenian, birth
382515_en_52	20	Female	Los Angeles, CA	Farsi, birth Hebrew, 5
383260_en_53	20	Female	Tehran, Iran	Farsi, birth

Table 4: Language and demographic information of the English participants.

Participant	Age	Gender	Birthplace	Other languages acquired + approx. age of acquisition
412389_be_02	30	Female	Kolkata	English, 3 Hindi, 10
424843_be_07	26	Male	Kolkata	English, 1 Hindi, 1
446946_be_08	23	Female	Near Kolkata	English, 3 Hindi, 5 Sanskrit, 12
449053_be_10	22	Female	Kolkata	English, 5 Hindi, 6
476993_be_18	21	Male	Kolkata	English, 7 Hindi, 8
487510_be_19	30	Female	Kolkata	English, 6 Hindi, 6
489487_be_20	29	Male	Kolkata	English, 1 Hindi, 1 French, 22
491610_be_22	29	Female	Kolkata	English, 2 Hindi, 11 German, 27 French, 29

Table 5: Language and demographic information of the Bengali participants.

Participant	Age	Gender	Birthplace	Other languages acquired + approx. age of acquisition
395319_uy_02	47	Female	Korla	Mandarin, 7 English, 26 Turkish, 28
406628_uy_03	33	Female	Ürümqi	Mandarin, 12 English, 20
423723_uy_06	36	Female	Ürümqi	Mandarin, 9 English, 12
433461_uy_08	28	Female	Qeshqer	Mandarin, age unknown English, age unknown
472570_uy_10	49	Male	Ghulja	Mandarin, 8 English, 30
520256_uy_11	38	Male	Hotan	Mandarin, ~11 (6th grade) English, ~19 (college sophomore)
520300_uy_12	48	Female	Ürümqi	Mandarin, 10 English, 18
551530_uy_14	73	Male	Ili Kazakh Autonomous Province	English, age unknown

Table 6: Language and demographic information of the Uyghur participants.

3.1.3. F0 Annotation

Each recording was annotated in *Praat* for syllables, words, PWords, and F0 turning points. F0 annotation was a multistep process. First, the pitch tracks were stylized using the annotation process described in Mennen, Schaeffler, and Docherty (2012) to create a simplified representation of the F0 contours without fluctuations caused by pitch estimation errors and microprosody. To stylize F0, a Manipulation Object was created for each file, and all the original F0 points were deleted. Next, the initial and final F0 points were added, and then points were added for each F0 minimum and maximum, excluding perturbations due to microprosody. Additional points were added whenever the interpolation between points differed substantially from the original contour, such as when the F0 was in a steady state or plateau before rising or

falling to the next tonal target. The stylized contour was saved as a separate Pitch Object. The audio files and corresponding Pitch Objects were then fed into SPPAS (Bigi, 2015), an open-source annotation tool that automatically detected the F0 turning points using the MOMEL (Modeling Melody) algorithm (Hirst & Espesser, 1993) and labelled the points using the INTSINT (INternational Transcription System for INTonation) system (Hirst & de Cristo, 1999). This labelling system annotates F0 values of pitch targets as either absolute tones that reflect the speaker's pitch range within the utterance (T = Top, M = Middle, B = Bottom) or as relative tones that refer to the value of the preceding tonal target (H = Higher, S = Same, L = Lower, U = Up-stepped, D = Down-stepped). The labels were compared to the stylized pitch contour and visually inspected for accuracy. The labels and their associated time and frequency information were then extracted and consolidated into L (low tonal target), H (high tonal target), and S (same as previous tonal target, i.e., a F0 plateau) labels to simplify the analysis. The final L or H point was excluded from analysis because of its association with the IP boundary. Figures 13-15 show examples of the annotation for each language. Tier 1 marks the F0 labels; Tier 2 marks the syllable boundaries, labelled in IPA; Tier 3 marks the words (transliterated into Latin script for Bengali and Uyghur); Tier 4 marks the MacR_Freq Index, in which each PWord was labeled either with a 1 if an H tonal target was present or a 0 if no H target was present; and Tier 5 labels the order of the PWord within the utterance. The final F0 target was labelled with the additional % label to indicate the IP boundary and was excluded from analysis.

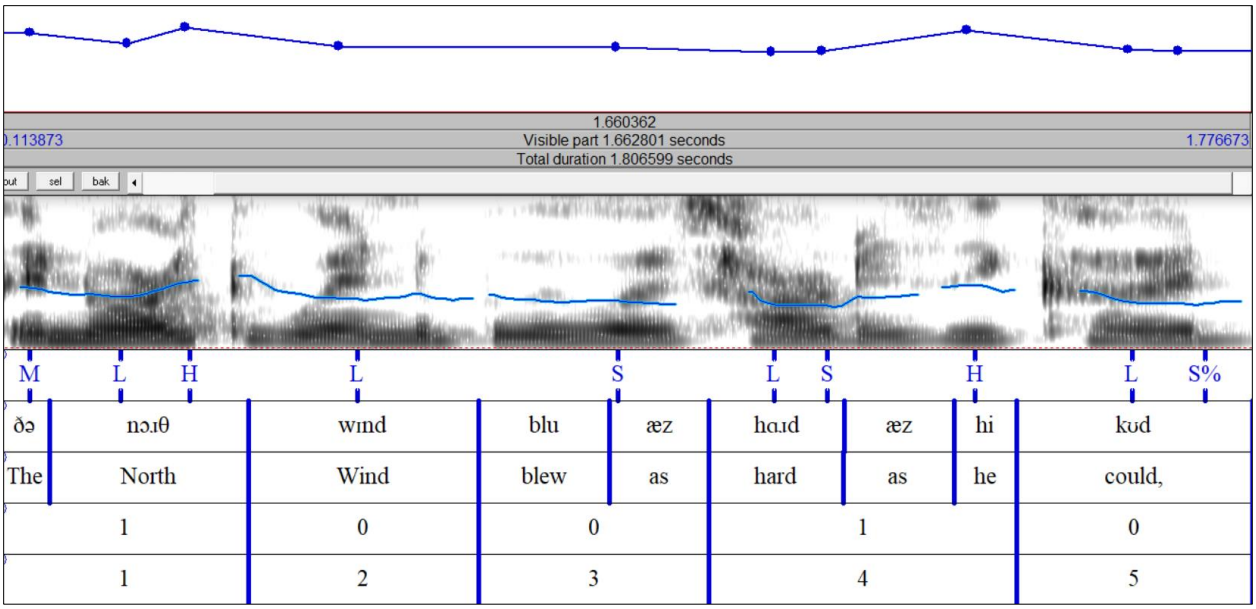

Figure 13: Example of pitch stylization + annotation of an IP produced by a female English speaker. Tier 1 = F0 labels, Tier 2 = syllables, Tier 3 = words, Tier 4 = MacR_Freq Index, Tier 5 = PWord order within IP.

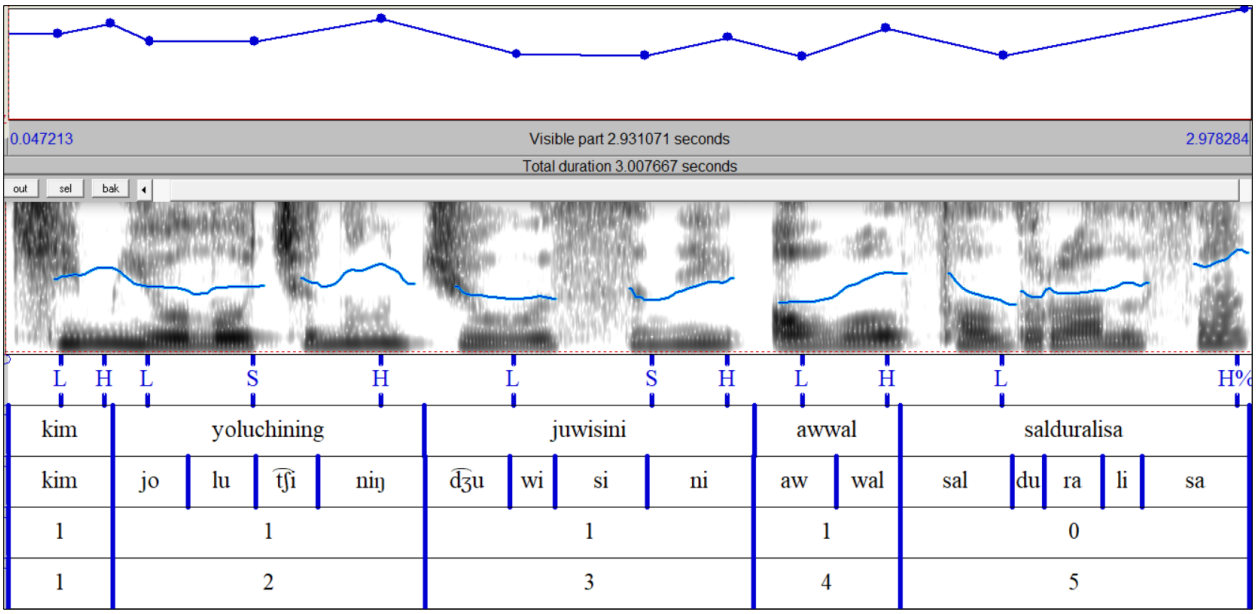

Figure 14: Example of pitch stylization + annotation of an IP produced by a female Uyghur speaker. Tier 1 = F0 labels, Tier 2 = syllables, Tier 3 = words, Tier 4 = MacR_Freq Index, Tier 5 = PWord order within IP.

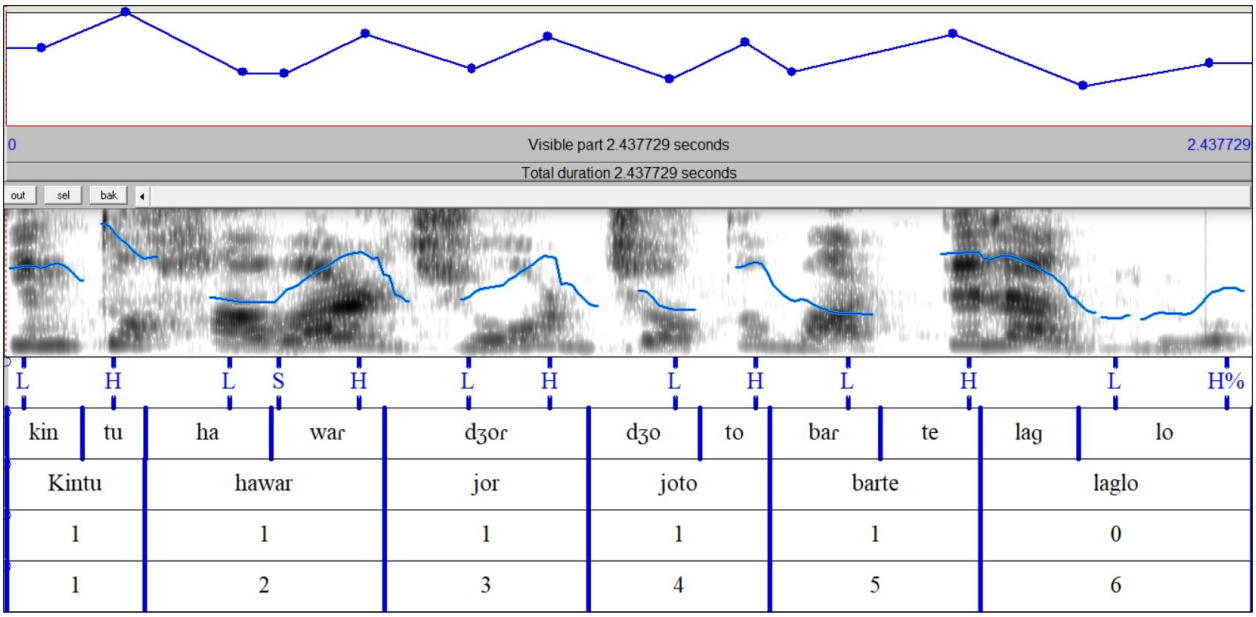

Figure 15: Example of pitch stylization + annotation of an IP produced by a female Bengali speaker. Tier 1 = F0 labels, Tier 2 = syllables, Tier 3 = words, Tier 4 = MacR_Freq Index, Tier 5 = PWord order within IP.

3.1.4. MacR Measures

As discussed in Chapter 1, numerous measures have been proposed to quantify MacR strength. While all of them capture some aspect of regularity or variability in F0 movement, most of them are limited in what they can tell us about tonal rhythm. Therefore, this study will take a holistic approach and choose measures that capture information about F0 alternations in both the temporal (time) and frequency (magnitude) domains. The following measures will be included for analysis: MacR Frequency (MacR_Freq) Index (Jun, 2014:538), the magnitude or height difference of F0 displacement between tonal targets (adapted from Polyanskaya et al., 2019), and the Contour Length Increase (CLI) measure (Kaland, 2022). Together, these measures are intended capture cross-linguistic differences in both the regularity of L/H intervals and the variability of F0 excursions.

The MacR_Freq Index will be used to calculate MacR in the temporal domain. As a reminder, this measure is calculated by dividing the number of F0 peaks per sentence by the number of PWords in the sentence. A language with stronger MacR will have a MacR_Freq ratio value close to 1, meaning that each PWord has one F0 peak. This has been the most reliable measure for quantifying L/H regularity so far (e.g., Polyanskaya et al., 2019; Prechtel, 2020) because it refers directly to the PWord or AP domain.

The magnitude of F0 displacement is calculated by taking the average of the height difference between successive tonal targets within an utterance. A larger average displacement indicates two things. First, it suggests that there are more phonological L targets, which are expected to be realized with lower F0 targets than phonetically low tonal targets resulting from downstepping or "sagging" between H targets. Second, larger average displacement implies that successive pairs of tonal targets are more likely to alternate between L and H, as opposed to a plateau or downstepping. Therefore, this measure can indirectly capture the presence of L/H alternations, and it complements the MacR_Freq measure, which captures the regularity of F0 peaks per PWord but does not capture information about the tonal alternation itself.

The CLI measure is calculated by using Pythagoras' theorem ($a^2 + b^2 = c^2$) to get the F0 contour length in each phrase, where *a* represents the duration difference between consecutive F0 points, *b* represents the F0 height difference between the points, and *c* (the hypotenuse) represents the length of the contour, as shown in Figure 16 (reproduced from Figure 4 in Chapter 1). The difference between the interval length (sum of all *a*) and contour length (sum of all *c*) is expressed as a percentage of length increase. In other words, it captures how much longer the movement F0 is compared to a flat line.

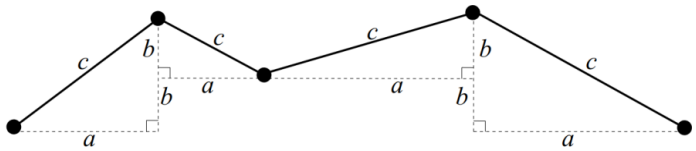

Figure 16: Example of a stylized f0 contour (solid line), where each dot represents an F0 point, and each line between points is the hypotenuse (c) of a right triangle (shown with dotted lines). From Kaland (2022:5235).

The formula is given in (7) (reproduced from (4) in Chapter 1), where L_C is the sum of all *c* and L_i is the sum of all *a* within the same utterance. The average contour length increase is calculated per IP to get the ratio of the average length increase within an utterance.

(7) Contour length increase (CLI) = $\frac{L_C}{L_i/100} - 100$

As mentioned in Chapter 1, CLI assumes that languages with stronger MacR will have more F0 alternations than languages with weaker MacR, and thus have a greater percentage increase in F0 length compared to the length of F0 with no perturbations.

