

JOURNAL ARTICLE

Effects of Musicality on the Perception of Rhythmic Structure in Speech

Natalie Boll-Avetisyan¹, Anjali Bhatara^{2,3} and Barbara Höhle¹

¹ Department of Linguistics, Cognitive Sciences Unit, Universität Potsdam, Potsdam, DE

² CNRS, Laboratoire Psychologie de la Perception, Paris, FR

³ Université Paris Descartes, Paris, FR

Corresponding author: Natalie Boll-Avetisyan (nboll@uni-potsdam.de)

Language and music share many rhythmic properties, such as variations in intensity and duration leading to repeating patterns. Perception of rhythmic properties may rely on cognitive networks that are shared between the two domains. If so, then variability in speech rhythm perception may relate to individual differences in musicality. To examine this possibility, the present study focuses on rhythmic grouping, which is assumed to be guided by a domain-general principle, the iambic/Trochaic law, stating that sounds alternating in intensity are grouped as strong-weak, and sounds alternating in duration are grouped as weak-strong. German listeners completed a grouping task: They heard streams of syllables alternating in intensity, duration, or neither, and had to indicate whether they perceived a strong-weak or weak-strong pattern. Moreover, their music perception abilities were measured, and they filled out a questionnaire reporting their productive musical experience. Results showed that better musical rhythm perception ability was associated with more consistent rhythmic grouping of speech, while melody perception ability and productive musical experience were not. This suggests shared cognitive procedures in the perception of rhythm in music and speech. Also, the results highlight the relevance of considering individual differences in musicality when aiming to explain variability in prosody perception.

Keywords: Musical ability; rhythm; grouping; iambic/Trochaic law; speech; speech perception; musicality; prosody; domain-general; German

1. Introduction

The rhythmic properties of music and language share some notable features: Both music and language are grouped into phrases that are marked by pauses as well as by differences in tone height and duration of beats and syllables (Patel, 2003). Given the many parallels, it has often been proposed that shared cognitive or perceptual mechanisms are active in the acquisition (e.g., McMullen & Saffran, 2004) and/or processing (Patel & Iversen, 2007) of music and language. If shared mechanisms are active, then training or aptitude in either music or language should facilitate the other (Patel, 2011).

The current study focuses on one rhythmic similarity between music and language: An asymmetry of cue distribution between the beginning and ends of larger units. Across musical cultures, initial beats are marked by higher intensity, and final notes are marked by longer duration in musical phrases (Lerdahl & Jackendoff, 1983; Narmour, 1990; Todd, 1985). Across languages, a similar distribution of rhythmic cues is found in metrical feet (i.e., the smaller rhythmic units consisting of one or more syllables that make up words): If metrical stress is trochaic, the prominent initial syllable of the (un-accented) foot is typically marked by increased intensity, whereas if metrical stress is iambic, the

prominent final syllable of the foot is typically marked by longer duration. Moreover, if a language has trochaic metrical stress and is weight-sensitive, then the default foot will be trochaic, but any long syllable will be a monosyllabic foot assigned finally (Hayes, 1995; see also Hyde, 2011, p. 1054).

It has been proposed that an innate domain-general auditory principle (Bolton, 1894; Hayes, 1985, 1995; Woodrow, 1909), referred to as the Iambic/Trochaic law (ITL, Hayes, 1985; Hayes, 1995), might account for this asymmetry in the distribution of rhythm cues in language and music. Furthermore, the ITL postulates preferences in auditory rhythmic grouping with variation in intensity leading to the perception of a strong-weak grouping and variation in duration leading to a weak-strong grouping. Hayes (1985, 1995) proposed the ITL to offer an account for the typological similarities regarding metrical stress. Many studies carried out over the past century have provided evidence for this law (Bhatara et al., 2016; Bhatara et al., 2013; Boll-Avetisyan et al., 2016; Bolton, 1894; Hay & Diehl, 2007; Rice, 1992; Woodrow, 1909, 1951). Nespors and colleagues (Langus et al., 2016a; Nespors et al., 2008) extended the ITL to account for typological similarities regarding phrasal stress, where trochaic phrasal stress (triggered by head-complement word order) is marked by pitch, while final (iambic) phrasal stress (triggered by complement-head word order) is marked by lengthening. Their proposal also receives support from acoustic analyses of speech productions (Nespors et al., 2008) and from certain rhythmic grouping studies (Abboub et al., 2016; Bion et al., 2011). There is, however, also evidence from phonetic studies and grouping studies that the asymmetry in the distribution of rhythmic cues in production (for an overview, see Fletcher, 2010) and perception (particularly for pitch: Bhatara et al., 2013; Kusumoto & Moreton, 1997; Rice, 1992; Woodrow, 1911) is not as consistent as postulated by Hayes (1995) and Nespors et al. (2008), suggesting that a more nuanced view of the ITL is needed.

