32 Individual and Language Differences in Rhythm Grouping Preferences: The Iambic–Trochaic Law Revisited

AMALIA ARVANITI AND HAE-SUNG JEON

32.1 Introduction

32.1.1 Background

Psychological experiments have long shown that when humans hear stimuli sequences alternating in duration or intensity, they tend to group them as iambs (short-long) and trochees (loud-soft), respectively (Bolton, 1894; Woodrow, 1909). The effect of intensity and duration on grouping biases is known as the iambic–trochaic law (henceforth ITL). It is said to reflect universal cognitive tendencies that lead to subjective rhythmisation and, according to Hayes (1985, 1995), shape the typology of linguistic stress and consequently rhythm (for alternative views, see Fuchs (Chapter 30) and Barros et al. (Chapter 34)). In particular, Hayes (1985, 1995) has argued that in unmarked cases, alternating stress will group syllables in accordance with the ITL, and should favour iambs when feet are sensitive to quantity and the positions of heavy syllables (see Hyde, 2011, for a review of the ITL in phonology). The ITL is also said to play a role in language acquisition, as newborn infants exhibit perceptual biases towards the predominant grouping of their native language (e.g., Allen, 1975; Abboub et al., 2016; Molnar et al., 2016). Finally, the ITL is also said to have a hand in processing music (e.g., Drake and Bertrand, 2001; Hannon and Trehub, 2005; see Crowhurst, 2020, for a review).

Recent experimental studies, however, indicate that grouping preferences are not always easy to elicit, resulting in responses that remain only slightly above chance and rarely exceed 70% (Hay and Diehl, 2007; Iversen et al., 2008; Bhatara et al., 2013; Crowhurst and Teodocio, 2014). Studies also report asymmetries, with preferences for trochees in response to intensity alternations being stronger and more consistent than duration-based preferences for iambs (Trainor and Adams, 2000; Iversen et al., 2008; de la Mora et al., 2013; Crowhurst, 2016; Crowhurst et al., 2016; Molnar et al., 2016). Finally, responses may be shaped by experience, including musicality (Boll-Avetisyan et al., 2017) and exposure to an L2 (Boll-Avetisyan et al., 2016), but mostly relating to L1 (first-language) prosody, such as a language's dominant stress pattern (Crowhurst and Teodocio, 2014), the presence of significant pre-boundary lengthening (Molnar et al., 2016), and the position of the phrase head (Iversen et al., 2008). For instance, unlike native English speakers, Japanese and Zapotec speakers group stimuli alternating in duration into trochees (Iversen et al., 2008; Crowhurst and Teodocio, 2014, respectively). Cross-linguistic differences also apply to the strength of agreement among participants: in Iversen et al. (2008), Japanese participants agreed less with each other than English participants; similarly, in Bhatara et al. (2013), French participants agreed less with each other than German participants.

Cross-study discrepancies may be also due to experimental manipulations: responses to linguistic stimuli are more influenced by the participants' native language (Crowhurst, 2016; Molnar et al., 2016), while studies using large ratios between alternating stimuli yield stronger preferences (cf., Iversen et al., 2008, versus Hay and Diehl, 2007). Here, we explore possible reasons for such disparities by means of two experiments with native speakers of English, Greek, and Korean.

32.1.2 Prosodic Features of the Tested Languages and Related Predictions

English is a stress accent language (Beckman, 1986). Stressed syllables are hyperarticulated (leading to changes in segmental quality, duration, and intensity), while unstressed syllables are markedly reduced (de Jong, 1995). Stressed syllables are often (but not necessarily) accompanied by changes in pitch as the outcome of intonation (Ladd, 2008). In English, most content words start with a stressed syllable (Cutler and Carter, 1987; Clopper, 2002; Ernestus and Neijt, 2008), a pattern exploited in perception (e.g., Donselaar et al., 2005). Stress adjustments, such as the rhythm rule, lead to trochees and a regular alternation of strong and weak metrical constituents (Hayes, 1995), for example, Chìnése éxpert > Chínèse éxpert. In short, English speakers are used to hearing trochees based on large acoustic differences between stressed and unstressed syllables. They are also sensitive to longer duration being associated with phrase-finality (e.g., Turk and Shattuck-Hufnagel, 2000; Fletcher, 2010; among many; cf., Iversen et al., 2008). Based on the above, we expected English participants to show a preference for trochees and, given the preponderance of trochees in the language, we expected that only large differences in duration would override this preference and lead to iambic grouping.

Greek has stress accent (Arvaniti, 2000). Stressed syllables are longer and louder, but the vowels of unstressed syllables are not significantly centralised (Fourakis et al., 1999; see Arvaniti, 2007, for a review). Each word, independently of length, has one stress on one of the last three syllables (Revithiadou and Lengeris, 2016).[1] Penultimate stress is the dominant pattern (approximately 45% of words with two or more syllables have penultimate stress; Protopapas, 2006), giving rise to frequent trochees. Greek speakers are sensitive to lexical stress during processing (Arvaniti and Rathcke, 2015; Protopapas et al., 2016). Phrase-final lengthening is limited (Arvaniti, 2007). In short, Greek speakers are not used to stress-based binary alternations at phrase level but are accustomed to a preponderance of trochees at word level. Based on the above, we expected Greek participants to show a preference for trochees

[1] Nespor and Vogel (1989) argue that Greek has rhythmic stresses. However, production, perception, and phonological evidence do not support this postulation (Arvaniti, 2007, for a review; Arvaniti and Rathcke, 2015; Andrikopoulou et al., 2021).

with both intensity and summation sequences, which reflect the integral of duration and intensity (see Section 32.1.3 for details), and a weak preference for iambs with duration sequences, since iambic patterns are not frequently found in the language.

Seoul Korean (the standard variety) has neither lexical stress nor lexical pitch accent (Jun, 2005). Its accentual phrase (AP) is primarily demarcated by the pitch contour (see Jun, 2005, and Jeon, 2015, for reviews). APs are on average 3.2 syllables long and contain on average 1.2 content words, often followed by particles that do not undergo phonetic reduction (Jun and Fougeron, 2000). The articulation of domain-initial consonants gets stronger at higher levels of the prosodic hierarchy, a phenomenon known as 'articulatory strengthening' (Cho and Keating, 2001; Keating et al., 2004). It is not clear, however, that articulatory strengthening leads to greater intensity (Cho and Keating, 2001). Domain-final lengthening, on the other hand, is extensive at the intonational phrase level (see Jeon, 2015, for a review) and listeners exploit it for speech segmentation (Jeon and Arvaniti, 2017). As native Korean speakers show uncertainty in detecting metrical prominence (Lim, 2001; Guion, 2005) and the intensity cues to the AP are unreliable, they may not show strong grouping preferences with respect to the intensity and summation sequences that reflect metrical strength differences; on the other hand, Korean speakers should prefer iambs with duration sequences, since long duration is a strong group-final cue in Korean.