3.2. Results

A total of 185 IPs were included for analysis (62 Bengali, 61 Uyghur, 62 English). IPs were excluded if they contained less than 3 Prosodic Words or contained disfluencies. Table 7 summarizes the average number of syllables and PWords per utterance for each language. On average, English had the most PWords per utterance (5.6), followed by Bengali (5.2), followed by Uyghur (5.0). In contrast, Uyghur had the largest average number of syllables per utterance (14.5), followed by Bengali (13.9), followed by English (13.6). Tables 8-10 summarize the number of IPs and mean number of PWords by speaker for each language.

	Language		
	English	*Uyghur*	*Bengali*
Mean number of syllables	13.6 (4.7)	14.5 (4.2)	13.9 (3.4)
Mean number of PWords	5.6 (1.1)	5.0 (1.1)	5.2 (1.1)

Table 7: Mean number of syllable and PWords per utterance in each language. Standard deviations are in parentheses.

Speaker	**Number of IPs**	**Mean number of PWords**
355010_en_27	7	6.0 (1.5)
355440_en_31	7	5.9 (1.3)
359757_en_39	8	5.8 (0.9)
359790_en_40	8	5.1 (0.8)
362322_en_41	8	5.6 (1.2)
381996_en_46	8	5.1 (1.0)
382515_en_52	8	5.8 (1.3)
383260_en_53	8	5.5 (1.2)
Average	*7.8 (0.5)*	*5.6 (1.1)*

Table 8: Number of IPs and mean number of PWords produced by each English speaker. Standard deviations are in parentheses.

Speaker	Number of IPs	Mean number of PWords
395319_uy_02	8	5.6 (0.9)
406628_uy_03	7	5.0 (1.0)
423723_uy_06	7	4.6 (0.8)
433461_uy_08	9	4.7 (0.5)
472570_uy_10	9	5.4 (1.4)
520256_uy_11	9	5.4 (1.3)
520300_uy_12	6	5.2 (1.0)
551530_uy_14	6	4.2 (0.4)
Average	*7.6 (1.3)*	*5.0 (1.1)*

Table 9: Number of IPs and mean number of PWords produced by each Uyghur speaker. Standard deviations are in parentheses.

Speaker	Number of IPs	Mean number of PWords
412389_be_02	7	5.3 (1.4)
424843_be_07	7	5.0 (1.0)
446946_be_08	8	5.3 (1.2)
449053_be_10	8	5.1 (1.0)
476993_be_18	8	5.0 (0.9)
487510_be_19	8	5.3 (1.6)
489487_be_20	8	5.4 (1.1)
491610_be_22	8	5.0 (1.1)
Average	*7.8 (0.5)*	*5.2 (1.1)*

Table 10: Number of IPs and mean number of PWords produced by each Bengali speaker. Standard deviations are in parentheses.

Tables 11-13 summarize the pitch range, F0 minimum, and F0 maximum values of all speakers in each language. Since the male speakers had smaller pitch ranges than the female speakers, the averages were calculated by gender.

Speaker	Gender	Pitch range (Hz)	F0 Min (Hz)	F0 Max (Hz)
355010_en_27	Female	182	88	270
362322_en_41	Female	229	85	314
381996_en_46	Female	110	215	325
382515_en_52	Female	157	163	320
383260_en_53	Female	176	92	268
Female Average		*170.8 (43.1)*	*128.6 (62.9)*	*299.4 (28.0)*
355440_en_31	Male	47	85	132
359757_en_39	Male	94	104	198
359790_en_40	Male	55	82	137
Male Average		*65.3 (25.1)*	*90.3 (11.9)*	*155.7 (36.7)*

Table 11: Average F0 range, minimum, and maximum of the English speakers. Standard deviations are in parentheses.

Speaker	Gender	Pitch range (Hz)	F0 Min (Hz)	F0 Max (Hz)
395319_uy_02	Female	190	159	349
406628_uy_03	Female	146	149	295
423723_uy_06	Female	167	171	338
433461_uy_08	Female	167	183	350
520300_uy_12	Female	120	171	291
Female Average		*158.0 (26.3)*	*166.6 (13.0)*	*324.6 (29.3)*
472570_uy_10	Male	48	94	142
520256_uy_11	Male	100	90	190
551530_uy_14	Male	112	90	202
Male Average		*86.7 (34.0)*	*91.3 (2.3)*	*178.0 (31.7)*

Table 12: Average F0 range, minimum, and maximum of the Uyghur speakers. Standard deviations are in parentheses.

Speaker	Gender	Pitch range (Hz)	F0 Min (Hz)	F0 Max (Hz)
412389_be_02	Female	209	119	328
446946_be_08	Female	355	143	498
449053_be_10	Female	155	133	288
487510_be_19	Female	139	158	297
491610_be_22	Female	261	125	386
Female Average		*223.8 (87.7)*	*135.6 (15.4)*	*359.4 (86.5)*
424843_be_07	Male	92	93	185
476993_be_18	Male	83	89	172
489487_be_20	Male	86	108	194
Male Average		*87.0 (4.6)*	*96.7 (10.0)*	*183.7 (11.1)*

Table 13: Average F0 range, minimum, and maximum of the Bengali speakers. Standard deviations are in parentheses.

The MacR_Freq Index was calculated for each utterance, and the distribution of the ratios is shown in Figure 17. As expected, Bengali had the largest MacR_Freq ratios, followed by Uyghur, and English had the smallest ratios, consistent with the predicted ranking. A linear mixed effects model was run with MacR_Freq ratio as the dependent variable, language as the predictor, and speaker and IP as random intercepts. The results found that Bengali had significantly larger ratios than Uyghur ($\beta = 0.14$, $t = 3.71$, $p = 0.004$) and English ($\beta = 0.27$, $t = 7.44$, $p < 0.001$), and a post-hoc pairwise comparison found that Uyghur had significantly larger ratios than English ($\beta = 0.13$, $t = 3.72$, $p = 0.001$), supporting the predicted ranking (Bengali > Uyghur > English).

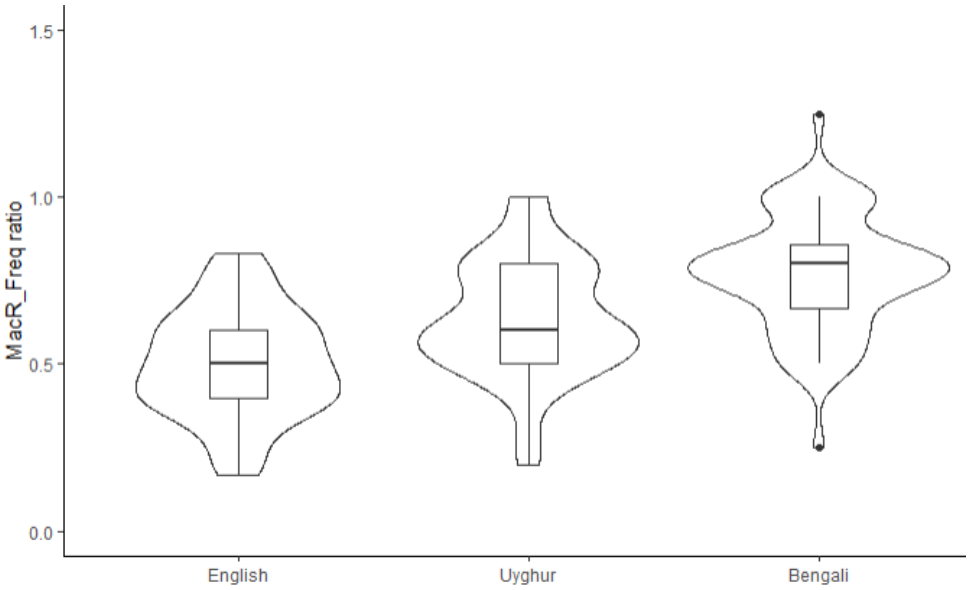

Figure 17: Distribution of MacR_Freq ratios in each language.

For the F0 analysis, the data were normalized to semitones to analyze male and female speakers together, using the following formula in (8), where f1 is a given F0 point and f2 is the mean F0 (Hz) of each speaker:

(8) F0 (semitone) = 12*(log2 (f1/f2))

Tables 14 summarizes the average overall F0 range (i.e., the difference between F0 minimum and maximum values per speaker) and F0 displacement (i.e., the difference in height between successive tonal targets within an utterance) in each language. Bengali speakers had the largest F0 range (14.6), followed by English (13.4), followed by Uyghur (11.5). As for the magnitude of F0 displacement between tonal targets, the languages behaved as expected: Bengali had the largest average F0 displacement (3.9), followed by Uyghur (3.1), followed by English (2.3). These results suggest that while the Uyghur speakers in this study had smaller pitch ranges on average than English speakers, the magnitude of F0 displacement between tonal targets was

greater, indicating that Uyghur had both more phonological L tonal targets and thus more consistent L/H alternations than English, consistent with the predicted MacR strength ranking. Bengali speakers had the largest average F0 displacement, indicating both the presence of phonological L targets and larger differences in F0 excursions between L and H targets than both Uyghur and English.

	Language		
	English	*Uyghur*	*Bengali*
Mean F0 range (semitones)	13.4 (1.5)	11.5 (1.4)	14.6 (2.0)
Mean F0 displacement (semitones)	2.3 (2.2)	3.1 (2.3)	3.9 (2.9)

Table 14: Summary of average F0 range and F0 displacement (i.e., the difference between successive tonal targets) for each language. Standard deviations are in parentheses.

Figure 18 shows the distribution of F0 displacement for each language. English utterances had more periods with little or no change in F0 (i.e., F0 "plateaus") than Uyghur and Bengali, as indicated by the larger concentration of values near or at 0 in the English distribution. In contrast, Bengali had the fewest F0 displacement values at or near 0. Both English and Bengali had more outliers than Uyghur, indicating that there were a few values that had a particularly large displacement between L and H tonal targets in these languages but not in Uyghur. A linear mixed effects model was run with F0 displacement as the dependent variable, language as the predictor, and speaker as the random intercept. The results found that that Bengali had a significantly larger magnitude of displacement than English ($\beta = 1.64$, $t = 3.84$, $p = 0.001$), but only marginally larger displacement than Uyghur ($\beta = 0.76$, $t = 1.78$, $p = 0.09$). Post-hoc pairwise comparisons found no significant difference between Bengali and Uyghur or between Uyghur and English. To summarize, the results of F0 displacement yielded the strength

ranking Bengali > English; Bengali ≈ Uyghur; Uyghur ≈ English. In other words, Bengali and English, the two languages on opposite ends of the predicted MacR strength spectrum, were significantly different from each other, but Uyghur was not different from either language.

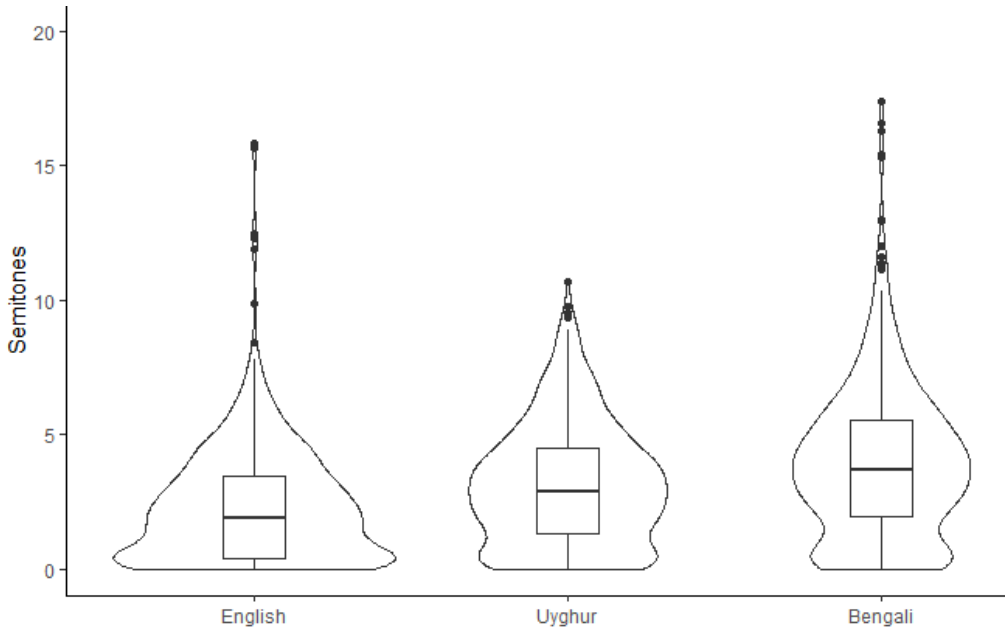

Figure 18: Average difference in F0 displacement (semitones) between successive tonal targets in each language.

The Contour Length Increase (CLI) was calculated for each utterance, and the distribution of the percentage increase is shown in Figure 19. English utterances had the largest proportion of length increase percentages near 0, while Bengali had the smallest proportion near 0, and the distribution of Uyghur percentages was between the two, although closer to English than Bengali. In addition, Bengali had the largest CLI with an average percent increase of 0.028 (standard deviation: 0.02), followed by Uyghur with an average percentage increase of 0.013 (0.01), and English with an average percentage increase of 0.009 (0.01). A linear mixed effects model was run with CLI as the dependent variable, language as the predictor, and speaker and IP as random intercepts. The results of the model found that Bengali had a significantly larger

percentage increase in contour length than Uyghur ($\beta = 0.02$, $t = 5.06$, $p = 0.0001$) and English ($\beta = 0.02$, $t = 6.36$, $p < 0.001$). Although Uyghur had a numerically larger mean value of CLI than English, post-hoc pairwise comparisons found no significant difference between these two languages. In summary, the results of the model yielded the strength ranking Bengali > Uyghur ≈ English.

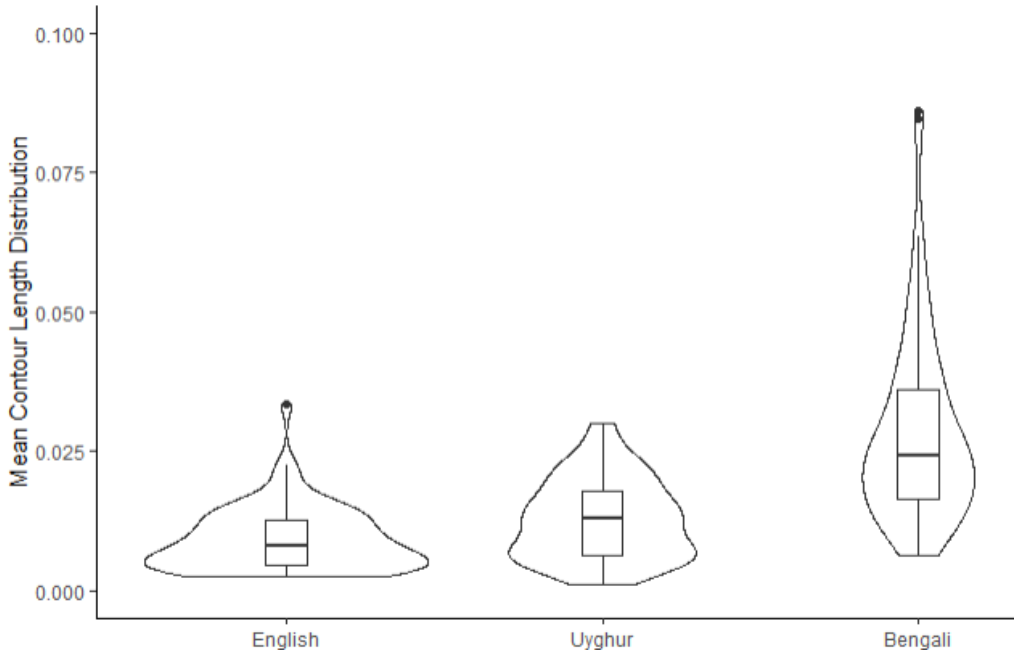

Figure 19: Distribution of the Contour Length Increase (in semitones) in each language.

