

RETHINKING RHYTHM IN LANGUAGE AND MUSIC:
CAN DURATION PRIME RHYTHM IN METRICAL SPEECH?

ZÜHEYRA TOKAÇ

BOĞAZİÇİ UNIVERSITY

2019

RETHINKING RHYTHM IN LANGUAGE AND MUSIC:
CAN DURATION PRIME RHYTHM IN METRICAL SPEECH?

Thesis submitted to the
Institute for Graduate Studies in Social Sciences
in partial fulfillment of the requirements for the degree of

Master of Arts
in
Psychological Sciences

by
Züheyra Tokaç

Boğaziçi University

2019

DECLARATION OF ORIGINALITY

I, Züheyra Tokaç, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date 11.25.2019.....

ABSTRACT

Rethinking Rhythm in Language and Music: Can Duration Prime Rhythm in Metrical Speech?

Rhythmic qualities of languages have been explored and discussed primarily in relation to the much-debated proposal of rhythm classes. Moreover, research on linguistic rhythm has been led by two premises: isochrony and stress. Predominated by a limited set of languages, research in this area fell short of examining remaining languages in compatible ways. The present study aims at taking an alternative approach by building its hypotheses in a language-specific way. As native speakers' perception of linguistic rhythm becomes most evident in their metrical speech, Turkish poetic meters were examined to identify possible constituents of rhythm in Turkish. Based on a poetic meter that creates rhythmicity through an organization of closed and open syllables, this study examines phonemic duration between vowels of subsequent syllables as such a candidate. Experiment 1 explores the ways in which musical primes with long- and short-interval pulses can prime metrical structure in bisyllabic and trisyllabic linguistic stimuli. Results suggest primes with solely durational contrasts cannot prime the hypothetical durational relationships in linguistic stimuli but can prime syllable number under certain conditions. However, primes with intensity accents were successful in priming syllable number and syllable structure in bisyllabic nonwords, although independently. Experiment 2 examines whether priming syllable number and duration can be differentiated and yields null results. Taken together, results lean towards an understanding of rhythm in Turkish that is based on syllable number while not also rejecting a duration-based alternative.

ÖZET

Müzik ve Dilde Ritmi Yeniden Düşünmek:

Ses Süresi Dilde Ölçüyü İlkleyebilir mi?

Dillerin ritmik özellikleri, tartışmalı bir iddia olan ritim sınıflandırmaları bağlamında incelenmeye tabi tutulmuştur. Dahası, dilsel ritim üzerine yürütülen araştırmalar iki kavram üzerine yoğunlaşmıştır: eşsürelilik ve vurgu. Kısıtlı sayıda dil tarafından yönlendirilen bu araştırma alanı, kalan dilleri uygun şekillerde araştırabilmek noktasında yetersiz kalmıştır. Bu çalışma, incelenen dile özgü hipotezler geliştirerek alternatif bir yaklaşım sunmayı hedeflemektedir. Bir dili anadili olarak konuşan kimselerin dillerindeki ritim algısının en somut yansıması dilin ölçülü kullanımında ortaya çıktığı için, Türkçede kullanılan şiirsel ölçüler, Türkçe ritmin muhtemel yapıtaşlarını bulmak üzere incelenmiştir. Ritimselliği açık ve kapalı hecelerin düzenlenmesiyle oluşturan bir ölçüden yola çıkarak, sesliler arasındaki fonem süreleri Türkçe ritmin kaynağı olmaya bir aday olarak tespit edilmiştir. Deney 1, kısa ve uzun aralıklarla düzenlenmiş müziksel uyaranların, iki ve üç heceli dilsel uyaranlardaki ölçüsel düzeni ilkleme imkanını incelemektedir. Sonuçlar, sadece süre farklarına dayanan müziksel uyaranların, bazı koşullarda hece sayısını ilkleme dışında, dilsel uyaranlarda ölçüsel düzeni ilkeleyemediğini göstermektedir. Ancak, ses yüksekliği ile belirginleştirilen müziksel uyaranlar, iki heceli uydurma sözcüklerde hece sayısını ve hece yapısını birbirinden bağımsız olarak ilkeleyebilmiştir. Deney 2 ise hece sayısını ve süresini ilklemeyi birbirinden ayırabilirmeyi test etmekte ancak istatistiksel olarak anlamlı bir ayrışma tespit edememektedir. Deney sonuçları, genel olarak Türkçede hece sayısı temelinde bir ritim anlayışını desteklemekle birlikte, süre temelinde geliştirilen alternatifi de tümüyle reddetmemektedir.

TABLE OF CONTENTS

CHAPTER1: INTRODUCTION	1
1.1 Historical background on linguistic rhythm.....	2
1.2 Isochrony, perception, and linguistic rhythm.....	6
1.3 Linguistic and psycholinguistic studies on linguistic rhythm classes.....	9
1.4 Rhythm in language and music	15
1.5 Rhythm and meter in Turkish speech and poetry.....	24
1.6 Present study	34
CHAPTER 2: EXPERIMENT 1	35
2.1 Experiment 1A	35
2.2 Experiment 1B	43
2.3 Experiment 1C	48
CHAPTER 3: EXPERIMENT 2	52
3.1 Method	52
3.2 Results	54
CHAPTER 4: DISCUSSION.....	58
4.1 Effects of number match	59
4.2 Effects of beat	64
4.3 Effects of metrical match	68
4.4 The ‘trisyllabic advantage’	74
4.5 Nonwords and pseudowords	76
4.6 Limitations	78
4.7 For future research	80
4.8 Conclusions	84

APPENDIX A: CAESURA POSITIONS IN TURKISH SYLLABIC METER	85
APPENDIX B: STRESS ORGANIZATION IN TURKISH SYLLABIC METER..	86
APPENDIX C: LIST OF LINGUISTIC STIMULI.....	87
APPENDIX D: ETHICS COMMITTEE APPROVAL	89
APPENDIX E: SPEED-ACCURACY INTEGRATED RESULTS.....	90
APPENDIX F: CORRECT ‘NO’ RESPONSES	93
REFERENCES.....	95

LIST OF APPENDIX TABLES

Table E1. Summary of LISAS Results for Interaction of Prime and Number in Nonwords	92
Table E2. Summary of LISAS Results for Interaction of Prime and Number in Pseudowords	92
Table F1. Summary of Results for Correct ‘No’ Responses, Interaction of Prime and Number in Nonwords.....	94
Table F2. Summary of Results for Correct ‘No’ Responses, Interaction of Prime and Number in Pseudowords	94

LIST OF FIGURES

Figure 1. Experimental design of Cason and Schön (2012).	23
Figure 2. ΔC and %V values for 14 languages.	27
Figure 3. P-center intervals according to syllable structure.....	33
Figure 4. Experimental design of Experiment 1A.	38
Figure 5. Experimental flow	39
Figure 6. Mean reaction times in phoneme detection for nonwords.....	41
Figure 7. Mean reaction times in phoneme detection for pseudowords.	42
Figure 8. Experimental design of Experiment 1B.....	45
Figure 9. Mean reaction times in phoneme detection for nonwords.....	47
Figure 10. Mean reaction times in phoneme detection for pseudowords.	48
Figure 11. Mean reaction times in phoneme detection.	51
Figure 12. Experimental design of Experiment 2.	53
Figure 13. Mean reaction times in phoneme detection for each condition.	55
Figure 14. Mean reaction times in phoneme detection for each condition.	56
Figure 15. Gestalt-driven perception of 8-loop binary primes.....	62
Figure 16. Metrical composition of a verse in classical Turkish music.....	73

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors, Esra Mungan for her willingness to embark on this journey with me, where our compass was our enthusiasm; I am truly indebted to her guidance, encouragement, support and kindness, and Stefano Canalis for helping me navigate a field so foreign to me.

I would like to thank my extended thesis committee, Gaye Soley, Kadir Gökgöz, Fuat Balçı, İnci Ayhan and especially Sumru Özsoy. I am grateful that my master's thesis brought me together with such inspiring people from various disciplines. Your perspectives and contributions were invaluable, and your interest in this humble project was very heartening.

I must thank Kim Fortuny and İpek Kısmet Bell, first for accepting a psychology senior to their classes and teaching me about iambic pentameter, and also for their continued support and encouragement during the time. Had it not been for them, I would not have taken interest in poetic meter and linguistic rhythm.

I am thankful to Alperen Karan and Emre Oral for their friendship and support when I needed it the most or even at times at which I had not yet mentioned needing help. I admire their skills and their goodwill. Also, I am thankful to my undergraduate squad. It warms my heart to think that we started as undergraduate students and have come this far while supporting each other throughout each step. I am so proud of each and every one of you.

Lastly, I am immensely thankful to my family. I am thankful to my parents for everything but especially for providing us a childhood embellished with poetry. I cannot imagine how I would know anything about the rhythm of the Arud meter if you had not taken the 12-year-old me to those boring Diwan poetry nights. I would

also like to thank my sisters for their friendship that I cherish for my life. It has been great pretending our family is a grad house and weirdly synchronizing our graduate milestones.

This thesis is a part of the project “In Search of Influences of Music and Language on Rhythm Perception” supported by the Boğaziçi University Scientific Research Fund Grant No. 14324. I would like to thank Boğaziçi University for supporting my research and the Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting me throughout my master’s education as part of the 2211 National Scholarship Programme for MSc Students.

J.

CHAPTER 1

INTRODUCTION

Languages are perceived to be rhythmically different (Arvaniti, 2009; Ramus, 2002; Ramus, Nespor, & Mehler, 1999). Yet, the decades-long debate concerning rhythm typology of languages presented numerous, albeit inconclusive, theoretical and practical approaches raised to understand rhythmicity in languages. Research on rhythm perception in language took into consideration the notion of rhythmical differences across languages and attempted to generalize its findings according to a hypothesized set of rhythm classes (e.g., Falk, Volpi-Moncorger, & Dalla Bella, 2017). Yet an interesting common aspect of rhythm perception research in language is that studies focus primarily on stress (Benadon, 2009), disregarding other ways in which rhythm can be generated.

It has been shown that stress may not be the single, definitive rhythm-inducing parameter of all languages (Patel, 2008; Ramus, Nespor, & Mehler, 1999). The aforementioned rhythmical differences suggest that other linguistic components need to be explored as potential rhythm-inducing factors to reach a more comprehensive understanding of rhythm (cf. Arvaniti, 2009; Ramus, 2002). Metrical speech and especially poetic meter might be the right place to turn to for this process of identification, as in these forms language is organized rhythmically as opposed to daily speech and hence might give an estimate of the unit relevant to a given language's rhythmical structure (Patel, 2008). Accordingly, based on an analysis of the poetic meters in Turkish, I will propose vocalic organization as an alternative candidate for exploring rhythm in language.

1.1 Historical background on linguistic rhythm

That speech contains rhythm is an ancient idea, as is evident in certain cases of ancient prose and verse that make use of rhythmical properties of speech. Yet the scientific inquiries concerning linguistic rhythm appear in the last centuries, along with a turn from purely intuitive approaches to a search for objective measures. The theoretical grounds for the studies extending over centuries display striking similarities, with first known works already pointing out the key themes for the years to come.

In one of the earlier studies in the field, in an attempt to refute Lord Monboddo's (1773/1968) claims that unlike Greek, English does not possess rhythm, Steele (1775) argues that once melody and rhythm are more accurately defined and examined, it will become evident that English has its own rhythm. Steele proposes a musical notation system for speech that can be used for all languages, annotating five properties of speech that he suggests to constitute melody and rhythm in speech, namely (1) accent, or pitch, (2) quantity, or duration, (3) pause, or silence, (4) emphasis, or stress, and (5) force, or intensity. According to his conceptualization, emphasis on syllables correspond to syllables being either heavy or light, and syllables in combination with pauses constitute cadences that always begin with a heavy emphasis. As a result of our inclination and natural tendency to produce and perceive time in pulses, Steele suggests that the succession of heavy and light syllables and pauses occur in cadences of equal duration. Hence, as in other languages, it is this order of the heavy beats that cause rhythm in English, while syllables are independently subject to differences in accent, quantity and force as well. Although Steele only remarks how changing the syllable emphasis on the English word 'impossible' would result in the French word 'impossible', his notation

of the French word also seems to suggest that each syllable would take equal amount of time in the French case, as opposed to differing syllable lengths in the English case.

Technical advancements in measurement enabled empirical investigations of previously posited perspectives on language rhythm. Initially, calculations of feet durations to inspect possible regularities relied on perception to mark the boundaries of rhythmical units, and consequently results of such studies reflected the nature of *perceived* rhythm (see Scripture, 1902 for a review). In one such study, Scripture (1902) calculated feet durations in French verse based on his perception of the points at which stresses occurred, and claimed that despite variability in the feet durations, line durations are made constant. As for his analysis of English stanzas, Scripture claims that there is a duration rhythm, a pitch rhythm and an intensity rhythm, yet what he actually suggests is that stress is an amalgamation of duration, pitch and intensity. Scripture claims that the idea of rhythm as induced by a syllable length ratio of 1:2 (as originating from Greek verse) is not supported in the data. On the other hand, he claims that in verse, intervals between *stresses* are “approximately constant”, but he does not present empirical evidence for this.

The quest of what gives rise to the perception of rhythm in language could not be resolved with the tools and subjective measurements of the time. Technical advancements enabling measurements that do not depend on subjective raters provided opportunities to explore regularities in speech itself rather than its perception. In this respect, Brown (1908) acquired sound recordings of actual and nonsense English verses with various types of feet read by different readers and calculated durations of syllables, feet and verses from syllable onsets. He identified six metrical accounts for rhythm in English verse and evaluated these with

calculations made accordingly when possible. Two of the accounts suggested syllable lengths to be of fixed units and ratios, yet Brown's calculations suggest that syllable durations are not fixed but variable depending on semantic, syntactical and rhythmical conditions. A third account suggested that beats on stressed syllables occurring with intervals of equal duration generates rhythm in verse. Brown conceptualized the stress beats as the points of maximal stress, and noted that the apparatus and methods of this particular study did not suffice to determine the beats. He thus proceeded to calculate the intervals between syllable onsets and commented that these intervals suggest a lack of isochrony but a precise detection of the maximal stress points would reveal an underlying regularity of feet durations. Brown's conclusions were that duration is not sufficient to comprehend rhythm and factors such as stress and prosody should be included in both theory and research for a complete understanding.

The idea that languages can be represented in rhythmically contrastive terms, first presented by Lloyd James in 1940 (as cited in Nespor, Shukla, & Mehler, 2010), was based on a perceptual observation that English had a Morse code-like rhythm whereas rhythm in Spanish sounded like a machine gun's. This distinction was made more concrete when Pike (1945) coined the terms 'stress-timed rhythm' and 'syllable-timed rhythm' to describe the rhythmical qualities speech can acquire. Pike's (1945/1958a; 1947/1958b) definitions for the terms were based on whether vowel reduction is employed in unstressed syllables and whether stresses or syllables occur at somewhat regular time points. In his work, he cited Classe's (1939) measurements of English speech which suggested that stress groups with differing numbers of syllables may have equal durations due to syllable contractions. However, Pike (1958a) also provided examples in which English is organized in

syllable-timed rhythm, implying that the two rhythmic types are not strictly exclusive in a given language (cf. Cumming, 2010).

While Pike (1958a) was the one to propose the terms that were well-received in the linguistic rhythm field, it was Abercrombie's (1967) interpretation of the terms 'stress-timed' and 'syllable-timed' that more precisely defined their contents.

According to Abercrombie, rhythm is an integral property of speech that arises from the muscle movements producing syllables and stresses, i.e., the chest- and stress-pulses, respectively. The production of these pulses is periodic in nature, and gives rise to two types of linguistic rhythm whereby all languages belong to either the syllable- or the stress-timed kind. In syllable-timed languages, syllables occur at isochronous intervals whereas in stress-timed languages, stresses occur isochronously. French, Telugu and Yoruba are examples of the former type; English, Russian and Arabic are of the latter. Abercrombie clearly states that the two types of periodicity and hence rhythmicity are mutually exclusive, and to demonstrate his claim he provides one English and one French sentence, asking the reader to tap at each stress or syllable while reciting the sentence to discover the isochrony or the lack thereof. Also in relation to the unit at which isochrony occurs, Abercrombie argues that syllable durations are approximately constant in syllable-timed languages, but more variable in stress-timed languages, hence providing a second distinctive feature for the two types of rhythm.

While Abercrombie (1967) argued for only two types of rhythm for all languages, the rhythm of Japanese had been described as a distinct case (e.g., Bloch, 1950; Jinbo, 1927/1980, as cited in Warner & Arai, 2001). Japanese recognizes speech units below the syllable level; the phonological inventory of Japanese consists of a limited number of such units, the subsyllabic units form the basis of the

line organization in Japanese poetry and the symbols in two of the three Japanese orthographies correspond to these units (Otake, Hatano, Cutler, & Mehler, 1993). Most of these units consist of a consonant and a vowel, yet other units consist of a single vowel, a single long consonant, or two consonants and a vowel. In a manner similar to that of Lloyd James (1940), Bloch (1950) claims the “staccato rhythm” to be the most prominent perceptual aspect of the Japanese syllable structure. According to his description, this rhythm is a consequence of the subsyllabic units with distinct boundaries that are attributed equal durations on the part of the listener. Bloch suggests this speech unit to be a durational unit, and states that he prefers the term “Japanese syllable” in contrast to the term “mora” which had been in use to describe the durational relations in Latin and Greek syllables (Ladefoged, 1975/2000). Yet, Japanese is still more commonly described as a “mora-timed” language, and Bloch’s (1950) emphasis on perceptual equality has been replaced by Ladefoged’s claim of durational equality in pronunciation.

1.2 Isochrony, perception, and linguistic rhythm

The idea of rhythmic classes coupled with the isochrony proposal that was thus introduced led to numerous empirical studies searching for the acoustic correlates of isochrony on the mora, syllable or foot levels for different languages, which did not provide supporting evidence (Arvaniti, 2009; Nespor et al., 2010; Ramus, Nespor, & Mehler, 1999; Warner & Arai, 2001). Interstress isochrony that was hypothesized to be the basis of stress-timing in English was not found in speech data as interstress interval duration variations were too large to justify isochrony (e.g., Lea, 1974). Similarly, Roach (1982) tested Abercrombie’s (1967) hypotheses by analyzing speech from the six languages he claimed to belong to the proposed rhythmic classes.

The languages could not be differentiated along the syllable duration variability dimension, contrary to Abercrombie's hypothesis that syllable-timed languages had less variability in syllable lengths. As for his hypothesis that syllable-timed languages having irregular interstress intervals, percentage deviation variations of the interstress intervals were actually greater in languages claimed to be stress-timed compared to syllable-timed ones. Beckman's (1982) duration and ratio measurements for the Japanese morae also did not match the values derived from the hypothesis that morae are isochronous.

In an interesting study of feet isochrony in English, Lehiste (1973) attempted to explore the perceptual rather than acoustic correlates of isochrony. Determining the feet durations of 17 sentences consisting of 4 feet each from a total of 20 repetitions of the sentences, Lehiste first showed that durations of feet with same structure and position were highly similar. Yet there was no isochrony across feet types with same position, except for the final feet where isochrony seemed to be present regardless of foot type. She also showed that different readings of syntactically ambiguous sentences resulted in changes in feet durations and suggested that syntax interacts with isochrony through duration. Lehiste then presented the 17-sentence corpus with non-isochronous feet to listeners, half of whom were linguists. Listeners marked the longest and shortest feet on a written version of the sentences. Results indicated that participants performed above chance in only certain cases, which might imply an inability to perceive durational variability of feet. As a control, Lehiste repeated this procedure with noise sequences synthesized from the rhythmic structures extracted from the same sentences. When judging for the longest and shortest intervals in the nonspeech stimuli, listeners performed better compared to the speech stimuli judgements in most of the

analogous conditions. Lehiste interprets these results as suggesting that listeners impose isochrony to speech stimuli and hence evaluated isochrony in linguistic rhythm to be a perceptual rather than acoustic phenomenon. However, it is possible that listeners do not perceive speech stimuli as parsed in feet and thus are unable to determine the feet boundaries and durations. Speech perception might be more driven by syntax and semantics rather than estimations of inter-stress intervals. The noise sequences lack such a complexity and their durational characteristics might hence be perceptually more salient. As such, Lehiste's results might be reflecting the dissimilarities in the sets of stimuli rather than in the perception of the two types of stimuli.

Further exploring the discrepancy between perceived isochrony and actual asynchrony, Morton and colleagues (1976) suggested that the acoustic measures that had been investigated might have been misguided. Based on their observation that regularity of the acoustic onsets of words presented successively do not correspond to a perception of regularity per se, the researchers calculated the intervals between pairs of words that listeners adjusted for perceptual isochrony (Morton, Marcus, & Frankish, 1976). The adjusted intervals for pairs of stimuli were found to be consistent across listeners, yet the variability between different pairs of stimuli suggest the perceptual regularity is influenced by complex characteristics of the stimuli. When pairs of words from the same corpus were presented binaurally, the adjusted intervals but not regular onsets resulted in perceived simultaneity.

Furthermore, Fowler (1979) showed that when asked to produce isochronous syllables, speakers made adjustments that deviated from actual isochrony. Listeners, in turn, indicated that sequences with intervals adjusted by speakers were more rhythmical compared to manipulated versions of the sequences with regular onset

intervals. Hence, the authors from both studies suggested the perceived moments of occurrence of events to be distinct from their actual onsets or peaks; the ‘perceived centers’ (P-centers) of words or syllables are distinguished from the onsets of the words or the stressed vowels as well as the moment at which acoustic cues of stress are the highest (Fowler, 1979; Morton et al., 1976). Fowler’s further analyses on stressed [CV] syllables presented a variant acoustic influence of vowels on the preceding consonant and she suggested that these anticipatory coarticulation cues rather than simple acoustic measures might be markers of the locations of P-centers.