32.1.3 Experimental Manipulation and Hypotheses

Our hypotheses relate to language-specific influences, discussed in Section 32.1.2, and experimental manipulations that would apply in the absence of language-related differences among groups. Our manipulations are listed below.

First, we used complex tones with three harmonic components and manipulated their duration, intensity, and summation. In the summation sequences, decreases in duration were compensated for by concomitant increases in intensity (see Section 32.2.1.2). This manipulation was based on the observation that amplitude and duration are perceptually integrated, so that a shorter stimulus sounds softer than a longer stimulus of equal average intensity (e.g., Woodrow, 1909; Beckman, 1986; Moore, 2012). There are indications that temporal summation may affect ITL responses: Bolton (1894) found that sequences in which short but loud sounds alternated with long but soft sounds resulted in a preference for trochees (cf., Crowhurst and Teodocio, 2014). Crucially, temporal summation suggests that typical manipulations of duration in ITL experiments – in which intensity stays intact – lead to the longer elements sounding not only longer but also louder. The summation manipulation used here aimed at minimising the auditory boost provided by intensity. If alternating elements in duration sequences must sound both longer and louder to elicit iambic responses, then we would expect typical duration sequences to yield stronger iambic preferences than summation sequences.

Second, we included two inter-stimulus interval (ISI) conditions, such that the duration of the silent interval between tones was either 20 ms (as in Iversen et al., 2008) or 200 ms (as in Hay and Diehl, 2007). Hay and Diehl (2007) did not find cross-linguistic differences, while Iversen et al. (2008) did. The difference could be due to the short ISI being temporally more similar to running speech. This is supported by studies using successions of syllables without breaks in between, which also report cross-linguistic differences (e.g., Crowhurst, 2016). Thus, we expected to find greater cross-linguistic differences with short ISI.

The third manipulation related to steps, that is, the stepwise increase in acoustic contrast between the alternating tones in a sequence. Based on previous work (e.g., Woodrow, 1909; Iversen et al., 2008; Bhatara et al., 2013; Crowhurst, 2016), we anticipated stronger preferences with increased differences between tones.

32.2 Experiment 1

32.2.1 Methods

32.2.1.1 Participants

The analysis is based on responses from 28 speakers of Southern Standard British English (13 females; age mean = 20.27, sd = 1.79), 25 speakers of Athenian Greek (19 females; age mean = 24.08, sd = 4.97), and 30 speakers of Seoul Korean (16 females; age mean = 24.67, sd = 4.25). The data of 13 participants (10 English, two Greek, one Korean) who did not meet the recruitment criteria (e.g., they turned out to be early bilinguals or had language impairments) were excluded. British participants had limited exposure to languages other than English. Greeks and Koreans had learnt at least one other language (mostly English) through formal instruction, as is the norm in Greece and South Korea, respectively. However, none had prolonged contact (> six months) with any language other than their L1. No participants had professional musical training or reported problems with speaking, hearing, or motor control. All participants gave informed consent and were modestly remunerated.

32.2.1.2 Stimuli and Experimental Procedures

The stimuli were sequences of complex tones involving a 'standard' alternating with a 'comparison' (see Figure A, Supplementary Materials, at https://osf.io/sw3c5/). The tones were generated in Praat (Boersma and Weenink, 2014) with a 44.1 kHz sampling rate. As shown in (1), the standard was a complex tone of 200 ms duration and 65 dB intensity, with a rise time of 15 ms, composed of the fundamental frequency (f0) (250 Hz) and the next two odd harmonics.

(1) $1/2 \times (\sin(2 \times pi \times 250 \times x) + \sin(2 \times pi \times 750 \times x) + \sin(2 \times pi \times 1250 \times x))$

The comparison tones differed from the standard in duration, intensity, or their summation, as shown in Table 32.1. For summation sequences, decreases in duration were compensated for by increases in intensity, using an approximate -3 dB slope for the doubling of duration (Moore, 2012). This set-up resulted in five types of tone sequences per acoustic parameter: sequences in which standards alternated with a comparison step, and sequences of standards (controls). Controls were included to investigate grouping biases (see Hay and Diehl, 2007; Crowhurst, 2018).

The ISI manipulation created an 'ISI-Short' (ISI = 20 ms) and an 'ISI-Long' (ISI = 200 ms) condition. Each sequence was 11–12 s long. To reduce order effects (due to sequences starting or ending in a prominent tone), the intensity in each sequence was gradually increased according to a raised cosine function over the first 2.5 s and decreased over the last 2.5 s. Additionally, for each step, both sequences starting with the standard and ending with the comparison tone and sequences starting with the comparison and ending with the standard

Table 32.1 Experimental stimuli

Acoustic parameters of comparison tones for Intensity (I), Duration (D), and Summation sequences (S) for Steps 1–4 in Experiment 1; *f0 was held constant at 250 Hz. The standard (Step 0) was 200 ms in duration, with 65 dB intensity. Items in bold were used in Experiment 2 as well (see* Section 32.3.1*)*

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

were used (referred to below as the 'standard-comparison order'). Finally, two practice stimuli were constructed per sequence type, using larger differences between standard and comparison.

The stimuli were presented in three blocks (Duration, Intensity, and Summation) in counterbalanced order across participants, resulting in six block orders. Each block started with a practice session of eight trials (two steps × two ISIs × two standard-comparison orders) followed by the test session. Each test session included 54 trials: 48 test trials (four steps × two ISIs × two standard-comparison orders × three repetitions) and six controls (two ISIs × three repetitions), for a total of 162 trials. Trial order was pseudo-randomised per participant so the same sequence was not heard twice in a row.

The experiment ran on DMDX. Participants were tested individually in a quiet room using the same laptop and headphones. Before the experiment, the participants' 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.