3.3. Discussion

Overall, the results support the predicted MacR strength ranking (Bengali > Uyghur > English). As expected, the MacR_Freq Index showed that Bengali had the highest peak-to-PWord ratio, followed by Uyghur, followed by English. Regarding differences in F0 displacement across the three languages, Bengali had a significantly larger average F0 displacement between tonal targets than English, and was marginally larger than Uyghur, but there was no significant difference between Uyghur and English (Bengali > English; Bengali ≥ Uyghur; Uyghur ≈

English). As for CLI, Bengali had a larger length increase percentage than both Uyghur and English, but there was no significant difference between Uyghur and English (Bengali > Uyghur ≈ English).

Although Uyghur behaved as predicted in the temporal domain, that is, had a significantly larger average MacR_Freq Index than English but a smaller one than Bengali, it behaved differently in the frequency domain. The results of the F0 displacement measure found that Uyghur had only marginally smaller displacement between tonal targets than Bengali in the main model, but the effect went away in the pairwise comparison. Comparing all three languages, the difference appears to be gradient, in the order of Bengali > Uyghur > English, but only significantly differs between languages on the opposite ends of the strength spectrum (i.e., Bengali and English). That is, Uyghur is not significantly different from Bengali or English, but Bengali has significantly larger F0 displacement than English. Regarding the CLI measure, Uyghur behaved more similarly to English than Bengali. This is a surprising finding, considering the results of the MacR_Freq Index. However, since the CLI measure reflects the difference between the length of F0 excursions and the distance intervals between successive F0 points within an utterance, it is possible that the measure is sensitive to factors such as pitch range and declination, or that English utterances with multiple downstep sequences were similar enough in length compared to Uyghur L/H alternations with small pitch range that the differences between the two languages were collapsed.

Taken together, these results reflect a nuanced look at MacR strength across the three languages, and they support the predicted strength ranking. Bengali had the peak-to-PWord ratio closest to 1, the largest average displacement between tonal targets, and the longest CLI while English had the smallest peak-to-PWord ratio, smallest F0 displacement, and a smaller CLI than

Bengali. While the Uyghur results varied by measure, they also support its predicted intermediate strength ranking. Of the three types of measurements, the MacR_Freq ratio results best align with the predicted strength ranking, which is the measure that most closely captures MacR strength. As for the measures in the frequency domain, Uyghur did not have significantly different F0 displacement from the other two languages, although Figure 18 shows that the average distribution of F0 displacement is between English and Bengali, which further supports its intermediate status. Regarding contour length increase, Uyghur did not have a significantly larger CLI percentage than English, although Figure 19 shows that the distribution of CLI values in Uyghur is again between English and Bengali.

These results have some interesting implications for the perception of differences in MacR strength. Although the MacR_Freq Index measure captured the predicted difference in strength ranking, the extent to which listeners can detect small differences in the regularity of L/H alternations is unknown. Regarding F0 displacement, the magnitude of F0 excursions will certainly play an important role in the perceived salience of L/H alternations; that is, large contrasts are easier for listeners to hear than small ones. The next chapter will test the perception of MacR strength in each language and determine how well listeners can hear differences in the regularity and magnitude of F0 movement.

CHAPTER 4

Macro-Rhythm Perception Experiment

This chapter presents the details of the MacR perception experiment. The purpose of this experiment was to test whether differences in MacR strength across the three languages were perceptible to listeners. Based on the results of the MacR production experiment, Bengali was predicted to have the strongest perceived MacR, followed by Uyghur, followed by English (Bengali > Uyghur > English). To test this prediction, participants rated how rhythmic or "melodic" utterances in each language sounded. The MacR of the stimuli were also phonetically quantified to compare the participants' perception ratings with the phonetic realization of MacR.

4.1. Methods

4.1.1. Stimuli

The stimuli consisted of utterances taken from read speech corpora for each language. The Bengali utterances were taken from the SHRUTI Bangla Speech Corpus (Das, Mandal, & Mitra, 2011), which consists of sentences read from news articles taken from Bengali-language newspaper *Anandabazar Patrika*, and the topics include sports, politics, general news, and geographical news. A total of 34 speakers were recorded (26 male, 8 female), ranging from 20-40 years old, and all but 2 of them reported university-level education (2 undergraduate, 30 graduate). All speakers grew up in West Bengal and spoke Standard Bengali. The Uyghur utterances were taken from the THUYG-20 corpus (Rouzi et al., 2017), which consists of sentences read from novels, newspaper articles, and various types of books. A total of 348 speakers were recorded (163 male, 185 female), ranging from 19-28 years old. They came from 30 counties within Xinjiang Uyghur Autonomous Region, and they were all university students

at the time of recording. The English utterances were taken from the ST-AEDS-20180100_1 Free ST American Speech Corpus (Surfing Technology Ltd, 2018), which is a subset of a larger American English corpus by Surfingtech (surfingtech.ai). This subset contained a total of 10 speakers (5 male, 5 female), and while the corpus description did not provide details about speaker age or the source(s) of the read utterances, the sentences and phrases appear to be from transcripts of TED talks. The primary purpose of all three corpora was to develop and train automatic speech recognition systems.

The stimuli in this experiment were created from 30 utterances (10 per language) that were 13-20 syllables long and contained a single Intonational Phrase (IP). All utterances were produced by female speakers, and multiple speakers were included for each language. Tables 15-17 show the details about each speaker. The number of speakers included in each language (2 English, 6 Uyghur, 4 Bengali) was uneven across languages because the total number of utterances per speaker differed across corpora, and the criteria for inclusion in this experiment were based on utterance duration, syllable number, and number of IPs, summarized in Table 15. See Appendix C for the utterances used for each language.

Speaker	**Pitch range (Hz)**	**F0 Min (Hz)**	**F0 Max (Hz)**
f0002	198	169	285
f0003	263	78	341
Average	*189.5 (103.9)*	*123.5 (64.3)*	*313 (39.6)*

Table 15: Average F0 range, minimum, and maximum of the English speakers included in the stimuli. Standard deviations are in parentheses.

Speaker	Pitch range (Hz)	F0 Min (Hz)	F0 Max (Hz)
F011	128	200	328
F016	111	207	318
F023	94	178	272
F053	210	104	314
F060	84	236	320
F064	135	199	334
Average	*112.3 (20.0)*	*202 (19.3)*	*314.3 (22.0)*

Table 16: Average F0 range, minimum, and maximum of the Uyghur speakers included in the stimuli. Standard deviations are in parentheses.

Speaker	Pitch range (Hz)	F0 Min (Hz)	F0 Max (Hz)
msm	273	158	431
punam	242	81	323
ritwika	205	218	423
suranjana	137	204	341
Average	*214.3 (58.5)*	*165.3 (61.7)*	*379.5 (55.4)*

Table 17: Average F0 range, minimum, and maximum of the Bengali speakers included in the stimuli. Standard deviations are in parentheses.

Table 18 summarizes the average number of syllables and PWords per utterance for each language. English had the most syllables and largest number of PWords per utterance while Bengali and Uyghur were more comparable to each other, having a similar number of syllables and PWords per utterance.

	Language		
	English	*Uyghur*	*Bengali*
Mean number of syllables	18.4	17.5	17.3
Mean number of PWords	7	6.1	6.2
Mean utterance duration (ms)	3.7	4.0	4.3

Table 18: Mean number of syllables and PWords per utterance in each language.

The utterances were manipulated and presented in two conditions: Filtered and F0-only. In the Filtered condition, which was presented first, the voiceless segments were extracted using the Praat Vocal Toolkit Plugin (Corretge, 2020), and the utterances were low-pass filtered to 500 Hz to remove segmental information, but they still retained some information about syllable structure. That is, information about syllables could still be heard because of the silent portions from the extracted voiceless segments. Finally, the stimuli were pitch-normalized using the Praat Vocal Toolkit to have a median F0 of 240 Hz, which is the median of minimum and maximum values of each file for all languages. This was done to minimize differences between speakers and languages. In the F0-only condition, the filtered utterances were resynthesized as a continuous hum, retaining only the F0 of the original utterance. To do this, the author recorded a continuous hum, which was normalized to match the duration and intensity of the filtered stimuli, and then the pitch tier of the hum was replaced by the pitch tier of each filtered utterance.

Within each condition, participants heard each utterance twice, or 60 utterances per condition (30 sentences x 2), and a total of 120 utterances over the entire experiment. In both conditions, the stimuli were duration-normalized to 2960 ms, which was between the mean and median of the duration of all sound files, and then intensity-normalized to 75 dB.

4.1.2. Participants

Participants were recruited through the UCLA psychology subject pool and were compensated with course credit. All participants were native speakers of American English. Before starting the experiment, participants filled out a background questionnaire, which asked basic demographic information such as age, gender, birthplace (city or general region), places lived for more than 12 months, language background, and formal language education. They were also asked about their music background, specifically whether they had any formal music training (e.g., played an instrument or sang in a choir). This question was included because previous studies have found that listeners with musical training can identify and discriminate pitch contours more accurately than non-musicians (e.g., Alexander, Bradlow, & Wong, 2005; Wayland, Herrera, & Kaan, 2010; Chen, Zhu, Wayland & Yang, 2020), and listeners with strong musical rhythm perception are better at rhythmically grouping speech (Boll-Avetisyan, Bhatara, & Höhle, 2017). Previous pilot experiments on MacR perception suggested that participants' rating responses differed depending on their musical background. In addition, Prechtel (2022b) found that participants who reported formal music training perceived Bengali utterances as significantly more macro-rhythmic than English and Uyghur in both the Filtered and F0-only conditions, while participants who reported no music background only rated Bengali as marginally more macro-rhythmic than Uyghur in the Filtered condition and marginally more macro-rhythmic than English in the F0-only condition. However, the analysis did not account for speaker variability in the stimuli, so the results from the pilot experiment need to be confirmed.

55 participants completed the experiment, but 8 were excluded for being non-native English speakers or having spent a significant portion of their life living outside of the U.S., so a total of 47 participants were analyzed (41 female, 6 male). The average participant age was 21.2

years old. 20 participants were monolingual (27 were bilingual from birth), and 22 reported having a background in music while 25 reported no formal music background.

4.1.3. Procedure

As with the production experiments, the perception experiment was conducted online through *LabVanced* (Finger et al., 2017). In each trial, participants were instructed to wear headphones and rate how "melodic" each utterance sounded on a 1-5 scale, where 1 was defined as "not melodic" and 5 defined as "very melodic," as shown in Figure 20. Before starting the experiment trials, participants first listened to an example of a "more melodic" utterance and a "less melodic" utterance, and then completed a few practice trials to familiarize them with the task. Based on feedback from the pilot experiment, participants took a break every 19 trials to mitigate listener fatigue. The whole experiment took about 25 minutes to complete.

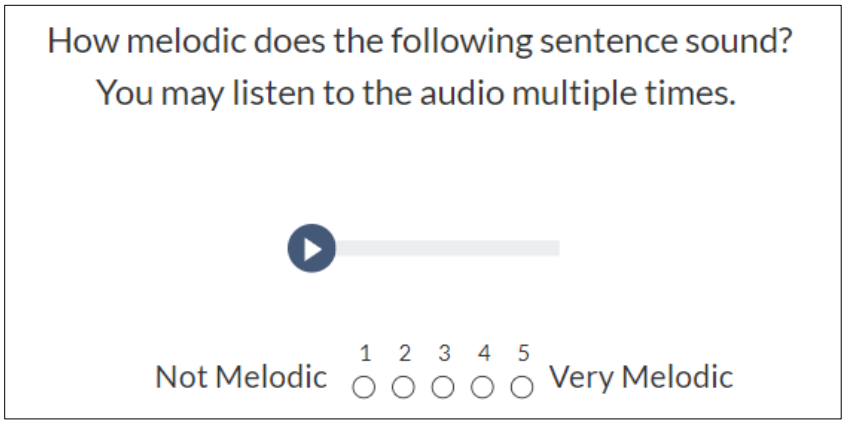

Figure 20: Example of an individual trial.

4.1.4. MacR Quantification of the Stimuli

In addition to testing perception, MacR quantification of the raw stimuli was done to directly compare the perception and production of MacR strength. The same methodology and quantification measures used in Chapter 3 were included here. The data were transformed into

semitones and the following MacR quantification measures were calculated: MacR_Freq Index, magnitude of F0 displacement, and Contour Length Increase (CLI).

4.2. Perception Results

The results were analyzed within each condition. Linear mixed effects models were run with rating response as the dependent variable, language and music background as predictors, and participant and speaker as random intercepts.

4.2.1. Filtered Condition

The model found no main effect of music background on rating responses. Bengali was rated significantly more melodic than both English ($\beta = 0.58$, $t = 2.68$, $p = 0.03$) and Uyghur ($\beta = 0.46$, $t = 2.77$, $p = 0.02$), and post-hoc pairwise comparisons found no significant difference between English and Uyghur ratings. The lack of significance is reflected in Figure 21, which shows that the ratings for both languages were very similar, although Uyghur utterances had slightly more 4 and 5 responses and slightly fewer 1 responses than English utterances. In contrast, Bengali utterances had the largest proportion of 4 and 5 responses and the fewest 1 and 2 responses. Therefore, the strength ranking in this condition is Bengali > Uyghur ≈ English.

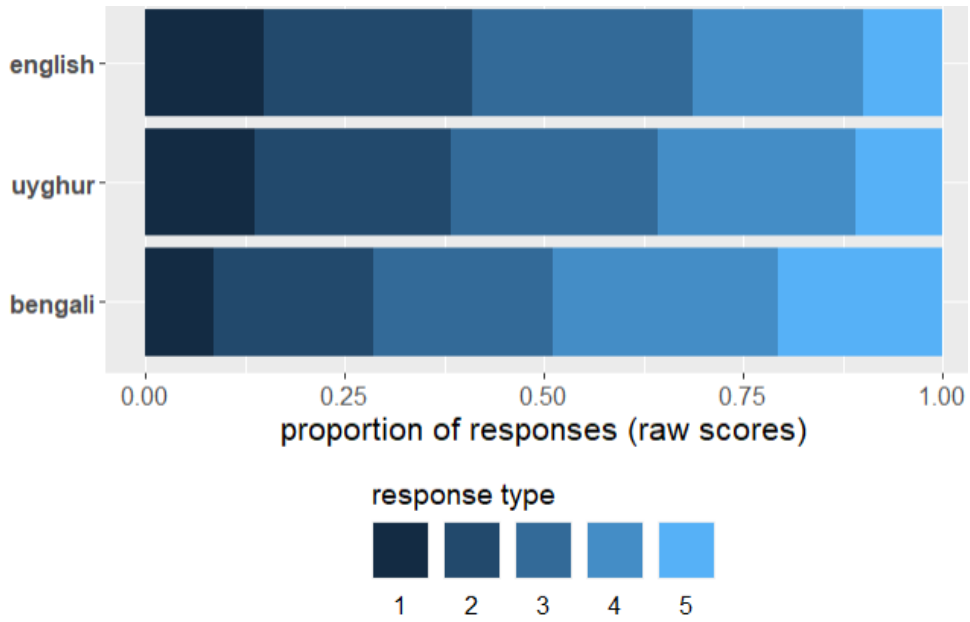

Figure 21: Proportion of rating responses for each language in the Filtered condition. 1 is "not melodic" and 5 is "very melodic."

