Rhythmic grouping in English, Greek and Korean:
testing the iambic-trochaic law

Hae-Sung Jeon[1], Amalia Arvaniti[2]

[1] University of Central Lancashire, UK
[2] University of Kent, UK

HJeon1@uclan.ac.uk, a.arvaniti@kent.ac.uk

Abstract

The iambic-trochaic law (ITL) states that repeating sounds with an intensity contrast are perceived as binary groups with initial prominence (trochees) and those with a durational contrast with final prominence (iambs). Although the ITL has been empirically supported, it is not clear whether it is due to universal cognitive mechanisms or whether language-specific prosodic properties affect listeners' grouping preferences. We tested the law with speakers of English, Greek and Korean who heard strings of tones varied in duration and/or intensity. The results revealed neither significant differences among languages nor a strong bias shared by speakers of all languages. Significantly, listeners' grouping preferences were influenced by the duration of the inter-tone interval (ITI), with long ITI (200 ms) resulting in stronger trochaic preferences than short ITI (20 ms), indicating that specific experimental conditions may be responsible for cross-linguistic differences in listener responses across experiments testing the ITL.

Index Terms: iambic-trochaic law, meter, grouping, rhythm, perception, English, Korean, Greek

1. Introduction

Humans tend to impose higher-level structure even on sequences of identical tones or *cadences* [7], [22]. That is, a metrical relation emerges when listening to such sequences, so that the tones are heard as binary groups of strong-weak (trochee) or weak-strong (iamb). This ability (or *bias*) to impose structure to the acoustic signal is considered important in processing speech and music [19].

A particular manifestation of this bias is the iambic-trochaic law (henceforth ITL) which states that when subjects listen to a sequence of sounds in which every other sound is louder, listeners preferentially group them into trochees, i.e. with the louder sound being interpreted as group initial; if every other sound is longer, they preferentially group them into iambs, i.e with the longer sound being interpreted as group final [7], [22]. The ITL has often been considered a cognitive universal and research has shown it to apply to various types of stimuli and to participants from different linguistic and cultural backgrounds, e.g. [9]. On the other hand, cross-linguistic differences in grouping preferences, as well as a lack of such preferences, have also been reported, e.g. [5], [8], [21]. Of the studies showing cross-linguistic differences with respect to the ITL, Iversen, Patel & Ohgushi (2008) [12] are of particular interest here because they report differences in the grouping preferences of speakers of English and Japanese. In their study, both English and Japanese speakers preferred a trochaic grouping for tone sequences with an intensity contrast. However, [12] also found that while English speakers preferred iambic grouping for tone sequences with a duration contrast, Japanese listeners as a group did not show a strong preference with these sequences. Closer examination of listeners' responses revealed that unlike the English group, in which most listeners strongly preferred the iambic grouping with the duration contrast, Japanese listeners were still biased towards trochaic grouping. Based on these results, the central claim of [12] is that language acquirers implicitly learn their native language's rhythms which then exert top-down influence on the processing of basic auditory signals such as tones. A particular explanation put forward in [12] is that differences in (prosodic) head direction are the source of the phrase-level rhythmic differences across languages: [12] claim that English speakers hear iambs frequently because content words follow functors in English (e.g. 'the book'), while Japanese speakers hear mostly trochees because content words precede functors (e.g. *hon-ga*, 'book + subject marker').

The aim of the present study was to test this hypothesis with English, Greek and Korean speakers. In Korean, content words are followed by function words like in Japanese; e.g. *tɕʰɛk-i*, 'book + subject marker'. If the 'head-direction' hypothesis is correct, the results in [12] should be replicated, with English and Greek speakers behaving similarly (in both languages function words precede content words [18]), and Korean speakers behaving similarly to Japanese speakers. In addition, it is worth noting the prosodic differences across the languages studied here. Unlike English, which has primarily word-initial stress and uses several strong stress cues, Korean has no lexical stress or other word-level prominence; see [13], [14]. Greek is in between English and Korean; it has lexical stress (on one of the *last* three syllables of a word), but it is also quantity insensitive and uses less salient cues to stress than English; e.g. stress in Greek does not involve grammaticalized changes in vowel quality [1], [2].