1.3 Linguistic and psycholinguistic studies on linguistic rhythm classes

While the acoustic isochrony account was being undermined, the rhythmic classes account continued receiving interest and being employed in research separately (Arvaniti, 2009). Important supporting evidence came from developmental research. Several findings built the foundations of the idea that the fetus learns from the neonatal auditory environment: the human fetus gives responses to auditory stimuli as early as 19 weeks and characteristics of the speech signal are available to the fetus, which are in turn utilized by the fetus, and later on, the newborn, in habituation-dishabituation paradigms to distinguish speech stimuli (Busnel, Granier-Deferre, & Lecanuet, 1992; Hepper & Shahidullah, 1994).

Neonatal auditory learning was also tested in relation to rhythmical classes. Infants as young as 5 days to 5 months discriminated between low-pass filtered sentences from languages belonging to different rhythmic classes, whereas there was no evidence of differentiation for languages belonging to the same rhythmic class, with the exception of 5-month-olds discriminating their native language from a nonnative dialect or language (Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Jusczyk, &

Johnson, 2000). Results from these infant studies were interpreted as supporting the rhythmic class proposal, however, other interpretations are possible. The studies include only a small set of languages that are ‘prototypical’ of the rhythmic classes, hence this pattern of results might be limited to the prototypical languages. Moreover, differentiation of the prototypical languages does not necessarily indicate that there are rhythmic classes. One possibility is that the languages in these studies are taken from two extreme ends of a continuum (cf. Arvaniti, 2009; Ramus, Nespor, & Mehler, 1999).

In a study with adults, Ramus and Mehler (1999) introduced an alternative method to the one used in the infant studies by Nazzi and colleagues (1998; 2000). From a standpoint similar to that of Nazzi and colleagues (1998), Ramus and Mehler (1999) hypothesize that the temporal organization of vowels, consonants and the according syllable structure is a constituent of rhythm in languages. However, they diverge from Nazzi et al. (1998) by arguing that low-pass filtering that limits segmental information is not sufficient to conclude that language discrimination is achieved based on intact prosodic information. In turn, Ramus and Mehler (1999) propose that resynthesizing speech would enable greater control over acoustic, phonetic, phonotactic and prosodic information present in speech. They resynthesized the sentences from the Nazzi et al. study and showed that French native speakers discriminated between English and Japanese sentences resynthesized with the ‘flat sasasa’ transformation in which all consonants of each sentence were replaced with /s/, all vowels with /a/ and pitch was fixed at a fundamental frequency value of 230 Hz. Since the only cue for discrimination in the ‘flat sasasa’ condition was syllabic rhythm, the authors concluded that languages can be discriminated

rhythmically on this basis, while also acknowledging the need for further studies with more languages to support classes or continuum views.

In an attempt to resolve the linguistic rhythm debate, rhythm typology was revisited with revised methods based on metrics (Arvaniti, 2009). The constitution of these metrics relied on a rationale paralleling the suggestions made by Roach (1982) and Dauer (1983, as cited in Ramus, Nespor, & Mehler, 1999) that stress-timed languages might have more complex syllable structure and display greater variability due to vowel reduction, as was also suggested by Lea (1974). Ramus, Nespor and Mehler (1999) interpreted the infant studies as suggesting that even newborns can perceive rhythmic properties in speech signal, and considered it convenient to shift the focus from complex phonological units to simpler phonetic properties of speech. They introduced several metrics for vocalic and consonantal quantities that would reflect syllable structure and vowel reduction. These metrics are the proportion of vocalic intervals (%V), standard deviations of vocalic intervals (ΔV), and standard deviations of consonantal intervals (ΔC) within an utterance. Ramus and colleagues showed that when plotting languages on a %V and ΔC plane languages clustered according to suggested rhythm typology. Stress-timed languages, for instance, were low in the %V dimension but high in the ΔC dimension. They also suggested that %V by itself served as a predictor of the language discrimination results obtained with both adults and infants. Yet the authors acknowledge that as the number of languages examined in terms of these metrics increase, additional classes or a continuous distribution might emerge (Ramus, Nespor, & Mehler, 1999; Ramus, 2002).

Grabe and Low (2002) propose another index of acoustic variability to account for the still unresolved issue of perceived rhythmic diversity with metrics

based on the variability in successive vocalic and intervocalic intervals. The mean absolute duration difference between each two successive intervals is calculated for strings of vowels and consonants, resulting in the raw Pairwise Variability Index (rPVI). In order to control for speech rate, the duration differences are divided by the mean duration of the two intervals to calculate the normalized mean absolute duration differences, giving the normalized Pairwise Variability Index (nPVI). The normalized index is preferred for vocalic intervals since it better reflects the influence of vowel contractions, whereas the non-normalized index is preferred for consonantal intervals since it better reflects syllable structure. As these metrics are devised to reflect whether a language employs vowel contraction and has complex syllable structure as the basis of the rhythmic diversity, there emerges specific predictions as to where a language would be posited on a graph of vocalic nPVI and consonantal rPVI. Grabe and Low acquired speech recordings of speakers of 18 languages and compared their predictions to the calculated results. The five languages that were prototypically stress- and syllable-timed were differentiated according to the differences in the vocalic nPVI values, yet the consonantal rPVI values failed to predict or differentiate classes for these languages. The prototypically mora-timed Japanese language acquired a vocalic nPVI value that was very similar to that of syllable-timed languages. As for the remainder of the non-prototypical or unclassified languages in the data, the distinguishing characteristic of the vocalic nPVI was lost. The authors interpreted this to suggest that languages might be situated on a continuum of stress- or syllable-timing rather than on a plane of distinct classes.

Arvaniti (2009) discusses several problems concerning the metrics approach. While both $\%V/\Delta C$ (Ramus, Nespor, & Mehler, 1999) and PVI's (Grabe & Low,

2002) success is founded on the clustering of the prototypical languages according to the rhythm typology hypothesis, the two metrics can yield conflicting results for a given language. Aside from some nonprototypical languages being classified differently by the two methods, prototypical languages acquire different positions in a rhythmic class. One of the languages that is classified differently in the two studies is Japanese. Ramus and colleagues (1999) argue that Japanese has a rhythmic class of its own while Grabe and Low (2002) argue that Japanese does not differ from syllable-timed languages and reject the mora-timed rhythmic class. Arvaniti points out that this approach is not at all limited to languages that are ‘harder to classify’. She shows that these scores cannot be a criterion of inclusion or exclusion for the rhythmic classes since in some language pairs, intraclass score differences were much higher than interclass score differences.

Arvaniti (2009) alternatively argues for a conceptualization of rhythm that is language-specific. She maintains that duration and timing are not objective measures of rhythm and that language-specific properties contribute to ‘prominence’ and its organization. She claims that speakers of different languages rely on their native languages’ prosodic parameters when listening to other languages and that this has been the primary mode of hypothetical and empirical classification so far. She argues that speakers of stress-timed languages identify languages employing full vowels in unstressed syllables as syllable-timed languages because they are unaccustomed to such occurrences. However, speakers of these ‘syllable-timed’ languages do not always agree with this classification. Arvaniti argues for placing greater importance to native speakers and how they experience rhythm in their languages with their own parameters. In addition, she emphasizes that studying nonprototypical languages that

challenge the previously suggested classifications would clarify which subjective parameters contribute to different rhythm perceptions.

Arvaniti's (2009) suggestions had been implemented in part by earlier research of Cutler and colleagues (e.g., Cutler, Mehler, Norris, & Segui, 1986; Cutler & Otake, 1994; Kim, Davis, & Cutler, 2008; Murty, Otake, & Cutler, 2007). These series of cross-linguistic research suggested that listeners applied their speech segmentation strategies to both their native and nonnative languages, and that these segmentation strategies were shared across speakers of languages that are hypothesized to be 'rhythmically similar'. Thus, evidence suggests that speakers of English and Dutch segment speech in a stress-based manner (Cutler & Butterfield, 1992; Vroomen, van Zon, & de Gelder, 1996), speakers of French, Spanish, Catalan and Korean segment speech in a syllable-based manner (Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992; Cutler et al., 1986; Kim, Davis, & Cutler, 2008), whereas speakers of Japanese and Telugu segment speech in a mora-based manner (Murty et al., 2007). A similar approach had been employed by Dupoux and colleagues (e.g., 2008) in their research exploring the impact of native language on processing of 'prominences'. Their results suggested that native speakers of French were unable to encode contrastive stress, a phenomenon named 'stress deafness' which arises from the native language not making use of stress in signaling phonological contrasts (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008). Together, speech segmentation and contrastive stress processing studies might provide insight into the perceptual grounds of the long-discussed rhythmic differences as a function of the rhythmic properties of listeners' native languages.

The research of linguistic rhythm, isochrony and rhythmic type emerged from perceptual observations, which received objection on acoustic grounds. However, as

much as the perceptual reality of these phenomena is established, their sources in the external signal are still unknown. To summarize, what seems to be critical are not so much the acoustic properties of the external stimulus but how these are perceived by the listener depending on their linguistic enculturation. For that reason, given the immensity of linguistic diversity, the number of studies that cover this diversity is still scarce and too much centered on Western languages.

1.4 Rhythm in language and music

Speech and music both involve temporal, accentual and phrasal organization of acoustic events, which enables a comparative exploration of prosody and especially rhythm (cf. Lerdahl, 2001; Patel, 2008). Indeed, this proposal had been implemented in the earliest work on linguistic rhythm reviewed above that introduced a system for annotating speech in musical terms and explored the constituents of melody and rhythm in speech from a musical perspective (Steele, 1775). However, as Steele also observed, speech sounds have discrete articulatory boundaries that interact with continuous pitch contours of speech whereas tones in music correspond to preorganized discrete pitches (Jackendoff, 2009). A pattern similar to that of the accentual comparison is observed in the durational comparison as reviewed above. Periodicity is more characteristic of music than language, spoken language lacks the degree of organization in terms of timing that is present in music (Patel, 2008). Yet, it has been seen as plausible to compare rhythm in language and music in terms of meter, defined as the organization of temporal events into abstract patterns. For a more adequate comparison, Patel (2008) argues for a dissociation of ‘temporal periodicity’ and ‘configurational periodicity’ as the means by which meter is constructed in music and language, respectively. Jackendoff, on the other hand,

claims poetry and song to be “hybrids” where language and music exchange principles and approach each other rhythmically by way of meter, an idea that is also maintained by Patel (2008).

Patel (2008) defines configurational periodicity in speech as recurrence of a prosodic unit. He contrasts this with the temporal periodicity of music which is brought about by a more simple, linear sequence of stressed and unstressed events within a more or less isochronous durational grid. In an earlier work, Patel (2003) suggested that linguistic rhythm is brought about via configurational periodicity, defined as organization of stressed and unstressed syllables, as well as via syllable durations and phrase boundaries. In explaining the concept in his later work, Patel (2008) refers to the bases of poetic meters in different languages and remarks that rhythmic configuration need not emerge exclusively from stress. The suggestion is that for any given language, the patterns of poetic meter reflect the rhythmical properties of the language, and thus reveal the prominent prosodic unit of rhythmical configuration. Patel’s (2008) examples mirror the three long-discussed rhythm types and the prototypical languages associated with them. Meter in English poetry involves organization of stresses principally; either disregarding the number of syllables and determining the number of stresses per line or by way of organizing lines in distinct numbers of feet. French and Chinese, on the other hand, have poetic meters that make use of the number of syllables per line. Japanese poetry presents poetic meters based on the number of morae per line. Patel’s (2008) position also appears to be in line with Arvaniti’s (2009) demand for prioritizing how native speakers experience their languages. The diversity in the preferred prosodic units in creating rhythmically organized utterances might indeed be seen as evidence supporting their rhythmical diversity.

Patel collaborated with Daniele (e.g., 2003; 2004) to find evidence for possible influences of linguistic prosody on musical compositions. Specifically, they compared nPVI values measured from sampled speech of English and French to nPVI values calculated from written music of English and French instrumental music compositions. As expected, they found higher vowel variability and an equivalent higher durational variability in musical notes for English compared to French speech and music, respectively (Patel & Daniele, 2003). Further research also provided supporting evidence that when exempt from other factors that override the influence of linguistic prosody, relative variability of note durations in music might be reflective of relative variability of vowel durations, and hence linguistic rhythm (e.g., Daniele & Patel, 2004; Huron & Ollen, 2003).

Jackendoff's (2009) claim of 'hybrid' forms of language and music, on the other hand, depends on what he argues to be the single formal characteristic they share, namely their representability in a metrical grid. The metrical grid system, as we know from music, was first applied to language by Liberman (1975). He plotted syllables as merely sequential beats without other temporal constraints, and built upon it a hierarchical order of relative importance. According to Liberman, the metrical grid implies a metrical organization that is based on different levels of stress superimposed on each other and their order need not depend on isochrony to be perceived rhythmically. In the simplest terms, the first level of hierarchy assigns a beat to each syllable, the second to each stressed syllable, the third to each syllable carrying the primary word stress, and the fourth to the syllable carrying the main stress of each phonological phrase (Nespor et al., 2010; Patel, 2008).

Liberman's (1975) linguistic metrical grid resembles the metrical grid of Western classical music, presenting a common ground but also limiting its

universality. However, Liberman developed the linguistic metrical grid and the rules regulating it specifically for English and admitted that other languages might differ in their metrical organization systems. Still, many studies that will be reviewed below rely on the applicability of the prosodic hierarchy to all languages, revealing levels of differential stress. Even when there is no direct referral to Liberman's metrical grid, it is the applicability of such a structure to both language and music that enables such a comparison. However, as in Patel's (2008) perspective on prosodic units of rhythmical configuration reviewed above, a valid argument is that the commonality of stress hierarchies is not an indication that rhythm in all languages emerges exclusively through the organization of stress beats.

As stated above, the durational irregularity of beats in spoken language raises concerns to the extent to which the metrical grids and stress hierarchies in language and music can be the equivalent of each other (Jackendoff, 2009). One observational evidence for the incompatibility of rhythm in language and music is that listeners often move in synchrony with different beat hierarchies of music, whereas speech does not seem to trigger such synchronized movements (e.g., Patel, 2003).

Accordingly, one line of research comparing rhythm perception in music and language explores how entrainment to speech differs from entrainment to music as measured by participants' tapping patterns to both stimuli. In a recent study that focuses on language and explores hierarchical tapping, monolingual speakers of French and English as well as balanced bilingual French-English speakers were asked to tap to the metrically constructed sentences in English and French. The stimuli consisted of an equal number of monosyllabic words in each sentence with an alternating stress pattern of strong and weak syllables. Across all language groups, participants tapped increasingly more regularly to stimuli with more regular

intervocalic intervals (Lidji, Palmer, Peretz, & Morningstar, 2011). In simplest terms, this finding suggests that listeners can treat speech rhythm like musical rhythm. Yet, there were also some differences between participant groups. While English sentences received more regular tapping in general, monolingual speakers of French tapped less regularly to English sentences. Monolingual speakers of English and bilinguals, on the other hand, tapped more regularly to both languages. More importantly, while French monolinguals mostly tapped to every syllable, English monolinguals mostly tapped at higher hierarchical levels such as every 2 or 4 syllables, with the bilinguals being equally likely to tap at low or high hierarchical levels. These results suggest that abstract stress patterns can be extracted hierarchically from speech with native speakers of stress-timed languages having a heightened ability to track rhythmic properties carried by stress in comparison to native speakers of syllable-timed languages. Another important point is that the findings are in line with phenomena such as stress deafness, as observed in French monolinguals, regarding the influence of native language on rhythm perception reviewed above (Lidji et al., 2011).

The results from the Lidji et al. (2011) study can be interpreted as syllable-based languages making less use of rhythmical properties of speech since stress is a less salient cue in these languages. Another possibility is that it is not necessarily stress per se that forms the metrical regularity in all languages. Since French monolinguals did not perform better with English sentences which obviously had very salient stress cues, these results suggest that a person's language may sensitize or desensitize her to more hierarchical stress cues. In this regard, the French monolinguals' tendency to tap at each syllable might reflect a more or less equal distribution of attention to all syllables regardless of stress; the perceptual salience of

the syllable level might be overriding the salience of higher-level stress. Thus, the results might call into question the validity of a metrical configuration defined by stress in this language. If such patterns built on stress are perceptually unavailable to native speakers of French, the focus should be more on the perceiving person rather than the acoustic properties of the external stimulus.

A method similar to tapping is to ask listeners to reproduce the rhythmic structures of presented linguistic and musical stimuli. In his work exploring rhythmical representation of spoken language, Benadon (2009; 2014) recently pointed out that both theory and research of meter in language are usually interested in stress while leaving out duration. According to his account, metrical compositions of syllables as the rhythmic units of speech also influence temporal positioning of the P-centers of syllables, which vary according to factors such as the preceding consonantal onsets and vowel durations (Benadon, 2014). He devised a method to assign metrical values to syllables according to their durations relative to preceding syllables and derived theoretically plausible musical notation patterns to account for how such stimuli might be perceived (Benadon, 2009). In a following study, Benadon (2014) asked English speaking participants to reproduce rhythms of four-syllable phrases with strong-weak-weak-strong stress patterns and four-tone piano sequences matching the syllabic phrases metrically and temporally. Findings suggested that listeners perceived speech intervals as having more even ratios and assimilated the syllable durations while this was not the case for musical stimuli.

A related way in which researchers addressed the question at stake is to compare metrical speech and music as distractors in a synchronized tapping task. Two recent studies employing this paradigm asked Polish speaking participants to tap to target isochronous tones with constant pitch and intensity, shortly after which

distractor stimuli were presented (Białuńska & Dalla Bella, 2017; Dalla Bella, Białuńska, & Sowiński, 2013). The studies measured the extent to which distractor stimuli interfered with target-synchronized tapping. Music distractors had the same inter-beat interval while speech distractors had the same inter-stress interval with the target's inter-onset interval. Results suggested that stressed syllables significantly interfered with target-synchronized tapping, causing taps to diverge from the target and approach the stressed syllable. However, this interference was not as strong as that caused by the beats in the music distractors. While providing important supporting evidence for shared rhythmic processing of music and metrical speech (cf. Patel, 2003), further studies are required for generalizing these results beyond the Polish language.

All of the studies reviewed in this section either explored the impact of language on music in terms of rhythm (Daniele & Patel, 2004; Patel & Daniele, 2003) or examined the extent to which meter in language and music are equivalent or similar (Benadon, 2009; 2014; Białuńska & Dalla Bella, 2017; Dalla Bella et al., 2013; Lidji et al., 2011). A third line of research goes from music to language and asks whether music can influence perception of speech. In a recent study, Falk and Dalla Bella (2016) explored how auditory-motor coordination can facilitate speech processing. Native German speakers were asked to detect a change in a metrically constructed German sentence with a stress pattern of weak-strong syllables. The comparison sentences were the same except for an occasional single word change (e.g., “heult” changed to “jault”). Participants in the tapping condition heard the sentences while tapping to an isochronous metronome tone. Participants in the cueing-without-tapping condition heard the metronome tones along with the sentences but did not tap to the tones. Participants in the baseline condition heard the

original and manipulated sentences in isolation. Compared to cueing-without-tapping and baseline conditions, an alignment of the tones with stressed syllables resulted in faster reaction times and heightened sensitivity when a change occurred in these. In an attempt to test the generalizability of their results, Falk and colleagues (2017) adapted the same procedure for native French speakers and French language and compared their performances to that of German speakers and stimuli. There was no cueing-without-tapping condition in this study and compared to the German stimuli, the French sentences had a more complex rhythmic structure with a pattern of five syllables with stresses on the second and fifth syllables, keeping inter-stress intervals constant. Compared to baseline, tapping congruently to stressed syllables resulted in enhanced change detection in French speakers as well. In comparison to French speakers, German speakers were better at detecting changes, but there was no information about listeners' reaction times. Yet the heightened performance in French speakers suggests that stress might be processed in similar metrical terms in speakers of both German and French when motor coordination is present.

The studies by Falk and colleagues (2016; 2017) described above were in fact inspired by Cason and Schön's (2012) study, in which French speakers heard isochronous prime tones with an either binary or ternary stress pattern. The rhythmic patterns were constructed such that stressed tones had higher pitch and amplitude compared to unstressed tones. Following four consecutive presentations of the primes, either a bisyllabic or a trisyllabic pseudo-word was presented auditorily and participants were asked to indicate whether a target phoneme presented visually throughout the trial was present in the last syllable of the word or not. The last stressed syllable of the pseudo-words either occurred in the primed position or in mismatch (see Figure 1). While their hypothesis was that temporal and metrical

matching would result in faster detection times, results were mixed. It was temporal (e.g., “on beat”) rather than metrical (e.g., binary/bisyllabic) matching that led to faster reaction times.