Participants were told they would hear tone sequences lasting approximately 10 seconds each. They had to decide how the tones were grouped by selecting one of two pictures, presented to them in counterbalanced order across trials and used to minimise cross-linguistic differences in how terms such as short, long, soft, and loud are expressed and understood (see Figure B, Supplementary Materials, at https://osf.io/sw3c5/; see Bhatara et al., 2013). Iambs were illustrated with two repetitions of a small circle followed by a large circle in a group, trochees by the reverse circle order.

After listening to each stimulus, participants pressed a labelled key on the keyboard to indicate their choice. Choice order was counterbalanced across trials. Following Iversen et al. (2008), participants could respond only after they heard the entire tone sequence. Upon registering their response, the experiment automatically proceeded to the next screen, where participants rated their confidence (3 = completely certain; 2 = somewhat certain; 1 = guessing). If they thought they had chosen the wrong grouping, they could press 0 instead of rating their confidence level. The experiment was self-paced.

After the experiment, the participants completed questionnaires on their musical training and linguistic background. The linguistic background questionnaire was adapted from LEAP-Q (the Language Experience and Proficiency Questionnaire; Marian et al., 2007) to fit linguistic and cultural expectations of each group. The questionnaires were used to ensure that participants did not have extensive exposure to an L2 or musical training to a professional standard. All instructions and questionnaires were in the speakers' native language and administered by research assistants who were native speakers of that language.

32.2.2 Results

The experiment yielded 13,446 data points (83 participants × 162 trials) of which 13,373 (English, N = 4496; Greek, N = 4037; Korean, N = 4840) were analysed. The other 73 data points were removed (English, N = 40; Greek, N = 13; Korean, N = 20), as participants had pressed 0 to indicate their choice was invalid. The differences in confidence ratings were minimal, and the only clear trend was that ratings increased along the steps (see supplementary Table A at https://osf.io/sw3c5/). Thus, no data were excluded based on the confidence score.

Mixed-effect logistic models were fitted to the response data using the glmer function in the lme4 package (Bates et al., 2015b) with R version 3.5.1 (R Core Team, 2018). The models estimated the maximum likelihood of the positively coded trochee response (0: iamb, 1: trochee). Separate models were fit to the data based on STIMULUS TYPE (Intensity, Duration, Summation); each model included LANGUAGE (English, Greek, Korean), ISI (Short, Long), and STEP as predictors. STEP was coded as a continuous variable ranging from 0 to 4 to examine the trend in responses associated with the change of one Step; step 0 corresponded to the control sequences.

We started with the simplest model, which included the aforementioned predictors and random intercepts for participants; interactions and random slopes for ISI and STEP were subsequently added (Bates et al., 2015a). The effects of these were determined by model comparison (anova function, $p < 0.05$*, $p < 0.01$**, $p < 0.001$***). When a convergence error occurred, we used the allFit() function (package optimx, Nash and Varadhan, 2011) to check whether all optimisers produced similar values. In identifying the best-fitting model, we added an interaction term to the model only when it significantly improved the model fit from a lower-order model and also referred to the AIC (Akaike information criterion) value. Lower AIC values indicate a better fit.[2] Below, we focus only on predictors that were statistically significant. Due to space limitations, the figures are provided as online supplementary materials (see Figures C1–C3, https://osf.io/sw3c5/). The model outputs are summarised in Table 32.2.

For Intensity, only ISI (est. = 0.33, $p < 0.001$) reached statistical significance: participants were more likely to choose trochees for ISI-Long than ISI-Short, for which responses were around chance level in all Steps and for all languages (see Figure C1). For Duration, STEP was significant, with increasingly larger differences between the standard and comparison leading to decreases in trochaic responses (est. = −0.20, $p < 0.001$). However, the iambic grouping preference was shown only for Step 4 for English and Greek participants (see Figure C2). Korean participants showed an overall preference for trochees (est. = 0.38, $p < 0.04$), which

[2] We also tested models that included two factors from our music experience questionnaire: (a) the number of hours of listening to music per day, and (b) whether participants played one (or more) musical instruments. Neither factor improved model fit, so they are not included.

waned as Step increased. For Summation, only STEP was significant, such that increasingly larger differences between the standard and comparison led to decreases in trochaic responses (est. = -0.13, $p < 0.01$), though not to a switch to iambs (see Figure C3).

32.2.3 Interim Discussion

The results weakly support the ITL as well as our prediction of a diminished effect of duration when counterbalanced by amplitude changes in the Summation sequences. The results partially support our prediction that preferences would be stronger with larger acoustic differences between standard and comparison; this applied particularly with the Duration sequences. On the other hand, short ISI did not lead to the hypothesised stronger and language-based preferences. However, Language did have an effect: the English and Greek participants, as expected, showed a preference for iambs with the Duration sequences (with Step 4, 100 ms in Figure C2), while Korean participants, against our predictions, did not.

These results are not as striking as those of earlier studies (e.g., Iversen et al., 2008, or Crowhurst, 2016), even when the stimuli were comparable. A possible reason could be that our experiment lasted approximately an hour and may have fatigued participants or led to habituation. To address this issue, we conducted Experiment 2, which included only sequences with the larger acoustic contrasts from Experiment 1 (Steps 3 and 4).

32.3 Experiment 2

32.3.1 Methods

Recruitment criteria and methods were the same as in Experiment 1. Below, we present only changes made to the experiment.

32.3.1.1 Participants

The results are based on responses from 36 English (19 females; age mean = 23.36, sd = 4.58), 39 Greek (19 females; age mean = 24.03, sd = 3.33), and 35 Korean participants (20 females; age mean = 22.47, sd = 2.64). Data from an additional 15 participants (five English, two Greek, eight Korean) who did not meet the recruitment criteria were discarded.

32.3.1.2 Stimuli and Experimental Procedures

Experiment 2 included the same controls (Step 0) as Experiment 1, plus Steps 3 and 4 of Experiment 1, referred to here as Steps 1 and 2, respectively (see Table 32.1).