Following up on the proposal of the ITL as a domain-general perceptual bias, the present study investigates whether musical aptitude influences rhythmic grouping of speech. If rhythm perception in general and rhythmic grouping in particular draw on shared cognitive resources between the domains of music and language, then musical ability may predict some of the variability between individuals' speech rhythm perception. We focused on native listeners of German—a group that has, to this point, been found to show strong grouping preferences consistent with the ITL (Bhatara et al., 2016; Bhatara et al., 2013). Recent research suggests that the perceptual effects of the ITL can be modulated by language experience as well as by musical experience. Studies that have investigated an effect of language experience on rhythmic grouping preferences presented cross-linguistic comparisons of two groups of speakers of different native languages or with different amounts of L2 experience. These studies have compared the rhythmic grouping preferences of speakers of languages that differ at the level of phrasal stress (Iversen et al., 2008; Langus et al., 2016b; Molnar et al., 2016; Molnar et al., 2014; Yoshida et al., 2010) or at the level of lexical stress (Bhatara et al., 2016; Bhatara et al., 2013; Boll-Avetisyan et al., 2016; Crowhurst, 2016; Crowhurst & Teodocio Olivares, 2014). Iversen et al. (2008) investigated potential effects of phrase-level prosodic knowledge on grouping by comparing native listeners of Japanese and English. Participants were presented with sequences of non-speech tones alternating in intensity or duration. As predicted by the ITL, both groups indicated a preference for trochaic groupings when listening to intensity-varied sequences. However, differences emerged for the grouping of duration-varied sequences: Only native listeners of English consistently showed a preference for iambic groupings as predicted by the ITL. Native listeners of Japanese, however, did not show a consistent pattern as a group: Almost half preferred trochaic groupings, about 26% preferred iambic groupings, and the rest had no consistent preference. Iversen and colleagues (2008)

interpret their data in line with Nespor et al.'s (2008) account, as English has iambic phrasal stress and Japanese has trochaic phrasal stress, suggesting that English listeners have more experience with duration as a grouping cue. An earlier study by Kusumoto and Moreton (1997), who studied native listeners of English and two Japanese dialects, found comparable results. Meanwhile, these findings have been extended to native listeners of Turkish and Persian (both languages with trochaic phrasal stress), who perform like Japanese listeners, and native listeners of Italian (iambic phrasal stress), who perform like English listeners (Langus et al., 2016b) as well as Basque-Spanish bilinguals (Molnar et al., 2014; Molnar et al., 2016), whose performance resembled that of Japanese listeners if their dominant language was Basque (trochaic phrasal stress) but that of English listeners, if their dominant language was Spanish (iambic phrasal stress). Note, however, that another study did not find any cross-linguistic differences: These experiments showed that both Japanese and English listeners were facilitated in their segmentation of rhythmically structured speech if duration was a cue to word endings but not if it was a cue to word beginnings (Frost et al., 2016).

Results from studies that tested an influence of experience with lexical stress on rhythmic grouping are less consistent. The first study addressing this issue (Hay & Diehl, 2007) compared grouping by listeners of French (no lexical stress) and English (contrastive lexical stress with metrical stress being trochaic plus weight-sensitive) of streams of tones or streams of repetitions of a single syllable that varied in either intensity, duration, or neither. Grouping preferences were consistent with the predictions of the ITL with no differences between the language groups. Later studies, however, found cross-linguistic differences in grouping tasks that used more complex material with either mixed syllables or mixed tones. For example, native listeners of Betaza Zapotec preferred trochaic groupings when hearing duration-varied sequences of syllables, which may relate to the fact that metrical stress in Betaza Zapotec is trochaic and weight-insensitive, and its prominence is acoustically cued by duration (Crowhurst & Teodocio Olivares, 2014). Moreover, native listeners of Spanish showed no preference for a specific grouping, which may be due to Spanish (its stress being trochaic with weight-sensitivity) having relatively fewer words with iambic patterns than English (Crowhurst, 2016), and indeed, English listeners displayed a preference for iambic groupings when tested on the same material (Crowhurst & Teodocio Olivares, 2014; Crowhurst, 2016), which replicates previous findings and is consistent with English stress.