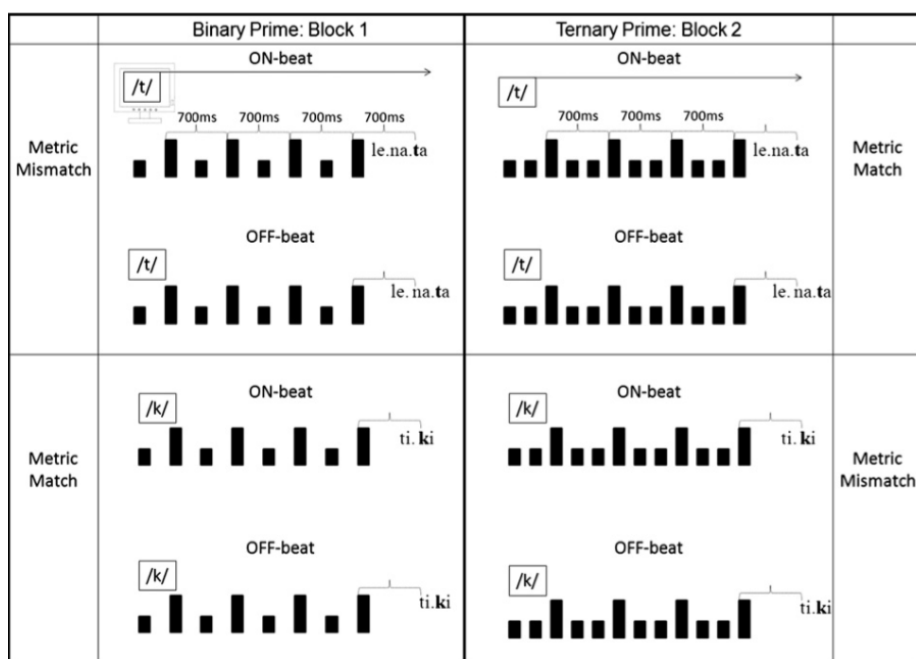


Figure 1. Experimental design of Cason and Schön (2012)

In a follow up study, metrically more complex primes consisting of six or seven isochronous tones were followed by French utterances (Cason, Astésano, & Schön, 2015). The utterances were constructed such that either a match or a mismatch in terms of stress and/or number of tones and syllables could occur. The hypothesis that metrical priming would enhance phoneme detection was not uniformly supported across prime and target combinations. However, the results suggested that primes with a ternary meter could have advantages over the remaining primes that had more complex patterns. The lack of a uniform enhancement of speech perception via musico-rhythmic primes in these two studies suggests that

French speakers do not necessarily perceive linguistic and musico-rhythmic stresses in same terms.

1.5 Rhythm and meter in Turkish speech and poetry

The currently proposed study will be examining standard modern Turkish of Turkey, a language that has not been sufficiently studied with respect to rhythm. In the majority of studies reviewed above, the languages of interest were Western European languages, which limits the generalizability of their findings unless the findings are tested in cross-cultural studies of auditory perception (Patel, 2008). In contrast, Turkish is classified as an Altaic language. Korean and Japanese, to which we referred earlier when discussing linguistic rhythmic categorizations, are two languages that have been arguably claimed to belong to the same linguistic family as Turkish. More definitive and perhaps more important is that Turkish diverges from the most studied Indo-European languages such as the Germanic languages English, German and Dutch, and the Romance languages French, Spanish and Italian. In that sense, the Turkish language will allow us to test claims on music and language interaction in a unique way since it has a different historical ancestry and, more importantly, a unique prosodic, phonological, phonotactic, syntactic structure when compared to any of the Western European languages used so far in linguistic rhythm research.

The Turkish sound system consists of 8 vowels and 20 consonants, from the combinations of which six syllable types emerge: [V], [VC], [CV], [CVC], [VCC] and [CVCC]. As Turkish only allows for a single consonant in the syllable-initial position and for two consonants in the syllable-final position, almost no consonant clusters with more than three consonants exists. In many cases, consonant clusters

are dissolved by way of inserting a vowel before one of the consonants. This strategy is a means of syllabifying the word in accordance with the Turkish syllable structure as consonants unlike vowels do not constitute syllables by themselves. On the other hand, Turkish syllable structure does not allow for multiple vowels in a syllable as well; each vowel corresponds to a single syllable. When two vowels co-occur in a word, a consonant is often inserted between the vowels (Ergenç & Bekar Uzun, 2017). Hence, the Turkish syllable structure is relatively simple and does not present much variability. The simple syllable structure is preserved by way of phonological precautions.

In addition to the aforementioned 8 fundamental vowels which are short, Turkish contains long vowels which occur in both Turkish words and loanwords. Long vowels in loanwords are not pronounced long in all cases; lengthening or shortening occurs based on word-specific morphological rules (Ergenç & Bekar Uzun, 2017; Özsoy, 2004). As for the Turkish words with long vowels, a controversial phoneme is the grapheme <ğ> which has been suggested to have a vowel lengthening function in various contexts. First, given they occur in the same syllable, vowels preceding <ğ> are pronounced as long vowels due to <ğ> being a voiceless phoneme (Ergenç & Bekar Uzun, 2017; Inkelas & Orkun, 2003; Özsoy, 2004; Topbas, 2006). The second case is when <ğ> occurs between two vowels that are essentially the same, causing the vowels to be merged into one long vowel. A third condition is an extension of the first case; when <ğ> in the syllable-final position with a preceding vowel is followed by a consonant, vowel lengthening also applies but results in a shorter vowel compared to the other conditions. Finally, <ğ> occurring between two different vowels results in diphthongs. Aside from <ğ>, an /h/ in the first two conditions also results in vowel lengthening and diphthongs also

emerge when the semivowel /y/ is preceded by a vowel in a syllable (Ergenç & Bekar Uzun, 2017). However, vowel lengthening and diphthongs are rather uncommon, and vowel lengths in Turkish present relatively little variability.

The combination of simple syllable structure and relatively constant vocalic duration observed in Turkish situates Turkish as a syllable-timed language according to several suggestions concerning rhythm typology reviewed above. In fact, Turkish has been claimed to be a syllable-timed language on more intuitive than empirical grounds (e.g., Ramus, 2002; Topbaş, 2006). To my knowledge, there are three studies that calculated rhythm metrics for Turkish in respect to rhythm classification (Mairano & Romano, 2011; Nespor et al., 2010; Romano, 2009). While the results of several metrics from Romano (2009) seem to suggest Turkish could not be classified as either syllable- or stress-timed, the %V value of Turkish in the Nespor et al. (2010) study situates Turkish midway between languages that are claimed to be syllable- and mora-timed (see Figure 2). The ΔC values in the same study, on the other hand, suggests that Turkish could equally likely be categorized along with syllable-timed or even stress-timed languages. Another important characteristic of Turkish that emerged in this study is that for both measures, Turkish has much less standard error compared to the other languages as marked by its small diameter in Figure 2. Mairano and Romano (2011) employ a Control and Compensation Index (CCI), a modified version of the rPVI, and conclude that Turkish is a controlling, i.e., a syllable-timed language.

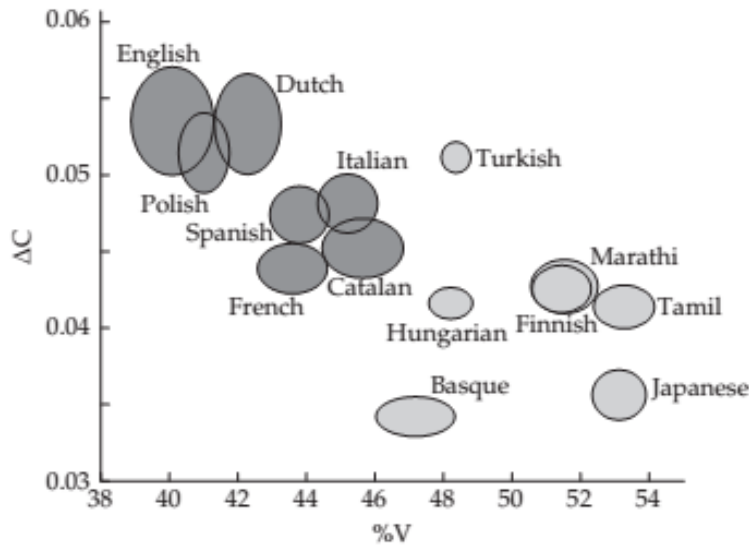


Figure 2. ΔC and %V values for 14 languages

Note: The widths of the ellipses along the two axes represent standard errors of the mean along the axes. Dark ellipses represent head-initial languages, and light ellipses head-final languages (from Nespor et al., 2010).

Also of concern here is the use of stress in Turkish. Turkish is an agglutinative language, with roots accepting a multitude of suffixes in a pre-specified order (Topbaş, 2006). Word stress in Turkish is highly influenced by agglutination; each word has only one main stress which is in word-final position by default. However, when clashed with exceptionally stressed roots or suffixes, word-final stress is replaced by the exceptional stress, e.g., root stress: *pénalti-lar* ‘penalty-PL’; stressed suffix: *bırak-arak* ‘leave-ADV – by leaving’; pre-stressing suffix: *bırak-^ma-dı* → *bırák-ma-dı* ‘leave-NEG-PAST’ (from Inkelas & Orgun, 2003). Exceptional non-final root stress exists in place names, foreign proper names and some loanwords. In compounds that are not phrase-like in their structure, the first member of the compound is stressed whereas the second member has no stress, e.g., *ye-mek* ‘eat-INF’+ *oda* ‘room’ → *yemék oda-sı* ‘dining room’ (from Inkelas & Orgun, 2003). The said roots and compounds preserve their inherent stress and do not comply with word-final stressing upon suffixation. The suffixes that present

exceptional stress are either pre-stressing or multisyllabic with stress on their non-final syllable. Such suffixes go against the word-final stress rule and retain their unique stress location even in case of additional suffixation (Inkelas & Orgun, 2003). Inkelas and Orgun also argue that exceptionally stressed roots and suffixes present a trochaic ('strong-weak') foot. However, exceptions aside, Turkish morphology allows for long strings of syllables without any stress as a result of the morphological nature of the language combined with the word-final stress rule. Moreover, in most cases, word stress is not lexically predetermined and subject to changes through suffixation. Hence, stress in Turkish diverges from that of prototypical stress-timed languages. Additionally, Inkelas and Orgun claim that on the phrase level, the main stress tends to occur on the final word or the preverbal word typically.

The usage of poetic meters in Turkish poetry throughout history is also worth investigating as a possible source of insight regarding rhythm in Turkish. In line with Patel's (2008) claim that poetic meter reflects rhythmical qualities of a given language, there are claims that the poetic meter of Turkish is syllabic. Meters based on syllable numbers have been in use for centuries in Turkish poetry. Folk poets had an important part in keeping syllabic meter alive in Turkish, followed by the revival of syllabic meters by poets favoring Turkish nationalism in the last decades (Aksan, 2013; Dilçin, 2009; Onay, 1996).

Syllabic meter in Turkish has two features; (1) each line has the same number of syllables in a poem, reaching up to 20 syllables and (2) caesuras that are employed to further organize the meter cannot occur mid-word or mid-compound. In fact, Turkish syllabic meter requires the caesuras to occur at phrase boundaries and there are claims that this is the defining feature of the Turkish syllabic meter (Aksan, 2013; Dilçin, 2009; Onay, 1996). There can be up to 5 caesuras in a line, segmenting

the line to syllable groups of 1 to 10 syllables (Dilçin, 2009). However, there are different approaches to the positions of caesuras being regular or variable. While Onay argues that the Turkish syllabic meter dictates that caesuras should always occur in the same position, Aksan and Dilçin suggest that caesuras occurring in different positions might add to the rhythmicity of the poem (see Appendix A). The latter claim might reflect irregularity becoming more and more favorable with the influence of contemporary poetry compared to the 1920s, the years in which Onay first made his claim.

Caesuras occurring at phrase boundaries enable making use of the stresses on the phrase level as the phrase-final words tend to be stressed (Inkelas & Orgun, 2003). An interesting claim in regard to the prominence of stress in the Turkish syllabic meter was posited by Mirşan (1966) in his attempt to manifest the ‘Turkish metrics’. He presents stress analyses of numerous poems written in Turkish as well as several Turkic languages from Central Asia. His claim is that caesuras not only mark the subdivisions of equal number of syllables in a line, but also that these subdivisions are organized to have equal number of stressed and unstressed syllables at each recurrence. However, Aksan (2013) argues that Mirşan’s stress analyses are not completely accurate and that while such regularities in terms of stress are evident in various examples, stress has not been an essential component of the Turkish meter (see Appendix B). While he suggests that the outstanding poets also make use of stress in creating rhythm, especially by way of repetitions, he argues that a more thorough examination of stress patterns in Turkish is required for positing claims concerning the exact role of stress in Turkish poetic meters. Aksan also analyzes several poems and points out that in rhythmically sophisticated poems, poets most

probably take into consideration prosody, phoneme harmony and repetitions of phonemes, morphemes, words or word phrases.

Aksan (2013) points to vowel length as a way of creating more sophisticated rhythmical organization in his analyses of Turkish poems written in syllabic meter. Aksan exemplifies poems in which long vowels occur in regular positions, mostly due to loanwords that provide the suitable material as Turkish vowels are typically short. Vowel length is one way in which syllable length is determined, a characteristic of syllables that is disregarded in syllabic meters. Syllabic meter does not treat syllables differently according to their vocalic or syllabic lengths, and this gives the poets the liberty to make use of vowel length without any constraints. However, there are other poetic meters that are based on an organization of syllables according to syllable length. In fact, one such meter based on vowel length as well as a distinction of open and closed syllables has been adapted to Turkish poetry, the ‘arud meter’. The adaptation of this metrical system to Turkish poetry resulted in a metrical organization that is based on syllable length contrasts rather than number of syllables.

Arud meter was devised in Arabic poetry and was already in use when Al-Khalil ibn Ahmad al-Farahidi systematically presented its governing rules in his 8th century writings (Onay, 1996). Arud meter is a quantitative metrical system that is based on an organization of syllables according to their durations. There are two main syllable categories in arud which are defined by their durations. The syllables that end with a short vowel are named open syllables and can be roughly described as short syllables. The syllables that either contain a long vowel or end with a consonant are named closed syllables and are considered to be long syllables (Dilçin, 2009). When coming into contact with the Arabic poetic meter, Persian poets

acquired its basic set of rules along with the most suitable meters for their language. The syllable types of the Arabic syllable structure can be seen as a subset of the Persian syllable structure, hence as Persian poets started writing in arud meters, new syllable types were introduced to the two essential syllable categories.

Arud meter was introduced to Turkish poetry through Persian poems written in arud meters and it was transmitted along with a vast number of loanwords from both Arabic and Persian, making it possible for Turkish poetry to fully incorporate the arud meter. As in Persian poetry, Turkish poets took the meters that were most suitable for Turkish and invented new meters according to their own language. For centuries, poets strictly applied the rules of arud meter with the help of the loanwords or modified vowel lengths when necessary. However, an era of nationalism in the 20th century led to changes in language and poetry, marked by the strict use of Turkish words and syllabic meters (Aksan, 2013; Dilçin, 2009; Onay, 1996). Interest in arud meters nonetheless continued and poets kept writing poems in arud meters without using loanwords or modifications. This was made possible through limiting the rules of arud meter to those that applied to the Turkish syllable structure, which meant the elimination of the syllable types with long vowels in most part. Thus, open syllables were restricted to the syllable types [V] and [CV] and closed syllables to the syllable types [VC], [CVC], [VCC] and [CVCC]. The resulting metrical structure is essentially based on the syllable structure of Turkish yet goes beyond a simple regulation of the number of syllables per line.

While there are claims that the arud meter and Turkish are not compatible due to the relative infrequency of long vowels in Turkish, long vowels are not actual building blocks of any given line written in arud meter other than occurring in a position specified for a closed syllable. The metrical system is built on the open and

closed syllable distinction with no indication of specific syllable types; on the contrary, numerous syllable types can be used interchangeably as long as the metrical pattern is attained. Thus, theoretically, a line can consist of only the syllable types that do not contain a long vowel and present a valid metrical configuration in terms of arud meter. In fact, Turkish syllable structure, having a limited set of syllable types, might even present a more regular pattern that bypasses the perceptual variability introduced by long vowels such as in Arabic or Persian (cf. Çetin, 1991). The sole determinant of membership to one of the two configurational unit categories in Turkish arud is the syllable ending either with a vowel or a consonant, a quality that is perceptually available to the Turkish speakers as it coincides with the strategies of speech segmentation and syllabification.

In the picture described above, the question then arises as to whether the categorical difference between a syllable ending with a vowel or a consonant is indeed perceived differently. One possible explanation is that the arud meter is a temporal configuration of the P-centers of syllables such that an open syllable with its subsequent syllable gives way to the two P-centers occurring relatively near in time whereas a closed syllable with its subsequent syllable results in the two P-centers being more distant. The increase in the P-center interval duration might emerge as a result of vowel length; the interval to the right of the P-center in a long vowel might be greater than that of a short vowel in terms of duration. Alternatively, an additional consonant that comes in the syllable-final position might add an amount of time to the right of the P-center of a vowel. On the other hand, when the preceding syllable is open, the time interval between two P-centers might be relatively constant as the syllable structures of Arabic, Persian and Turkish do not allow for consonant clustering in the syllable-initial position. Figure 3 depicts how

the syllable structure of the first syllable influences the interval between two subsequent P-centers. Compared to (a) in which the first syllable has a short vowel, the interval is longer in (b) due to the vowel being long and in (c) due to the vowel being followed by a consonant in the first syllable. Listeners might be able to track the P-center intervals and extract the patterns that the arud meters introduce as described.

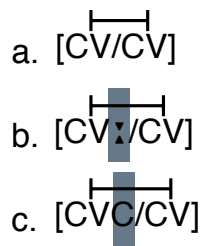


Figure 3. P-center intervals according to syllable structure

It had been suggested that Turkish is a syllable-timed language (e.g., Nespor et al., 2010; Topbaş, 2006), and syllable-timed languages are typically attributed poetic meters based on number of syllables per line (e.g., Patel, 2008), as is the case for Turkish (e.g., Onay, 1996). On the grounds that rhythm is alternation, linguistic rhythm has been defined as alternation of stressed and unstressed syllables. There are claims that syllables of same prosodic quality in syllable-timed languages recur but do not alternate and hence syllable-timed languages lack rhythmicity (see Arvaniti, 2009; Patel, 2008). Others claim that there is rhythm in syllable-timed languages but that this rhythm is also organized by means of stress (e.g., Brown, Pfodresher, & Chow, 2017; Nespor et al., 2010). However, the tailored arud meter used in Turkish as described above suggests that alternation might not necessarily emerge from stress and that the alternation of open and closed syllables might also be a means of introducing rhythmicity in the purportedly syllable-timed Turkish.

1.6 Present study

Accepting the view that rhythm in language is a perceptual phenomenon that is primarily experienced by the native speakers of a given language, the present study focuses on poetic meter as an expression of this experience. Research up to date is rather limited in the variety of languages studied and stress as the prosodic component of interest predominates studies of rhythm in language. Given this picture, bringing alternatives into question might be promising. Turkish is a language that is reportedly syllable-timed and that has not been widely studied in terms of rhythm. Importantly, Turkish has in its metrical inventory an originally quantitative poetic meter that is based on an organization of the syllables according to their structure, namely being open or closed. This metrical system presents an intriguing combination and an alternative to the typically analyzed components of linguistic rhythm, i.e., stress, amplitude and pitch. In order to explore this alternative comparatively with preexisting studies of stress in the same context, the rhythmic priming paradigm employed in Cason and Schön's phoneme detection study (2012) will be modified to reflect the hypothesized rhythmical configuration of arud meter in Turkish.

CHAPTER 2

EXPERIMENT 1

2.1 Experiment 1A

2.1.1 Method

2.1.1.1 Participants

Participants were 48 undergraduate students enrolled in psychology courses at Boğaziçi University who participated in return for course credits. All participants were native speakers of Turkish who were raised in monolingual environments and had normal hearing and normal or corrected-to-normal vision. Data from 6 of 48 participants were excluded due to bilingual development ($N = 1$), tapping or noise during the experimental procedure ($N = 2$) and missing data ($N = 3$). Participants provided informed consents prior to the experimental procedure.

All participants were also speakers of English and the mean age at which they began acquiring English was 8.88 years old. On a scale of 10, participants' mean self-ratings for their current levels of speaking and understanding English were 6.66 and 8.2 points, respectively. Participants also reported a third language ($N = 30$) or a fourth language ($N = 10$) at which they were less proficient compared to English. Eleven participants reported having received music education and/or having practiced an instrument. Mean age at which participants began musical activity was 10.18 years old and mean duration of music education and practice were 3.53 and 4 years, respectively.

2.1.1.2 Stimuli

Target linguistic stimuli consisted of 140 bisyllabic and 140 trisyllabic configurations (see Appendix C). The bisyllabic configurations had a closed-open syllable structure ([CVC/CV]) while the trisyllabic configurations had a closed-closed-open syllable structure ([CVC/CVC/CV]). The last syllable of the stimuli contained a target plosive consonant (/b/, /d/, /g/, /k/, /p/, /t/) as in the Cason & Schön (2012) study, but only in half of the experimental corpus. The remaining half of the corpus contained a nontarget consonant (/c/, /ç/, /l/, /m/, /r/, /s/) in the last syllable. Each consonant, including the target consonants in the final position, occurred only once in a given configuration.