There were two additional aims to the study. First, we aimed to investigate the effect of inter-tone interval (ITI) duration, since differences in ITI could create different temporal impressions and thus they could be the reason behind differences in the results reported in studies like [9] and [12] which relied in markedly different ITIs. Second, we aimed to examine the role of summation between duration and intensity (or *temporal integration* [17]) in the perception of rhythmic grouping. It is known that a longer stimulus sounds louder than a stimulus of equal average amplitude but shorter duration [4], [15]. This could mean that when assessing the role of duration

variation with respect to the ITL, it is not possible to eliminate the role of loudness (cf. [8]). Here we report the results of preliminary analyses demonstrating the effect of ITI and stimulus type. Due to lack of space the effect of Step is not discussed.

2. Methods

2.1. Participants

English, Greek and Korean participants were recruited. They were all monolingual speakers of one of the following dialects: Southern Standard British English [N = 38], Standard Greek [N = 27], and Standard Seoul Korean [N = 31]. They were all between 18 and 35 years of age and had not spent more than 6 months outside their native country. None of the participants had professional musical training and there were no self-reported problems in speaking or hearing. English speakers were tested in Canterbury and Manchester, UK. Greek speakers were tested in Canterbury, UK, and in Missolonghi and Athens, Greece. Korean speakers were recruited in London and in Seoul, South Korea. Data from participants who did not meet the recruitment criteria (e.g. they turned out to be bilingual) were excluded; the data analyzed are based on responses from 28 English, 25 Greek and 30 Korean speakers.

2.2. Experimental stimuli

The stimuli were strings of complex tones involving a *basic* and an *alternating* tone. The basic tone was a complex tone of 200 ms duration and 65 dB intensity composed of the F0 (250 Hz) and the next two odd harmonics (i.e. of the 1st, 3rd and 5th harmonic). The alternating tones differed from the basic tone in duration, intensity or their summation as shown in Table 1. Thus, in addition to the control stimuli, there were strings of three Stimulus Types, Duration, Intensity and Summation. For the Summation strings, an increase in duration along the steps (see below) was compensated for by a decrease in intensity. This set-up resulted in five steps in each Stimulus Type: strings with only basic tones (controls), and strings with basic tones and one of four additional steps per acoustic parameter, each involving an increasingly large difference between the basic and alternating tone from Step 4 (smallest difference) to Step 1. Stimuli were generated using Praat [6] with a 44.1 kHz sampling rate.

The tone strings were created by concatenating the basic tone, silence and an alternating tone. The silence or inter-tone interval (ITI) was either 20 ms or 200 ms creating a "short" and a "long" condition respectively. Each string was 11-12 sec long. In the "short" strings, there were 27-35 pairs of tones; in the "long" strings, there were 16-17 pairs. To reduce order effects known to affect responses in ITL experiments [22], the tone presentation order was counterbalanced; in addition, the intensity of each string was gradually increased according to a raised cosine function over the first 2.5 sec and decreased over the last 2.5 sec in each string.

Two practice stimuli (two steps) were also constructed for each Stimulus Type using different values from those in the test stimuli. This resulted in 24 practice trials (3 Stimulus Types × 2 Steps × 2 ITIs × 2 presentation orders). In the main experiment, there were in total 162 trials, 144 test trials [3 Stimulus Types × 4 Steps × 2 ITIs × 2 presentation orders × 3 repetitions] and 18 controls [3 Stimulus Types × 2 ITIs × 3 repetitions].

2.3. Experimental procedure

Participants were tested individually with a laptop and headphones in a quiet room. The same laptop and headphones were used for all participants. The experiment ran on DMDX. Before the experiment, listeners' hearing was tested by examining whether they could detect a 250 Hz tone of 200 ms duration at 25 dB. All participants passed the test and proceeded to the experiment. The stimuli were presented in blocks by Stimulus Type (Duration, Intensity, Summation). The order of the blocks was counterbalanced across subjects.