4.2.2. F0-Only Condition

The results of the F0-only condition differed from the Filtered condition, although there was still no main effect of music background. Bengali was rated as significantly more melodic than Uyghur ($\beta = 0.38$, $t = 2.37$, $p = 0.04$), but only marginally more melodic than English ($\beta = 0.47$, $t = 2.23$, $p = 0.06$). Post-hoc pairwise comparisons found no significant differences between any language pairs, suggesting that participants may have found this condition more difficult. Curiously, Figure 22 shows that the rating responses for each language generally reflect the predicted ranking more clearly than in Figure 21. This apparent discrepancy between the model and the graph is probably due to large variability in pitch range between English speakers. Despite the marginal significance in the model, Bengali utterances appear to have a larger proportion of 4 and 5 responses and fewer 1 and 2 responses compared to English utterances. Similarly, Uyghur utterances appear to have slightly more 4 and 5 responses and slightly fewer 1 responses than English utterances. Based on the results of the model, the strength ranking in this

condition is Bengali (>) Uyghur ≈ English. In other words, Bengali had stronger perceived MacR than Uyghur, but only marginally stronger perceived MacR than English, and there was no significant difference between Uyghur and English.

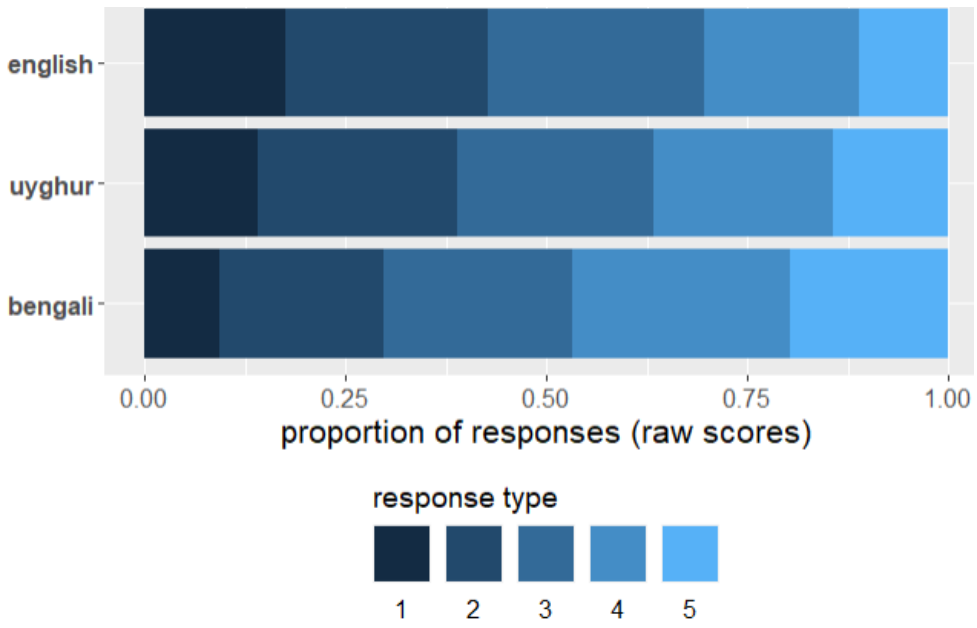

Figure 22: Proportion of rating responses for each language in the F0-only condition.

4.3. MacR Quantification Results

MacR was quantified over the original raw utterances before they were manipulated and resynthesized for the Filtered and F0-only conditions. The MacR_Freq Index was calculated for each utterance, and the distribution of the peak-to-PWord ratios is shown in Figure 23. As expected, English utterances tended to have the smallest MacR_Freq ratios of the three languages, although a few utterances had strong MacR. In contrast, both Uyghur and Bengali utterances had MacR_Freq ratios close to 1. Due to the small number of data points, a simple linear regression was run to test if language group significantly predicted MacR_Freq ratios. The overall regression was statistically significant ($R^2 = 0.34$, $F(2,27) = 6.98$, $p = 0.004$), and the

model found that Bengali had significantly larger ratios than English (t(27)=3.03, p=0.005), but there was no difference between Bengali and Uyghur. Post-hoc pairwise comparisons found that Uyghur had significantly larger ratios than English (t(27)=3.41, p=0.006), which reflect the distributions in Figure 23. The results suggest that Bengali and Uyghur have more consistent peaks per PWord compared to English, as expected, but the ratios were comparable to each other. Thus, the strength ranking based on this measure is Bengali ≈ Uyghur > English.

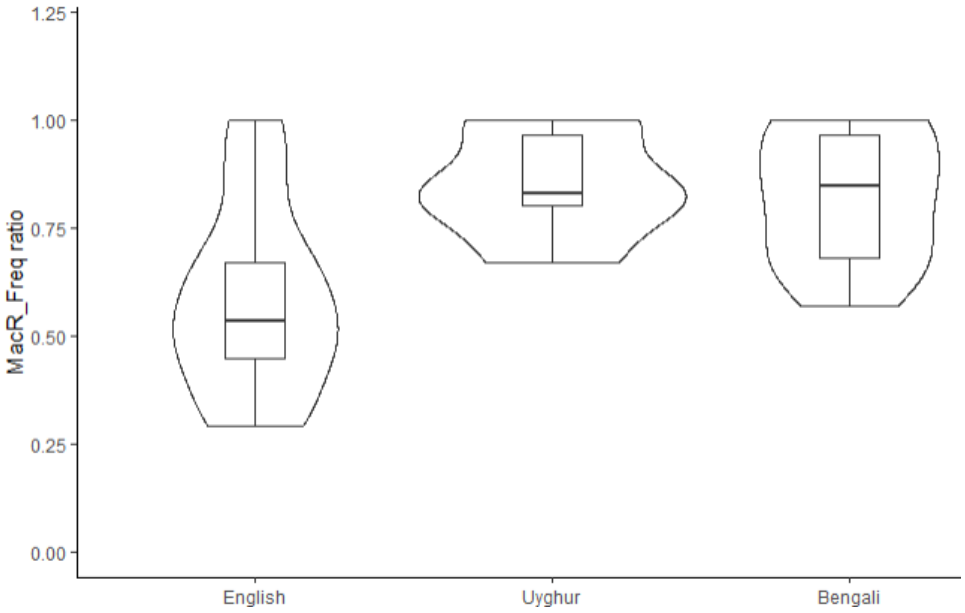

Figure 23: Distribution of the MacR_Freq ratios in each language.

Table 19 summarizes the average pitch range and F0 displacement between successive tonal targets. Uyghur utterances had the smallest average pitch range (7.7) while Bengali utterances had the largest (12.4), and the English pitch range was closer to Bengali than Uyghur. Regarding the average F0 displacement between tonal targets, English had the smallest difference (2.5) while Bengali had the largest difference (4.0). These differences are also reflected in the distributions of average values in Figure 24.

	Language		
	English	*Uyghur*	*Bengali*
Mean F0 range (semitones)	11.3 (3.3)	7.7 (1.3)	12.4 (3.6)
Mean F0 displacement (semitones)	2.5 (1.8)	2.7 (1.7)	4.0 (2.4)

Table 19: Summary of average F0 range and F0 displacement between successive tonal targets in each language. Standard deviations are in parentheses.

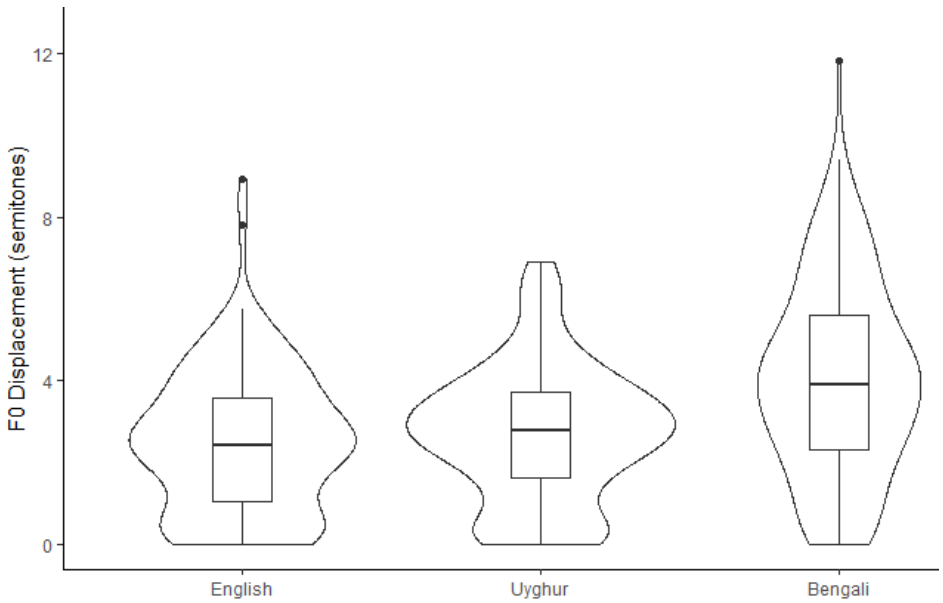

Figure 24: Average magnitude of F0 displacement (semitones) between successive tonal targets in each language.

A linear mixed effects model was run with F0 displacement as dependent variable, language as predictor and speaker as random intercept. Utterance was not included as a random intercept due to the small number of observations. The results found that Bengali had significantly larger F0 displacement than Uyghur ($\beta = 1.47$, $t = 3.36$, $p = 0.009$) and English ($\beta = 1.69$, $t = 3.06$, $p = 0.02$), and post-hoc pairwise comparisons found no significant difference

between Uyghur and English. Therefore, in terms of F0 displacement, the strength ranking is Bengali > Uyghur ≈ English, which matches the perception results in the Filtered condition.

The Contour Length Increase (CLI) measure yielded similar results to F0 displacement. A linear mixed effects model was run with the average CLI as the dependent variable, language as the predictor, and speaker as the random intercept. The results found that Bengali had a significantly larger average contour length increase than Uyghur ($\beta = 0.02$, $t = 4.14$, $p = 0.003$) and English ($\beta = 0.02$, $t = 3.98$, $p = 0.01$), but post-hoc pairwise comparisons found no difference between English and Uyghur. These results are reflected in Figure 25. In terms of CLI, the strength ranking is Bengali > Uyghur ≈ English, which matches both the F0 displacement results and the perception results in the Filtered condition.

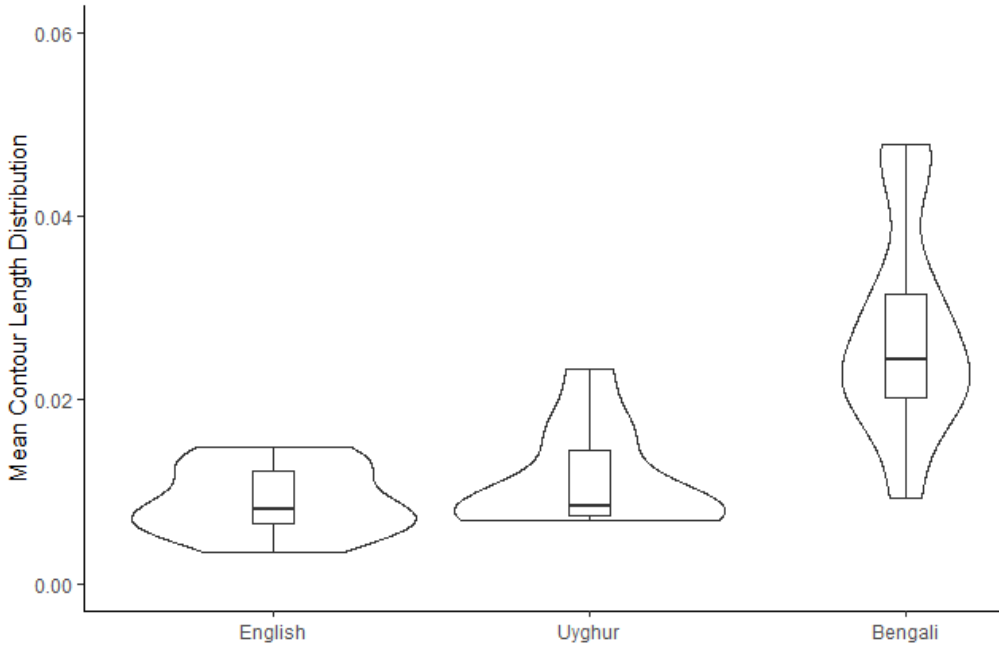

Figure 25: Distribution of the Contour Length Increase (in semitones) in each language.

4.4. Discussion

The results of the perception experiment suggest partial support for the perceptibility of the predicted MacR strength ranking. Bengali was rated significantly more melodic than English and Uyghur in the Filtered condition, and significantly more melodic than Uyghur but only marginally more melodic than English in the F0-only condition. However, there was no significant difference between English and Uyghur ratings in either condition. This lack of difference aligns with the results of the F0 displacement and CLI quantification measures of the stimuli, suggesting that F0 movement and pitch range informed listener perception more than the regularity of L/H alternations across PWords.

Indeed, within the temporal domain of MacR quantification, the MacR_Freq Index results found that both Bengali and Uyghur utterances had peak-to-PWord ratios significantly closer to 1 compared to English, suggesting that the Uyghur utterances had stronger MacR than English as predicted. Additionally, Uyghur and Bengali utterances did not have significantly different ratios, although this is not surprising given that both languages have tonally marked AP units. However, despite the similarity in MacR strength between Bengali and Uyghur and despite the difference between Uyghur and English, listeners did not seem to perceive this difference in either condition. It is likely that the experimental design and the nature of the stimuli affected listeners' perception of tonal rhythm, especially in the F0-only condition where listeners had no other prosodic cues besides F0 to suggest PWord boundaries. Since the stimuli purposely obfuscated word boundaries within the utterances, the perceptibility of MacR strength differences may have been dampened. Therefore, although Uyghur utterances were not rated as significantly more macro-rhythmic than English utterances, the results and the numerical

differences do not contradict the predicted strength ranking. In other words, there is no evidence to suggest that English tended to be perceived as more macro-rhythmic than Uyghur.

Within the frequency domain of MacR quantification, Bengali utterances had a larger magnitude of F0 displacement between tonal targets than both Uyghur and English, but there was no significant difference between the latter two languages (Bengali > Uyghur ≈ English), which is consistent with the perception experiment results. The larger F0 displacement in Bengali presumably made the L/H alternations more salient and "melodic" than the other two languages to listeners. The lack of significance in F0 displacement between Uyghur and English is a surprising result because Uyghur marks APs with a rising tonal pattern like Bengali while English marks H tonal targets with frequent downstepping and fewer phonological L targets. Since the MacR_Freq Index results confirmed that Uyghur had significantly larger peak-to-PWord ratios than English, it is likely that there are other factors such as pitch range that affected the results of both the F0 displacement and CLI measures, which yielded the same ranking of Bengali > Uyghur ≈ English. These results are similar to the ones in the MacR production experiment in Chapter 3. In both F0 displacement and CLI, Uyghur was not significantly different from English.