Half of the linguistic stimuli were nonwords constructed from random strings of phonemes violating one or more rules regulating vowels and consonants in Turkish. The other half of the linguistic stimuli were pseudowords complying to the rules regulating vowels and consonants. Pseudowords were generated from Turkish words by substituting phonemes with plausible candidates derived from the existing phonotactical patterns in the Turkish lexicon (e.g., *serpilme* > *serçilme*). In this process, the Turkish module developed for the pseudoword generator software Wuggy (<http://crr.ugent.be/programs-data/wuggy>, Keuleers & Brysbaert, 2010) by Erten, Bozsahin, & Zeyrek (2014) was used whenever possible. In other cases, the pseudowords were generated manually from Turkish words. When the original word fulfilled the specified phonological structure, the logic explained above was followed (e.g., *bildir+ge* > *bilmirge*). When the original word had a mismatching phonological structure, compatible phonemes were inserted to required positions to attain the specified phonological structure (e.g., *bileşke* > *birleşke*).

The experimental corpus was fed to the Microsoft Text-to-Speech software's Turkish speaker at a normal speech rate and the output was recorded at 44.1 Hz sample rate and 32 bit resolution. The linguistic stimuli had a falling pitch contour, as is natural of the software's reading style.

Rhythmic primes consisted of exact repetitions of a bongo timbre sample with long or short inter-onset intervals (IOI) that mimic the intervals between closed and open syllables. These durations were derived from Şayli's (2002) measurements of vowel and consonant durations in Turkish words. Long intervals were matched to the two types of intervals that followed closed syllables in our corpus. For closed syllables followed by an open syllable, the average durations of the vowel (125 ms) and the consecutive consonants (101 ms and 74 ms) in single word environments give an average duration of 300 ms. For closed syllables followed by an open syllable, the average durations of the vowel (125 ms) and the consecutive consonants (101 ms) in single word environments give an average duration of 327 ms. Short intervals were matched to the interval between an open syllable followed by a closed syllable. The average durations of the vowel (129 ms) and the consecutive consonant (74 ms) gave an average duration of 203 ms. As a simplified version of these average durations, an IOI of 300 ms constituted a long interval whereas an IOI of 200 ms constituted a short interval.

Binary primes consisted of two bongo timbres whereas ternary primes consisted of three bongo timbres, each repeated four times. 'Binary on-beat' primes had a 'long-short' interval structure whereas 'ternary on-beat' primes had a 'long-long-short' interval structure, hypothetically mimicking the 'closed-open' and 'closed-closed-open' syllable structures of the bisyllabic and trisyllabic linguistic stimuli, respectively. 'Binary off-beat' primes had a 'short-long' interval structure

Response keys were counterbalanced across participants to control for handedness.

Figure 5 shows the approximate experimental flow.

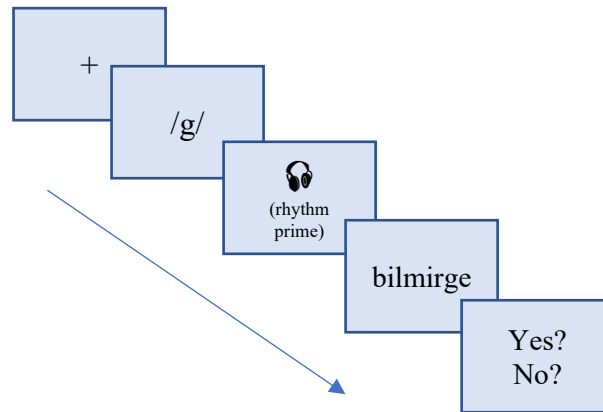


Figure 5. Experimental flow

Trials were blocked according to prime (baseline, binary on-beat, binary off-beat, ternary on-beat and ternary off-beat) and target phoneme (/b/, /d/, /g/, /k/, /t/, /p/). Aside from the baseline blocks which were completed in the beginning of the experimental trials, the prime blocks were presented in randomized order. For each block, half of the linguistic stimuli were bisyllabic and the other half were trisyllabic. Combining binary primes with bisyllabic stimuli and ternary primes with trisyllabic stimuli resulted in the ‘number match’ conditions whereas combining binary primes with trisyllabic stimuli and ternary primes with bisyllabic stimuli resulted in the ‘number mismatch’ conditions.

In each target phoneme block, the linguistic stimuli contained the target phoneme in the last syllable in half of the trials. Such trials constituted the ‘target’ condition. In the other half of the trials, linguistic stimuli with a nontarget consonant in the final syllable were presented (‘nontarget’ condition). There were four trials corresponding to four conditions in each target phoneme block (match/target,

match/nontarget, mismatch/target and mismatch/nontarget) and these occurred in a randomized order. Before each target phoneme block, reminding instructions were presented visually to prepare the participant to a new target phoneme. Each linguistic stimulus was presented only once in a fully balanced design across all experimental trials and conditions.

After the experimental trials, which lasted approximately 35 minutes, participants provided information regarding their language, music and literature backgrounds. Language background and multilingualism were assessed through the Turkish version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007; DBVY-A; Yüksel-Sökmen & Şerifoğlu, n.d.). There were additional questions regarding participants' exposure to and experience with literary works and poems for each language they know. Musical experience was assessed through the Turkish version of the Ollen Musical Sophistication Index (Ollen, 2006) translated by the researcher. The entire experiment took approximately 1 hour. The study was approved by the Ethics Committee of Boğaziçi University (Appendix D).

2.1.1.4 Analysis

Reaction times for correct 'yes' responses in the target conditions were analyzed separately for nonwords and pseudowords. A three-way repeated measures 2 (prime: binary and ternary) x 2 (number: match and mismatch) x 2 (beat: on-beat and off-beat) ANOVA was run on SPSS.

2.1.2 Results

Participants responded correctly to 93.9% of all trials. Correct ‘yes’ responses ($M = 555$ ms, $SE = 33$ ms) were significantly faster than correct ‘no’ responses ($M = 585$ ms, $SE = 36$ ms, $F(1, 41) = 5.19$, $p = .03$, $\eta_p^2 = .11$). Prime trials ($M = 543$ ms, $SE = 32$ ms) resulted in faster reaction times compared to baseline trials ($M = 608$ ms, $SE = 43$ ms, $F(1, 41) = 9.74$, $p = .003$, $\eta_p^2 = .19$). Remaining analyses were conducted on correct ‘yes’ responses for nonwords and pseudowords separately.

2.1.2.1 Nonwords

For nonwords (Figure 6), there was an interaction of prime and number, $F(1, 41) = 16.03$, $p < .001$, $\eta_p^2 = .28$. Post-hoc tests revealed that in binary primes, number mismatch ($M = 565$ ms, $SE = 32$ ms) resulted in faster reaction times compared to number match ($M = 504$ ms, $SE = 31$ ms, $p = .007$, $\eta_p^2 = .17$) whereas in ternary primes, number match ($M = 529$ ms, $SE = 35$ ms) resulted in faster reaction times compared to number mismatch ($M = 589$ ms, $SE = 38$ ms, $p = .002$, $\eta_p^2 = .20$). All remaining effects were nonsignificant ($ps > .05$).

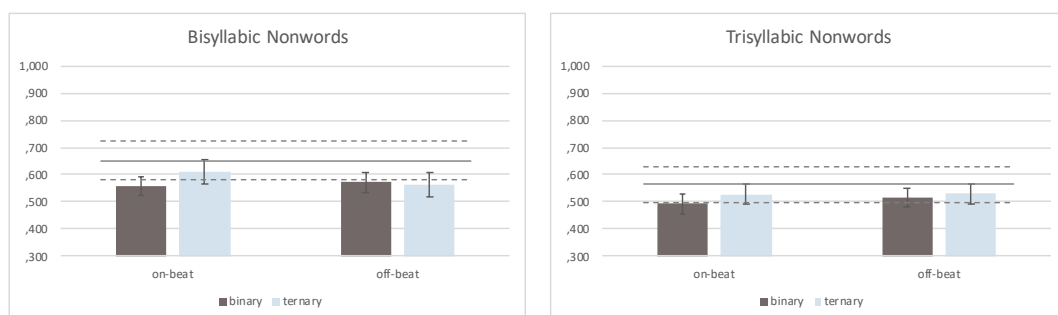


Figure 6. Mean reaction times in phoneme detection for nonwords

Note: Error bars represent ± 1 standard error of the mean. Baseline reaction times are represented with 95% confidence intervals.

The interaction of prime and number indicates a ‘trisyllabic nonword advantage’ as conditions that result in faster reaction times correspond to trisyllabic nonwords for both primes (i.e., number mismatch in binary primes and number match in ternary primes). In order to clarify the priming effects that could be obscured by this ‘trisyllabic nonword advantage,’ bisyllabic and trisyllabic nonwords were also analyzed separately. Two 2 (prime: binary and ternary) X 2 (beat: on-beat and off-beat) repeated measures ANOVAs were run. All effects for both bi- and trisyllabic nonwords were nonsignificant ($ps > .05$).

2.1.2.2 Pseudowords

For pseudowords (Figure 7), there was an interaction of prime and number, $F(1, 41) = 25.84, p < .0001, \eta_p^2 = .39$. Post-hoc tests revealed that in binary primes, number mismatch ($M = 495$ ms, $SE = 36$ ms) resulted in faster reaction times compared to number match ($M = 563$ ms, $SE = 38$ ms, $p < .0001, \eta_p^2 = .34$) whereas in ternary primes, number match ($M = 515$ ms, $SE = 40$ ms) resulted in faster reaction times compared to number mismatch ($M = 589$ ms, $SE = 34$ ms, $p = .005, \eta_p^2 = .18$). All remaining effects were nonsignificant ($ps > .05$).

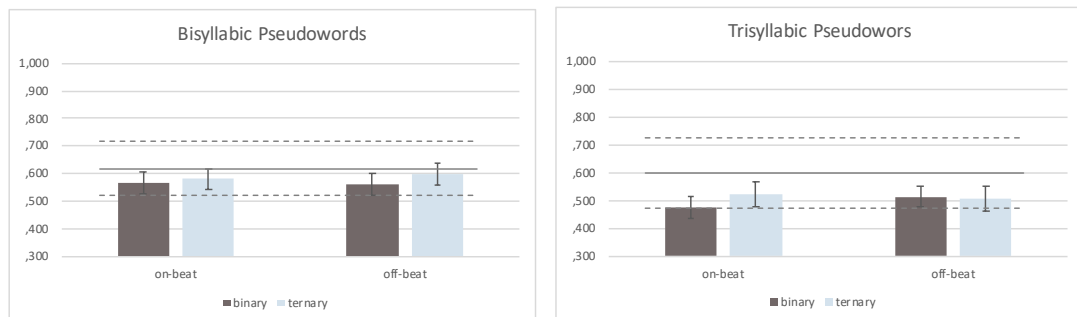


Figure 7. Mean reaction times in phoneme detection for pseudowords

Note: Error bars represent ± 1 standard error of the mean. Baseline reaction times are represented with 95% confidence intervals.

The interaction of prime and number indicates a ‘trisyllabic pseudoword advantage’ as was observed with nonwords. Similarly, bisyllabic and trisyllabic pseudowords were also analyzed separately. Two 2 (prime: binary and ternary) X 2 (beat: on-beat and off-beat) repeated measures ANOVAs were run with no significant findings for either set of pseudowords ($ps > .05$).

2.1.3 Summary

Experiment 1A employed purely durational contrasts in primes to induce expectations about the metrical structures of subsequent linguistic stimuli. Results revealed a ‘trisyllabic advantage’ indicating all conditions yielding only faster reaction times for trisyllabic stimuli, which remain constant across conditions. These findings suggest that the current presentation of primes is insufficient for the intended priming. Two follow-up experiments were designed to explore ways in which primes of Experiment 1A can be made more salient.

2.2 Experiment 1B

2.2.1 Method

2.2.1.1 Participants

Participants were 45 undergraduate students enrolled in psychology courses at Boğaziçi University who received course credit for their participation in the study. All participants were native speakers of Turkish who were raised in monolingual environments and had normal hearing and normal or corrected-to-normal vision. Data from 4 of 45 participants were excluded due to bilingual development ($N = 2$) and tapping during the experimental procedure ($N = 2$). Participants provided informed consents prior to the experimental procedure.

All participants were also speakers of English and the mean age at which they began acquiring English was 9.59 years old. On a scale of 10, participants' mean self-ratings for their current levels of speaking and understanding English were 6.81 and 8.61 points, respectively. Participants also reported a third language ($N = 19$), a fourth language ($N = 6$) or a fifth language ($N = 1$) at which they were less proficient compared to English. Eighteen participants reported having received music education and/or having practiced an instrument. Mean age at which participants began musical activity was 9.08 years old and mean duration of music education and practice were 3.47 and 3.18 years, respectively.

2.2.1.2 Stimuli

Target linguistic stimuli consisted of the same 140 bisyllabic and 140 trisyllabic configurations prepared for Experiment 1A (see Appendix C), except that the stimuli were controlled for unnatural phoneme or syllable durations and were corrected through Audacity or Praat when needed.

Rhythmic primes were repetitions of a bongo timbre organized in long or short intervals as in Experiment 1A and primes were repeated four times. Importantly, in order to make grouping patterns more salient to the listeners, bongo pulses followed by long intervals were accented by 10 dB. The baseline configuration was the same as in Experiment 1A except that the bongo pulses had an intensity level that is intermediary of the long- and short-interval bongo pulses in the primes. Figure 8 depicts the stimuli and the conditions emerging from their combinations.

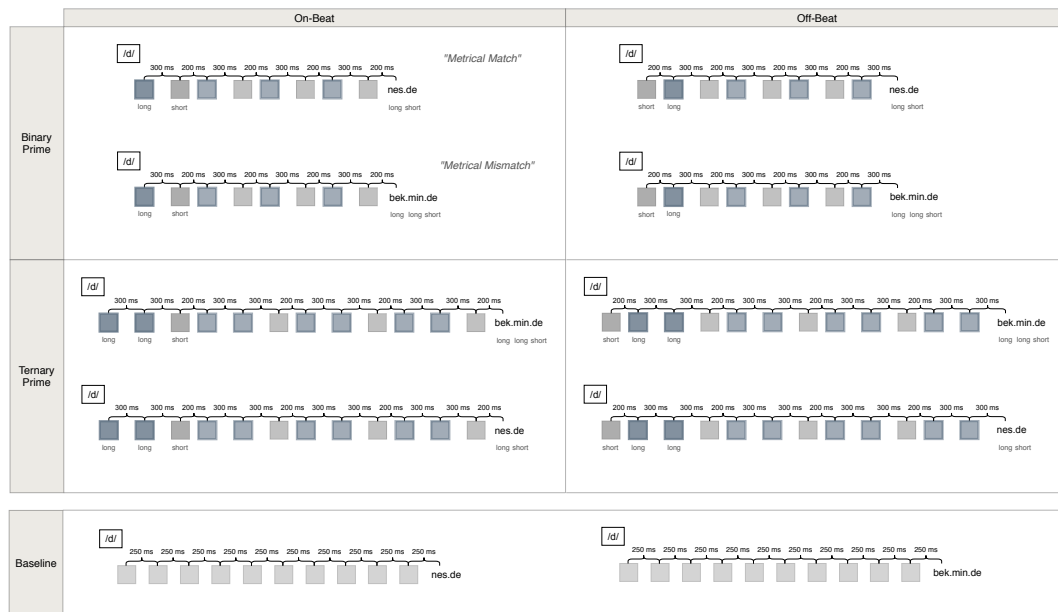


Figure 8. Experimental design of Experiment 1B

2.2.1.3 Procedure and design

Participants completed a phoneme detection task consisting of 12 practice trials and 240 experimental trials, lasting approximately 35 minutes. The remaining procedure was the same as in Experiment 1A. After the experimental trials, participants provided information regarding their language, music and literature backgrounds. The entire experiment took approximately 1 hour. The study was approved by the Ethics Committee of Boğaziçi University (Appendix D).

2.2.1.4 Analysis

Reaction times for correct ‘yes’ responses in the target conditions were analyzed separately for nonwords and pseudowords. A three-way repeated measures 2 (prime: binary and ternary) x 2 (beat: on-beat and off-beat) x 2 (number match: match and mismatch) ANOVA was run on SPSS.

2.2.2 Results

Participants ($N = 41$) responded accurately to 93.69% of all trials. Correct “yes” responses ($M = 637$ ms, $SE = 30$ ms) resulted in faster reaction times compared to correct “no” responses ($M = 661$ ms, $SE = 33$ ms, $F(1, 40) = 5.83$, $p = .02$, $\eta_p^2 = .13$). Primes ($M = 623$ ms, $SE = 30$ ms) resulted in faster reaction times compared to baseline ($M = 694$ ms, $SE = 35$ ms, $F(1, 40) = 12.59$, $p = .001$, $\eta_p^2 = .24$). Remaining analyses were conducted on nonwords and pseudowords separately.

2.2.2.1 Nonwords

For nonwords (Figure 9), there was a main effect of beat, $F(1, 40) = 6.90$, $p = .01$, $\eta_p^2 = .15$. On-beat condition ($M = 614$ ms, $SE = 29$ ms) resulted in faster reaction times compared to off-beat condition ($M = 656$ ms, $SE = 33$ ms). There was a main effect of number, $F(1, 40) = 9.88$, $p = .003$, $\eta_p^2 = .20$. Number match ($M = 620$ ms, $SE = 30$ ms) resulted in faster reaction times compared to number mismatch ($M = 650$ ms, $SE = 31$ ms). There was an interaction of prime and number, $F(1, 40) = 15.81$, $p < .001$, $\eta_p^2 = .28$. In ternary primes, number match ($M = 603$ ms, $SE = 33$ ms) resulted in faster reaction times compared to metric mismatch ($M = 692$ ms, $SE = 37$ ms, $F(1, 40) = 30.51$, $p < .0001$, $\eta_p^2 = .43$). Metric match and mismatch did not differentiate binary primes. All remaining effects were nonsignificant ($ps > .05$).

Bisyllabic and trisyllabic nonwords were also analyzed separately. Separate 2 (prime: binary and ternary) X 2 (beat: on-beat and off-beat) repeated measures ANOVAs were run for each set of stimuli. For bisyllabic nonwords, there was a main effect of prime, $F(1, 40) = 7.07$, $p = .01$, $\eta_p^2 = .15$. Binary primes ($M = 636$ ms, $SE = 31$ ms) resulted in faster reaction times compared to ternary primes ($M = 692$ ms, $SE = 37$ ms). There was also a main effect of beat, $F(1, 40) = 7.41$, $p = .01$, $\eta_p^2 = .16$.

On-beat condition ($M = 634$ ms, $SE = 32$ ms) resulted in faster reaction times compared to off-beat condition ($M = 694$ ms, $SE = 37$ ms). The interaction term was nonsignificant. For trisyllabic nonwords, none of the effects were significant ($ps > .05$).

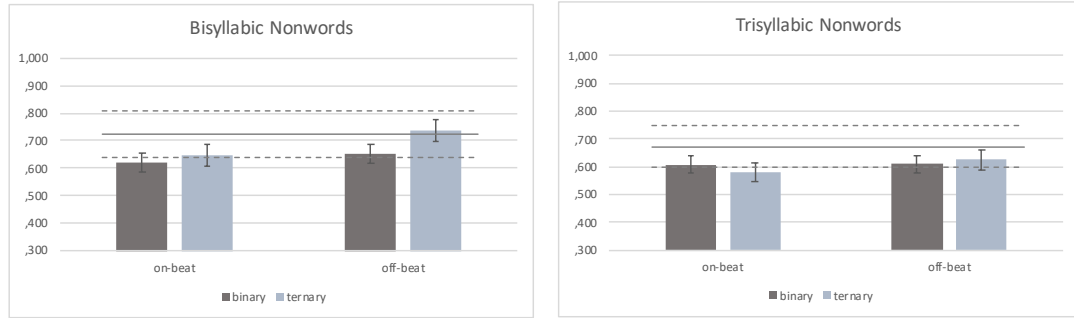


Figure 9. Mean reaction times in phoneme detection for nonwords

Note: Error bars represent ± 1 standard error of the mean. Baseline reaction times are depicted with 95% confidence intervals.

2.2.2.2 Pseudowords

For pseudowords (Figure 10), there was an interaction of prime and number, $F(1, 40) = 23.49$, $p < .0001$, $\eta_p^2 = .37$. In binary primes, metric mismatch ($M = 581$ ms, $SE = 35$ ms) resulted in faster reaction times compared to metric match ($M = 649$ ms, $SE = 36$ ms, $F(1, 40) = 10.55$, $p = .002$, $\eta_p^2 = .21$). In ternary primes, metric match ($M = 589$ ms, $SE = 34$ ms) resulted in faster reaction times compared to metric mismatch ($M = 644$ ms, $SE = 35$ ms, $F(1, 40) = 7.24$, $p = .01$, $\eta_p^2 = .15$). All remaining effects were nonsignificant ($ps > .05$).

Bisyllabic and trisyllabic pseudowords were also analyzed separately.

Separate 2 (prime: binary and ternary) X 2 (beat: on-beat and off-beat) repeated measures ANOVAs were run with no significant findings for either set of pseudowords ($ps > .05$).

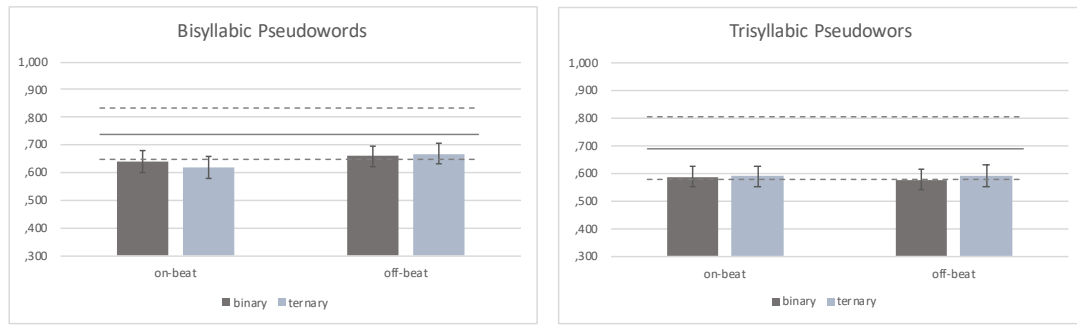


Figure 10. Mean reaction times in phoneme detection for pseudowords
Note: Error bars represent ± 1 standard error of the mean. Baseline reaction times are depicted with 95% confidence intervals.