Table 1. *Acoustic parameters of basic and alternating tones for Duration (D), Intensity (I) and Summation (S) for Steps 1-4*

Step	Duration (ms)	Intensity (dB)	F0 (Hz)
basic	200	65	250
D1	100	65	250
D2	125	65	250
D3	150	65	250
D4	175	65	250
I1	200	53	250
I2	200	56	250
I3	200	59	250
I4	200	62	250
S1	100	68.2	250
S2	125	67.4	250
S3	150	66.6	250
S4	175	65.8	250

Each block started with a practice session of 8 trials (2 Steps × 2 ITIs × 2 presentation orders) followed by the test session which included 54 trials: 48 test trials [4 Steps × 2 ITIs × 2 presentation orders × 3 repetitions] and 6 controls [2 ITIs × 3 repetitions]. Stimulus order was pseudo-randomized for each subject. After listening to a stimulus, listeners pressed a labelled key on the keyboard to indicate their grouping choice; the choices were visually presented on the screen as shown in Fig. 1; the order of the two choices was counterbalanced across trials. Listeners could not respond to a stimulus and move onto the next screen until they heard the whole stimulus; there was no time limit for providing a response. Once they registered their response to grouping, the experiment automatically proceeded onto the next screen where participants rated their confidence using a three-point scale (3: completely certain; 2: somewhat certain; 1: guessing). At this stage, if listeners realized that they had pressed the wrong button about their grouping choice, they could press 0 instead of rating their confidence level (these responses were later discarded). After the experiment, the listeners completed questionnaires on their musical training and on their linguistic background (based on LEAP-Q [16]), and received a small payment. All instructions and questionnaires were in the participants' native language and the experiment was administered by a native speaker of that language.

3. Results

After mistakes indicated by listeners pressing '0' were removed (40 data points for English, 15 for Greek and 20 for Korean), the number of data points submitted to analysis was N = 4496

for English, N = 4034 for Greek and N = 4840 for Korean. The frequency of iambic responses was the dependent variable in all analyses.
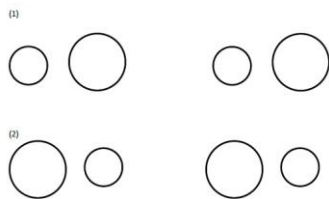

Fig. 1. *Listeners' choice, (1) iambic grouping, (2) trochaic grouping*

3.1.1. Control stimuli

In order to examine listeners' response bias, the percentages of iambic responses for the control stimuli (separately for each language, ITI and stimulus type) were tested against chance (one-sample *t*-test, α level adjusted following the Bonferroni procedure; see Fig. 2). There was no evidence that English participants were biased towards a particular response. Greek participants showed an above chance preference for trochaic grouping in the following conditions: Short-Summation [$t(24) = -2.74$, $p < .05$], Long-Summation [$t(24) = 03.21$, $p < .01$] and Long-Intensity [$t(24) = 03.21$, $p < .01$]. Korean speakers showed the same preference for Long-Duration [$t(29) = -3.3$, $p < .01$] and Long-Intensity [$t(29) = -2.65$, $p < .05$].

In addition, the responses to the control stimuli were tested by means of a three-way ANOVA with Language, Stimulus Type and ITI as factors. Language and ITI were not significant, but listeners, independently of native language, chose the trochaic grouping significantly more often in stimuli with Long than Short ITI, supporting the results presented immediately above [Long ITI mean = 38.09, sd = 30.09; Short ITI mean = 45.95, sd = 31.66; $F(1, 79) = 6.31$, $p < .05$].
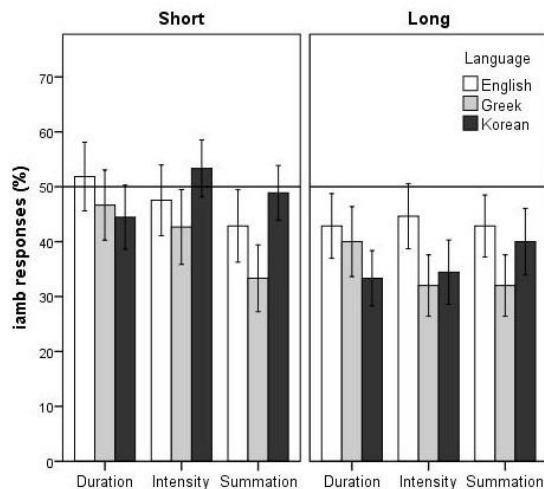