The procedures were the same as in Experiment 1 except for two points: Experiment 2 was conducted using PsychoPy version 1.83.01, and participants could change their grouping choice, if they wished, before answering the confidence-rating question. There were 24 practice trials (three sequence types × two steps × two ISIs × two standard-comparison orders). In the main

Table 32.2 Statistical model summary for Experiment 1

Experiment 1, best-fitting model summary for Intensity (n = 4460, AIC = 6005.1, PARTICIPANT with random slopes for ISI), Duration (n = 4456, AIC = 5911.6, PARTICIPANT with random slopes for STEP and ISI), and Summation (n = 4455, AIC = 5953.4, PARTICIPANT with random slopes for ISI). Reference level: Language-English, ISI-Short, and Step-Control (0).

		est.	*SE*	*z*	*p*
Intensity	Intercept	−0.19	0.12	−0.16	0.87
	Step	−0.01	0.02	−0.61	0.54
	Language-Greek	0.23	0.14	1.64	0.1
	Language-Korean	0.13	0.13	0.97	0.33
	ISI-Long	**0.33**	**0.09**	**3.6**	**< 0.001*******
Duration	Intercept	0.26	0.15	1.73	0.08
	Step	**−0.20**	**0.04**	**−5.18**	**< 0.001*******
	Language-Greek	−0.06	0.20	−0.30	0.76
	Language-Korean	**0.38**	**0.19**	**2.02**	**0.04***
	ISI-Long	0.07	0.18	0.36	0.72
	ISI-Long × Language-Greek	0.38	0.27	1.43	0.15
	ISI-Long × Language-Korean	−0.06	0.25	−0.25	0.80
Summation	Intercept	0.16	0.15	1.06	0.29
	Step	**−0.13**	**0.04**	**−3.17**	**< 0.01****
	Language-Greek	0.25	0.23	1.13	0.26
	Language-Korean	0.01	0.21	0.05	0.96
	ISI-Long	0.23	0.17	1.33	0.18
	Step × Language-Greek	−0.07	0.06	−1.09	0.28
	Step × Language-Korean	0.11	0.06	1.91	0.06
	ISI-Long × Language-Greek	0.39	0.25	1.54	0.12
	ISI-Long × Language-Korean	−0.04	0.24	−0.15	0.88

Significance codes: $p < 0.05$*, $p < 0.01$**, $p < 0.001$***.

experiment, there were in total 90 trials, 72 test trials (three sequence types × two steps × two ISIs × two standard-comparison orders × three repetitions) and 18 controls (three sequence types × two ISIs × three repetitions). The experiment lasted approximately 30 minutes.

32.3.2 Results

The analysis is based on 9,770 data points; there were 130 missing responses (English, N = 99; Greek, N = 11; Korean, N = 20), yielding 3,141 data points for English, 3,499 for Greek, and 3,130 for Korean. Differences in confidence ratings were minimal across conditions and languages (see supplementary Table B, at https://osf.io/sw3c5/). Thus, no data were excluded based on the confidence score. The modelling procedure was the same as in Experiment 1. The model outputs are summarised in Table 32.3 (see also supplementary Figures D1–D3, at https://osf.io/sw3c5/).

For Intensity, listeners were more likely to choose trochees with ISI-Long (est. = 0.38, $p < 0.01$). In addition, the Greek listeners showed an increase in trochee responses with increased STEP (est. = 0.54, $p < 0.01$), but Korean and English listeners did not (see Figure D1). For Duration, larger STEP differences led to a decrease in trochaic responses (Table 32.3, est. = −0.61, $p < 0.001$), but this effect depended on LANGUAGE. For English, the effect of STEP was minimal, as participants preferred iambs (Figure D2). For Greek, there was a decrease in trochee responses with each step, particularly for ISI-Long (est. = −1.46, $p < 0.001$), while the Korean group retained a preference for trochees (est. = 0.7, $p < 0.05$). For Summation, responses were influenced by LANGUAGE in interaction with STEP and ISI, with the Greek participants showing a decrease in trochaic responses with each step, particularly with ISI-Long (est. = −0.89, $p < 0.001$). However, the English and Korean participants' responses showed no consistent and strong group preferences (Figure D3).

32.3.3 Interim Discussion

The expectation that the shorter experiment would lead to stronger preferences was not strongly supported. For Intensity and Summation, the results were largely comparable to those of Experiment 1, although Experiment 2 showed more significant effects related to LANGUAGE. For Duration, English participants showed the expected preference for iambs, as did the Greek group (with the largest step only), while the Korean group retained a preference for trochees. As we found discrepancies in the results of Experiments 1 and 2, we examined individual variation in responses to determine the extent to which it drives aggregate results.

32.4 Individual Differences in Responses

The individual response data (Figure 32.1) revealed variation across participants but overall chance-level responses dominated. When a preference was shown, participants slightly preferred trochees (16% on average) over iambs (12% on average); see Table 32.4 (for raw counts, see Table B, Supplementary Materials, at https://osf.io/sw3c5/).

A preference for trochees and individual variation were evident for control sequences as well. We ran binomial tests to determine whether individual participants' responses were significantly different from chance. The results showed that responses to controls (supplementary Figure E) were significantly different from chance for all three groups in Experiment 1 (for English, $p < 0.05$; for Greek, $p < 0.001$; for Korean, $p < 0.001$), but only for Koreans in Experiment 2 (for English, $p = 0.32$; for Greek, $p < 0.24$; for Korean, $p < 0.001$). However, at the individual level, only a small number of participants, mostly Greek and Korean, showed clear preferences, largely for trochees (Table 32.4; see also supplementary Table C and Figure F). This may explain the preference for trochees with the experimental sequences as well.

Table 32.3 Statistical model summary for Experiment 2

Experiment 2, the best-fitting model summary for Intensity (n = 3255, AIC = 4252), Duration (n = 3250, AIC = 4004.3), and Summation (n = 3265, AIC = 4201.3); for all models, PARTICIPANT with random slopes for ISI and Step. Reference level: Language-English, ISI-Short, and Step-0.