2.2.3 Summary

Experiment 1B employed durational as well as amplitudinal contrasts in primes to induce expectations about metrical structures of subsequent linguistic stimuli.

Results revealed a ‘trisyllabic advantage’ in pseudowords and faster reaction times under number match and on-beat conditions in bisyllabic nonwords. These findings suggest emphasizing prime structure through intensity might enable metrical priming.

2.3 Experiment 1C

2.3.1 Method

2.3.1.1 Participants

Participants were 39 undergraduate students enrolled in psychology courses at Boğaziçi University and received course credit for their participation in the study.

All participants were native speakers of Turkish who were raised in monolingual environments and had normal hearing and normal or corrected-to-normal vision.

Data from 3 of 39 participants were excluded due to multilingual experience as determined by LEAP-Q. Participants provided informed consents prior to the experimental procedure.

All participants were also speakers of English and the mean age at which they began acquiring English was 9.19 years old. On a scale of 10, participants' mean self-ratings for speaking and understanding English were 6.81 and 8.22 points, respectively. Participants also reported a third language ($N = 22$), a fourth language ($N = 8$) or a fifth language ($N = 2$) at which they were less proficient compared to English. Eleven participants reported having received music education and/or having practiced an instrument. Mean age at which participants began musical activity was 11.36 years old and mean duration of music education and practice were 4.78 and 3.67 years, respectively.

2.3.1.2 Stimuli

Target linguistic stimuli were the same as in Experiment 1B, except only nonwords were used in this experiment. Pseudowords were not employed in order to reduce the length of the experimental procedure and due to their robustness to priming in Experiments 1A and 1B. Rhythmic primes were exact repetitions of a bongo pulse organized in long or short intervals as in Experiment 1A but were repeated eight rather than four times to enhance entrainment. The baseline sequence consisted of a total of twenty bongo pulses with a constant intermediary IOI of 250 ms.

2.3.1.3 Procedure and design

Participants completed a phoneme detection task consisting of 8 practice trials and 120 experimental trials, which lasted approximately 25 minutes. The procedure was the same as in Experiment 1A. After the experimental trials, participants provided information regarding their language, music and literature backgrounds. The entire

experimental procedure lasted approximately 50 minutes. The study was approved by the Ethics Committee of Boğaziçi University (Appendix D).

2.3.1.4 Analysis

Reaction times for correct ‘yes’ responses in the target conditions were analyzed separately for nonwords and pseudowords. A three-way repeated measures 2 (prime: binary and ternary) x 2 (number: match and mismatch) x 2 (beat: on-beat and off-beat) ANOVA was run on SPSS.

2.3.2 Results

Participants ($N = 36$) responded accurately to 89.19% of all trials. Reaction times in correct “yes” responses ($M = 784$ ms, $SE = 36$ ms) and correct “no” responses ($M = 785$ ms, $SE = 38$ ms) were not differentiated, $F(1, 35) = .001$, $p = .97$. Primes ($M = 754$ ms, $SE = 37$ ms) resulted in faster reaction times compared to baseline ($M = 931$ ms, $SE = 42$ ms, $F(1, 35) = 22.23$, $p < .0001$, $\eta_p^2 = .42$).

There was a main effect of prime, $F(1, 35) = 4.78$, $p = .04$, $\eta_p^2 = .12$. Binary primes ($M = 732$ ms, $SE = 40$ ms) resulted in faster reaction times compared to ternary primes ($M = 776$ ms, $SE = 37$ ms). All remaining effects were nonsignificant ($ps > .05$). Mean reaction times grouped by condition are presented in Figure 11.

Bisyllabic and trisyllabic nonwords were also analyzed separately. Separate 2 (prime: binary and ternary) X 2 (beat: on-beat and off-beat) repeated measures ANOVAs were run for each set of stimuli. For bisyllabic nonwords, there was a main effect of number, $F(1, 35) = 6.46$, $p < .05$. Number match ($M = 740$ ms, $SE = 41$ ms) resulted in faster reaction times compared to number mismatch ($M = 802$ ms, $SE =$

42 ms). All remaining effects were nonsignificant ($ps > .05$). For trisyllabic nonwords, all effects were nonsignificant ($ps > .05$).

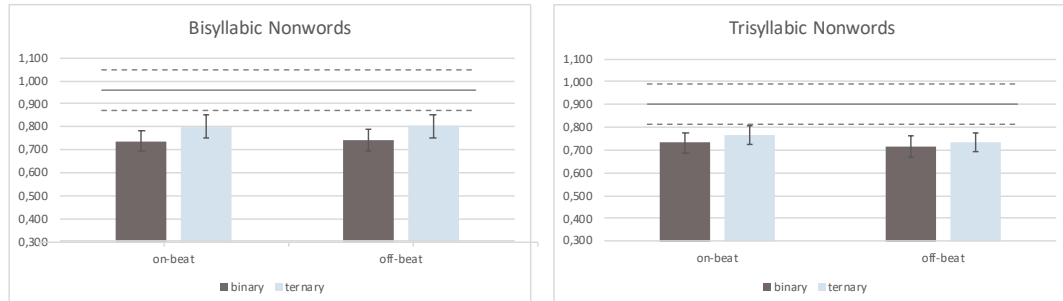


Figure 11. Mean reaction times in phoneme detection

Note: Error bars represent ± 1 standard error of the mean. Baseline reaction times are depicted with 95% confidence intervals.

2.3.3 Summary

Experiment 1C employed durational contrasts in primes and doubled their number of repetitions to induce expectations about the metrical structures of subsequent linguistic stimuli. Results revealed faster reaction times under number match conditions in bisyllabic nonwords. These findings suggest emphasizing prime structure through increased number of repetitions might enable metrical priming, however the priming effect is not as strong as in Experiment 1B.

CHAPTER 3

EXPERIMENT 2

3.1 Method

3.1.1 Participants

Participants were 43 undergraduate students enrolled in psychology courses at Boğaziçi University who received course credit for their participation in the study. All participants were native speakers of Turkish who were raised in monolingual environments and had normal hearing and normal or corrected-to-normal vision. Participants provided informed consents prior to the experimental procedure.

All participants were also speakers of English and on average, they began acquiring English when they were 9.5 years old. On a scale of 10, participants' mean self-ratings for their current levels of speaking and understanding English were 6.23 and 8.13 points, respectively. Participants also reported a third ($N = 26$) or a fourth language ($N = 7$) at which they were less proficient compared to English. Thirteen participants reported having received music education and/or having practiced an instrument. Mean age at which participants began musical activity was 11.22 years old and mean duration of music education and practice were 3.77 and 3.89 years, respectively.

3.1.2 Stimuli

Target linguistic stimuli were the same as in Experiment 1B. Rhythmic primes were exact repetitions of a bongo pulse organized in long or short intervals as in Experiment 1A and primes were repeated four times. However, the organization of primes and the emerging grouping patterns were revised and gestalt grouping

conditions were introduced. ‘Binary meter’ primes had a ‘long-short’ interval structure whereas ‘ternary meter’ primes had a ‘long-long-short’ interval structure as in Experiment 1A. ‘Binary gestalt’ primes had a ‘short-long’ interval structure while ‘ternary gestalt’ primes had a ‘short-short-long’ interval structure, giving rise to a gestalt grouping after every second and third bongo pulse, respectively (Figure 12). The baseline configuration was the same as in Experiment 1A.

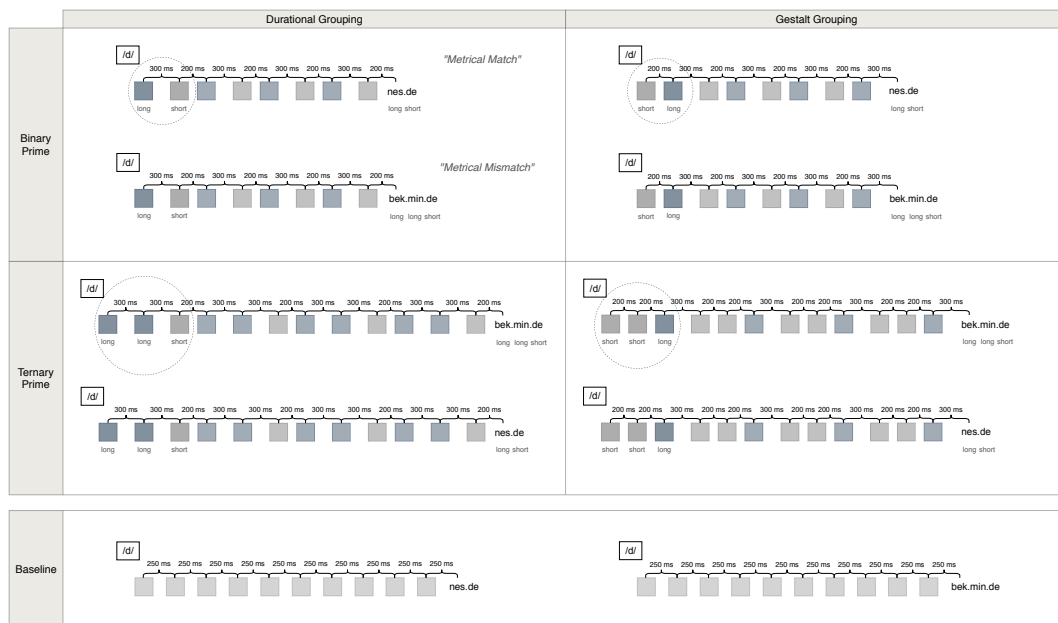


Figure 12. Experimental design of Experiment 2

3.1.3 Procedure and design

Participants completed a phoneme detection task consisting of 12 practice trials and 240 experimental trials, lasting approximately 35 minutes. The remaining procedure was the same as in Experiment 1A. After the experimental trials, participants provided information regarding their language, music and literature backgrounds. The entire experimental procedure took approximately 1 hour. The study was approved by the Ethics Committee of Boğaziçi University (Appendix D).

Trials were blocked according to lexicality (nonword and pseudoword), prime (baseline, binary meter, binary gestalt, ternary meter and ternary gestalt) and target phoneme (/b/, /d/, /g/, /k/, /t/, /p/). Meter primes constituted the ‘duration’ conditions whereas gestalt primes constituted the ‘gestalt’ conditions. For each block, half of the linguistic stimuli were bisyllabic and the other half were trisyllabic.

Combining binary primes with bisyllabic stimuli and ternary primes with trisyllabic stimuli resulted in the ‘number match’ conditions whereas combining binary primes with trisyllabic stimuli and ternary primes with bisyllabic stimuli resulted in the ‘number mismatch’ conditions. The grouping conditions of gestalt and duration along with the number conditions enabled examining whether the hypothesized metrical organization of open and closed syllables merely emerge from the number of syllables or have a durational basis in addition.

3.1.4 Analysis

Reaction times for correct ‘yes’ responses in the target conditions were analyzed separately for nonwords and pseudowords. A three-way repeated measures 2 (prime: binary and ternary) x 2 (number match: match and mismatch) x 2 (grouping: duration and gestalt) ANOVA was run on SPSS.

3.2 Results

Participants ($N = 43$) responded accurately to 92.71% of all trials. Correct “yes” responses ($M = 686$ ms, $SE = 33$ ms) resulted in faster reaction times compared to correct “no” responses ($M = 731$ ms, $SE = 35$ ms, $F(1, 42) = 9.44$, $p = .004$, $\eta_p^2 = .18$). Primes ($M = 667$ ms, $SE = 34$ ms) resulted in faster RT compared to baseline

($M = 760$ ms, $SE = 35$ ms, $F(1, 42) = 14.44$, $p < .001$, $\eta_p^2 = .26$). Remaining analyses were conducted on nonwords and pseudowords separately.

3.2.1 Nonwords

For nonwords (Figure 13), there was an interaction of prime and number ($F(1, 42) = 10.28$, $p = .003$, $\eta_p^2 = .20$). In binary primes, number mismatch ($M = 629$ ms, $SE = 35$ ms) resulted in faster reaction times compared to number match ($M = 698$ ms, $SE = 38$ ms, $F(1, 42) = 11.76$, $p = .001$, $\eta_p^2 = .22$). Number match and mismatch did not differentiate ternary primes. All remaining effects were nonsignificant ($ps > .05$).

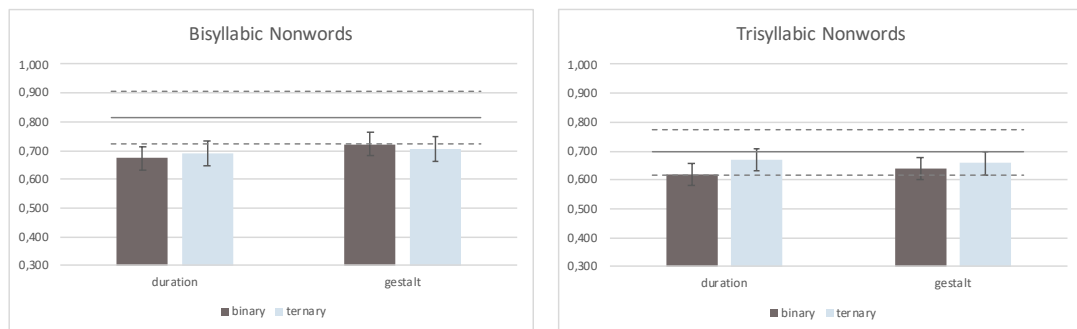


Figure 13. Mean reaction times in phoneme detection for each condition
Note: Error bars represent ± 1 standard error of the mean for reaction times. Baselines are depicted with 95% confidence intervals.

Bisyllabic and trisyllabic nonwords were also analyzed separately. Separate 2 (prime: binary and ternary) X 2 (beat: on-beat and off-beat) repeated measures ANOVAs were run no significant findings for either set of nonwords ($ps > .05$).

3.2.2 Pseudowords

For pseudowords (Figure 14), there was an interaction of prime and number ($F(1, 42) = 57.27$, $p < .0001$, $\eta_p^2 = .58$). In binary primes, number mismatch ($M = 621$ ms, $SE = 37$ ms) resulted in faster reaction times compared to number match ($M = 708$

ms, $SE = 42$ ms, $F(1, 42) = 19.55$, $p < .0001$, $\eta_p^2 = .32$). In ternary primes, number match ($M = 623$ ms, $SE = 44$ ms) resulted in faster reaction times compared to number mismatch ($M = 718$ ms, $SE = 46$ ms, $F(1, 42) = 24.00$, $p < .0001$, $\eta_p^2 = .36$). All remaining effects were nonsignificant ($ps > .05$).

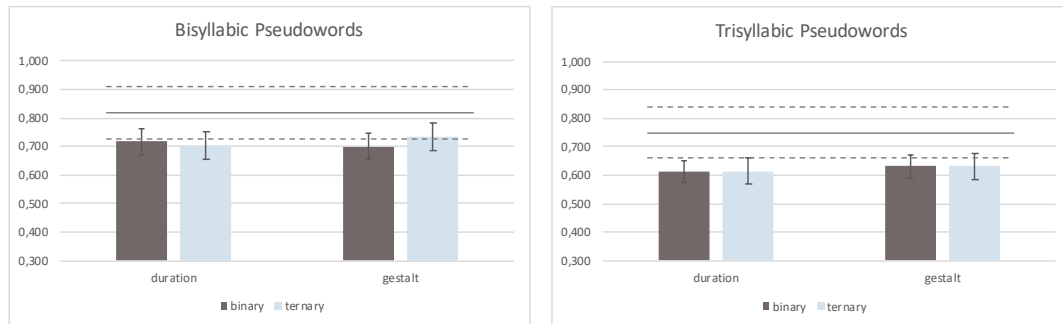


Figure 14. Mean reaction times in phoneme detection for each condition
Note: Error bars represent ± 1 standard error of the mean for reaction times. Baselines are depicted with 95% confidence intervals.

Bisyllabic and trisyllabic pseudowords were also analyzed separately.

Separate 2 (prime: binary and ternary) X 2 (beat: on-beat and off-beat) repeated measures ANOVAs were run no significant findings for either set of pseudowords ($ps > .05$).

3.2.3 Summary

Experiment 2 employed durational and gestalt grouping patterns in primes to induce expectations about the metrical structures of subsequent linguistic stimuli. Results revealed a 'trisyllabic advantage' in nonwords and pseudowords. As in Experiment 1A, these findings suggest that the current presentation of primes is insufficient for the intended priming. Table 1 presents a summary of results from Experiments 1A, 1B, 1C and 2.

Table 1. Summary of ANOVA Results for Experiments 1A, 1B, 1C and 2

	Main effects				Interactions			
	Primes	Prime (binary and ternary)	Beat (on-beat and off-beat)	Number (match and mismatch)	Prime *Beat	Prime *Number	Number *Beat	Prime *Beat *Number
Experiment 1A	LS/SL – LLS/SLL	ns.	ns.	ns.	ns.	'trisyllabic advantage'	ns.	ns.
Experiment 1B (1A with accented primes)	LS/SL – LLS/SLL	ns.	on-beat < off-beat	match < mismatch	ns.	'trisyllabic advantage'	ns.	ns.
Experiment 1C (1A with repeated primes)	LS/SL – LLS/SLL	binary < ternary	ns.	ns.	ns.	ns.	ns.	ns.
Experiment 2	LS/SL – LLS/SSL	ns.	ns.	ns.	ns.	'trisyllabic advantage'	ns.	ns.

CHAPTER 4

DISCUSSION

A total of four experiments were conducted to explore possible ways in which durational constituents of linguistic rhythm can be primed in Turkish. The rhythmic primes were constructed with the aim of inducing expectations concerning syllable numbers and durations, i.e., possible counterparts of meter in Turkish. Reaction times were measured in a phoneme detection task and were analyzed in relation to different levels of metrical match or mismatch between primes and target linguistic stimuli in order to explore whether metrical match enhances speech processing.

In Experiment 1, Turkish native speakers were presented primes consisting of bongo pulses with long-short or long-long-short intervals or their reversals, followed by nonwords or pseudowords with [CVC/CV] or [CVC/CVC/CV] patterns. In Experiment 1A, primes were repeated four times prior to presentation of target linguistic stimuli. Results suggest a trisyllabic nonword and pseudoword advantage as well as no significant influence of metrical match or mismatch on reaction times. Experiments 1B and 1C revised Experiment 1A in order to make metrical organization of primes more evident since primes in the latter were hard for listeners to group (e.g., long-long-short or short-long-long presented four times in a loop). In Experiment 1B, bongo pulses with long intervals were emphasized through an increase in intensity. Results were similar to Experiment 1A's findings with faster reaction times for number match and on-beat conditions. In Experiment 1C, the number of prime repetitions were doubled (i.e., each binary or ternary prime was presented eight instead of four times in a loop). Results revealed faster reaction times for prime match conditions for bisyllabic nonwords. In Experiment 2, Turkish native

speakers were presented primes with durational (long-short or long-long-short intervals) or gestalt (short-long or short-short-long intervals) grouping patterns and nonwords or pseudowords with [CVC/CV] or [CVC/CVC/CV] patterns. Similar to Experiment 1A, results suggest a trisyllabic nonword and pseudoword advantage and no significant influence of metrical match or mismatch on reaction times.

In all experiments except for Experiment 1C, correct ‘yes’ responses received faster reaction times compared to correct ‘no’ responses (see Appendix F for additional results from correct ‘no’ responses). The same pattern of reaction times was also evident in the Cason and Schön (2012) study. The authors’ interpretation was that rejections require an exhaustive search for the target phoneme in the phoneme detection task. This explanation also befits the pattern that emerged in the current study. The exceptional case of Experiment 1C will be discussed in detail in Section 4.5. Additionally, in all experiments, primes received faster reaction times compared to baseline, suggesting that the durational contrasts employed in primes present an advantage in processing the linguistic stimuli. The exact nature of this advantage will be discussed in relation to different prime structures in the upcoming sections.

4.1 Effects of number match

One of the main research questions in the present study was whether syllable number in linguistic stimuli can be primed rhythmically. Aside from the ‘trisyllabic advantage’ that emerged primarily in pseudowords, which will be discussed in Section 4.4, the most often observed effect was the effects of a match between the number of units that constitute primes and subsequent linguistic stimuli. In the overall analyses, this corresponds to the main effect of number, with number match

resulting in faster reaction times compared to number mismatch. In the separate analyses, this corresponds to the main effect of prime, with binary primes resulting in faster reaction times compared to ternary primes in bisyllabic stimuli, and vice-versa for trisyllabic stimuli. For the ease of interpretation, main effect of number in the overall analyses and main effect of prime in separate analyses will be denoted as ‘number match’ conditions.

Number match resulted in faster reaction times in nonwords in the overall analysis for Experiment 1B and in bisyllabic nonwords in the separate analyses for Experiments 1B and 1C. Taken together, these results suggest that under certain conditions, primes consisting of two or three bongo timbres can generate expectations for a subsequent linguistic stimulus consisting of two or three syllables, respectively. In comparison to Cason and Schön (2012)’s null findings regarding such a match, which they termed ‘metrical match,’ these results might be first to show successful numerical priming as indexed by a behavioral measure. Hence, it is possible to suggest that syllable number information is perceptually available to speakers of Turkish and can be primed cross-modally (i.e., from a pure rhythmic to a phonemic modality). A closer look at the nature of primes in Experiments 1A, 1B and 1C will provide a better understanding of the exact influence of number match.