In both the production and perception of the stimuli, Uyghur behaved differently from the predicted language ranking in the magnitude and salience of F0 displacement and overall movement. While the numerical average of F0 displacement was higher in Uyghur than English (Table 16), there was no significant difference in the perception of macro-rhythm between the two languages. Notably, Uyghur had the smallest mean F0 range of the three languages (Table 16), which may have made Uyghur sound more "monotonous" compared to both Bengali and English. The greater F0 range in the English stimuli compared to the Uyghur stimuli also

suggests that English had more noticeable differences between F0 peaks and valleys across the utterance. Even though English had less frequent alternation of peaks and valleys than Uyghur, listeners could have perceived the larger overall pitch differences within English utterances as being more melodic than Uyghur utterances. Therefore, the perception results may not fully reflect the MacR strength of each language.

Taken together, the results of the perception experiment and the MacR quantification of the stimuli demonstrate that the magnitude of F0 displacement and overall increase of contour length may have been a more salient cue for rating how "melodic" an utterance was than the regularity of L/H alternations. In other words, bigger differences in F0 movement were more noticeable than the regularity of L/H alternations. Regardless of musical background, participants were especially attuned to the larger pitch differences between Bengali utterances and the utterances of the other two languages in both conditions. They may have paid particular attention to F0 magnitude because the instructions asked participants to rate the "melody" of the sentence rather than its "tonal rhythm." An early pilot of this experiment had explicitly instructed participants to listen for tonal rhythm (i.e., patterns of L/H alternations) and rate how rhythmic the alternations sounded. However, participant feedback indicated that these instructions were confusing, especially for those with formal music training, because listeners tended to associate 'rhythm' with regular alternations of 'beats' (e.g., amplitude) rather than pitch. Therefore, while the perception experiment results do support the prediction that differences in MacR strength are perceptible to listeners, they also reflect some of the limitations of the study, and they may not fully reflect perceptible differences in MacR strength across languages.

CHAPTER 5

General Discussion and Conclusion

5.1. Summary of Results

The results of all three experiments plus the MacR quantification of the perception stimuli are summarized in Table 20, which compares the predicted strength ranking in each experiment to the actual strength ranking results for each condition and individual measure.

Experiment	Predicted Ranking	Actual Ranking
Lexical Stress Production	English > Uyghur > Bengali	English > Bengali > Uyghur
MacR Production	Bengali > Uyghur > English	**MacR_Freq:** Bengali > Uyghur > English **F0 displacement:** Bengali > English; Uyghur **CLI:** Bengali > Uyghur ≈ English
MacR Perception	Bengali > Uyghur > English	**Filtered:** Bengali > Uyghur ≈ English **F0-Only:** Bengali (>) Uyghur ≈ English
MacR Perception Stimuli Quantification	Bengali > Uyghur > English	**MacR_Freq:** Bengali ≈ Uyghur > English **F0 displacement:** Bengali > Uyghur ≈ English **CLI:** Bengali > Uyghur ≈ English

Table 20: Summary of predicted vs actual strength ranking across the three languages. (>) represents marginal significance between languages.

In the lexical stress production experiment, English had the largest vowel duration differences between the first (stressed) and second (unstressed) syllables and therefore the strongest stress, as predicted. However, Bengali had larger duration differences than Uyghur, contrary to the predicted ranking. Since all the Bengali participants acquired English at a young age, it is possible that their realization of stress on nonce words in Bengali was influenced by English, which contrasts with the Uyghur participants who acquired English as adults. However, as the MacR production results show, the potential influence of English did

not seem to affect Bengali speakers' phrasal prosody, further suggesting that the use of nonce words affected stress realization.

In the MacR production experiment, each quantification measure yielded a slightly different result, but the overall pattern supported the predicted ranking, with Bengali on the stronger end of the MacR spectrum and English on the weaker end. The results of the MacR_Freq Index match the predicted ranking, which is consistent with the previous studies quantifying MacR strength using this measure (Prechtel, 2019, 2020, 2021) or an equivalent measure (Polyanskaya et al., 2019). The other macro-rhythm measures (F0 displacement and CLI) also showed the ranking of the languages in the predicted direction, though the difference between languages was not always significant. Of these three measures, the MacR_Freq Index is the measure that most closely captures macro-rhythm because it accounts for the regularity of F0 alternations within the PWord/AP domain. The other quantification measures included in this study do not capture the temporal regularity of L/H alternations. Although F0 displacement was intended to reflect the presence and variability of L/H alternations and was informed by the intonational phonology of the languages in question (Polyanskaya et al., 2019), it is a strictly phonetic measure that captures phonetic differences in F0 alternations that do not necessarily reflect the presence of phonological L targets. Similarly, CLI is a phonetic measure with an additional level of abstraction: it treats F0 as a metaphorical length of string, and the total amount of F0 movement is calculated as the total percentage increase of the length of a string compared to a hypothetical flat F0 contour with no movement and thus no increase in length. While the results of Kaland's (2022) experiment support the predicted MacR strength ranking between Greek, German, and European Portuguese, CLI does not necessarily capture the regularity or rhythmicity of

F0 movement. By only measuring the increase in contour length with no reference to some unit of regularity such as a PWord interval, it cannot directly capture MacR strength. In addition, this measure appears to be influenced by pitch range. That is, English speakers produced higher F0 maxima and lower F0 minima than Uyghur, but the overall F0 contours of Uyghur speakers had more L/H alternations than English. The current CLI measure therefore cannot differentiate between contour length increase due to consistent L/H alternations and length increase due to large differences in F0 minima and maxima.

As for the MacR perception experiment, Bengali was rated as more melodic than Uyghur and English in both the Filtered and F0-Only conditions, although the difference was only marginally significant between Bengali and English in the F0-Only condition. In both conditions, Uyghur and English were not rated significantly differently from one another. However, these perception ratings only partially aligned with the results of the MacR quantification of the stimuli. Bengali and Uyghur had a significantly larger MacR_Freq Index ratios, and thus a peak-to-PWord ratio closer to 1, than English, indicating that the English utterances had weaker MacR than both Bengali and Uyghur utterances. In contrast, both the magnitude of F0 displacement and the CLI measure results found that Bengali had significantly larger F0 excursions and larger contour length increase than Uyghur and English, but there was no difference between Uyghur and English, consistent with the perception experiment results. As discussed in Chapter 4, the perception results probably aligned more closely with the quantification measures capturing F0 movement because the stimuli obfuscated and excluded information about word boundaries. In other words, listeners probably paid particular attention to factors such as pitch range and magnitude of F0 displacement to inform their judgements.

The perception results showed that the melodicity ratings were not significantly influenced by the participants' music background, which differs from the results of previous studies on musicality and prosodic perception (e.g., Boll-Avetisyan, Bhatara, & Hoehle, 2017; Gregori & Kügler, 2021). Compared to the previous studies, the background questionnaire used in the current experiment did not capture detailed information about music experience. For example, we do not know how much the listeners with formal music training varied in level of experience and musical ability. Studies have shown that there is individual variation in both rhythm perception (e.g., McAuley et al., 2006; Iversen & Patel, 2008; Grahn & McAuley, 2009) and in pitch discrimination (e.g., Morrill, McAuley, Dilley & Hambrick, 2015). In any case, the current results demonstrate that the differences in melodicity between languages are perceptible to listeners, and the degree of music experience that individual participants had did not have a significant effect.

Together, these results support the predicted inverse relationship: English had stronger lexical stress than both Bengali and Uyghur, and Bengali had stronger MacR than English in both production and perception. Uyghur behaved somewhere in the middle and the results were more variable. In the MacR production experiment, the MacR_Freq Index for Uyghur was significantly larger than English and smaller than Bengali, and in the MacR quantification of the perception stimuli, the MacR_Freq Index for Uyghur was significantly larger than English but there was no difference from Bengali. While the results of the F0 displacement and CLI measures found no significant difference between Uyghur and English, these measures did not capture the temporal regularity of L/H alternations. Indeed, even though the differences were not significant, Uyghur had larger numerical differences in F0 displacement and larger CLI percentages, suggesting that Uyghur tended to behave as

predicted. In other words, there is no evidence to contradict the predicted MacR strength ranking. The lack of significant differences between Uyghur and English in MacR perception was surprising given that Uyghur and Bengali have similar phrasal prosody patterns; that is, they both use F0 to mark the edge of AP units. However, the lack of significance seems to be related to F0 displacement and contour length rather than the peak-to-PWord relationship. Uyghur utterances had a smaller average pitch range than Bengali and English in both the production experiment and the perception stimuli, and this probably contributed to the perception results. Again, these results do not contradict the predicted MacR strength ranking. Therefore, while the results do not completely align with the predicted strength ranking across languages, they do support the general inverse relationship between lexical stress strength and MacR strength. That is, English, a Head-prominent language, had stronger phonetic realization of lexical stress than the two Head/Edge-prominent languages, and both Uyghur and Bengali had consistently stronger MacR than English in the temporal domain (i.e., MacR_Freq Index). Furthermore, differences in MacR strength are perceptible to a certain degree, especially between Bengali and English, the two languages on opposite ends of the MacR strength spectrum in this study.

5.2. Study Limitations

While the results of these experiments provide support for the predicted inverse relationship between lexical stress and MacR strength, they are also constrained by the limitations of experimental design and the nature of the stimuli. In the lexical stress production experiment, the target stimuli were nonce words, and no real words were included for comparison. As discussed in Chapter 2, the exclusion of real words was based on the results of a pilot experiment (Prechtel, 2021) that included both real and nonce words and found no significant

difference between productions. However, the pilot only included one participant per language, so it is possible that differences would have emerged with more speakers. The more English-like production of nonce words by Bengali speakers may have been facilitated by the context in which they were introduced. Indeed, the experiment was designed to introduce the target non-words as though they were real words that the speaker was previously unfamiliar with, which implicitly framed them as borrowings from a different language. Since the Bengali participants acquired English at a young age and frequently use English in their personal and/or professional lives, their lexical stress production may be more influenced by English when presented with novel or unfamiliar words. As for the Uyghur speakers, despite explicit instruction to put the stress on the first syllable, speakers frequently had small durational differences between the first and second vowel. Most of the tokens that were excluded from analysis were due to greater duration on the second syllable, suggesting that the Uyghur speakers are more likely to put stress on the second syllable. Given that there are relatively few stress-based minimal pairs in Uyghur, the default location of lexical stress is the word-final syllable, and stress location is weight-sensitive, this result is perhaps not as surprising as the result for Bengali.

Testing the perception of tonal rhythm strength was a very challenging task, and the experimental design imposed multiple challenges and limitations. To determine the contribution of alternating F0 to the perception of speech rhythm, the stimuli were stripped of all other acoustic cues that could affect rhythm perception. However, using this type of stimuli to test MacR strength presents a few issues. First, rhythm perception seems strongly tied to the “beat” rather than to pitch movement, at least for L1 English speakers. A pilot version of the perception experiment was run in which participants were asked to give

feedback upon completion. Multiple participants with formal music training reported that being instructed to associate pitch movement with rhythm was confusing to them because their intuition and musical training had taught them to associate pitch with melody and rhythm with a metrical beat. Therefore, listeners reported varying levels of difficulty imposing a beat over the patterns of F0 alternations. Based on this feedback, the instructions were changed to ask participants to rate how "melodic" the sentence was. However, although participants listened to example stimuli that had stronger and weaker MacR, their judgements based on "melody" may have focused on other aspects of pitch movement in addition to the regularity of F0 alternations, such as the magnitude of F0 displacement and overall pitch range of the utterance. This relates to the second issue, which is that the proposed function of MacR is to facilitate word segmentation (Jun, 2014), but listeners in the experiment were presented with stimuli without words or any other phonetic or segmental information. With nothing to "anchor" the F0 alternations to specific intervals on a metrical grid, it was left up to the listeners to impose rhythm on the F0 movement. While rating the melodicity of utterances stripped of segmental context informs our understanding of the salience and perceptibility of cross-linguistic differences in F0 alternations, this shifts the question from "do listeners perceive differences in tonal rhythm strength across languages?" to "do listeners perceive differences in pitch movement across languages?" One way to potentially offset these issues is to run an experiment that tests the perception of tonal alternations using longer utterances or even short paragraphs that retain some segmental information so that listeners have more input upon which to base their judgements. Alternatively, tonal rhythm perception could be tested using an artificial language with varying degrees of MacR strength that reflect differences in real languages.

Another limitation of the perception experiment was that all the participants were L1 English speakers. Since previous studies have shown that native language influences the perception of prosodic cues and their use in word segmentation (e.g., Bhatara et al., 2013; Ordin & Nespor, 2013, 2016; Molnar, Carreiras, & Gervain, 2016), future work on MacR perception should expand upon this experiment by including participants who natively speak a strong MacR language, such as Bengali, and compare their ratings to the English listeners' ratings. If the function of MacR is to facilitate word segmentation, and languages with stronger MacR have relatively consistent PWord-sized L/H alternations, then one might predict that L1 listeners of a strong MacR language will be more sensitive to L/H tonal alternations. On the other hand, if the perception task does not involve word segmentation, where L/H alternations would provide cues to word boundaries, participants with a strong MacR L1 may perform similarly to English participants.

Finally, the potential dialectal differences between Uyghur speakers may have affected the results in both production and perception. Due to the nature of the data collection, the speakers who participated in the MacR production experiment were originally from various regions within Xinjiang Uyghur Autonomous Region (XUAR). While there are some known prosodic differences between Uyghur varieties spoken in Kazakhstan and China (Major & Mayer, 2019), there is not much research on prosodic differences between the dialects spoken within XUAR. However, based on anecdotes from multiple Uyghur speakers who grew up in China, different dialect regions have perceptibly different intonation patterns, and some of the speakers in the production experiments came from distinct dialect regions. As for the perception experiment, precise information about region/dialect for individual speakers was not included in the THUYG-20 corpus from which the stimuli originated.

Therefore, it is possible that dialectal differences may have affected the results, especially if some dialects have weaker MacR than others.

5.3. Implications for MacR, Prosodic Typology, and Speech Rhythm

Despite the limitations of the experiments, the results are consistent with or otherwise complement previous literature on MacR and sentence melody perception. The perception experiment was the first of its kind to directly compare perceived MacR strength across languages, and the findings that Bengali was rated significantly more melodic than English (and Uyghur) aligns with the findings of previous MacR production studies.

The results of the lexical stress production experiment were partially consistent with previous studies. English had the largest differences between stressed and unstressed syllables, which corroborates the findings of Mairano, Santiago, & Romano's (2015) cross-linguistic comparison of lexical stress strength between English, German, French, Spanish, and Italian. However, Bengali had the next largest duration differences between stressed and unstressed syllables, which is contrary to previous literature on Bengali stress (e.g., Hayes & Lahiri, 1991).

As for the MacR production experiment, the results of the magnitude of F0 displacement and the MacR_Freq Index were consistent with the results of Polyanskaya et al. (2019), who found that Italian had more frequent L/H alternations and larger magnitude of F0 excursions than English, supporting the predicted strength ranking. The MacR_Freq Index results were also consistent with Prechtel (2019, 2020), who found that Spanish had more consistent peak-to-PWord ratios than English, supporting the predicted strength ranking. However, the results of CLI in this study differed from Kaland's (2022) study, which found that the CLI measure supported the predicted MacR strength ranking between Greek,

German, and European Portuguese (i.e., Greek > German > European Portuguese). In contrast, the results of both the MacR production experiment and the quantification of the perception stimuli found that although Bengali had larger CLI than English and Uyghur, there was no significant difference between Uyghur and English. As previously discussed, the lack of significance could be due to differences in average pitch range between the language groups, and it is likely that CLI collapses the distinction between F0 displacement due to pitch range and displacement due to L/H alternations. Since it only captures how long the contour length of the whole utterance is relative to a straight line, a large contour length increase cannot differentiate between an utterance with large F0 displacement for a few words and an utterance with smaller but consistent F0 displacement for every word. Therefore, CLI cannot capture the regularity of L/H alternations, nor does it reflect the link between L/H alternations and the PWord. This is the primary weakness of CLI as a measure of MacR strength, and it should be revised to calculate the contour length increase per PWord, which could capture the regularity of L/H alternations more directly. Given this issue, any future studies on MacR strength should not use CLI as the sole quantification measure but rather use it in conjunction with the MacR_Freq Index, which does capture the regularity of L/H alternations.