		est.	*SE*	*z*	*p*
Intensity	Intercept	0.10	0.18	0.55	0.58
	Step	−0.13	0.13	−1.05	0.29
	Language-Greek	−0.40	0.24	−1.64	0.10
	Language-Korean	0.22	0.25	0.88	0.38
	ISI-Long	**0.38**	**0.13**	**2.93**	**< 0.01****
	Step × Language-Greek	**0.54**	**0.18**	**3.06**	**< 0.01****
	Step × Language-Korean	−0.11	0.18	−0.59	0.56
Duration	Intercept	−0.11	0.24	−0.46	0.64
	Step	**−0.61**	**0.16**	**-3.94**	**< 0.001*****
	Language-Greek	−0.09	0.33	−0.29	0.77
	Language-Korean	**0.70**	**0.35**	**2.00**	**< 0.05***
	ISI-Long	−0.21	0.35	−0.6	0.55
	Step × Language-Greek	**0.68**	**0.21**	**3.31**	**< 0.001*****
	Step × Language-Korean	0.05	0.21	0.23	0.81
	Step × ISI-Long	0.32	0.20	1.65	0.10
	ISI-Long × Language-Greek	**2.06**	**0.49**	**4.23**	**< 0.001*****
	ISI-Long × Language-Korean	0.13	0.50	0.25	0.80
	Step × Language-Greek × ISI-Long	**−1.46**	**0.27**	**-5.41**	**< 0.001*****
	Step × Language-Korean × ISI-Long	−0.06	0.28	−0.23	0.82
Summation	Intercept	0.20	0.22	0.92	0.36
	Step	−0.21	0.15	−1.46	0.15
	Language-Greek	−0.27	0.30	−0.88	0.38
	Language-Korean	0.18	0.21	0.86	0.39
	ISI-Long	0.21	0.31	0.69	0.49
	Step × Language-Greek	**0.50**	**0.20**	**2.48**	**0.01***
	Step × Language-Korean	0.18	0.21	0.86	0.39
	Step × ISI-Long	−0.21	0.18	−1.14	0.26
	ISI-Long × Language-Greek	**1.24**	**0.43**	**2.86**	**< 0.01****
	ISI-Long × Language-Korean	0.29	0.44	0.66	0.51
	Step × Language-Greek × ISI-Long	**−0.89**	**0.26**	**-3.48**	**< 0.001*****
	Step × Language-Korean × ISI-Long	−0.14	0.26	−0.54	0.59

Significance codes: $p < 0.05$*, $p < 0.01$**, $p < 0.001$***.

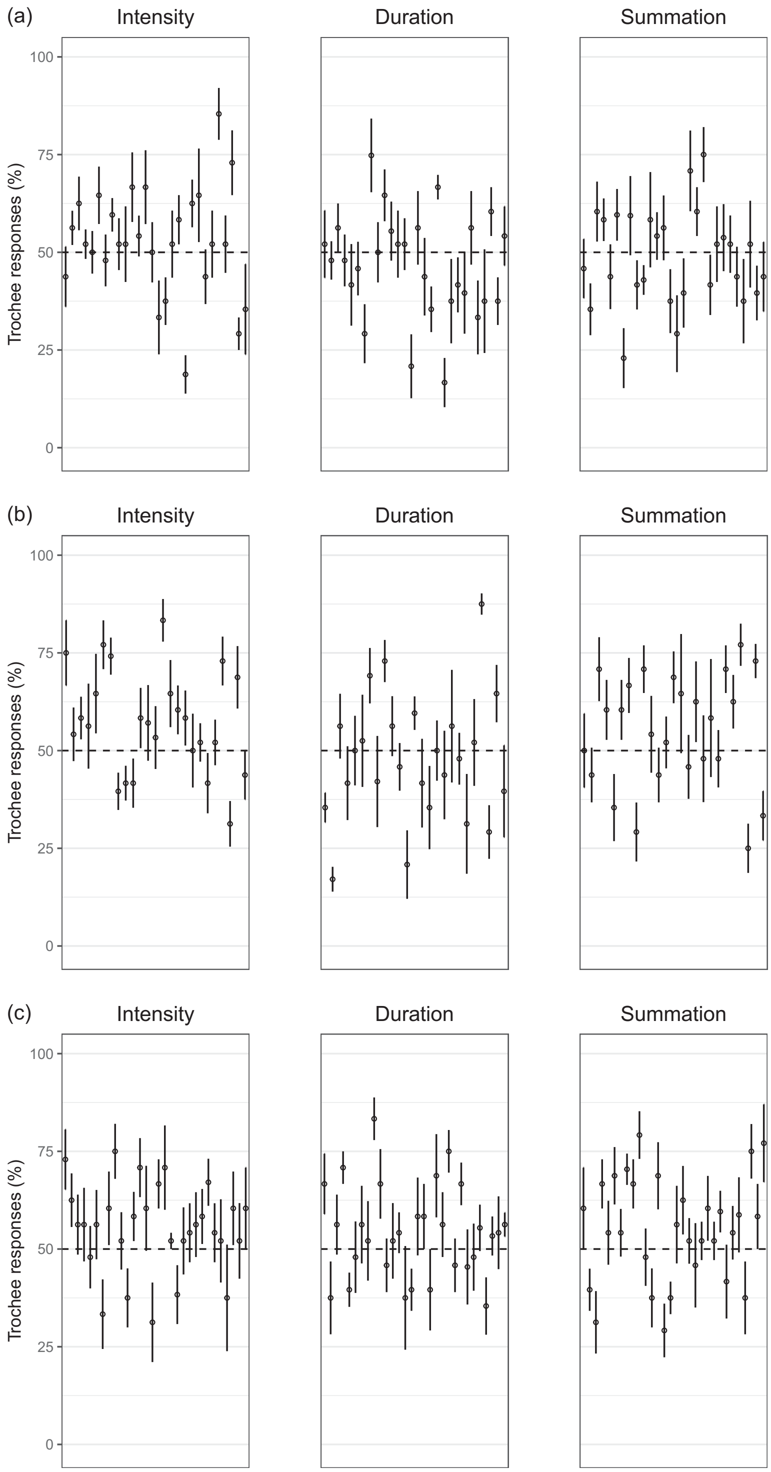

Figure 32.1 Individual responses by language and sequence type.
Individual participant responses by language and sequence type in Experiment 1 (a: English, b: Greek, c: Korean) and Experiment 2 (d: English, e: Greek, f: Korean). Each dot represents the mean percentage of trochee responses from one participant. The error bars show ±1 standard error. Responses to controls excluded.