Fig. 2. *Means and standard errors of listeners' iambic responses (%) for control stimuli.*

3.1.2. Overall responses

The percentage of listeners' iambic responses is shown in Fig. 3. Generalised linear mixed-effect models were fitted to the data in R [20] using lme4 [3]. The dependent variable was listener's choice (0: trochee, 1: iamb). Subject was a random factor in all models; fixed factors in the full model included Stimulus Type, Language, ITI, and Step. The best-fitting model was identified using log-likelihood χ^2 tests; *p*-values were corrected for multiple pairwise comparisons using the mcp function in the multComp package [11].

In the model including all language groups, the significant fixed factors were: Stimulus Type [$\chi^2(2) = 21.02$, $p < .001$], ITI [$\chi^2(1) = 52.33$, $p < .001$], and Step [$\chi^2(4) = 57.16$, $p < .001$]. Language was not statistically significant [$\chi^2(2) = 4.87$, ns]. Tukey contrast tests revealed no significant differences between Stimulus Types either.

3.1.3. Within Language effects

Additional models were constructed separately for each language with Stimulus Type, ITI and Step (see Table 2). With respect to Stimulus Type, Tukey tests (shown in Table 3) indicated that English speakers were more likely to choose iambs for Duration than Intensity. Greek speakers were more likely to choose iambs for Duration than both Intensity and Summation, while there was no effect for the Korean speakers.
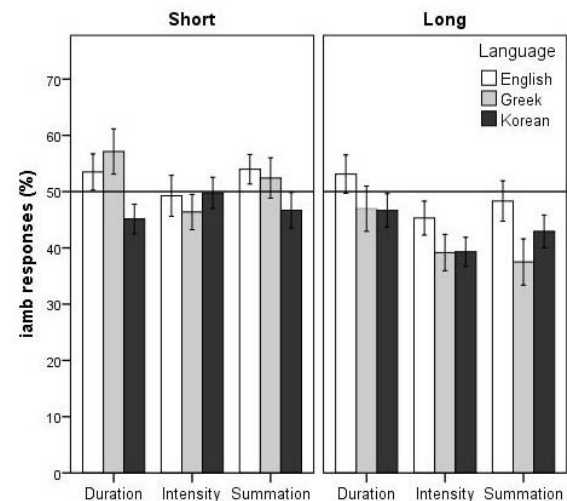

Fig. 3. *Listeners iambic reaponses (%) by ITI, language, and Stimulus Type. Responses for control stimuli are excluded. Error bars indicate ± one standard error.*

Table 2. *Summary of fixed effects (Stimulus Type (ST), ITI, Step) for each language.*

		English χ^2	Korean χ^2	Greek χ^2
ST	2	9.6***	0.1	24.4***
ITI	1	5.1**	13.5***	47.0***
Step	4	13.3*	14.6**	43.0***

3.1.4. Duration vs. Summation strings

In order to compare Duration and Summation directly, the data were split by ITI and models with Language, Stimulus Type and Step as fixed factors were constructed. The Language effect was not statistically significant ($\chi^2(2) = 5.25$, ns). Stimulus Type was not statistically significant for Short ITI either (est. = -0.04, SE = 0.06, z = -0.73, ns); it was, however, significant for Long ITI (est. = -0.22, SE = 0.06, z = -3.53, $p < .001$), suggesting that listeners were more likely to choose iambs for Duration than

Summation stimuli when ITI was long (though they generally did not show a strong preference for iambs).