This study contributes to our understanding of both the prosodic typology and speech rhythm in the following ways. First, it is the first study of its kind to directly test the hypothesized inverse relationship between lexical prominence and phrasal prominence (Jun, 2014:537), and the results support the predicted strength ranking. Specifically, the results of the production experiments found that Bengali and Uyghur (the two Head/Edge-prominence languages) had stronger MacR than English (the Head-prominence language), and English

had stronger realization of lexical stress than Bengali and Uyghur. The perception experiment supported what was observed in the MacR production experiment, with Bengali being rated as more melodic than English and Uyghur. Second, the results suggest that listeners can perceive differences in MacR strength between languages, which complements previous findings that listeners can differentiate between languages based only on F0 (e.g., Vicenik & Sundara, 2013) and that F0 is an important prominence cue for speech rhythm perception (e.g., Niebuhr, 2009; Barry et al., 2009).

The results of this study set up multiple future directions for experiments on MacR strength perception. Since the perception experiment only tested L1 English participants, who have strong associations between rhythm and lexical stress correlates (i.e., duration, loudness), it would be interesting and informative to run a similar experiment with L1 speakers of a strong MacR language, especially one which does not mark lexical stress, such as Korean. Tonal rhythm helps Korean listeners find word boundaries (Kim, 2004; Kim & Cho, 2009), so Korean listeners are predicted to perform the MacR perception rating task with greater accuracy than English listeners. Since the function of MacR is to help with word segmentation, another future study could test the perception of MacR strength by including segmental information in the stimuli. As with previous studies that used an artificial language to test prominence cues in word segmentation (e.g., Bhatara et al., 2013; Ordin & Nespor, 2013, 2016; Molnar, Carreiras, & Gervain, 2016), a future study could test how differences in MacR strength affect word segmentation, and what effect the participant's L1 has on the results.

5.4. Conclusion

The goals of this dissertation were to 1) test Jun's hypothesis that there is an inverse relationship between lexical stress strength and MacR strength (2014:537), and 2) test whether differences in MacR strength are readily perceptible to listeners. The results of all three experiments provide partial support for the first goal. The languages predicted to be on both ends of the strength spectrum, Bengali and English, were found to be significantly different from each other in both lexical stress strength realization and in MacR strength realization in the predicted direction. Uyghur was more variable in the strength ranking, but crucially behaved as predicted in the production of MacR as determined by the MacR_Freq Index, and no language data showed an exception to the predicted strength ranking.

Listeners were also able to perceive differences in tonal rhythm, or melodicity, between languages, supporting the second goal. Specifically, listeners rated Bengali utterances as more melodic than English and Uyghur utterances, although the latter two languages were not rated significantly different. Overall, the results support the inverse relationship between lexical (i.e., stress) and phrasal (i.e., tonal rhythm) prominence cues, and contribute to the growing body of literature on MacR strength differences across languages. More broadly, the results add to our understanding of how the interplay of prominence cues affects speech rhythm perception.

Appendix A: Lexical Stress Production Stimuli by Language

English Stimuli

1. I baked a pastry called **bada** today. It's a type of sweet pastry filled with cheese and fruit. I heard that **bada** should be very tasty. I'm excited to eat it!

2. I bought a new fruit called **baga** today. It grows in tropical rainforests. I heard that **baga** should be very tasty. I'm excited to try it!

3. I baked a bread called **dama** today. It's a type of flatbread made with sesame seeds. I heard that **dama** should be very tasty. I'm excited to eat it!

4. I bought a beverage called **gama** today. It's a yogurt-based beverage mixed with fruit. I heard that **gama** should be very tasty. I'm excited to drink it!

5. I bought a new vegetable called **lada** today. It's a vegetable that grows in cold climates. I heard that **lada** should be very tasty. I'm excited to try it!

6. I found a recipe for a soup called **lana** today. It's a spicy noodle soup with garlic and eggplant. I heard that **lana** should be very tasty. I'm excited to make it!

7. I found a recipe for a dish called **maba** today. It's a dish made of rice, lentils, and onions. I heard that **maba** should be very tasty. I'm excited to make it!

8. I bought a dessert called **naba** today. It's a large pastry made with honey and berries. I heard that **naba** should be very tasty. I'm excited to eat it!

Uyghur Latin Script Stimuli

English translation in italics

1. Bügün **bada** dëgen pëchinini pishurdum. U pishlaq bilen mëwe arilashturulup pishurulidighan tatliq pëcine. Men **bada**ning bek temlik ikenlikini anglighan, uni yëgenlikimdin xushalmen.
   Today I baked bada pastry. It's a sweet pastry baked with a mixture of cheese and fruit. I heard that bada is very tasty, so I'm happy that I ate it.

2. Bügün **baga** dep atilidighan bir yëngi mëwe sëtiwaldim. U issiq belwagh ormanliqida ösidiken. Men **baga**ning bek temlik ikenlikini anglighan, uni yep bek hayajanlandim.
   Today I bought a new fruit called baga. It grows in tropical forests. I heard that baga is very tasty, so I was very excited to eat it.

3. Bügün **dama** dëgen nanni pishurdum. U künjüt bilen yasalghan nëpiz nanning bir türi. Men **dama**ning bek temlik ikenlikini anglighan, uni yëgenlikimdin xushalmen.
   Today I baked bread called dama. It is a type of thin bread made with sesame seeds. I heard that dama is very tasty, so I'm happy that I ate it.

4. Bügün **gama** dep atilidighan ichimlik sëtiwaldim. U qëtiqqa mëwe arilashturup yasalghan ichimlik. Men **gama**ning bek temlik ikenlikini anglighan, uni ichkenlikimdin xushalmen.
   Today I bought a drink called gama. It is a drink made with fruit mixed with yogurt. I heard that gama is very tasty, so I'm happy that I drank it.

5. Bügün **lada** dep atilidighan bir yëngi köktat sëtiwaldim. U soghuq këlimatta ösidighan köktat iken. Men **lada**ning bek temlik ikenlikini anglighan, uni yep bek hayajanlandim.
   Today I bought a new vegetable called lada. It is a cold climate vegetable. I heard that lada is very tasty, so I was very excited to eat it.

6. Bügün **lana** dep atilidighan shorpining rëtsëpini taptim. U samsaq we tuxum bilen ëtilidighan achchiq - chüchük shorpa. Men **lana**ning bek temlik ikenlikini anglighan, uni etkenlikimdin xushalmen.
   Today I found a recipe for a soup called lana. It is a hot and spicy soup with garlic and egg. I heard lana is very tasty, so I'm happy that I made it.

7. Bügün **maba** dëgen tamaqning rëtsëpini taptim. U gürüc, chilan we piyazdin yasalghan tamaq. Men **maba**ning bek temlik ikenlikini anglighan, uni etkenlikimdin xushalmen.
   Today I found a recipe for a dish called maba. It is a dish made of rice, green chilies, and onions. I heard that maba is very tasty, so I'm happy that I made it.

8. Bügün **naba** dëgen bir tatliq - türümni sëtiwaldim. U hesel we mëwe-chiwe bilen yasalghan cong pëchine. Men **naba**ning bek temlik ikenlikini anglighan, uni yëgenlikimdin xushalmen.
   Today I bought a sweet called naba. It is a large cake made with honey and fruit. I heard that naba is very tasty, so I'm happy that I ate it.

Uyghur Perso-Arabic Script Stimuli

1. بۇگۈن **بادا** دېگەن پېچىننى پىشۇردۇم. ئۇ پىشلاق بىلەن مېۋە ئارىلاشتۇرۇلۇپ پىشۇرۇلىدىغان تاتلىق پېچىنە. مەن **بادا**نىڭ بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى يېگەنلىكىمدىن خۇشالمەن.

2. بۇگۈن **باگا** دەپ ئاتىلىدىغان بىر يېڭى مېۋە سېتىۋالدىم. ئۇ ئىسسىق بەلۋاغ ئورمانلىقىدا ئۆسىدىكەن. مەن **باگا**نىڭ بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى يەپ بەك هاياجانلاندىم.

3. بۇگۈن **داما** دېگەن ناننى پىشۇردۇم. ئۇ كۈنجۈت بىلەن ياسالغان نېپىز ناننىڭ بىر تۈرى. مەن **دامانىڭ** بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى يېگەنلىكىمدىن خۇشالمەن.

4. بۇگۈن **گاما** دەپ ئاتىلىدىغان ئىچىملىك سېتىۋالدىم. ئۇ قېتىققا مېۋە ئارىلاشتۇرۇپ ياسالغان ئىچىملىك. مەن **گاما**نىڭ بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى ئىچكەنلىكىمدىن خۇشالمەن.

5. بۇگۈن **لادا** دەپ ئاتىلىدىغان بىر يېڭى كۆكتات سېتىۋالدىم. ئۇ سوغۇق كېلىماتتا ئۆسىدىغان كۆكتات. مەن **لادا**نىڭ بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى يەپ بەك هاياجانلاندىم.

6. بۇگۈن **لانا** دەپ ئاتىلىدىغان شورپىنىڭ رېتسېپىنى تاپتىم. ئۇ سامساق ۋە توخۇم بىلەن ئېتىلىدىغان ئاچچىق - چۈچۈك شورپا. مەن **لانا**نىڭ بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى ئەتكەنلىكىمدىن خۇشالمەن.

7. بۇگۈن **مابا** دېگەن تاماقنىڭ رېتسېپىنى تاپتىم. ئۇ گۈرۈچ، چىلان ۋە پىيازدىن ياسالغان تاماق. مەن **مابا**نىڭ بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى ئەتكەنلىكىمدىن خۇشالمەن.

8. بۇگۈن **نابا** دېگەن بىر تاتلىق - تۈرۈمنى سېتىۋالدىم. ئۇ هەسەل ۋە مېۋە-چېۋە بىلەن ياسالغان چوڭ پېچىنە. مەن **نابا**نىڭ بەك تەملىك ئىكەنلىكىنى ئاڭلىغان، ئۇنى يېگەنلىكىمدىن خۇشالمەن.

Bengali Stimuli

Note: English translation in italics

1. আজ আমি **বাদা** নামের একটি মিষ্টি বানিয়েছি। এটি গুড় আর ফল দিয়ে ভরা এক ধরনের মিষ্টি। শুনেছি **বাদা** খুব সুস্বাদু। এখনই খেতে ইচ্ছে করছে!

Today I made a sweet named bada. It is a sweet filled with jaggery and fruits. I heard bada is very tasty. (I) want to eat (it) now!

2. আজ আমি **বাগা** নামের একটি নতুন ফল কিনেছি। এই ফলটি শুধুমাত্র অতি গরম দেশে হয়। শুনেছি **বাগা** খুব সুস্বাদু। এখনই খেতে ইচ্ছে করছে!

Today I bought a new fruit named baga. This fruit only grows in very hot countries. I heard baga is very tasty. (I) want to eat (it) now!

3. আজ আমি **দামা** নামের একটি রুটি বানিয়েছি। এটি তিলের বীজ দিয়ে তৈরি এক ধরনের পাতলা রুটি। শুনেছি **দামা** খুব সুস্বাদু। এখনই খেতে ইচ্ছে করছে!

Today I made a bread called dama. It is a type of thin bread made with sesame seeds. I heard the dama is very tasty. (I) want to eat (it) now!

4. আজ আমি **গামা** নামের একটি শরবত কিনেছি। এটি দই আর ফল একসঙ্গে মেশানো একটি শরবত। শুনেছি **গামা** খুব সুস্বাদু। এখনই খেতে ইচ্ছে করছে!

Today I bought a sherbet called gama. It is a sherbet made with yogurt and fruit mixed together. I heard gama is very tasty. (I) want to eat (it) now!

5. আজ আমি **লাদা** নামের একটি নতুন সবজি কিনেছি। এটি খুব ঠান্ডা দেশে পাওয়া যায় এমন একটি সবজি। শুনেছি **লাদা** খুব সুস্বাদু। এখনই রাঁধতে ইচ্ছে করছে!

Today I bought a new vegetable named lada. It is a vegetable found in very cold countries. I heard Lada is very tasty. (I) want to cook (it) now!

6. আজ আমি **লানা** নামের একটি রেসিপি খুঁজে পেয়েছি। এটি বেগুন আর রসুনের ফালি দিয়ে তৈরি বেশ ঝাল একটি নুডল স্যুপ। শুনেছি **লানা** খুব সুস্বাদু। এখনই বানাতে ইচ্ছে করছে!

Today I found a recipe called lana. It is a very spicy noodle soup made with eggplant and garlic slices. I heard lana is very tasty. (I) want to make (it) now!

7. আজ আমি **মাবা** নামের একটি রেসিপি খুঁজে পেয়েছি। এটি চাল, মসুর, আর পেঁয়াজ দিয়ে তৈরি একটি খাবার। শুনেছি **মাবা** খুব সুস্বাদু। এখনই বানাতে ইচ্ছে করছে!

Today I found a recipe called maba. It is a dish made of rice, lentils, and onions. I heard that maba is very tasty. (I) want to make (it) now!

8. আজ আমি **নাবা** নামের একটি কেক কিনেছি। এটি মধু এবং বেরি দিয়ে তৈরি একটি বড় পেস্ট্রি। শুনেছি **নাবা** খুব সুস্বাদু। এখনই খেতে ইচ্ছে করছে!

Today I bought a cake named naba. It is a large pastry made with honey and berries. I heard naba is very tasty. (I) want to eat (it) now!

Appendix B: Macro-Rhythm Production Experiment Stimuli

The North Wind and the Sun story taken from Aesop Language Bank Team (2010)

English

The North Wind and the Sun were arguing about which one of them was stronger, when a traveler came by wearing a heavy coat. They agreed that whoever got the traveler to take off his coat first would be considered stronger. The North Wind blew as hard as he could, but the harder he blew, the tighter the traveler wrapped his coat around him, and finally the North Wind had to give up. Then the sun began to shine, and the traveler immediately took off his coat. And so the North Wind had to admit that the Sun was stronger.

Uyghur

Latin orthography:

Shimal shamili bilen quyash qaysimiz tëximu küchlük dep, munazirilishiwatqanda qëlin juwiliq bir yoluchi këlip qa(l)ptu. Ular kim yoluchining juwisini awwal salduralisa shu küchlük hësablansun dep këlishiptu. Shimal shamili jënining bariche chiqiptu, emma u chiqqansëri yoluchi juwisigha tëximu mehkem orinptu. Axiri shimal shamili waz këchishke mejbur boluptu. Emdi quyash parlighan iken, yoluchi derhal juwisini sëlip tashlaptu. Shuning bilen shimal shamili quyashning küchlüklikini ëtirap qilishqa mejbur boluptu.