In Experiment 1A, primes consisted of exact four-time repetitions of bongo pulses which had binary or ternary groupings regarding inter-stimulus time gaps. Under these conditions, a combination of binary primes with bisyllabic nonwords or ternary primes with trisyllabic nonwords did not result in the predicted enhancement in reaction times. In Experiment 1A, prime boundaries that cued the number of units might have been indefinite due to acoustically invariant presentations or an insufficient number of repetitions and hence might have hindered number match or

mismatch influencing reaction times. Consequently, primes revised to address these issues in Experiments 1B and 1C resulted in more evident priming

In Experiment 1B, long-interval bongo pulses were emphasized through an increase in intensity levels compared to short-interval bongo pulses, thus accentuating the binary or ternary grouping. In this experiment, both binary-bisyllabic and ternary-trisyllabic nonword combinations presented more rapid reaction times. As intended, the intensity emphasis placed on long-interval bongo pulses might have made prime boundaries more evident and thus made information regarding the number of units in each prime more salient to listeners. In comparison to the combinatorial contrast employed in Experiment 1B, Experiment 1A targeted duration in isolation, and number match did not result in priming in the latter experiment. Hence, it is possible to attribute the enhanced reaction times in Experiment 1B to the intensity contrast rather than the duration contrast. However, the fact that the binary on-beat prime in Experiment 1B preserves the durational pattern of Experiment 1A also suggests that this pattern does not disrupt the facilitation presented by the intensity emphasis. Additionally, as on-beat and off-beat primes collectively receive faster reaction times, it is possible to conclude that both primes are equally successful at priming syllable numbers, lending support to a possible durational as well as numerical priming.

On the other hand, Experiment 1C preserves the solely durational contrast of Experiment 1A but doubles the number of prime repetitions to achieve heightened perceptual salience of prime structure through enhanced entrainment. As with Experiment 1B but in contrast to Experiment 1A, the separate analyses for Experiment 1C yielded faster reaction times in binary-bisyllabic combinations. The finding that number match enhanced performance without an intensity emphasis

might suggest that priming syllable number need not necessitate an additional intensity contrast (i.e., the solely durational contrast was sufficient). However, the fact that, in Experiment 1C, main effect of number reached significance only when bisyllabic nonwords were analyzed separately also suggests that the effect might reflect another property of the binary-bisyllabic combination.

One such possibility is that while primes were repeated eight times instead of four to enhance entrainment, listeners might have perceived the eight repetitions of the binary primes in a way that overrides their durational contrasts. That is, while the gestalt grouping of the binary off-beat prime (short-long) might have become more evident with eight repetitions, the durational grouping of the binary on-beat prime (long-short) repeated eight times might have acquired a perception such that aside from the first and the last bongo pulse, the remaining bongo pulses being perceived as seven repetitions of gestalt-grouped bongo pulses of two (Figure 15).

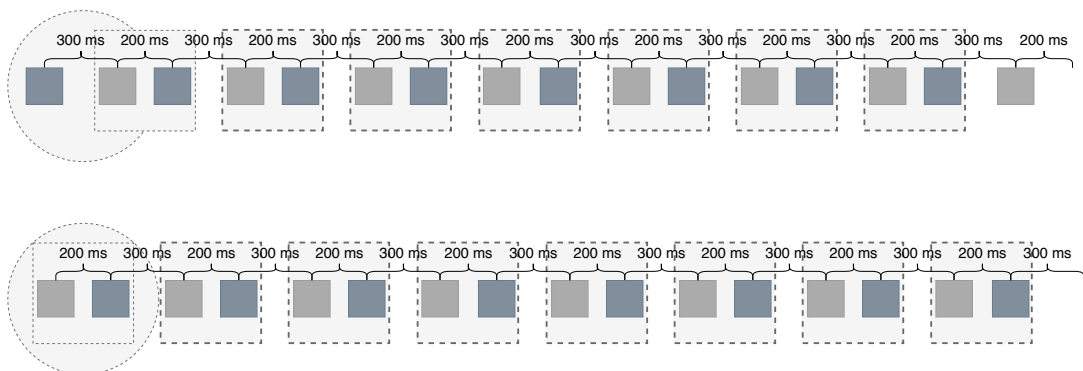


Figure 15. Gestalt-driven perception of 8-loop binary primes

Note: Circles depict actual prime structures whereas squares depict bongo pulses grouped according to gestalt principles. As opposed to the 8-loop binary off-beat prime (bottom) in which prime structure and gestalt-driven perception are aligned, gestalt grouping might override the prime structure in the 8-loop binary on-beat prime (top).

As gestalt grouping that emerges between repetitions of the on-beat prime would become more evident with the increase in number of repetitions, the

perceptual shift described here might be specific to the prime presentation in Experiment 1C. If both primes were to acquire a dominant perception of a binary gestalt, it might be that this gestalt-induced number match results in reaction time enhancement rather than a number match that could reflect both durational and numerical priming. In fact, such a shift in perception would increase the perceptual discrepancy between binary and ternary primes due to the ease of gestalt processing specific to binary primes. Such a differentiation would also provide an explanation to the main effect of prime emerging in the overall analysis for Experiment 1C with binary primes resulting in faster reaction times compared to ternary primes.

In conclusion, emphasizing prime structure through intensity or increased number of repetitions might have made number of units in primes more salient. The resulting facilitation of processing number-matched linguistic stimuli provides support for a cross-modal priming of syllable number. This number match facilitation, which was found in two experiments with differently presented primes, might suggest that syllable number is a salient feature of speech and can be primed in various ways. Moreover, the possibility of a gestalt perception in Experiment 1C also suggests that the durational contrast might be irrelevant to this priming. This interpretation leans towards a syllable-based account of rhythm in Turkish that depends on syllable number but not duration or structure. On the other hand, both binary on-beat and off-beat primes yielding similar facilitation in Experiment 1B lends support to a possible durational priming along with a broader number priming, which will be discussed in the next section.

4.2 Effects of beat

A major goal of these series of experiments was to explore whether metrical organization of speech as defined by syllable number as well as structure can be primed rhythmically. The previous section examined the numeric counterpart of this metrical conceptualization. The other one was the on-beat/off-beat component in relation to fulfilling durational expectations generated by the primes. Although both the present study and the Cason and Schön (2012) study refer to ‘fulfilling beat expectations’ and yield significant results in this respect, it is important to note that these results are not directly comparable. In order to emphasize the distinction between the conceptualization of beat in Cason and Schön (2012) and the present study, I will first explain the beat conditions in more detail.

In the Cason and Schön (2012) study, beat alignment refers to temporally priming the exact point at which the target syllable will occur. To this end, both binary and ternary primes are constructed with stressed beats occurring every 700 ms. In the on-beat conditions, target syllable occurs at the expected point, whereas in the off-beat conditions, the penultimate syllable occurs at the expected point.

In contrast, in Experiments 1A, 1B and 1C, on-beat primes were generated to mimic the durational organization of syllable structures of linguistic stimuli (i.e., long-short to mimic [CVC/CV] and long-long-short to mimic [CVC/CVC/CV]), whereas off-beat primes reversed the durational organizations while keeping the internal relationships of any given prime constant (i.e., short-long and short-long-long). The number of units in a prime interacts with the ‘beat’ structure in determining the metrical organization of the prime. This resulted in binary and ternary on-beat primes (or binary and ternary off-beat primes) having different metrical organizations but nonetheless being assigned a common function in the

analyses. This is made possible through the combinations of the primes with the linguistic stimuli. In on-beat primes, the first bongo pulse is followed by a long interval while the last bongo pulse is followed by a short interval. The predicted expectation following such a prime is another stimulus beginning with a long-interval unit. On the other hand, in off-beat primes, the first bongo pulse is followed by a short interval while the last bongo timbre is followed by a long interval. The predicted expectation following such a prime is another stimulus beginning with a short-interval unit. However, the linguistic stimuli employed in the current experiments always begin with a [CVC] syllable, hypothetically fulfilling the durational expectations generated only by the on-beat primes. Hence, if linguistic stimuli are perceived in terms of durational relationships between syllables, this might become evident in on-beat conditions resulting in shorter reaction times.

Among Experiments 1A, 1B and 1C, Experiment 1B was the only experiment in which there was a main effect of beat. On-beat condition in Experiment 1B resulted in faster reaction times in both the overall nonword analysis and the separate analysis for bisyllabic nonwords. As Experiment 1B employed intensity contrasts in addition to duration contrasts, it might be possible that it is intensity but not duration that enables priming, as was also discussed in the previous section. However, considering that the linguistic stimuli employed in the experiments did not consist of identical syllables, but rather consisted of syllables differing in syllable structure (i.e., [CV] and [CVC] syllables), it is also possible that the intensity emphasis resulted in primes that better reflect the auditory features of linguistic stimuli. The durational differences between closed and open syllables might be accompanied by intensity differences in the speech signal. Different syllable structures might not necessarily be pronounced with identical intensity levels. Moreover, it is possible

that listeners perceive syllables with different structures to have different levels of intensity. Ruling out of these possibilities requires examining acoustical and perceptual correlates of closed and open syllables in Turkish first.

Nonetheless, the interpretation of the main effect of beat relies on the durational expectation that is only relevant to the first syllable of the linguistic stimulus. The main effect does not distinguish between binary or ternary primes, which can be combined with linguistic stimuli in number match or mismatch conditions (i.e., (long-)long-short with [(CVC)/CVC/CV]). Hence, a linguistic stimulus fulfilling the first expectation as well as the following ones (e.g., bisyllabic nonword with binary on-beat prime) and a linguistic stimulus fulfilling the first expectation but violating the following ones (e.g., bisyllabic nonword with ternary on-beat prime) receive a similar treatment under the main effect of beat in the absence of an interaction effect of beat by number. Such a picture raises the question why the violation of the durational expectations is only influential on reaction times when it occurs in the first syllable but not in the following ones.

One possible explanation is that the intensity emphasis on long-interval bongo pulses results in an attentional pulse due to the prime having acquired a more evident stress contrast, i.e., a joint accent structure (Jones, 1987; 2016), in comparison to the solely durational primes. When the first syllable occurs on the point at which there is an attentional pulse, this might present an advantage for processing the nonword. More specifically, a bisyllabic nonword presented after a ternary on-beat prime will have its first syllable as well as its final, target syllable on the potential attentional pulses generated by the ternary prime. Hence, the processing of the second and final syllable might display an advantage derived from the attentional pulse originally designated for the second long-interval bongo pulse. This

additional advantage might compensate for the possible exact metrical priming (i.e., priming both syllable number and structure) achieved by the binary on-beat prime, thus yielding similar facilitation in both binary and ternary on-beat primes. Such a picture would also suggest that even though the target syllable of a bisyllabic nonword following a binary on-beat prime occurs on a non-attended point (i.e., less intense bongo pulse followed by a short interval), it is the metrical priming which results in the enhanced speech processing, as the facilitation in this case cannot be attributed to an attentional pulse.

Alternatively, it might be that the phoneme detection task requires predicting when a nonword will begin more so than when a nonword will end, as in any given prime block, bisyllabic and trisyllabic nonwords occurred in a random order, making the prediction of syllable numbers practically improbable for the most part. Hence, listeners might be making use of having attended to the first syllable of the nonword due to the attentional pulse in the on-beat conditions, but shift their attention to the nonword with no regard to the prime once they detect the beginning of a nonword. However, given that in Experiment 1B there was also a main effect of number indicating faster reaction times in number match conditions, the assumption in this alternative explanation that listeners could not predict syllable numbers is challenged.

Lastly, the reason why Experiments 1A or 1C did not produce differential reaction times in response to the beat conditions might be explained by the prime presentations employed in these experiments. Primes in Experiment 1A not being sufficiently salient might have also hindered possible effects of beat in these experiments. As for Experiment 1C, and especially in regard to the findings from bisyllabic nonwords, the binary gestalt perception discussed in the previous section

might have also overridden the influence of beat structure, as both on-beat and off-beat primes might have been perceived in a single beat structure. On the other hand, ternary primes do not present such a perceptual shift. However, even eight repetitions might have been insufficient to make their patterns more salient, and their beat structure might have been too hard for listeners to process and differentiate.

To conclude, although ‘beat’ did not reveal an influence over reaction times except for Experiment 1B, the findings do not contradict a possible influence of fulfilling durational expectations. In fact, these results might be the first preliminary findings suggesting that durational relationships might be influential in metrical priming (e.g., binary on-beat primes coupled with bisyllabic nonwords). Plausibility of this suggestion will be evaluated in more detail in the following section.

4.3 Effects of metrical match

Building on top of number and beat, the ultimate research question in the present study was to explore whether metrical organization can be primed and whether durational meters and priming can be differentiated from numerical meters and priming. Experiments 1A, 1B and 1C were conducted in regard to the first part of the question, whereas Experiment 2 targeted the second part of the question.

First turning to the results of Experiments 1A, 1B and 1C, despite significant main effects of number and beat on certain cases, the analyses yielded null findings in their interaction terms, suggesting that the procedure employed in these experiments could not prime metrical structure in linguistic stimuli. In fact, the Cason and Schön (2012) study that paved the way to the present experiments also yielded similar results. In the mentioned study, the authors found a main effect of beat, i.e., temporal alignment of stressed beats in primes and target syllables in

linguistic stimuli on reaction times with on- compared to off-beat alignment leading to faster reaction times. However, this effect was not accompanied by a main effect of metrical match, i.e., the match between the number of units constituting a prime and a linguistic stimulus. Despite the lack of reaction time differences between metrical match and mismatch, an N100 component was identified in metrical match conditions. Hence, the authors claim that it is the attentional processes relating to stress and temporal structure that predominantly guide perception and behavioral results. Moreover, there was no interaction of beat and meter, the latter being predicted to exert influence on top of the temporal match presented by beat. The lack of interaction is interpreted in relation to distinct structures involved in processing of temporal and metrical information, and the authors suggest that the two components might be processed independently.

In similar terms, the current results might also suggest that the main component showing priming is number of units, as beat did not interact with number to yield an interaction that represents metrical structure. However, it is also important to note that as opposed to the only significant finding being the main effect of beat in Cason and Schön, main effects of both number and beat were significant in Experiment 1B. As results revealed a main effect of both components simultaneously, the present study might be closer to priming metrical structure than Cason and Schön. In addition, although there was no electrophysiological measure in the present experiments, the independence of processing proposed by Cason and Schön might also apply to number and beat in the present case. As priming of number yielded more successful results compared to priming of beat in the current studies, their independence might be interpreted in favor of the syllabic account of meter in Turkish as opposed to the durational one, suggesting rhythm and meter in

Turkish predominantly depend on number and not duration or syllable structure.

Another possibility is that participants in the current experiments responded to either number or beat, reflecting two different strategies of processing employed by two subsets of participants.

In the context of the present experiments, the null interaction results might be attributed to several reasons. One such reason as suggested by previous sections relates to prime structures. Perceptual salience of primes appeared as a critical factor based on findings in Experiment 1A. The solutions provided in Experiments 1B and 1C might have been insufficient in achieving metrical priming. Although both follow-up experiments improved priming in at least one aspect compared to Experiment 1A, these improvements need not yield a higher-level improvement and reveal significant interaction terms. The possible reasons for the revised primes to selectively improve priming, such as a perceptual shift that might be in favor of a gestalt perception, were discussed in previous sections. Such an interpretation requires exhausting alternative primes in further research in order to reject the hypothesized metrical priming.

At this point, it is also important to evaluate the overall plausibility of current primes in priming metrical structure. As stated above, one of the main objectives of this study is to prime metrical structure through interactions of syllable number and structure. To this end, rhythmical primes that mimic syllable number but not syllable structure were compared to those that mimic both syllable number and syllable structure. In primes that hypothetically mimic syllable number and structure, bongo pulses that constitute the primes are likened to vowels in syllables, and hence, prime vowel positions temporally in relation to each other. Interestingly, such a configuration of bongo pulses to represent vowels and their temporal positions in

syllables has a real-life example in the Amazonian Bora drummed language documented by Seifart, Meyer, Grawunder and Dentel (2018). I will discuss the plausibility of metrical priming in the present experiments in relation to the drummed Bora language, as the priming rationale in the former parallels the structure of the latter.

The spoken Bora language has a relatively simple syllable inventory comparable to that of Turkish, but differs from Turkish in its use of non-lexical phonological tone. In comparison, in the drummed Bora language, each beat corresponds to a syllable and beats represent vowel-to-vowel intervals in speech and not phonological tone. The analyses in the Seifart et al. study revealed four durationally distinct vowel-to-vowel intervals in both drummed and spoken Bora, i.e., $V < VC < VVC < VCC$. It appears that speakers of the drummed Bora language make use of rhythmical information in inter-beat durations and thus convey and comprehend fully intelligible speech messages within this acoustically reduced signal. The study of drummed Bora is interpreted in relation to a proposal regarding vowel-to-vowel intervals being the basic unit of rhythm in speech, which has found supporting evidence from rhyming domains of numerous languages (e.g., Steriade, 2012). The authors also suggest their data might be interpreted by taking a dynamic approach focusing on P-centers, which might present slight deviations from actual vowel onsets employed in determining vowel-to-vowel intervals (Seifart et al., 2018).

While the proposal of vowel-to-vowel intervals lend support to the current experiment theoretically, the case of the drummed Bora language suggests practical plausibility of the priming in the present studies. Drawing the latter parallel importantly depends on the similarities between phonological structures of the

spoken Bora language and Turkish. However, a point raised by Seifart et al. (2018) should be taken into consideration in theoretically extending their findings to Turkish. Although the drummed Bora language does not lexically represent phonological tone, the property nonetheless exists in the spoken Bora language. Interestingly, the authors argue that while both drummed and whistled languages acoustically reduce speech in their essence, drummed languages emerge almost exclusively in spoken languages that employ phonological tone. Hence, phonological tone might be one quality of spoken language that enables reducing speech to beats. In this respect, Turkish might be a language less suitable to being represented in drummed beats due to its lack of phonological tone, which presents a very interesting perspective to the lack of metrical priming in the present study.

In comparison, classical Turkish music which developed in close relation to classical Turkish poetry and arud meter provide an interesting perspective on the subject. Compositions in Turkish classical music are mostly vocal and are based on poems written in arud meter. In terms of musical meter, Turkish classical music makes use of a metrical system named *usul*, which provide predetermined rhythmical patterns of a given number of strong and weak beats. The beats are assigned to syllables in composing poetry, however, the assignment of strong and weak beats to syllables is determined by *usul* and is independent of the prosody introduced by poetic meters. On the other hand, the beat assignment takes into consideration syllable length as determined by arud meter (Tanrıkorur, 1990; 1991).

Tanrıkorur's (1990; 1991) analyses of compositions of Turkish arud poetry suggests that closed and open syllables are consistently assigned long and short beats, respectively, mostly following a 2:1 length ratio. Tanrıkorur argues that poems written in the same arud meters have been composed in the same *usuls* across

generations of composers, and that distinct arud meters require certain *usuls* and cannot be composed in ways that violate the durational relationships of syllables. Tanrıkorur illustrates his claim in Figure 16: He suggests that the verse in his example have been composed in the *aksak usul* by different composers (top), whereas it is improbable that a rhythmical pattern that is different from the syllabic structure would be employed in a composition (bottom). Tanrıkorur (1990) suggests that “(t)he syllabic structure of the Turkish language created by the alternation of long and short syllables is reflected harmoniously in the musical compositions in a similar manner by the long and short beats of the *usuls*.” Tanrıkorur’s claims suggest it is important to examine Turkish classical music compositions and performances. This might provide insight into an element of Turkish culture whereby linguistic and musical rhythm meet and interact. However, this endeavor is beyond the scope of the current study.



Figure 16. Metrical composition of a verse in classical Turkish music
Note: Beats above the line represent strong beats whereas beats below the line represent weak beats. Figure reproduced from Tanrıkorur, 1991.

Experiment 2 was run as an extension of the question of metrical priming and to further elaborate the plausibility of a metrical organization in Turkish that introduces duration as an addition to the more conventional syllable number

component. The ternary off-beat prime employed in Experiment 1 was replaced by a ternary gestalt prime in this experiment. As a result, gestalt primes represented number of units without entering into an interaction with duration, while duration primes represented both number of units and their durations. Unlike prime structures, prime presentations were exactly the same as in Experiment 1A, i.e., primes consisting of exact bongo pulses were repeated four times. Not surprisingly, the results from Experiment 2 also paralleled those of Experiment 1A. The only significant effect was the prime and number interaction indicating a ‘trisyllabic advantage’. When comparing Experiments 1A and 2, it is possible to suggest that primes in Experiment 2 were also insufficient in their perceptual salience and hence did not influence reaction times.

In conclusion, as was discussed in the previous section, Experiment 1B constitutes a case distinct from the remaining experiments. Experiment 1B reveals main effects of number and beat, both of which are hypothesized to be components of metrical match. Although metrical match did not emerge as a significant interaction term, the exploratory nature of the present study suggests optimizing the stimuli and procedure might lead to successful metrical priming in further studies. In this sense, while current findings do not provide concrete support for metrical priming, they might be argued to be more promising especially in comparison to previous studies.