Table 3. *Tukey test results for comparisons between Duration (D), Intensity (I) and Summation (S) for English and Greek; Stimulus Type effect was ns for Korean.*

	English			Greek		
	Est.	SE	z	Est.	SE	z
I-D	-0.23	0.07	-3.08**	-0.37	0.08	-4.61***
S-D	-0.1	0.07	-1.30	-0.31	0.08	-3.82***
S-I	0.13	0.074	1.78	0.06	0.08	0.8

4. Discussion

The present results do not reveal striking cross-linguistic differences. Participants' native language was not a significant factor in predicting grouping choice, while the most notable effect was from the ITI rather than the differences between the tones themselves (i.e. differences in duration, amplitude or their summation).

Listeners' responses plotted in Fig. 3 do not show grouping biases as strong as those reported in [9] or [12]. For example, the experimental conditions with short ITI (20 ms) in the present study are comparable to those in [12], yet with short ITI-Intensity, both our English and Korean listeners' responses were around chance, in contrast to the clear trochaic bias observed for both English and Japanese listeners in [12]. For a condition similar to our short ITI-Duration, [12] found a strong iambic bias in their English group (their iambic responses were at 70% or higher), but English participants in the present experiment showed average iambic responses below 60%.

The present results in Figs. 2 and 3 for the duration strings do show a weak preference for trochaic grouping among Korean speakers particularly in comparison to the other language groups. This is akin to the results from the Japanese participants in [12] and could at first indicate support for the 'head-direction' hypothesis. However, the Language effect was not statistically significant. In addition, the results of the Greek and Korean participants show more similarities, particularly for the long ITI condition, than the responses of Greek and English participants do. This is contrary to expectations based on the 'head-direction' hypothesis, since in English and Greek content words follow function words, while Korean has the opposite pattern.

A reason for the differences between our results and those of [12] could be that in [12] data from listeners who were confident in their grouping choice were highlighted in analysis, while the present analysis included all responses. This difference in itself could indicate that not everyone shares a strong bias. If so, then the present results suggest that the ITL may be related to listeners' *preferences*, as discussed in [10], than a general auditory bias.

Further, comparisons between the present results and [12] indicate that cross-linguistic differences may not generalize across experimental stimuli with different acoustic profiles. In [12] the alternating tones involved larger differences than those used here: e.g. in their duration strings, the ratio between short and long tones changed in three steps with the following ratios: 1:1.25, 1:1.75, 1:3. In contrast, the ratios were smaller in [9] and the present study: 1:1.14, 1:1.33, 1:1.6, 1:2. This could indicate that the perceptual integration of duration and amplitude (which is accentuated in higher ratios in the duration series) may be essential for a switch from trochees to iambs.

The above interpretation is supported by the comparison of the Duration and Summation strings used here which indicate that both duration and amplitude are needed to shift listeners' grouping preferences to iambs. Specifically, the results reported in §3.1.4 show that in Summation strings with long ITI listeners' preference for iambic grouping was weaker than it was in Duration strings, although tones in the Summation and Duration strings had the same durational profile. This suggests that if participants have any grouping preference, this would be a preference for trochees; they need the combined effects of longer duration and concomitant greater loudness to shift to a preference for iambs: duration alone does not easily accomplish this shift. Although cross-linguistic differences on this point, as reported in §3.1.3 and [8], are clearly worth investigating further, these results support the overall conclusion from the present experiment that the ITL may not be a *law* as such.

The possibility that the experimental set-up strongly influences listeners' performance is further supported by the ITI effect. This is notable because of the interaction of ITI with other factors such as Stimulus Type, and the effect the ITI had overall: listeners' responses to variations in acoustic parameters were more pronounced with long ITI (200 ms); participants were also more likely to choose trochees in strings with long than short ITI. This could be because the short ITI (20 ms) creates the impression that there is virtually no silence between the tones and the perceived continuity may hinder listeners from imposing a grouping structure in tone strings.