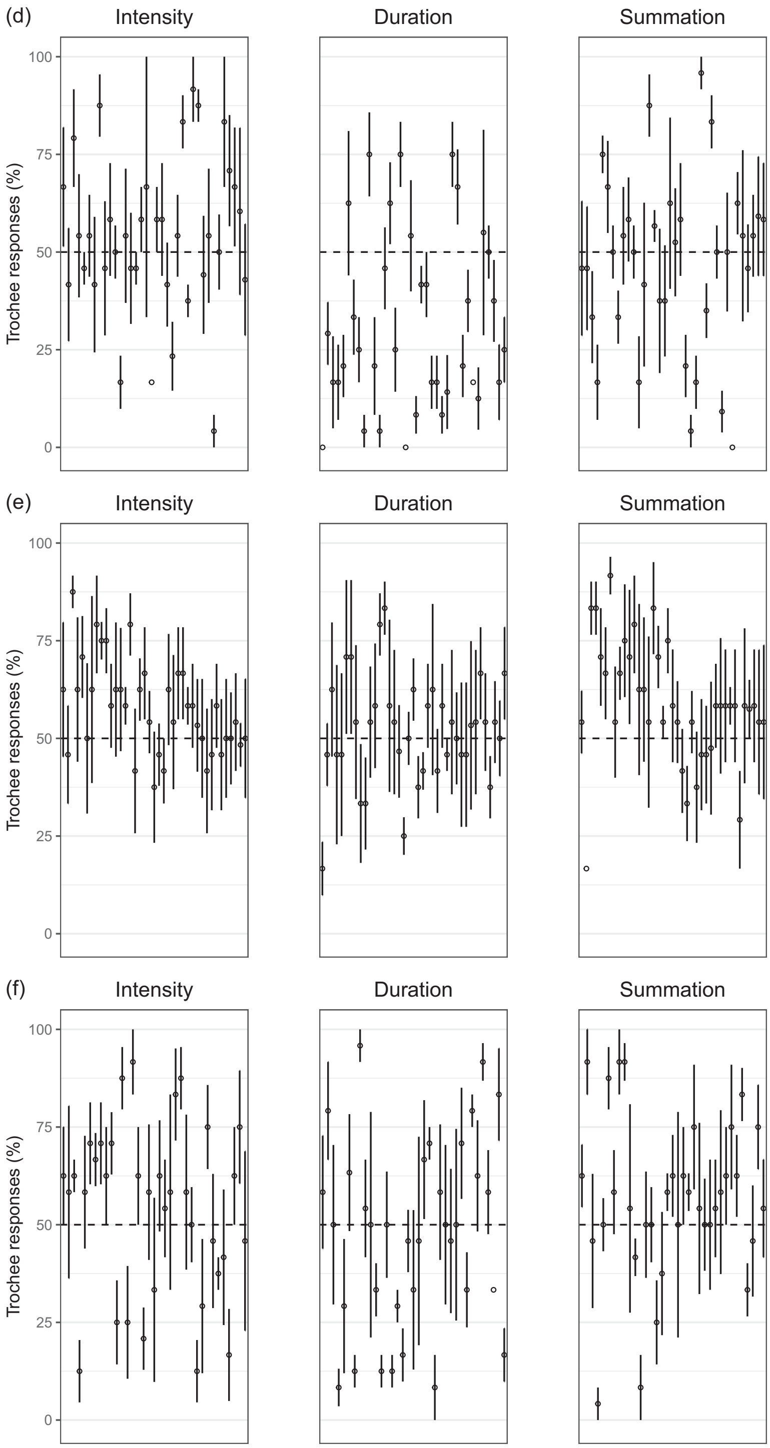

Figure 32.1 (*cont.*)

Table 32.4 Individual grouping preferences

Expressed as percentage of participants from each language showing a (significantly different from chance) preference for trochees, iambs, or no preference in each of the two experiments; data pooled over Step and ISI.

	Language	*Trochaic preference*		*Iambic preference*		*No preference*	
Sequence		Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
Intensity	English	14%	17%	4%	11%	75%	72%
	Greek	28%	13%	4%	0%	68%	87%
	Korean	20%	17%	7%	17%	73%	66%
Duration	English	4%	8%	14%	53%	82%	39%
	Greek	12%	5%	16%	5%	72%	90%
	Korean	23%	14%	0%	20%	77%	66%
Summation	English	7%	11%	7%	19%	86%	69%
	Greek	28%	18%	12%	3%	60%	80%
	Korean	27%	23%	7%	9%	67%	69%
Control	English	4%	17%	7%	11%	89%	72%
	Greek	24%	5%	0%	0%	76%	95%
	Korean	20%	20%	0%	11%	80%	69%

32.5 General Discussion

Our results supported the ITL, though not as strongly as those reported elsewhere (e.g., Iversen et al., 2008; Bhatara et al., 2013; Boll-Avetisyan et al., 2016, 2017). Some of the differences between our experiments and others may relate to specific experimental manipulations. For example, our results are comparable to those of Hay and Diehl (2007), a study using tones. Similarly, our prediction that larger steps would lead to stronger preferences was, to an extent, borne out similarly to Iversen et al. (2008). The weaker effect of Summation relative to Duration indicates that the preference for iambs with duration-varying sequences is reinforced by the perceptual integration of duration and intensity (Moore, 2012, inter alia). One way to interpret this finding is that in order to elicit a strong iambic preference, the combined effect of duration and loudness is needed; duration differences alone are not sufficient.[3] On the other hand, our expectation that short ISI would lead to stronger grouping preferences was not borne out, indicating that different experimental outcomes – such as Hay and Diehl (2007) versus Iversen et al. (2008) – cannot plausibly be attributed to this condition.

[3] Covarying manipulations of duration and intensity reported by Crowhurst and Teodocio (2014) and Crowhurst (2016) largely yielded results that differ from ours. However, it is not possible to directly compare those studies to ours, as the aim of their manipulations was to make duration and intensity compete with each other or work synergistically, whereas our aim was to eliminate a possible boost of intensity on to duration sequences.

Our results showed some cross-linguistic differences. English and Greek participants behaved mostly consistently with a weak version of the ITL. For English listeners, in particular, grouping and prominence relations went hand in hand (see London, 2012, Chapter 1), though their preferences were not strong, as in Iversen et al. (2008) or Crowhurst (2016). The Greek results were weaker compared to those of Crowhurst (2016) for Spanish, a language that is prosodically very similar to Greek. These discrepancies suggest that prosodic similarities between languages are not sufficient to predict responses in ITL experiments (see Moghiseh, et al., 2023, for a similar discussion, and also Chapter 4.3 on the role of prosody in processing).