4.4 The ‘trisyllabic advantage’

A recurring finding in Experiments 1A, 1B and 2 was the interaction of prime and number mainly in pseudowords and partially in nonwords. In general, the interaction term indicated an advantage of number *mismatch* in binary primes and number *match*

in ternary primes. Restating the interaction term in syllable numbers of linguistic stimuli simplifies the finding to a great extent. Number mismatch in binary primes and number match in ternary primes both denote trisyllabic stimuli. Hence, the interaction of prime and number is but an expression of trisyllabic stimuli receiving shorter reaction times across conditions compared to bisyllabic stimuli.

The same pattern emerged in the Cason and Schön (2012) study with trisyllabic stimuli processed faster than bisyllabic ones. The authors interpreted this finding to indicate the additional time period introduced by the third syllable in the trisyllabic stimuli presenting extra time for preparing to give a response in the phoneme detection task. Importantly, trisyllabic stimuli in the Cason and Schön study had an average duration of approximately 600 ms, and were approximately 200 ms longer than bisyllabic stimuli. In comparison, trisyllabic stimuli in the present study had an average duration of approximately 800 ms, and were approximately 300 ms longer than bisyllabic stimuli. Hence, Cason and Schön's interpretation of the 'trisyllabic advantage' seems applicable to the present case.

In the current study, the 'trisyllabic advantage' was bypassed through running separate analyses according to syllable count. Once bisyllabic stimuli were analyzed without trisyllabic stimuli obscuring fine-grained differentiations in reaction times, several significant findings emerged. However, the same was not true for trisyllabic stimuli: even when analyzed separately, trisyllabic stimuli yielded no significant differences in reaction times across conditions. This null finding indicates that trisyllabic stimuli not only obscures significant results in bisyllabic stimuli, but they are also robust against priming in most part. This might be due to the 'trisyllabic advantage' reducing reaction times to the shortest possible values and hence not allowing further enhancement.

The ‘trisyllabic advantage’ suggests that the nature of trisyllabic stimuli is inherently unsuitable to priming in a procedure such as the one employed in the present study. However, trisyllabic stimuli present a randomly occurring and equally likely alternative to bisyllabic stimuli. This prevents responses to linguistic stimuli depending on a knowledge or a prediction regarding the number of syllables to be encountered in any given trial and hence enables primes to be the only possible predictor. However, one might have to think of other types of linguistic stimuli that achieve this and still meaningfully interact with the main manipulations.

4.5 Nonwords and pseudowords

The Cason and Schön (2012) study and the present study differed in their linguistic stimuli. While the former had phonologically and orthographically valid ‘pseudowords’ that did not resemble existing words as its linguistic stimuli, the latter had both ‘pseudowords’ that differ from actual Turkish words by 1-3 phonemes but also ‘nonwords’ that were impossible constructions according to Turkish phonotactic rules and regularities. Hence, lexicality was introduced as a control variable in the present study. Here, I will discuss specific characteristics of each type of stimuli.

First, in order to clarify a point in the previous section, I will refer to the ‘trisyllabic advantage’ in relation to lexicality. As was stated above, the ‘trisyllabic advantage’ applied to pseudowords mainly and to nonwords to a lesser extent. As opposed to ‘trisyllabic pseudoword advantage’ that was evident in all experiments that employed pseudowords, the ‘trisyllabic nonword advantage’ did not emerge in Experiment 1B and in ternary primes in Experiment 2. More importantly, there was no interaction of prime and number and hence no ‘trisyllabic nonword advantage’ in Experiment 1C, in which only nonwords were employed. Hence, the relative

specificity of the ‘trisyllabic advantage’ to pseudowords constitute one way in which nonwords and pseudowords can be differentiated, suggesting that usage of trisyllabic stimuli might be justified to a greater extent in nonwords.

Another difference between nonwords and pseudowords emerged in Experiment 1C. In this experiment, participants were presented an experimental corpus consisting of only nonwords. This experiment yielded lower overall percentage of correct responses (89.19%) and longer mean reaction times (798 ms) compared to the other experiments. In a similar vein, in Experiment 1C, correct ‘yes’ and correct ‘no’ responses received similar reaction times, whereas in all other experiments correct ‘yes’ responses were significantly faster than correct ‘no’ responses. Taken together, these results suggest that nonwords might have had a special quality that presented a hardship in their processing. In fact, several participants indicated that they could not identify the target phoneme in nonwords in numerous trials and developed their own strategies of responding in such cases. Moreover, when several nonwords that received wrong answers were inspected, it became evident that the phonologically invalid nature of nonwords might have created an ambiguity in the resulting pronunciation of the text-to-speech software. The ambiguity in phonemes might be an issue with other nonwords as well, resulting in lesser confidence in participant responses and hence delaying reaction times to even correct ‘yes’ responses.

On the other hand, in the remaining experiments, several participants indicated that they noticed half of the linguistic stimuli resembled existing Turkish words and that such stimuli attracted their attention. It is likely that these were the intermixed pseudowords. Hence one could think that pseudowords, by attracting participants’ attention, distracted them from the phoneme detection task. In the

extreme case, perceiving pseudowords as actual words might have influenced participants' perception of phonemes, and more importantly, the perception of target syllables. In such a case, responses to pseudowords might be contaminated by these perceptions or expectations. Interestingly, a small number of participants indicated having heard actual words in several trials. Some of these participants even reported hearing proper names that had personal connotations for the participant.

Pseudowords acquiring such unintended semantic and personal meanings might have influenced reaction times and/or responses in unpredictable ways, possibly even confounding results. Such restraints argue against the use of pseudowords that are highly similar to actual words in future research.

To conclude, both nonwords and pseudowords in the present experimental corpus had certain defects that might have had unintended influence over responses and reaction times. It might be preferable to either control linguistic stimuli for such defects or opt for the use of phonotactically valid pseudowords that do not resemble existing words as in Cason and Schön (2012).

4.6 Limitations

Several limitations of the present experiments have been discussed above, such as those relating to the use of nonwords and pseudowords. Another related and important limitation is the linguistic stimuli being read by a text-to-speech software rather than a native speaker of Turkish. Employing a text-to-speech software might have certain advantages considering that half of the experimental corpus violated Turkish phonotactical rules and the other half resembled actual words to some extent. For nonwords, performing phonotactically invalid configurations might be easier for a software compared to a speaker of the language. Similarly, identifying

the original word from which a pseudoword was generated is improbable for a software, which might influence speakers' pronunciation of pseudowords. On the other hand, the artificial nature of the text-to-speech software might be distracting, as was stated by numerous participants. Also, the software employed in the present study consistently mispronounced certain phonemes or produced unnatural phoneme durations in certain cases. Although these were minimal defects or could be corrected, the stimuli still sounded artificial or ambiguous at times.

Other limitations might relate to our pool of participants. They were Turkish native speaking university students received their education in English. This, too, might have influenced their perception of linguistic rhythm. In a similar vein, a large proportion of the participants knew or were learning languages other than English. In fact, several participants noted how they sometimes treated target phonemes in terms of the corresponding sound in one of the languages they know (e.g., treating a visually presented /c/ as a /k/, as is possible in English). Such instances suggest a pool of strictly monolingual participants might provide data uninfluenced by languages other than Turkish.

Lastly, it is important to consider limitations relating to the exact temporal positioning of primes and linguistic stimuli. Primes in the current experiments were presented in blocks with no variation in their number of repetitions. These factors might have made identifying the onsets of linguistic stimuli extremely predictable, which in turn might have influenced phoneme detection in the temporally less predictable last syllable. While primes in the current study imitated the average intervals between vowel onsets of [CVC] closed and [CV] open syllables as an estimate of P-centers, target linguistic stimuli were temporally presented according to their syllable onsets. Employing vowel onsets of the experimental corpus to

temporally align the beats in the linguistic stimuli with the primes might have provided more precise results compared to syllable onsets. This issue might best be overcome by collecting data on P-centers of different phoneme configurations and syllable structures in Turkish. Also related to this is the lack of data on acoustical and perceptual properties of syllable types in Turkish. Obtaining data on these parameters might provide much more refined sets of stimuli and experimental procedure. For the time being, such studies can only rely on theoretical accounts on these issues or data from other languages.

4.7 For future research

The present study provides exploratory work examining possible ways in which metrical structure in speech can be cross-modally primed in Turkish. The first step in refining the current research to better target its aims would be to reevaluate the experimental linguistic stimuli. As was discussed in Section 4.4, trisyllabic stimuli did not provide meaningful results with respect to the manipulations. Further research might employ linguistic stimuli that have durations below the threshold that results in the ‘trisyllabic advantage’. For instance, bisyllabic stimuli that are differentiated in terms of their syllable structure, such as [V/CV], [CV/CV] or [CV/CVC] might be employed in addition to [CVC/CV]. Especially, using [CV/CVC] would present a similar durational contrast where the target phoneme occurs on a heavy syllable, thus providing an important alternative to the target detection that always occurred on a light syllable in the current study. Based on an analysis of perceptual qualities of long vowels in Turkish, the pool of linguistic stimuli might also include syllables with long vowels for closed syllables (e.g., [CV:]

compared to [CVC], see Figure 3). A comparison of reaction times to stimuli such as these might provide a more precise index of metrical priming.

Following Cason and Schön (2012), using linguistic stimuli that comply to phonotactical rules of Turkish while not resembling existing words might help overcome the issues regarding lexicality discussed in Section 4.5. Also important is to monitor and regulate acoustic features of syllables other than duration. Especially Experiment 1B might be replicated with syllables manipulated to have constant intensity to eliminate possible confounds introduced by text-to-speech-predicted intensity in the present experiments. Another possibility is to pilot linguistic stimuli that have word-final stress as documented by acoustic measures in the literature (e.g., Pycha, 2006).

Another important part of the present study to reconsider is its construction and presentation of primes. As suggested above, presentation of bongo pulses and primes might be optimized through basic acoustical and perceptual analyses of Turkish syllable types. These findings might be useful in generating primes that represent fundamental components of metrical structure of speech through an acoustically reduced signal. The number of prime repetitions might be controlled for confounding characteristics such as the gestalt-driven perception discussed in relation to Experiment 1C. Furthermore, the number of prime repetitions might be revised and made random to make onsets of linguistic stimuli less predictable. Moreover, in order to test whether this priming is specific to auditory modalities or reflects a more generalized rhythm processing, visual primes with similar temporal structures might be employed.

Also important is to include an accompanying electrophysiological measure to the experimental procedure. Although behavioral studies of the issue at hand

provide valuable information regarding the perception of metrical structures in Turkish, ERP measures, as used in Cason and Schön (2012), might be much more informative with respect to degree of entrainment as reflected in expectation-driven electrophysiological responses. Such data might also allow for differentiating more as opposed to less strongly entrained participants and hence a comparative reanalysis of their behavioral data.

Different tasks and methods might also be employed to obtain data on linguistic rhythm perception. Within the priming paradigm, one interesting task to use might be priming metrical structure for linguistic targets that have ambiguous word boundaries and collect responses regarding words heard by the participants as in Breen, Dilley, McAuley and Sanders (2014). Although Breen and colleagues employed linguistic primes, their study might also be revised for cross-modal musical priming of durational intervals in speech as in the current experiments. The methods that do not involve priming that might be employed include, but are not limited to, tapping to speech (e.g., Lidji et al., 2011), reproducing speech through tapping or tones (e.g., Benadon, 2014), or studies of memory or speech production within metrically consistent and inconsistent linguistic stimuli (e.g., Kimball, Yiu, & Watson, 2019; Myers & Watson, 2019).

The present study and the proposed future directions in this section focus on attaining a basic level understanding of linguistic rhythm in Turkish. However, an important future direction that might be built upon findings from such studies concerns developmental language disorders (DLD) and dyslexia. It has been suggested that in DLD and dyslexia, impairments in language processing emerge as a result of cascading effects. Rhythm processing has been proposed to be at the basis of these factors with disfluencies in temporal processing leading to deficiencies in

phonological processing, which in turn influence metrical, syntactic and semantic processing as well as literacy (e.g., Goswami, 2019; Overy, 2000; Richards & Goswami, 2019; Tallal & Gaab, 2006). Moreover, it has been suggested that by mimicking and emphasizing linguistic rhythm in a non-linguistic medium, music can be employed as an intervention mechanism in such cases (e.g., Goswami, 2019; Overy, 2000). For instance, research with French-speaking children with and without DLD and dyslexia shows that musical regular rhythmic sequences improve sensitivity in syntactical error detection tasks (Canette et al., 2020; Przybylski et al., 2013). Hence, one important question to pursue based on the present study is whether identifying potential language-specific cues to rhythm might help devising musical intervention programs for Turkish-speaking children with DLD or dyslexia.

Lastly, the current conceptualization of linguistic rhythm as temporal organization of P-centers might be extended to other languages and speakers of these languages. Such comparative studies might reveal whether P-centers are a reliable cue in rhythm perception for speakers of Turkish in processing speech. Moreover, it might be suggested that P-centers are a linguistic universal at least for languages that recognize syllables and not sub-syllabic units as a fundamental segmentation level. Importantly, P-centers might acquire and represent language-specific rhythmic qualities, such as P-center interactions with stress. This possibility might enable constructing comparable rhythmical representations for speakers of languages other than Turkish such as the one attempted in the current study. Hence, it is possible that through language-specific adaptations, P-centers might present a common ground for comparative studies on linguistic rhythm across languages proposed to be rhythmically distinct.

4.8 Conclusions

In three experiments, cross-modal priming of duration-based metrical structure of speech in Turkish was explored. Numerical and durational priming were compared in a fourth experiment. Results suggest that syllable number can be primed especially in bisyllabic nonwords and durational expectations can be generated when durational structure of primes are emphasized by intensity. To my knowledge, these results constitute the first documentation of a durational rather than stress-based approach to meter in language. Although results do not directly support a conceptualization of meter that emerges through syllable number and structure, they support accounts of syllable number and structure separately, which might be successfully brought together in further research. On the other hand, syllable number seems to have a predominance over syllable structure in priming in the current experiments, whereas the exact reasons of this is not yet clear. These results might be attributed to experiment-specific factors or interpreted in line with the account that argues metrical structure of Turkish is based on syllable numbers rather than syllable structure.

These four exploratory studies aimed to ask certain questions rather than providing immediate answers. These questions included whether stress-based approaches to rhythm and meter were applicable to all languages without any exceptions, whether we could increase the number of languages studied on a case-by-case basis without relying on broad generalizations, and whether we could explain this process in relation to native speakers' understanding and perception of rhythm and meter in their languages. These questions converge on the concept of native speaker driven perception, and only accumulation of further research with a similar approach can provide further answers to the questions at hand.

APPENDIX A

CAESURA POSITIONS IN TURKISH SYLLABIC METER

In Onay's (1999) view, Turkish syllabic meter is defined by invariant syllable numbers and caesura positions. He provides an example in which the two parameters are kept constant and hence meter is attained:

Bir kız ile / bir gelinin / cengi var (4+4+3)

Bilemedim / hangisinin/ dengi var (4+4+3)

However, Onay (1999) argues that varying caesura positions across lines violates meter. He presents a poem by Kuloğlu as an example in which despite the consistency in syllable numbers on the line level, varying caesura positions invalidate meter:

Uyan hey / ceylan bakışım / aç gözüün (3+5+3)

Cün sabah oldu / ağardı tan yeri (5+6)

On the other hand, Dilçin (2009) argues that syllabic meter is defined solely by number of syllables per line and that caesuras can occur in varying positions without effecting meter. In fact, Dilçin claims that variable caesura positions might add to the rhythmicity of a poem as in the following example by Tokatlı Nuri:

Nâ-dâna esrârı / etmeyiz ızhâr (6+5)

Nâ-ehline / ârifandan / sır çıkmaz (4+4+3)

Bezme kadem bassa / hezârân ağyâr (6+5)

Dilce söyler / tercümandan / sır çıkmaz (4+4+3)

APPENDIX B

STRESS ORGANIZATION IN TURKISH SYLLABIC METER

Mirşan (1966) claims that in Turkish syllabic meter, caesuras divide the line into “metrons” consisting of an equal number of stressed and unstressed syllables. In an analysis of a stanza by Köroğlu, Mirşan argues that each metron has two stressed and two unstressed syllables:

Selam verdim / almadın sen (.+.+ / +..+)

-özüm için / alam seni (.+.+ / .+.+)

Selam kadrin / bilmedin sen (.+.+ / +..+)

-aç yüzünü / görem seni (+..+ / .+.+)

Aksan’s (2013) analysis of the same stanza assigns stresses differentially. His analysis also reveals a regularity of stress patterns, with stress organization being implemented not as a necessary component of meter but as an aesthetic preference:

Selam verdim / almadın sen (.+.. / 4)

Özüm için / alam seni (4 / .+..)

Selam kadrin / bilmedin sen (.+.. / 4)

Aç yüzünü / görem seni (4 / .+..)

Aksan (2013) also argues that regularities in stress patterns or numbers emerge naturally through mastery in language, as in Yunus Emre’s following stanza:

Kimi yiğit / kimi koca (.+.+ / .+.+)

Kimi vezir / kimi hoca (.+.+ / .+.+)

Gündüzleri / olmuş gece (.+.+ / .+.+)

Ancılayın / çoklar yatur (+..+ / .+.+)

APPENDIX C

LIST OF LINGUISTIC STIMULI

Bisyllabic		Trisyllabic	
Nonwords	Pseudowords	Nonwords	Pseudowords
yenba	yurba	nerlodba	şışkimbe
tılbu	tezbe	şınlatbi	gerçembe
rasbi	kembi	gartüşbü	gönçerbe
ticba	zırba	löftüsbi	gersizbe
nerba	tolba	kövrelbü	tarsamba
cavbe	selbi	hörkısba	serçembe
komde	cerdi	çezyosda	karpında
vogde	nesde	sügçaşda	bekminde
ceşda	zolda	busrukdi	silnerdi
çördü	yümde	cölgrogdı	cislerde
şökdı	yufda	çuyşurdü	kelginde
höcdu	gezde	gilvüzdu	gerçende
hüfge	bünge	tobyazge	çeltirge
neçge	yırğa	nimtubgu	pünterge
catgü	tümge	bönpamgu	böstenge
lergu	tongu	sunbipge	bilmirge
möçge	şalgı	vetfirga	sildirgi
şüngu	serge	pobsızge	yirmilgi
cufki	hezki	çipnatka	çerleşki
çoske	temke	nubvedke	birlezke
codke	balka	nomdurkü	maprıska
diçku	runku	şüfpanka	bitlenke
tüzki	şalkı	lürçoskı	tüzleşke
hıfke	himki	fimroykı	yenletki
cöspa	katpa	livkampa	sundurpa
varpo	verpi	çiy мүkpe	sünterpe
zıdpe	kürpi	ralcönpü	kıstanpa
şidpa	yanpa	mizlürpe	kaytarpa
tuşpı	zerpe	zidsampa	tilkirpi
valpi	narpa	rudnaspa	taşçırpa
rözta	lartı	zölci ftu	sarlıntı
pücti	kelti	bempünta	gurlanta
kaytü	masta	cöygarta	birlenti
zunti	selte	nacmudtı	bimlikte
şoltü	şantı	vegrüçtu	çeyrelti
kıbtu	barta	ricboltı	yaslantı
neşcü	holca	fünkumcu	rulmunca
gapcu	şelci	sançözcü	bislerce
sovcü	tüzce	dülzürca	pirsince
parcu	yurcu	bolnamcü	soptuncu
lügca	kamca	büdvercü	günverci
voyco	bulca	roknoscı	şöylence

List of Linguistic Stimuli (continued)

Bisyllabic		Trisyllabic	
Nonwords	Pseudowords	Nonwords	Pseudowords
sezçü	zerçi	zıpmaşçe	dosluşçu
mazçe	sençi	hepsobçu	giftikçe
lödçi	görçü	rikyebçu	yerşekçi
lasçu	kapça	dorşimçi	sırdapça
ropçe	perçe	zemvarça	yanlakça
holçi	kırçı	höymünça	yamsırça
nödlü	famla	çimbuyle	kormusla
puflü	tüslü	şoznicle	çencemli
kiplu	rümle	nücseglu	göskemli
hotlı	mizli	zurşönli	bamtıklı
zöklü	şarlı	rışsupli	yürtüklü
şoglı	bemli	dazkenle	dirtekli
güfma	tözmü	suzrüpme	batırşma
puzmü	barma	zölhiçmi	tarkulma
kutmi	bözme	pidvermü	püzletme
kıdme	şevmi	çığcidme	serçilme
sovmü	devme	şenreymü	satruşma
tühmı	yaşma	liscüyma	bilgirmi
pebra	disre	şöpdöçrü	dımsanrı
sıbru	kayrı	hüzgadru	tendemri
navrü	keşre	lıpnosrı	kulbumru
möçra	çaprı	dosçinre	yüskenri
napru	çekre	çegşömre	talçayrı
fasre	payrı	neşmuyru	gönkemri
şinsü	himsi	tübducsa	merşeksi
tigsa	şarsı	nifbügse	kevrimsi
cugsi	palsa	kuncorsı	yuldişsı
kensü	temse	çirlumsa	tanlamsı
dacsü	heçsi	gaphebso	teçkimse
löcsa	yapsı	tufloçsi	dirliksi

APPENDIX D

ETHICS COMMITTEE APPROVAL

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ
İnsan Araştırmaları Kurumsal Değerlendirme Alt Kurulu

Sayı: 2018-40

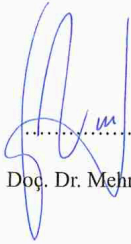
10 Ekim 2018

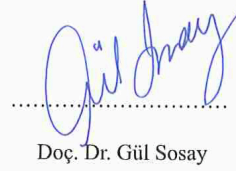
Züheyra Tokaç

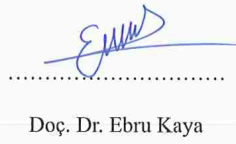
Psikoloji

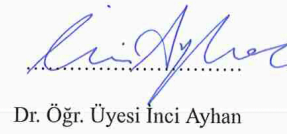
Sayın Araştırmacı,

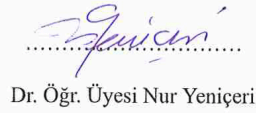
- ☐ "Rethinking Duration and Rhythm in Language and Music: Can Duration Prime Rhythm in Metrical Speech?" başlıklı projeniz ile ilgili olarak yaptığımız SBB-EAK 2018/40 sayılı başvuru İNAREK/SBB Etik Alt Kurulu tarafından 10 Ekim 2018 tarihli toplantıda incelenmiş ve uygun bulunmuştur.