Another group of researchers has focused on French and German listeners to explore the effects of experience with lexical stress in a native language (Bhatara et al., 2013; Bhatara et al., 2016) or in a second language (Boll-Avetisyan et al., 2016) as well as the effects of musical experience (Bhatara et al., 2016; Boll-Avetisyan et al., 2016) on rhythmic grouping preferences. As the current study immediately follows up on this work, these studies will be reviewed in somewhat more detail. Both French and German are similar at the level of phrasal stress, where both iambs and trochees can be formed (due to complement-head order in main clauses and head-complement order in subordinate clauses). The languages differ at the level of lexical stress: French has no lexical stress, while German has contrastive lexical stress with metrical stress being trochaic, but iambs can also be formed due to weight-sensitivity. Their first study (Bhatara et al., 2013) compared French and German monolinguals in two rhythmic grouping experiments using sequences of mixed syllables. They contrasted syllable sequences in which either intensity or duration was varied or there was no rhythmic variation. Both the French and German listeners' grouping preferences were as predicted by the ITL: They perceived iambs in the duration condition and trochees in the intensity condition. However, the French had significantly weaker grouping preferences than the German listeners. Moreover, the German

but not the French listeners perceived trochees in sequences without rhythmic variation. A trochaic perception of invariant structures is in line with the ITL (Hayes, 1995; Hyde, 2011; see also Bolton, 1894), which postulates that trochees should be the default grouping if no duration information is present.

The authors connected their results to earlier studies by Dupoux and colleagues (Dupoux et al., 1997; Dupoux et al., 2001; Peperkamp et al., 2010) that had shown that French listeners have weaker prosodic processing abilities than native listeners of languages with lexical stress. It seems that French listeners can perceive prosodic information in relatively simple tasks, but when task demands are high, their performance decreases. On these grounds, Dupoux and colleagues argue that French listeners lack the abstract symbolic representations of lexical stress, which facilitate prosodic perception in native listeners of lexical stress languages. Rather, they may process metrical information at an acoustic level. The same explanation can also account for the grouping results: When French listeners process simple rhythmic sequences consisting of one tone or just one syllable (hence, requiring little cognitive load), they do not differ from native listeners of a language with lexical stress such as English (Hay & Diehl, 2007). If, however, rhythmic sequences consist of multiple syllables (hence, introducing higher demands on processing), French listeners fall behind native listeners of a lexical stress language like German, as only the Germans receive facilitation from their abstract higher-level representations (Bhatara et al., 2013).

Following up on this work, Bhatara et al. (2016) predicted that German but not French listeners should transfer their ability to process rhythm at a higher level of processing to the perception of sequences of mixed complex tones. In their study, sequences of chimeric musical instrument sounds that varied in intensity, duration, or neither were used. They were either presented in a Low Variability condition, in which the sequences consisted of repeated exemplars of one sound (a chimera of two musical instruments) or in a High Variability condition, in which sequences consisted of multiple different instrument chimeras (mimicking the mix of syllables in Bhatara et al., 2013). As expected and in line with Hay and Diehl (2007), no differences were found between French and German listeners in the Low Variability condition. However, in the High Variability condition, only the German listeners showed the expected grouping preferences (iambics in the duration condition, trochees in the control and intensity condition), while the French listeners had no grouping preferences at all. The same study furthermore explored whether musical experience affects rhythmic grouping preferences. To this end, musical experience was measured using a composite score that combined information about the number of learned musical instruments, the age of acquiring a first instrument, and the total number of years of musical training. Results showed that musical experience predicted the rhythmical grouping by French but not by German listeners. More specifically, the more musical experience the French listeners had, the more they reported iambic groupings in the duration condition and trochaic groupings in the intensity condition. Grouping results in the control condition, however, were unaffected by the French listeners' musical experience.