The Korean groups, when they showed a preference, preferred trochees. This applied even with Duration sequences, contrary to our predictions. The Korean listeners' trochee bias for Duration sequences (although not systematic and strong) is reminiscent of Japanese participants' responses in Iversen et al. (2008). As Korean and Japanese do not have lexical stress, native speakers of these languages may exhibit weak association between acoustic prominence and grouping. Consequently, they may rely on general perceptual principles when engaging in prominence-related tasks (see de la Mora et al., 2013). This possibility is supported by the responses of the Korean participants to the control sequences: more participants showed trochee (20% for both Experiment 1 and Experiment 2) than iamb preferences (0% for Experiment 1; 11% for Experiment 2; Table 32.4). Trochee preferences are generally considered automatic, as they are easier to induce relative to iambic preferences (de la Mora et al., 2013).

Finally, both experiments show substantial individual variation, making replication hard to achieve. This is not the first time that ITL results have shown discrepancies. Woodrow (1909: 37) describes one participant (out of 13) who heard trochees even when increased duration was coupled with a long silent break between groups of stimuli. Crowhurst and Teodocio (2014) found different results with Zapotec speakers in two experiments. Such findings suggest that reported ITL results may be driven by some participants with strong preferences. This may partly explain the very strong results of Iversen et al. (2008), who analysed only the responses with the highest confidence rating. Alternatively, weak preferences may reflect the fact that experiments on the ITL, including ours, are based on the forced-choice paradigm. Seen in this light, the results may indicate that participants do not find either trochees or iambs an appropriate grouping for the stimuli, possibly reflecting differences in subjective rhythmisation (see Bolton, 1894; Woodrow, 1909; London, 2012, Chapter 1).

All in all, the present findings indicate that the ITL is at best a tendency, not a rule. Its effects may be strengthened or attenuated by language experience, while they are also susceptible to experimental manipulations (see Moghiseh et al., 2023). If so, then it is unlikely that the ITL directly governs language processing, and thus, its role in language acquisition may have been overestimated. Our brief descriptions on English, Greek, and Korean (Section 32.1.2) indicate that the relationship between prominence, grouping, and acoustic parameters differs across languages, and in Korean, which does not have lexical stress, the notion of binary strong–weak alternation is simply not applicable. The cross-linguistic differences question the validity of the ITL as a universal. Furthermore, the relationship between linguistic structure and acoustic parameters is ambiguous within a language; for instance, longer duration can signal both prominence and group finality. The individual differences within each language group in our results suggest that either the ITL experimental paradigm is not suitable for finding a grouping strategy shared by all speakers of a given language or that multiple strategies are employed by its speakers. Recent work has linked such differences to musicality (Boll-Avetisyan et al., 2017) and exposure to languages other than L1 (Boll-Avetisyan et al., 2016). Both hypotheses, with perhaps particular attention to participant rhythm-related abilities, are worth investigating further, as are other potential sources of individual variation (see Orrico et al., 2023). What is of importance, however, is that in all instances, the large individual differences indicate that the results are not easy to replicate. At the same time, individual differences are worth investigating further to fully understand the extent of variability within seemingly homogeneous populations.

32.6 Conclusions

We conducted two experiments with English, Korean, and Greek participants to examine whether the ITL is subject to cross-linguistic differences and susceptible to experimental manipulations, such as the duration of the silent intervals between alternating tones. Our results, though overall consistent with the predictions of the ITL, did not provide strong confirmation, in that the participants did not show pronounced grouping preferences. Similarly, the cross-linguistic differences were neither strong nor consistent across the two experiments. Close examination revealed substantial individual variation, which may be critical in explaining the gamut of results reported in the ITL literature, regarding the strength of the effect and cross-linguistic differences. All together, these results indicate that the ITL is susceptible to experimental manipulation and thus may not be readily replicable. In turn, this suggests that, while our findings for Korean highlight the need to test the ITL with more languages that do not have stress, the ITL may be a tendency exhibited by some individuals but is unlikely to hold the universal and central role in speech processing, acquisition, and the development of metrical systems often attributed to it.

32.7 Acknowledgements

We thank Taehong Cho for access to the Hanyang Phonetics and Psycholinguistics Laboratory, Natalie Fecher for assistance in experiment preparation, and Yuna Baek, Jiyoun Choi, Jiyoung Jang, Hyojin Kim, Miru Lee, Jinhee Park, Christina Kanouta, Deepthi Gopal, Heather Rolfe, and Louisa Salhi for assistance with data collection. A preliminary analysis of the first experiment was reported in Jeon and Arvaniti (2016). Both authors have contributed equally to this research, and their names are listed here in alphabetical order. The research was supported by an Academy of Korean Studies grant (#AKS-2014-R-13). This support is hereby gratefully acknowledged.

Box 32.1 Chapter Overview

Summary

We tested the ITL with English, Greek, and Korean speakers who responded to tone sequences varying in duration, intensity, or both. We found weak evidence for the ITL, with responses being influenced by the listeners' native language, as well as substantial inter-speaker variation that casts doubt on ITL replicability.

Implications

The ITL is said to reflect universal cognitive tendencies leading to subjective rhythmisation (moderated by language) and shaping the typology of linguistic (stress and) rhythm. Our findings indicate that not everyone is susceptible to the ITL; results can be driven by a few participants with strong preferences.

Gains

This chapter discusses discrepancies across studies related to the classic idea about binary grouping and how individual differences can influence experimental results and conclusions. Investigating interactions between stimulus properties and listeners' characteristics in processing complex sounds, rather than overemphasising the role of binary grouping, will lead to more fruitful outcomes.

REFERENCES

Abboub, N., Nazzi, T., and Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language*, 162, 46–59.

Allen, G. D. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3(2), 75–86.

Andrikopoulou, A., Protopapas, A., and Arvaniti, A. (2021). Lexical stress representation in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 47(6), 830–851.

Arvaniti, A. (2000). The acoustics of stress in Modern Greek. *Journal of Greek Linguistics*, 1, 9–39.

Arvaniti, A. (2007). Greek phonetics: The state of the art. *Journal of Greek Linguistics*, 8, 97–208.