Doç. Dr. Mehmet Yiğit Girdal


Doç. Dr. Gül Sosay


Doç. Dr. Ebru Kaya


Dr. Öğr. Üyesi İnci Ayhan


Dr. Öğr. Üyesi Nur Yeniçeri

APPENDIX E

SPEED-ACCURACY INTEGRATED RESULTS

Phoneme detection task has two measures of performance: speed and accuracy. Data analysis in the present study is run on reaction times and hence only represents one component of performance. Several methods that combine speed and accuracy into composite scores have been proposed, and one such method is linear integrated speed-accuracy scores (LISAS; Vandierendonck, 2017; 2018). LISAS integrates proportion of errors to reaction time by weighting the two components and yields scores that can be interpreted as reaction time corrected for errors. While there is more than one calculation method provided in different publications, guidelines by the author were implemented in choosing the suitable method:

$$\text{LISAS} = \text{RT}_C + \text{PE} \times \text{S}_{\text{RT}}/\text{S}_{\text{PE}}$$

Where RT_C is mean correct reaction time per condition, PE is proportion of errors per condition, S_{RT} is standard deviation of reaction times across conditions, and S_{PE} is standard deviation of proportion of errors across conditions per participant.

Vandierendonck (2018) suggests when there are few data points per condition, standard deviations can be calculated by pooling conditions per participant. Moreover, as calculating S_{RT} and S_{PE} per condition underestimates corrected reaction times, the author again suggests calculating standard deviations per participant (Vandierendonck, 2019). Since there were relatively few observations per condition (i.e., 12 observations) in the present experiments, in light of these suggestions, standard deviations were calculated by pooling conditions across nonwords or pseudowords. Lexicality was a control variable in the present study and

nonwords and pseudowords were analyzed separately. Moreover, there was a differential influence of lexicality on performance (see discussion section 4.5). Analyses of LISAS calculated as described above paralleled analyses run on reaction times. In all four experiments, overall analyses of LISAS only yielded significant results for interaction of prime and number, i.e., a much stronger ‘trisyllabic advantage’ (Tables E1 and E2), with the exception of Experiment 1B. As is natural of the ‘trisyllabic advantage’, all other effects were obscured. Moreover, due to the strength of the ‘trisyllabic advantage’, no effects were revealed when bisyllabic and trisyllabic stimuli were analyzed separately (all $ps > .05$). Furthermore, the ‘trisyllabic advantage’ was magnified differentially according to number of syllables in stimuli and lexicality, suggesting accuracy was dependent on these factors. These results provide further support to the discussion on ‘trisyllabic advantage’ (section 4.4) and on lexicality (section 4.5).

The only experiment that yielded additional significant results was Experiment 1B: For nonwords, there was a main effect of number, $F(1, 40) = 6.36$, $p = .02$, $\eta_p^2 = .14$. Number match ($M = 686$ ms, $SE = 32$ ms) resulted in faster reaction times compared to number mismatch ($M = 716$ ms, $SE = 33$ ms). For bisyllabic nonwords, there was a main effect of prime, $F(1, 40) = 4.90$, $p = .03$, $\eta_p^2 = .11$. Binary primes ($M = 719$ ms, $SE = 34$ ms) resulted in faster reaction times compared to ternary primes ($M = 770$ ms, $SE = 40$ ms). These results remain in favor of a possible numerical priming in the context of accented primes. For pseudowords, there was an interaction of beat and number, $F(1, 40) = 5.76$, $p = .02$, $\eta_p^2 = .13$. However, post-hoc tests revealed no significant differences between conditions.

In conclusion, LISAS results corroborate the discussion on ‘trisyllabic advantage’ and lexicality by making their influences even more pronounced, while limiting the support for all other effects except for a modest numerical priming.

Table E1. Summary of LISAS Results for Interaction of Prime and Number in Nonwords

	Test of Within-Subject Effects				Post-hoc comparisons			
					Binary prime		Ternary prime	
	Error df	F	p	η_p^2	Number match	Number mismatch	Number match	Number mismatch
Exp.1A	41	15.38	< .001	.27	<i>M</i> = 642 ms <i>SE</i> = 39 ms	<i>M</i> = 584 ms <i>SE</i> = 37 ms	<i>M</i> = 574 ms <i>SE</i> = 36 ms	<i>M</i> = 661 ms <i>SE</i> = 43 ms
Exp. 1B	40	27.19	< .0001	.41	<i>M</i> = 719 ms <i>SE</i> = 34 ms	<i>M</i> = 662 ms <i>SE</i> = 35 ms	<i>M</i> = 574 ms <i>SE</i> = 36 ms	<i>M</i> = 770 ms <i>SE</i> = 40 ms
Exp. 1C	35	9.88	= .003	.22	<i>M</i> = 863 ms <i>SE</i> = 47 ms	<i>M</i> = 802 ms <i>SE</i> = 46 ms	<i>M</i> = 835 ms <i>SE</i> = 48 ms	<i>M</i> = 916 ms <i>SE</i> = 47 ms
Exp. 2	42	19.13	< .0001	.31	<i>M</i> = 710 ms <i>SE</i> = 40 ms	<i>M</i> = 789 ms <i>SE</i> = 43 ms	<i>M</i> = 739 ms <i>SE</i> = 44 ms	<i>M</i> = 799 ms <i>SE</i> = 45 ms

Table E2. Summary of LISAS Results for Interaction of Prime and Number in Pseudowords

	Test of Within-Subject Effects				Post-hoc comparisons			
					Binary prime		Ternary prime	
	Error df	F	p	η_p^2	Number match	Number mismatch	Number match	Number mismatch
Exp.1A	41	112.73	< .0001	.73	<i>M</i> = 525 ms <i>SE</i> = 37 ms	<i>M</i> = 647 ms <i>SE</i> = 43 ms	<i>M</i> = 558 ms <i>SE</i> = 45 ms	<i>M</i> = 684 ms <i>SE</i> = 42 ms
Exp. 1B	40	75.12	< .0001	.65	<i>M</i> = 762 ms <i>SE</i> = 44 ms	<i>M</i> = 606 ms <i>SE</i> = 35 ms	<i>M</i> = 624 ms <i>SE</i> = 35 ms	<i>M</i> = 732 ms <i>SE</i> = 40 ms
Exp. 1C	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Exp. 2	42	84.88	< .0001	.67	<i>M</i> = 660 ms <i>SE</i> = 38 ms	<i>M</i> = 817 ms <i>SE</i> = 48 ms	<i>M</i> = 672 ms <i>SE</i> = 46 ms	<i>M</i> = 819 ms <i>SE</i> = 55 ms

APPENDIX F

CORRECT ‘NO’ RESPONSES

Participants’ correct responses to nontarget items (i.e., correct ‘no’ responses when the phoneme in the target position was /c/, /ç/, /l/, /m/, /r/ or /s/) were also analyzed. Correct ‘no’ responses were slower than correct ‘yes’ responses in all experiments except Experiment 1C. Additional 2 (prime: binary and ternary) x 2 (beat: on and off) x 2 (number: bisyllabic and trisyllabic) repeated measures ANOVAs were run on correct ‘no’ responses. Results revealed ‘trisyllabic advantage’ on certain cases. On the other hand, most significant results that were observed in the correct ‘yes’ responses were absent in the present analyses (cf. Table 1). Collectively, these results suggest correct ‘no’ responses were not differentiated with primes and are qualitatively different from correct ‘yes’ responses in this regard.

In Experiment 1A, there was a main effect of prime for pseudowords ($F(1, 41) = 5.46, p = .02, \eta_p^2 = .12$). Ternary primes ($M = 558$ ms, $SE = 36$ ms) resulted in faster reaction times compared to binary primes ($M = 595$ ms, $SE = 38$ ms). There was also an interaction of prime and number (see Table F2). All remaining effects were nonsignificant for pseudowords and all effects were nonsignificant for nonwords ($ps > .05$).

In Experiment 1B, there was an interaction of prime and number for nonwords (see Table F1). All remaining effects were nonsignificant for nonwords and all effects were nonsignificant for pseudowords ($ps > .05$).

In Experiment 1C, there was a main effect of prime ($F(1, 36) = 6.44, p = .02, \eta_p^2 = .15$). Binary primes ($M = 764$ ms, $SE = 41$ ms) resulted in faster reaction times compared to ternary primes ($M = 811$ ms, $SE = 44$ ms). The advantage of the 8-loop

binary prime was discussed in relation to a possible gestalt-driven perception in Section 4.1. There was also an interaction of prime and number in the overall analysis (see Table F1). All remaining effects were nonsignificant ($ps > .05$).

In Experiment 2, there was an interaction of prime and number in both nonwords and pseudowords (see Tables F1 and F2). All remaining effects were nonsignificant for both nonwords and pseudowords ($ps > .05$).

Table F1. Summary of Results for Correct ‘No’ Responses, Interaction of Prime and Number in Nonwords

	Test of Within-Subject Effects				Post-hoc comparisons			
					Binary prime		Ternary prime	
	Error df	F	p	η_p^2	Number match	Number mismatch	Number match	Number mismatch
Exp.1A	41	2.71	.11	.06	$M = 573$ ms $SE = 34$ ms	$M = 556$ ms $SE = 36$ ms	$M = 585$ ms $SE = 43$ ms	$M = 614$ ms $SE = 45$ ms
Exp. 1B	40	8.17	< .05	.17	$M = 644$ ms $SE = 37$ ms	$M = 619$ ms $SE = 36$ ms	$M = 638$ ms $SE = 41$ ms	$M = 664$ ms $SE = 37$ ms
Exp. 1C	36	32.98	< .0001	.48	$M = 791$ ms $SE = 45$ ms	$M = 701$ ms $SE = 38$ ms	$M = 769$ ms $SE = 41$ ms	$M = 853$ ms $SE = 50$ ms
Exp. 2	42	12.34	= .001	.23	$M = 729$ ms $SE = 38$ ms	$M = 686$ ms $SE = 37$ ms	$M = 684$ ms $SE = 40$ ms	$M = 748$ ms $SE = 41$ ms

Table F2. Summary of Results for Correct ‘No’ Responses, Interaction of Prime and Number in Pseudowords

	Test of Within-Subject Effects				Post-hoc comparisons			
					Binary prime		Ternary prime	
	Error df	F	p	η_p^2	Number match	Number mismatch	Number match	Number mismatch
Exp.1A	41	21.82	< .0001	.35	$M = 622$ ms $SE = 42$ ms	$M = 588$ ms $SE = 36$ ms	$M = 527$ ms $SE = 37$ ms	$M = 590$ ms $SE = 38$ ms
Exp. 1B	40	3.83	.06	.09	$M = 677$ ms $SE = 37$ ms	$M = 640$ ms $SE = 38$ ms	$M = 642$ ms $SE = 37$ ms	$M = 664$ ms $SE = 38$ ms
Exp. 1C	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Exp. 2	42	6.28	< .05	.13	$M = 754$ ms $SE = 46$ ms	$M = 715$ ms $SE = 44$ ms	$M = 693$ ms $SE = 41$ ms	$M = 715$ ms $SE = 36$ ms

REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Aksan, D. (2013). *Şiir dili ve Türk şiir dili: Dilbilim açısından bakış*. Ankara: Bilgi Yayınevi.
- Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica*, 66(1-2), 46-63.
- Beckman, M. (1982). Segment duration and the ‘mora’ in Japanese. *Phonetica*, 39(2-3), 113-135.
- Benadon, F. (2009). Speech rhythms and metric frames. In E. Chew, A. Childs, and C.-H. Chuan (Eds.), *Proceedings of the Second International Conference of the Society for Mathematics and Computation in Music* (pp. 22–31). Berlin, Germany: Springer-Verlag.
- Benadon, F. (2014). Metrical perception of trisyllabic speech rhythms. *Psychological Research*, 78(1), 113-123.
- Białyńska, A., & Dalla Bella, S. (2017). Music and speech distractors disrupt sensorimotor synchronization: effects of musical training. *Experimental Brain Research*, 235(12), 3619-3630.
- Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, 26(1), 86-125.
- Breen, M., Dilley, L.C., McAuley, J.D., & Sanders, L.D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, Cognition and Neuroscience*, 29(9), 1132-1146. doi:10.1080/23273798.2014.894642
- Brown, W. (1908). *Time in English verse rhythm*. New York: The Science Press.
- Busnel, M. C., Granier-Deferre, C., & Lecanuet, J. P. (1992). Fetal audition. *Annals of the New York Academy of Sciences*, 662(1), 118-134.

- Canette, L. H., Lalitte, P., Bedoin, N., Pineau, M., Bigand, E., & Tillmann, B. (2020). Rhythmic and textural musical sequences differently influence syntax and semantic processing in children. *Journal of Experimental Child Psychology*, 191, 1-19.
- Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica*, 155, 43-50.
- Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, 50(11), 2652-2658.
- Classe, A. (1939). *The rhythm of English prose*. Oxford: Blackwell.
- Cumming, R. E. (2010). *Speech rhythm: The language-specific integration of pitch and duration* (Unpublished doctoral dissertation). University of Cambridge, Cambridge, United Kingdom.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31(2), 218-236.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385-400.
- Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33(6), 824-844.
- Dalla Bella, S., Białuńska, A., & Sowiński, J. (2013). Why movement is captured by music, but less by speech: Role of temporal regularity. *PLoS ONE*, 8(8): 1-16.
- Daniele J. R., & Patel A. D. (2004). The interplay of linguistic and historical influences on musical rhythm in different cultures, in Lipscomb, S.D., Ashley, R., Gjerdingen, R.O., & Webster, P. (Eds.), *Proceedings of the International Conference on Music Perception and Cognition* (pp. 3-7). Sydney, Australia: Causal Publications.
- Dilçin, C. (2009). *Örneklerle Türk şiir bilgisi*. Ankara: Türk Dil Kurumu Yayınları.

- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: The case of French learners of Spanish. *Cognition*, 106(2), 682-706.
- Ergenç, İ., & Bekar Uzun, İ. P. (2017). *Türkçenin ses dizgesi*. Ankara: Seçkin Yayıncılık.
- Falk, S., & Dalla Bella, S. (2016). It is better when expected: aligning speech and motor rhythms enhances verbal processing. *Language, Cognition and Neuroscience*, 31(5), 699-708.
- Falk, S., Volpi-Moncorger, C., & Dalla Bella, S. (2017). Auditory-motor rhythms and speech processing in French and German listeners. *Frontiers in Psychology*, 8, 395. doi: 10.3389/fpsyg.2017.00395
- Fowler, C. A. (1979). “Perceptual centres” in speech production and perception. *Perception and Psychophysics*, 25(5), 375-388.
- Goswami, U. (2019). A neural oscillations perspective on phonological development and phonological processing in developmental dyslexia. *Language and Linguistics Compass*, 13(5), 1-21.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In N. Warner & C. Gussenhoven (Eds.), *Papers in laboratory phonology VII* (pp. 515-543). Berlin: Mouton de Gruyter.
- Hepper, P. G., & Shahidullah, B. (1994). The development of fetal hearing. *Fetal and Maternal Medicine Review*, 6(3), 167-179. doi:10.1017/S0965539500001108
- Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception: An Interdisciplinary Journal*, 21(2), 267-271.
- Inkelas, S., & Orgun, C. (2003). Turkish stress: A review. *Phonology*, 20(1), 139-161. doi:10.1017/S0952675703004482
- Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception: An Interdisciplinary Journal*, 26(3), 195-204.

- Jinbo, K. (1980). Kokugo no onseijou no tokushitsu (The top phonetic characteristics of Japanese). In Shibata, Kitamura & Kindaichi (Eds.), *Nihon no gengogaku (Linguistics of Japan)* (pp. 5-15). Tokyo: Taishukan.
- Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research *Perception & Psychophysics*, 41(6), 621-634. doi:10.3758/BF03210494
- Jones, M. R. (2016). Musical time. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 125-141). Oxford: Oxford University Press.
- Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech*, 51(4), 343-359.
- Kimball, A.E., Yiu, L.K., & Watson, D.G. (2019). Word recall is affected by surrounding metrical context. *Language, Cognition, and Neuroscience*, doi:10.1080/23273798.2019.1665190
- Ladefoged, P. (2001). *A course in phonetics* (4th ed.). Boston: Heinle and Heinle.
- Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, 54(5), 1228-1234.
- Lerdahl, F. (2001). The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences*, 930(1), 337-354.
- Liberman, M. Y. (1975). *The Intonational System of English* (Unpublished doctoral dissertation). MIT, Cambridge, MA.
- Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin & Review*, 18(6), 1035-1041.
- Lloyd James, A. (1940). *Speech signals in telephony*. London: Sir Isaac Pitman & Sons.
- Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4), 377-401.

- Mairano, P., & Romano, A. (2011). Rhythm metrics for 21 languages. In W. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1318–1321). Hong Kong, China: City University of Hong Kong.
- Mirşan, K. (1966). *Türk metriği*. İstanbul: Kutulmuş Matbaası.
- Monboddo, L. (1968). *Of the origin and progress of language*, vol. I. London: J. Balfour.
- Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers. *Psychological Review*, 83(5), 405- 408.
- Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora rhythm. *Language and Speech*, 50(1), 77-99.
- Myers, B. R. & Watson, D. G. (2019). Paying the meter: Effect of metrical similarity on word lengthening. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-019-01635-4
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756-766.
- 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), 1-19.
- Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell companion to phonology* (1147-1159). Oxford: Blackwell Publishing.
- Onay, A. T. (1996). *Türk şiirlerinin vezni*. Ankara: Akçağ Yayınları.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32(2), 258-278.
- Overy, K. (2000). Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children. *Psychology of Music*, 28(2), 218-229.

- Özsoy, A. S. (2004). *Türkçe'nin yapısı – I. Sesbilim*. Istanbul: Boğaziçi Üniversitesi Yayınevi.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7), 674-681.
- Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87(1), B35-B45.
- Patel, A. D. (2008). *Music, language, and the brain*. Oxford: Oxford University Press.
- Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Pike, K. 1947. *Phonemics: A technique for reducing languages to writing*. Ann Arbor: University of Michigan Press.
- Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., Kotz, S.A., & Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology*, 27(1), 121–131. doi:10.1037/a0031277.
- Pycha, A. (2006). A duration-based solution to the problem of stress realization in Turkish. *UC Berkeley Phonology Lab Annual Report*. 141–151.
- Ramus, F. (2002). *Acoustic correlates of linguistic rhythm: Perspectives*. In Bel, B., & Marlien, I. (Eds.), *Proceedings of Speech Prosody 2002* (pp. 115-120). Aix-en-Provence, France: Laboratoire Parole et Langage.
- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech synthesis. *Journal of the Acoustical Society of America*, 105(1), 512-521.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 263-292.
- Richards, S., & Goswami, U. (2019). Impaired recognition of metrical and syntactic boundaries in children with developmental language disorders. *Brain Sciences*, 9(2), 1-18.

- Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies: Essays in honour of F.R. Palmer* (pp. 73-79). London: Edward Arnold.
- Romano, A. (2009). Speech rhythm and timing: Structural properties and acoustic correlates. In S. Schmid, M. Schwarzenbach, & D. Studer (Eds.), *La dimensione temporale del parlato* (45-75). Torriana: EDK Editore.
- Scripture, E. W. (1902). *The elements of experimental phonetics*. New York: Charles Scribner's Sons.
- Sebastián-Gallés, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, 31(1), 18-32.
- Seifart, F., Meyer, J., Grawunder, S., & Dentel, L. (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society Open Science*, 5(4), 1-18.
- Steele, J. (1775). *Prosodia rationalis: Or, an essay towards establishing the melody and measure of speech*. London: J. Nichols.
- Steriade, D. (2012). *Intervals vs. syllables as units of linguistic rhythm*. [Class handout]. École d'Automne de Linguistique, Paris, France. Retrieved from <http://ealing.cognition.ens.fr/ealing2012/handouts/Steriade/EALING-DS-intervals1-hd.ppt.pdf>
- Şayli, Ö. (2002). *Duration analysis and modelling for Turkish text-to-speech synthesis* (Unpublished master's thesis). Boğaziçi University, Istanbul, Turkey.
- Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, 29(7), 382-390.
- Tanrıkorur, C. (1990). Concordance of prosodic and musical meters in Turkish classical music. *Turkish Music Quarterly*, 3(1), 1-7.
- Tanrıkorur, C. (1991). Türk mûsikîsinde usûl-vezin münasebeti. *Erdem*, 7(20), 709-732.
- Topbas, S. (2006). A Turkish perspective on communication disorders. *Logopedics Phoniatics Vocology*, 31(2), 76-88.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, 24(6), 744-755.

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, 58, 1-25.