Finally, the interaction of ITI with other factors points to some cross-linguistic differences. For example, Greek and Korean listeners' responses to the controls were affected by the context in which they were heard, i.e., whether they were heard together with Duration, Intensity or Summation strings (see §3.1.1), while English speakers responses were consistently at chance level (see [9] for a similar finding). However, these preferences were not consistent within each language group. For instance, while Korean listeners showed a trochee bias with control stimuli presented with Duration and Intensity strings, their responses to non-control stimuli were not affected by Stimulus Type (see §3.1.3). English and Greek listeners, on the other hand, showed a stronger preference for iambs with Duration than Intensity strings, as predicted by the ITL. This result may be related to the presence of stress in Greek and English and its absence from Korean (see §1). Since there is no word-level prominence in Korean, Korean participants may be more sensitive to context when making their rhythmic grouping judgement. English participants, on the other hand, are the least swayed by context, while Greek participants present an intermediate pattern (showing some influence of context), which is consistent with more sparse stress patterns and weaker stress cues present in Greek as compared to English.

5. Conclusions

Taking previous studies and the present results together, the iambic-trochaic law seems to be a tendency but not necessarily a cognitive requirement for the processing of tones. Speakers of English, Greek and Korean seemed to be affected to a different extent in their grouping choice by inter-tone interval and the duration and intensity profile of the tones, but participants' native language was never a significant factor in predicting their grouping choice. Therefore, listeners' strong grouping bias and cross-linguistic differences may or may not be found depending on the precise nature of the stimuli used in ITL experiments.

6. Acknowledgements

7. References

[1] Arvaniti, A. 2000. The acoustics of stress in Modern Greek. *Journal of Greek Linguistics* 1: 9-39.

[2] Arvaniti, A. 2007. Greek phonetics: The state of the art. *Journal of Greek Linguistics* 8: 97-208.

[3] Bates, D., Maechler, M., Bolker, B. & Walker, S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, <URL: https://cran.r-project.org/web/packages/lme4/index.html >.

[4] Beckman, M. E. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.

[5] Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T. & Höhle, B. 2013. Native language affects rhythmic grouping of speech. *Journal of the Acoustical Society of America* 134: 3828-3843

[6] Boersma, P. & Weenink, D. 2014. *Praat: doing phonetics by computer* [Computer program]. Version 5.4.04, retrieved from http://www.praat.org/

[7] Bolton, T. L. 1894. Rhythm. *American Journal of Psychology* 6: 145-238.

[8] Crowhurst, M. J. & Olivares, A. T. 2014 Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech. *Phonology* 31: 51-94.

[9] Hay, J. F. & Diehl, R. L. 2007. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics* 69: 113-122.

[10] Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

[11] Hothorn, T., Frank Bretz, F., & Peter Westfall, P. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3): 346–363.

[12] Iversen, J. R., Patel, A. D. & Ohgushi, K. 2008. Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America* 124: 2263-2271.

[13] Jeon, H.-S. 2015. Prosody. In Brown, L. & Yeon, J. (Eds.), *Handbook of Korean Linguistics*, pp. 41-58. New York: Wiley-Blackwell.

[14] Jun, S.-A. 2014. Prosodic typology: by prominence type, word prosody, and macro-rhythm. In: Jun, S.-A. (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing*, pp. 520-540. Oxford University Press.

[15] Lieberman, P. 1960. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America* 32: 451-454.

[16] Marian, V., Blumenfeld, H. K. & Kaushanskaya, M. 2007. The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research* 50: 940-967.

[17] Moore, B. C. J. 2012. *An Introduction to the Psychology of Hearing*. 6th edition. Bingley, UK: Emerald Group Publishing Limited.

[18] Nespor, M. & Vogel, I. 1986. *Prosodic Phonology*, Dordrecht: Foris Publications.

[19] Patel, A. D. 2008. *Music, Language and the Brain*. Oxford University Press.

[20] R Development Core Team. 2015. *R: A Language and Environment for Statistical Computing*, ver. 3.20 [computer software].

[21] Trainor, L. J. & Adams, B. 2000. Infants and adults use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics* 62: 333-340

[22] Woodrow, H. 1909. A quantitative study of rhythm: The effect of variations in intensity, rate and duration. *Archives of Psychology* 14: 1-66.