These results suggest the following: German listeners rely on abstract prosodic knowledge to parse rhythmic sequences into metrical feet. French listeners, however, do not (because they lack the relevant abstract representations), and only rely on lower level acoustic information. Musical experience helps French listeners' to improve their acoustic processing skills (hence their enhanced grouping preferences when acoustic cues from intensity or duration are present), but it does not make them process the rhythmic sequences at a higher level of processing (hence, the lack of a default trochaic grouping preference in the control condition). That is, we would argue that musical experience has an effect on French listeners' acoustic acuity, but not on their abstract representations of prosody.

A third study by the same group (Boll-Avetisyan et al., 2016), however, suggests that French listeners may establish abstract prosodic representations when they have learned

a foreign language that has lexical stress, but—importantly—this is modulated by musical experience. Using Bhatara et al.'s (2013) syllable sequences as material, their study assessed the rhythmic grouping preferences of French listeners who had learned German as a second language (L2). It was found that the L2 learners' preferences for grouping rhythmic speech were modulated not only by qualitative and quantitative aspects of their received L2 input, but also by their musical experience. The more musical experience L2 learners had, the more their grouping preferences resembled that of the native German speakers than that of the French monolinguals in all three conditions (i.e., more iambic groupings in the duration condition, and more trochaic groupings in both the intensity and control condition). When these data were compared to the data of the monolingual French and German listeners tested in Bhatara et al. (2013), it became evident that musical experience affected the L2 learners but did not affect the monolingual groups. To account for these differences between the monolingual speakers and the L2 learners, Boll-Avetisyan et al. (2016) proposed that musical experience did not influence rhythmic grouping of speech directly (hence the lack of an influence in the two monolingual groups). Instead, the improved perception of rhythmic structure resulting from musical experience may have helped the L2 learners to acquire L2 lexical stress. The possibility that the L2 learners have actually acquired lexical stress as opposed to merely enhanced acoustic processing abilities is specifically supported by the fact that the musical L2 learners significantly increased their trochaic perception in the control condition, in which no acoustic prosodic cues were provided, and, hence, a default metrical parsing strategy must have been applied.

The following question arises: Why has musical experience not been found to influence German listeners' grouping preferences? It could be that German listeners' grouping preferences are at ceiling because they speak a lexical stress language, and the fact that they draw on linguistic representations of prosodic knowledge statistically overshadows any influence of their musical experience when processing rhythmic speech. However, it is also possible that their musical experience is not directly linked to their processing of rhythm in speech, whereas other aspects of musicality would be. It is possible that *musical ability*, in particular rhythm perception ability, is a more direct link between music and language in speech rhythm processing.

There are multiple reasons why musical ability would be linked more directly to speech rhythm processing than musical experience. They are summarized here, but see Levitin (2012) for further discussion: First, musicality has at least partially biological origins. Hence, people with equal musical experience can still differ with regards to their sensitivity to music. Second, Boll-Avetisyan et al.'s (2016) and Bhatara et al.'s (2016) Musical Experience factor (which combined years of practicing, age of acquisition, and number of instruments) focused on experience with *producing* music, which excludes a large group of people with extensive musical *perception* experience, including “disc jockeys, music critics, recording engineers, film music supervisors, and record company talent scouts” (Levitin, 2012, p. 634). Third, it ignores the fact that experience with different musical instruments and with different musical styles may lead to different skills. Wallentin and colleagues (Wallentin et al., 2010), for example, speculate that experience with string instruments may lead to enhanced melody skills, while experience with drum instruments may lead to enhanced rhythm skills (see also Rauscher & Hinton, 2003). In any case, there is evidence that the processing of music is a highly complex auditory skill that requires processing many different components, including pitch, duration, loudness, and timbre, all of which must then be integrated into higher-order representations of melody, rhythm, tempo, meter, and phrases.