Perso-Arabic orthography:

شىمال شامىلى بىلەن قۇياش قايسىمىز تېخىمۇ كۈچلۈك دەپ، مۇنازىرىلىشىۋاتقاندا قېلىن جۇۋىلىق بىر يولۇچى كېلىپ قاپتۇ. ئۇلار كىم يولۇچىنىڭ جۇۋىسىنى ئاۋۋال سالدۇرالسا شۇ كۈچلۈك ھېسابلانسۇن دەپ كېلىشىپتۇ. شىمال شامىلى جېنىنىڭ بارىچە چىقىپتۇ، ئەمما ئۇ چىققانسېرى يولۇچى جۇۋىسىغا تېخىمۇ مەھكەم ئورىنىپتۇ. ئاخىرى شىمال شامىلى ۋاز كېچىشكە مەجبۇر بولۇپتۇ. ئەمدى قۇياش پارلىغان ئىكەن، يولۇچى دەرھال جۇۋىسىنى سېلىپ تاشلاپتۇ. شۇنىڭ بىلەن شىمال شامىلى قۇياشنىڭ كۈچلۈكلىكىنى ئېتىراپ قىلىشقا مەجبۇر بولۇپتۇ.

Bengali

উত্তর বায়ু এবং সূর্যের বিবাদ লেগেছে, কে বেশী শক্তিশালী। এই সময় এক পথযাত্রী সেই রাস্তা দিয়ে যাচ্ছিলো। তার গায়ে একটা গরম চাদর। উত্তর বায়ু আর সূর্য ঠিক করলো, যে ওই পথযাত্রীকে প্রথম চাদরটা খুলে নিতে বাধ্য করবে, তাকেই বেশী শক্তিশালী বলে মানা হবে। তখন উত্তর বায়ু ভীষণ বেগে বইতে শুরু করলো। কিন্তু হাওয়ার জোর যত বাড়তে লাগলো, সেই যাত্রী তার চাদরটা ততটাই শক্ত করে নিজের গায়ে জড়িয়ে নিলো। অবশেষে উত্তর বায়ু হাল ছাড়লো। তখন সূর্য প্রবল তেজের সঙ্গে আকাশে উঠলো, আর সেই যাত্রী সঙ্গে সঙ্গে তার গায়ের চাদরটা খুলে ফেললো। এই ভাবে উত্তর বায়ু স্বীকার করতে বাধ্য হল যে তাদের দুজনের মধ্যে সূর্যই বেশী শক্তিশালী।

Appendix C: Macro-Rhythm Perception Experiment Stimuli

File Name	Utterance Text	σ #	PW #
f0002_00003	So the incentives are much larger to produce drugs which treat more people	18	7
f0002_00050	Investments in education are increasing the supply of new ideas	20	6
f0003_00065	I just like to dive right in and become sort of a human guinea pig	18	8
f0003_00114	the terms of the second Fort Laramie Treaty had been violated	17	7
f0003_00134	And the story is that Leopold Auenbrugger was the son of an innkeeper	20	6
f0003_00149	It's a group of prop crazies just like me called the Replica Props Forum	18	9
f0003_00155	And west Antarctica cropped up on top some under-sea islands	16	6
f0003_00157	and he gave them one of his original plasters of the Maltese Falcon	18	6
f0003_00179	I had a reputation as being interested in patients with chronic fatigue	20	7
f0003_00208	These are supposed to simulate the actual form of a sprinter when they run	19	8

Table C1: English utterances taken from ST-AEDS-20180100_1 Free ST American English Corpus (Surfing Technology Ltd, 2018). σ # = number of syllables, PW # = number of Prosodic Words per utterance.

File Name	Utterance Text (Latin Orthography) + Translation	σ #	PW #
F011_006	Gülhépizem horazgha egeshken mikiyandek uninggha egiship mangdi *Gülhépizem followed him/her just like a hen follows a rooster*	20	7
F011_012	U herqandaq bir adem teripiden tonulushqa muhtaj emestur *S/he does not need to be known by anyone*	18	6
F011_025	Semet jüjang titrep turup Qing dobanning héliqi buyruqini tekrarlidi *Semet director trembled and repeated that order of Qing supervisor (chief)*	20	9
F016_002	Bu dewirde béshigha zibu-zinnet taqash ewj alghan idi *During this period, wearing jewelry on the head has come to its peak*	18	6
F023_020	Bu séliqning töttin bir qismi Qobtu qebilisige chüshti *A quarter of this levy fell on the Qobtu tribe*	18	6
F053_001	Men Tughulghan künümde nurghunlighan sowgha qobul qildim *I received a lot of gifts on my birthday*	17	5
F053_014	Yashanghan Dérbimu Ruseldek késeljchan adem idi *The old Derbimu was also a sick man like Rusul*	16	5
F060_013	Bu Yujifning dokilati asasida maqullighan qarar *This is a decision that has been made on the basis of Yujif's report*	17	5
F064_013	Eysa begning Zöhrexan dégen bir qizi bar idi *Master Eysa had a daughter named Zohrakhan*	15	6
F064_015	Tekebbur adem özini özi halaketke bashlaydu *The arrogant man begins to destroy himself*	16	6

Table C2: Uyghur utterances taken from THUYG20 corpus (Rouzi et al., 2017). σ # = number of syllables, PW # = number of Prosodic Words per utterance.

File Name	Utterance Text + Translation	σ #	PW #
msm_019	এখন ছবিটির ডিভিডি এবং ভিসিডি পাওয়া যাচ্ছে Ēkhon chobiṭi ḍiviḍi ēboṁ visiḍi pāōwā dācchē *Now one can find DVDs and VCDs of the show*	17	7
msm_111	পান খাওয়া দাঁত বের করে কেই একজন বলে দিচ্ছেন pān khāōwā dāḿ bēr korē kēi wēkjon bolē dicchēn *Revealing teeth (stained from) eating paan, someone is saying [it] (for the benefit of others)*	14	7
punam_000	প্রথম থেকেই বিজ্ঞাপনের জগতে তার`` চাহিদা তুঙ্গে prothom thekei biggaponer jôgote tar čahida tuṅge *From the very beginning he was in high demand in the world of advertising*	17	6
punam_042	এসএসকীম রাজ্যের অভিজাততম সরকারী হাসপাতাল es es ke em rajjer obhijoggotômo šôrkari hašpatal *SSKM (is) the state's most elite/noble government hospital*	17	5
punam_073	ঢাক` বাজানোর পরে` দেখালেন` ভাংরা নাচের নমুনা ḍhak bajanor pôre dêkhalen bhaṅra načer nomuna *After playing the ḍhak, he/she showed an example of Bhangra dance*	16	7
ritwika_001	আর তত দ্রুত চড়ছে আশঙ্কার পারদ ar tôto druto čorčhe ašoṅkar parod *And the mercury of fear is rising faster*	12	6
ritwika_279	ত্ড়ড়তীয় অনুষ্ঠানেও শোনা গেল যুগলবন্দি tritio onušṭhaneo šona gêlo jugôl bondi *Jugalbandi (music) was also heard in the third event*	15	5
ritwika_413	চলছে কর্মশালা চর্চা নতুন নতুন প্রযোজনা` čolčhe kôrmošala čôrča notun notun projojona *Workshops are being conducted and new productions are underway*	16	6
Suranjana_107	দ্রুতপায়ে সিঁড়ি ভেঙে আমি উপরে উঠে এসেছি druto pae širi bheṅe ami opore uṭhe ešečhi *Quickly having traversed the staircase I've come upstairs*	17	7
Suranjana_128	পয়সাকড়ির ব্যাপারে এদের মাথায় কত বুদ্ধি pôešakorir bêpare eder mathae kôto buddhi *On the matter of wealth in their heads (there's) so much intelligence*	15	6

Table C3: Bengali utterances taken from SHRUTI Bangla Speech Corpus (Das, Mandal, & Mitra, 2011). σ # = number of syllables, PW # = number of Prosodic Words per utterance.

References

Abbas, A. & Jun, S-A. (2022). Prosodic structure of Farasani Arabic: Accentual Phrase without a pitch accent. *Proceedings of the 1st International Conference on Tone and Intonation (TAI)*, 224-228. https://doi.org/10.21437/TAI.2021-46

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Aesop Language Bank Team (2010). *Aesop Language Bank*. http://www.aesoplanguagebank.com/bn.html

Aguilar, L., de-la-Mota, C., & Prieto, P. (2009). SP_ToBI training materials. http://prosodia.upf.edu/sp_tobi/en/index.php

Alexander, J., Bradlow, A., & Wong, P. (2005). Lexical tone perception in musicians and non-musicians. *Proceedings of Interspeech 2005*, 397–400. https://doi.org/10.21437/Interspeech.2005-271

Allen, G. D. (1972a). The location of rhythmic stress beats in English: An experimental study I. *Language and Speech, 15*(1), 72–100. https://doi.org/10.1177/002383097201500110

Allen, G. D. (1972b). The location of rhythmic stress beats in English: An experimental study II. *Language and Speech, 15*(2), 179–195. https://doi.org/10.1177/002383097201500208

Allen, G. D. (1975). Speech rhythm: Its relation to performance and articulatory timing. *Journal of Phonetics, 3*, 75–86. https://doi.org/10.1016/S0095-4470(19)31351-8

Allen, G. D. & Hawkins, S. (1978). The development of phonological rhythm. In A. Belt & J. B. Hooper (Eds.), *Syllables and segments* (pp. 173-185). Amsterdam: North-Holland.

Anderson, M., Pierrehumbert, J., & Liberman, M. (1984). Synthesis by rule of English intonation patterns. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84, 9*, 77-80. IEEE. https://doi.org/10.1109/ICASSP.1984.1172427

Arvaniti, A. & Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In S-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 84-117). Oxford University Press.

Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica, 66*(1-2), 46-63. https://doi.org/10.1159/000208930

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Avesani, C. & Vayra, M. (2005). Accenting, deaccenting and information structure in Italian dialogue. *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, 19-24. https://aclanthology.org/2005.sigdial-1.3

Aylett, M. & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31–56. https://doi.org/10.1177/00238309040470010201

Baker, R. E. & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech, 52*(4), 391–413. https://doi.org/10.1177/0023830909336575

Barry, W. J., Andreeva, B., & Koreman, J. (2009). Do rhythm measures reflect perceived rhythm? *Phonetica, 66* (1-2), 78-94. https://doi.org/10.1159/000208932

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Usinglme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Batterink, L. & Paller, K. (2017). Online neural monitoring of statistical learning. *Cortex, 90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Beckman, M. (1986). *Stress and non-stress accent*. Foris Pub USA.

Beckman, M. (1992). Evidence for speech rhythm across languages. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 457-463). IOS Press: Amsterdam.

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

Benguerel, A-P. & D'Arcy, J. (1986). Time-warping and the perception of rhythm in speech. *Journal of Phonetics, 14*, 231-246. https://doi.org/10.1016/S0095-4470(19)30665-5

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *Journal of the Acoustical Society of America, 134*(5), 3828–3843. https://doi.org/10.1121/1.4823848

Bloch, B. (1950). Studies in colloquial Japanese IV: phonemics. *Language, 26*, 86–125. https://doi.org/10.2307/410409

Bigi, B. (2015). SPPAS - Multi-lingual approaches to the automatic annotation of speech. *The Phonetician: A Publication of ISPhS, International Society of Phonetic Sciences, 111-112*, 54-69. https://hal.science/hal-01417876

Boersma, P. & Weenink, D. (2022). *Praat: doing phonetics by computer (Version 6.2.10)*. [Computer Program]. Retrieved from http://www.praat.org/

Bolinger, D. (1968). *Aspects of language*. New York: Harcourt, Brace & World, Inc.

Boll-Avetisyan, N., Bhatara, A., & Höhle, B. (2017). Effects of musicality on the perception of rhythmic structure in speech. *Laboratory Phonology, 8*(1), 9. https://doi.org/10.5334/labphon.91.

Botinis, A., Banert, R., Fourakis, M., & Pagoni-Tetlow, A. (2002). Prosodic effects and crosslinguistic segmental durations. *Proceedings of Fonetik, TMH-QPSR 44*(1), 77-80.

Boudreault, M. (1970). Le rythme en langue franco-canadienne. Prosodic feature analysis. In P. Leon et al., (Eds.), *Studia Phonetica 3* (pp. 21-30). Montreal: Didier.

Burdin, R. S. (2020). The perception of Macro-rhythm in Jewish English intonation. *American Speech, 95*(3), 263–296. https://doi.org/10.1215/00031283-7706542

Burdin, R., Phillips-Bourass, S., Turnbull, R., Yasavul, M., Clopper, C., & Tonhauser, J. (2014). Variation in the prosody of focus in head- and head/edge-prominence languages. *Lingua, 165*, 254-276. https://doi.org/10.1016/j.lingua.2014.10.001

Chatterji, S. K. (1921). Bengali phonetics. *Bulletin of the School of Oriental Studies, University of London, 2*(1), 1–25.

Chen, S., Zhu, Y., Wayland, R., & Yang, Y. (2020). How musical experience affects tone perception efficiency by musicians of tonal and non-tonal speakers? *PLOS ONE, 15*(5), e0232514. https://doi.org/10.1371/journal.pone.0232514

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

Clopper, C. G. & Turnbull, R. (2018). 2. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation* (pp. 25-72). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110524178-002

Corretge, R. (2020). *Praat Vocal Toolkit*. http://www.praatvocaltoolkit.com

Cumming, R. (2011a). The language-specific interdependence of tonal and durational cues in perceived rhythmicality. *Phonetica, 68*, 1–25. https://doi.org/10.1159/000327223

Cumming, R. (2011b). Perceptually informed quantification of speech rhythm in pairwise variability indices. *Phonetica, 68*, 256–277. https://doi.org/10.1159/000335416

Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language 25*, 385–400. https://doi.org/10.1016/0749-596X(86)90033-1

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133–142. https://doi.org/10.1016/0885-2308(87)90004-0

Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 157-166). London: Macmillan. https://doi.org/10.1007/978-1-349-12670-5_14

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24, 381–410. https://doi.org/10.1016/0010-0285(92)90012-Q

Cutler, A. & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33*, 824–844. https://doi.org/10.1006/jmla.1994.1039

Dainora, A. (2001). *An Empirically Based Probabilistic Model of Intonation in American English* [Doctoral dissertation, University of Chicago].

Dainora, A. (2006). Modelling Intonation in English. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory Phonology, 8* (pp. 107-132). Berlin: Mouton de Gruyter.

Das, B., Mandal, S., & Mitra, P. (2011). Bengali speech corpus for continuous automatic speech recognition system. *Proceedings of the 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA)*, 51-55. https://doi.org/10.1109/ICSDA.2011.6085979

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4

de-la-Mota, C., Butragueño, P.M., & Prieto, P. (2010). Mexican Spanish Intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 319-350). Munich: Lincom.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics, 4*, 183-198. https://doi.org/10.1515/iral.1966.4.1-4.183

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for delta C. In P. Karnowski & I. Szigeti (Eds.), *Language and language-processing* (pp. 231-241). Frankfurt am Main: Peter Lang. https://doi.org/10.5167/uzh-111789

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

Dilley, L. & Shattuck-Hufnagel, S. (1999). Effects of repeated intonation patterns on perceived word-level organization. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1487-1490.

D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication, 33*(4), 339-356. https://doi.org/10.1016/S0167-6393(00)00064-9

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*, 158–164. https://doi.org/10.1038/nn.4186

Domahs, U., Plag, I., & Carroll, R. (2014). Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics, 17*, 1–38. https://doi.org/10.1007/s10828-014-9063-9

Donovan, A. & Darwin, C. J. (1979). The perceived rhythm of speech. *Proceedings of the Ninth International Congress of Phonetic Sciences, 2*, 268-274. Copenhagen: Institute of Phonetics.