Arvaniti, A., and Rathcke, T. (2015). The role of stress in syllable monitoring. The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK. www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html

Bates, D., Kliegl, R., Vasith, S., and Baayen, H. (2015a). Parsimonious mixed models. arXiv. http://arxiv.org/pdf/1506.04967.pdf

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Dordrecht: Foris.
Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., and Höhle, B. (2013). Native language affects rhythmic grouping of speech. *Journal of the Acoustical Society of America*, 134, 3828–3843.
Boersma, P., and Weenink, D. (2014). Praat: doing phonetics by computer (computer program), Version 5.4.04. www.praat.org
Boll-Avetisyan, N., Bhatara, A., and Höhle, B. (2017). Effects of musicality on the perception of rhythmic structure in speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 9. http://doi.org/10.5334/labphon.91
Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., and Höhle, B. (2016). Effects of experience with L2 and music on rhythm grouping by French listeners. *Bilingualism: Language and Cognition*, 19(5), 971–986.
Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, 6, 145–238.
Cho, T., and Keating. P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190.
Clopper, C. G. (2002). Frequency of stress patterns in English: A computational analysis. *IULC Working Papers*, 2(1). https://pdfs.semanticscholar.org/3404/d56321a167b484a3daa0654ece56aa1b8aff.pdf
Crowhurst, M. J. (2016). Iambic–trochaic law effects among native speakers of Spanish and English. *Laboratory Phonology*, 7(1), 12, 1–41. http://dx.doi.org/10.5334/labphon.42
Crowhurst, M. J. (2018). The influence of varying vowel phonation and duration on rhythmic grouping biases among Spanish and English speakers. *Journal of Phonetics*, 66, 82–99.
Crowhurst, M. (2020). The iambic/trochaic law: Nature or nurture? *Language and Linguistic Compass*, 14(1), 1–16.
Crowhurst, M. J., and Teodocio, A. O. (2014). Beyond the iambic–trochaic law: The joint influence of duration and intensity on the perception of rhythmic speech. *Phonology*, 31, 51–94.
Crowhurst, M. J., Kelly, N. E., and Teodocio, A. (2016). The influence of vowel glottalisation and duration on the rhythmic grouping preferences of Zapotec speakers. *Journal of Phonetics*, 58, 48–70.
Cutler, A., and Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142.
de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97(1), 491–504.
de la Mora, D., Nespor, M., and Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law? *Attention, Perception, & Psychophysics*, 75(1), 92–100.
Donselaar, W. V., Koster, M., and Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, 58, 251–274.
Drake, C., and Bertrand, D. (2001). The quest for universals in temporal processing in music. *Annals of the New York Academy of Sciences*, 930, 17–27.
Ernestus, M., and Neijt, A. (2008). Word length and the location of primary word stress in Dutch, German, and English. *Linguistics*, 43(3), 507–540.
Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, and F. E. Gibbon (Eds.), *Handbook of Phonetic Sciences* (second edition) (pp. 523–602). New York: Wiley-Blackwell.
Fourakis, M., Botinis, A., and Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica*, 56, 28–43.
Guion, S. G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition*, 27, 503–533.
Hannon, E. D., and Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *PNAS*, 102(35), 12639–12643.

Hay, J. F., and Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, 69, 113–122.

Hayes, B. (1985). *A Metrical Theory of Stress Rules*. New York: Garland.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago and London: University of Chicago Press.

Hyde, B. (2011). The iambic–trochaic law. In M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1052–1077). Hoboken, NJ: Wiley.

Iversen J. R., Patel, A. D., and Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, 124, 2263–2271.

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

Jeon, H.-S., and Arvaniti, A. (2016). Rhythmic grouping in English, Greek and Korean: Testing the iambic–trochaic law. Proceedings of Speech Prosody, 31 May–3 June, Boston, USA, pp. 1134–1138.

Jeon, H.-S., and Arvaniti, A. (2017). Effects of rhythm and phrase-final lengthening on word-spotting in Korean. *Journal of the Acoustical Society of America*, 141, 4251–4263.

Jun, S.-A. (2005). Korean intonational phonology and prosodic transcription. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 201–229). Oxford: Oxford University Press.

Jun, S.-A., and Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, Modeling and Technology* (pp. 209–242). Dordrecht: Springer.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, and R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 145–163). Cambridge: Cambridge University Press.

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

Lim, B.-j. (2001). The production and perception of word-level prosody in Korean. *Indiana University Working Papers in Linguistics*, 1, 3. www.indiana.edu/~iulcwp/wp/article/view/01-03

London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter* (second edition). New York: Oxford University Press.

Marian, V., Blumenfeld, H. K., and 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(4), 940–967.

Molnar, M., Carreiras, M., and Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition*, 152, 150–159.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing* (sixth edition). Bingley: Emerald Group Publishing Limited.

Moghiseh, E., Sonderegger, M., and Wagner, M. (2023). The iambic–trochaic law without iambs or trochees: Parsing speech for grouping and prominence. *Journal of the Acoustical Society of America*, 153(2), 1108–1129.

Nash, J. C., and Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software*, 43(9), 1–14.

Nespor, M., and Vogel, I. (1989). On clashes and lapses. *Phonology*, 6(1), 69–116.

Orrico, O., Gryllia, S., Kim S., and Arvaniti. A. (2023). The influence of emphathy and autistic-like traits in prominence perception. Proceedings of the 20th International Congress of Phonetic Sciences, International Phonetic Association, pp. 1280–1284.

Protopapas, A. (2006). On the use and usefulness of stress diacritics in reading Greek. *Reading and Writing: An Interdisciplinary Journal*, 19, 171–198.

Protopapas, A., Panagaki, E., Andrikopoulou, A., Gutiérrez Palma, N., and Arvaniti, A. 2016. Priming stress patterns in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 42(11), 1739–1760.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

Revithiadou, A., and Lengeris, A. (2016). One or many? In search of the default stress in Greek. In J. Heinz, R. Goedemans, and H. van der Hulst (Eds.), *Dimensions of Phonological Stress* (pp. 263–290). Cambridge: Cambridge University Press.

Trainor, L. J., and Adams, B. (2000). Infants and adults use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics*, 62, 333–340.

Turk, A., and Shattuck-Hufnagel, S. (2000). Word-boundary-related durational patterns in English. *Journal of Phonetics*, 28, 397–440.

Woodrow, H. (1909). A quantitative study of rhythm: The effect of variations in intensity, rate and duration. *Archives of Psychology*, 14, 1–66.