Hence, we hypothesize that a measure of perceptive musical ability or experience would better capture an association between speech rhythm processing and musicality than a

measure of productive musical experience. In the current study, we chose to investigate German listeners. The reason for this choice was because of the lack of previous evidence of the influence of musical *experience* in this group. Hence, an effect of musical *ability* among German listeners would be more striking than the same effect among French listeners, where some effects of musical experience have already been shown. More specifically, we explored potential effects of both musical rhythm and melody perception abilities as well as productive musical experience. We hypothesized that musical rhythm perception ability would predict the perception of speech rhythm, and that it would be a better predictor than melody perception ability or productive musical experience. Furthermore, we raised the question of whether musical aptitude would affect lower-level acoustic or higher-order abstract processing, with no specific hypothesis about the direction of the effect.

To test these questions, we replicated Bhatara et al.'s (2013) grouping experiment with speech sequences varying in duration, intensity, or neither (control). In addition, we used the Musical Ear Test (MET, Wallentin et al., 2010), a standardized test for separately assessing melody and rhythm perception abilities with an equal emphasis on both. Moreover, we used a musical background questionnaire to obtain information related to (productive) musical experience.

We made the following predictions: We expected to replicate Bhatara et al.'s (2013) results in that the German listeners would show iambic grouping preferences when hearing duration-varied syllable sequences, and trochaic groupings when hearing intensity-varied or unvaried sequences, and we did not expect that productive musical experience would predict grouping preferences. Instead, we expected that perceptive musical aptitude—particularly for perceiving musical rhythm—as measured by the MET would predict grouping preferences.

Regarding the question of whether perceptive musical aptitude would affect listeners' lower- or higher-level processing of rhythm, we predicted the following: If musical aptitude is exclusively associated with listeners' sensitivity to acoustic prosodic information, we predicted that participants who are more musical (those with higher MET scores) would show more iambic groupings in the duration condition and more trochaic groupings in the intensity condition, while trochaic groupings in the acoustically invariant control condition should be uninfluenced by musicality. If, on the other hand, musical aptitude is associated with more abstract metrical grouping strategies as well as sensitivity to acoustic information, then, in addition to enhanced groupings in the intensity and duration condition, we also predicted that musical aptitude would alter grouping preferences in the acoustically invariant control condition.

2. Method

2.1. Participants

Twenty adult native speakers of German participated in this study (11 women, 9 men), who were raised monolingually by parents who were all native speakers of German. No participant reported any difficulties regarding their speech, language, or hearing. A summary of the participants' demographic information including their language and musical experience is given in **Table 1**. All procedures were performed in compliance with relevant laws and institutional guidelines and the appropriate institutional committee has approved them. Participants were compensated by a fee.

2.2. Material

2.2.1. Grouping experiment

In order to assess rhythmic speech grouping preferences, we used the stimuli from Bhatara et al. (2013), Experiment 1. There were 90 speech-like streams of different consonant-vowel syllables that were flat in F_0 . Each sequence consisted of 16 different

Participant background information		Mean	Range
	Participants' age	24	18–35
Language experience	Number of learned second languages	3	2–5
	Age of acquiring first second language	9	6–11
	Years spend learning second languages	12	8–22
Musical experience	Number of acquired musical activities (instruments, singing, and dancing)	2	0–6
	Age of acquiring first musical activity	9	4–20
	Years spent practicing a musical activity	10	0–30
Time spent in h/week	Singing	3	0–10
	Playing an instrument	3	0–20
	Dancing	2	0–7
	Listening to music	14	0–40
Self-estimated ability on scale from 0 = none to 10 = perfect	Playing an instrument	5	0–9
	Dancing	4	0–9
	Singing	5	0–9
MET scores	Melody perception ability	75%	52–92
	Rhythm perception ability	73%	52–83

Table 1: Summary of the participant background information as obtained by a questionnaire and their scores obtained in the Musical Ear Test (MET).