Engesæth, T., Yakup, M., & Dwyer, A. M. (2010). *Greetings from the Teklimakan: A handbook of modern Uyghur* (version 1.1). Retrieved from http://hdl.handle.net/1808/5624

Estebas-Vilaplana, E. & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto & P. Rosano (Eds.), *Transcription of intonation of the Spanish language* (pp. 17-48). Munich: Lincom Europa.

Ferguson, C. A. & Chowdhury, M. (1960). The phonemes of Bengali. *Language, 36*(1), 22–59. https://doi.org/10.2307/410622

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: a unified JavaScript framework for online studies. *International Conference on Computational Social Science (IC2S2)*, Cologne, 1-3. Retrieved from https://www.labvanced.com/static/2017_IC2S2_LabVanced.pdf

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

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31,* 307–319. https://doi.org/10.1177/002383098803100401

Frotà, S. (2014). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 6-42). Oxford University Press.

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

Fry, D. B. (1958). Experiments in the Perception of Stress. *Language and Speech 1*(2), 126–152. https://doi.org/10.1177/002383095800100207

Gordon, R. G. Jr. (Ed.) (2005). *Ethnologue: Languages of the World*, Fifteenth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com/

Gordon, M. & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard, 3*(1), 20170007. https://doi.org/10.1515/lingvan-2017-0007.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences, 99*(16), 10911–10916. https://doi.org/10.1073/pnas.122368599

Grahn, J. A. & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *NeuroImage, 47,* 1894–1903. https://doi.org/10.1016/j.neuroimage.2009.04.039

Gregori, A. & Kügler, F. (2021). An empirical investigation on the perceptual similarity of prosodic language types. *Proceedings of the 1st International Conference on Tone and Intonation (TAI),* 209-213. https://doi.org/10.21437/TAI.2021-43

Grice, M., Baumann, S., & Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In S-A. Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 55-83). Oxford University Press.

Hahn, R. (1991). *Spoken Uyghur*. University of Washington Press.

Hahn, R. (1998). Uyghur. In L. Johanson & É. Á. Csató (Eds.), *The Turkic languages* (pp. 379-396). Routledge.

Han, M. S. (1962). The feature of duration in Japanese. *Onsei no Kenkyuu 10*, 65–80.

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

Hirst, D. & Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de I'Institut de Phonetique d'Aix, 15*, 71–85.

Hirst, D. & de Cristo, A. (1999). *Intonation systems: A survey of twenty languages*. Cambridge, England: Cambridge University Press.

Hockett, C. (1955). *A manual of phonology*. Bloomington: Indiana University Publications in Anthropology and Linguistics.

Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope. *Perception and Psychophysics, 43*(1), 90–93. https://doi.org/10.3758/bf03208980

Hualde, J. I. & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frotà & P. Prieto (Eds.), *Intonation in Romance* (pp. 350-391). Oxford: Oxford University Press.

Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica, 35*, 86–105. https://doi.org/10.1159/000259924

Iversen, J. R. & Patel, A. D. (2008). The beat alignment test (BAT): surveying beat processing abilities in the general population. In Adachi et al. (Eds.), *Proceedings of the 10th International Conference on Music Perception and Cognition*, 465-468. Adelaide: Causal Productions.

Jun, S-A. (2005). Prosodic Typology. In S-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430-453). Oxford: Oxford University Press.

Jun, S-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 520-539). Oxford University Press.

Kaland, C. (2022). Bending the string: intonation contour length as a correlate of macro-rhythm. *Proceedings of Interspeech 2022*, 5233-5237. https://doi.org/10.21437/Interspeech.2022-185

Katz, J. & Selkirk, E. (2011). Contrastive focus vs. discourse-new: evidence from phonetic prominence in English. *Language*, *87*(4), 771-816. https://doi.org/10.1353/lan.2011.0076.

Kawasaki, H. & Shattuck-Hufnagel, S. (1988). Acoustic correlates of stress in four demarcative-stress languages. *Journal of the Acoustical Society of America*, *84*, S98. https://doi.org/10.1121/1.2026588

Khan, S. D. (2008). *Intonational phonology and focus prosody of Bengali* [Doctoral dissertation, University of California, Los Angeles].

Khan, S. D. (2014). The intonational phonology of Bangladeshi Standard Bengali. In S-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 81-117). Oxford University Press.

Khan, S. D. (2016). The intonation of South Asian languages: towards a comparative analysis. In M. Menon & S. Syed (Eds.), *Proceedings of Formal Approaches to South Asian Languages 6*, (pp. 23–36). Retrieved from https://ojs.ub.uni-konstanz.de/jsal/index.php/fasal/issue/view/17

Kim, S. (2004). *The role of prosodic phrasing in Korean word segmentation* [Doctoral dissertation, University of California, Los Angeles].

Kim, S. & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence form word-spotting experiments in Korean. *Journal of the Acoustical Society of America*, *125*(5), 3373-3386. https://doi.org/10.1121/1.3097777

Kohler, K. (2008). The perception of prominence patterns. *Phonetica*, 65, 257-269. https://doi.org/10.1159/000192795

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Ladd, D. R. (1996/2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*(3), 253–263. https://doi.org/10.1016/S0095-4470(19)31139-8

Li, A. & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition*, *36*(2), 223-255.

Lindström, E. & Remijsen, B. (2005). Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics*, *43*(4), 839-870. https://doi.org/10.1515/ling.2005.43.4.839

Lloyd James, A. (1940). *Speech signals in telephony* (pp. 16–27). London: Pitman and Sons.

Mairano, P., Santiago, F., & Romano, A. (2015). Cross-linguistic differences between accented vs unaccented vowel durations. *Proceedings of the 18th International Congress of Phonetic Sciences*. https://shs.hal.science/halshs-01440315

Major, T. & Mayer, C. (2018). Towards a phonological model of Uyghur intonation. *Proceedings of the 9th International Conference on Speech Prosody*, 744-748.

Major, T. & Mayer, C. (2019). A phonological model of Uyghur intonation [PowerPoint Slides]. http://socsci.uci.edu/~cjmayer/papers/major_mayer_uyghur_intonation_icphs_2019.pdf

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 397–408. https://doi.org/10.1037/0096-1523.30.2.397

McAuley, J. D., Frater, D., Janke, K., & Miller, N. S. (2006). Detecting changes in timing: evidence for two modes of listening. *Proceedings of the 9th International Conference on Music Perception and Cognition*, 566-573.

McAuley, J. D. (2010). Tempo and rhythm. In M. R. Jones, R. R. Fay, A. N. Popper (Eds.), *Music perception* (pp. 165–199). Springer Science + Business Media. https://doi.org/10.1007/978-1-4419-6114-3_6

McCollum, A. G. (2020). Sonority-driven stress and vowel reduction in Uyghur. *Proceedings of AMP 2019*. https://doi.org/10.3765/amp.v8i0.4693

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: a comparison of English and German. *Journal of the Acoustical Society of America*, *131*(3), 2249–2260. https://doi.org/10.1121/1.3681950

Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition, 152*, 150–159. https://doi.org/10.1016/j.cognition.2016.03.023

Moore-Cantwell, C. & Sanders, L. (2017). Effects of probabilistic phonology on the perception of words and nonwords. *Language, Cognition and Neuroscience, 33*, 1-17. https://doi.org/10.1080/23273798.2017.1376101

Moore-Cantwell, C. (2020). Weight and final vowels in the English stress system. *Phonology, 37*(4), 657-695. https://doi.org/10.1017/S0952675720000305

Morrill, T. H., Dilley, L. C., McAuley, J. D., & 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*(1), 69-74. https://doi.org/10.1016/j.cognition.2013.12.006

Morrill, T. H., McAuley, J. D., Dilley, L. C., & Hambrick, D. Z. (2015). Individual differences in the perception of melodic contours and pitch-accent timing in speech: Support for domain-generality of pitch processing. *Journal of Experimental Psychology. General, 144*(4), 730–736. https://doi.org/10.1037/xge0000081

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review, 83*(5), 405–408. https://doi.org/10.1037/0033-295X.83.5.405

Motz, B. A., Erickson, M. A., & Hetrick, W. P. (2013). To the beat of your own drum: Cortical regularization of non-integer ratio rhythms toward metrical patterns. *Brain and Cognition, 81*, 329–336. https://doi.org/10.1016/j.bandc.2013.01.005

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity. I. Mora rhythm. *Language and Speech, 50*(1), 77–99. https://doi.org/10.1177/00238309070500010401

Nagao, J. & Ortega-Llebaria, M. (2021). The interaction of micro- and macro-rhythm measures in English and Japanese as first and second languages. *Proceedings of the 1st International Conference on Tone and Intonation (TAI)*, 273-277. https://doi.org/10.21437/TAI.2021-56

Nadzhip, Ė. N. (1971). *Modern Uigur*. Nauka.

Nazrova, G. & Niyaz, K. (2013). *Uyghur: An elementary textbook*. Georgetown University Press.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766. https://doi.org/10.1037//0096-1523.24.3.756

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: effects of rhythm and familiarity. *Journal of Memory and Language*, 43, 1–19. https://doi.org/10.1006/jmla.2000.2698

Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41*, 233–243. https://doi.org/10.1016/S0167-6393(02)00106-1

Niebuhr, O. (2009). F0-based rhythm effects on the perception of local syllable prominence. *Phonetica*, 66, 95−112. https://doi.org/10.1159/000208933

Nolan, F. & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B Biological Sciences*, 369, 20130396. https://doi.org/10.1098/rstb.2013.0396

O'Connor, J. (1965). *The perception of time intervals (Progress Report 2)*. Phonetics Laboratory, University College London.

Ordin, M. & Nespor, M. (2013). Transition probabilities and different levels of prominence in segmentation. *Language Learning, 63*(4), 800–834. https://doi.org/10.1111/lang.12024

Ordin, M. & Nespor, M. (2016). Native language influence in the segmentation of a novel language. *Language Learning and Development, 12*(4), 461-481. https://doi.org/10.1080/15475441.2016.1154858

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition, 45*, 863–876. https://doi.org/10.3758/s13421-017-0700-9

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 358–378. https://doi.org/10.1006/jmla.1993.1014

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* [Doctoral dissertation, Massachusetts Institute of Technology].

Pike, K. (1946). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Polyanskaya, L., Busà, M. G., & Ordin, M. (2019). Capturing cross-linguistic differences in macro-rhythm: the case of Italian and English. *Language and Speech* https://doi.org/10.1177/0023830919835849

Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics, 17*, 175–192. https://doi.org/10.1016/S0095-4470(19)30428-0

Prechtel, C. (2019). Quantifying macro-rhythm in English and Spanish. *Proceedings of the 19th International Congress of Phonetic Sciences*, 2896-2900.

Prechtel, C. (2020). *Quantifying macro-rhythm in English and Spanish: A comparison of tonal rhythm strength* [Master's thesis, University of California, Los Angeles].

Prechtel, C. (2021). Lexical stress strength vs macro-rhythm strength: An inverse relationship between prominence cues. *Proceedings of the 1st International Conference on Tone and Intonation (TAI)*, 102-106. https://doi.org/10.21437/TAI.2021-21

Prechtel, C. (2022). Testing the inverse relationship between lexical stress strength and macro-rhythm strength. [Conference presentation]. *3rd International Conference "Prominence in Language"*, University of Cologne, 2-3 June 2022.

Prieto, P., Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm: Evidence from Catalan, English, and Spanish. *Speech Communication, 54*, 681–702. https://doi.org/10.1016/j.specom.2011.12.001

Ramus, F. & Mehler, J. (1999). Language identification with suprasegmental cues: a study based on speech resynthesis. *The Journal of the Acoustical Society of America, 105*(1), 512-21. https://doi.org/10.1121/1.424522.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265-292. https://doi.org/10.1016/S0010-0277(00)00101-3

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences*, 337–342. https://web-archive.southampton.ac.uk/cogprints.org/3079/1/ICPhS03.pdf

Rialland, A. & Robert, S. (2001). The intonational system of Wolof. *Linguistics, 39*(5), 893-939. https://doi.org/10.1515/ling.2001.038

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73-79). London: Edward Arnold.

Roettger, T. & Gordon, M. (2017). Methodological issues in the study of word stress correlates. *Linguistics Vanguard, 3*(1), 20170006. https://doi.org/10.1515/lingvan-2017-0006

RStudio Team (2020). RStudio: Integrated development for R. RStudio, PBC: Boston, MA. http://www.rstudio.com/

Rouzi, A., Yin, S., Zhang, Z., Wang, D., Hamdulla, A., & Zheng, F. (2017). THUYG-20: A free Uyghur speech database. *Qinghua Daxue Xuebao/Journal of Tsinghua University 57*, 182-187. https://doi.org/10.16511/j.cnki.qhdxxb.2017.22.012.

Schmerling, S. F. (1976). *Aspects of English sentence stress*. Austin: University of Texas Press.

Shattuck-Hufnagel, S. & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193-247. https://doi.org/10.1007/BF01708572

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54, 1-32. https://doi.org/10.1016/j.cogpsych.2006.04.002.

Surfing Technology Ltd. (2018). *ST-AEDS-20180100_1, Free ST American English Corpus*. Open Speech and Language Resources (OpenSLR). http://openslr.org/45/

Tilsen, S. & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, *124*(2), EL34–EL39. https://doi.org/10.1121/1.2947626

Tilsen, S. & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America*, *134*(1), 628-639. https://doi.org/10.1121/1.4807565

Tyler, M. D. & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America, 126*(1), 367–376. https://doi.org/10.1121/1.3129127

Vicenik, C. & Sundara, M. (2013). The role of rhythm and intonation in language and dialect discrimination by adults. *Journal of Phonetics, 41*(5), 297-306. https://doi.org/10.1016/j.wocn.2013.03.003

Vogel, I., Athanasopoulou, A., & Pinkus, 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: Cambridge University Press.

Warner, N., Otake, T., & Arai, T. (2010). Intonational structure as a word boundary cue in Japanese. *Language and Speech, 53*, 107-131. https://doi.org/10.1177/0023830909351235

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics, 38*(4), 654-662. https://doi.org/10.1016/j.wocn.2010.10.001

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication, 49*, 28-48. https://doi.org/10.1016/j.wocn.2010.10.001

Wehrle, S., Cangemi, F., Krüger, M., & Grice, M. (2018). Somewhere over the spectrum: Between singsongy and robotic intonation. In A. Vietti, L. Spreafico, D. Mereu, & V. Galatà (Eds.), *Il parlato nel contesto naturale. Proceedings of 14th Associazione Italiana Scienze della Voce Conference 2018* (pp. 179-194). Bolzano, Italien.

Wehrle, S., Cangemi, F., Hanekamp, H., Vogeley, K., & Grice, M. (2020). Assessing the intonation style of speakers with autism spectrum disorder. *Proceedings of the 10th International Conference on Speech Prosody*, 809-813. https://doi.org/10.21437/SpeechProsody.2020-165

White, L. & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*, 501-522. https://doi.org/10.1016/j.wocn.2007.02.003

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language, 66*(4), 665-679. https://doi.org/10.1016/j.jml.2011.12.010

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America, 127*, 1559–1569. https://doi.org/10.1121/1.3293004

Yakup, M. (2013). *Acoustic correlates of lexical stress in native speakers of Uyghur and L2 learners* [Doctoral dissertation, University of Kansas].

Yakup, M. & Sereno, J. A. (2016). Acoustic correlates of lexical stress in Uyghur. *Journal of the International Phonetic Association, 46*(1), 61-77. https://doi.org/10.1017/S0025100315000183