CV syllables, constructed by combining four long and tense vowels /e/, /i/, /o/, /u/ and four consonants of mixed manner and place of articulation /b/, /z/, /m/, /l/. Each of the syllables was presented twice, once in a strong and once in a weak position. This resulted in 32 syllables per sequence (e.g., /...zulebolilozimube.../), which were combined in a different order in each of the 90 stimuli. The streams were generated with a German voice (voice 'De5') using the text-to-speech software MBROLA (Dutoit et al., 1996). The onset of the sequences was masked over the first 3 seconds by a combination of white noise fading out and intensity of the stimulus fading in. Moreover, the sequences were counter-balanced for whether the initial syllable was strong or weak. There were three conditions: An intensity condition including 40 streams in which every second syllable had a higher intensity than the preceding one (a difference of 2, 4, 6, or 8 dB; 10 streams of each variation level), a duration condition including 40 streams in which every second syllable was longer than the preceding one (a difference of 50, 100, 150, or 200 ms; as with intensity, 10 streams of each variation level), and a control condition, in which all syllables were of equal intensity and duration. Further details on the acoustical properties of the material can be found in Bhatara et al. (2013).

2.2.2. Musical Ear Test

The Musical Ear Test (MET, Wallentin et al., 2010) was used to assess receptive musical abilities. This standardized test consisted of two parts, one for assessing rhythmic discrimination and one for assessing melodic discrimination. Each part consists of 52 trials. In each trial, participants hear two rhythmic or melodic phrases that are either the same or different. Rhythmic phrases consist of between 4 and 11 beats recorded with wood blocks. Melodic phrases consist of between 3 and 8 piano tones. All phrases have the same duration independent of the number of beats or tones.

2.3. Procedure

2.3.1. Grouping experiment

The procedure was analogous to that used by Bhatara et al. (2013). Stimuli were presented through headphones. Participants were instructed to listen to each of the nonsense speech streams carefully and to indicate by pressing one of two buttons as soon as they perceived a grouping pattern (even if it was before the end of the sequence) indicating whether this pattern consisted of strong-weak (trochaic) or weak-strong (iambic) disyllables.

2.3.2. Musical Ear Test

The procedure was analogous to that used by Wallentin et al. (2010). Participants heard prerecorded instructions (translated to German) and all trials through headphones. The task was to decide whether two phrases in a trial were the same or different. Participants indicated their response by checking a box on an answer sheet presented on a computer screen.

2.3.3. Questionnaire

After the experiment, participants completed the questionnaire (see **Table 1** for a summary of the results). Regarding the musical background, the first three questions asked 1) the number of acquired musical activities including playing an instrument, singing, and dancing, 2) their age of acquisition, and 3) the number of years practicing this musical activity (these were the questions on which Bhatara et al. (2016), Boll-Avetisyan (2016), and the current study based their Musical Experience variable). Furthermore, we added questions assessing their weekly time spent with singing, dancing, and practicing an instrument. We also added two questions assessing their perceptive musical behavior—one asked about the weekly time spent with listening to music, and the other one asked about their preferred musical styles. Lastly, they were required to estimate their abilities in singing, playing an instrument, and dancing on a scale from 0 (none) to 10 (perfect). Regarding the language background, the questions addressed the participants' foreign language knowledge, for which they had to list their foreign languages with the respective age of acquisition and the number of years they learned or were immersed in the language. Furthermore, they were asked to name their parents' native languages to verify that they had been raised monolingual.¹

2.4. Data processing and analysis

The dependent variable was response type (1 = trochaic versus 0 = iambic). Given the binomial distribution of the data, a logit generalized linear mixed-effects model with random factors for participants and items was applied.

To assess the influence of musical aptitude for rhythm and melody perception as well as of productive musical experience, we did the following: We generated two continuous fixed factors on the basis of the scores obtained in the MET. One of these factors represented the scores obtained in the rhythm part (METrhythm), and the other represented the scores obtained in the melody part (METmelody). To assess the influence of musical experience, we generated another continuous variable, which was a composite score based on the first three questions from the questionnaire (see above). This score was created the same way as in Boll-Avetisyan et al. (2016) and Bhatara et al. (2016), using a Principal Component Analysis. The first Principal Component, which represented all three variables to a comparable degree, captured 84% of their variance. These three continuous

¹ All participants were furthermore tested on a battery of tasks assessing general cognitive skills. This data is not considered here as it was part of a separate study.

variables (METrhythm, METmelody, and musical experience) were centered around their respective means to reduce collinearity.

For the condition factor, we used a successive difference contrast. This is an orthogonal contrast that puts the grand mean in the intercept and makes successive comparisons between conditions (that is, level 1 is compared to level 2, and level 2 to level 3 etc.; but no comparison can be made between level 1 and level 3, because this would give up the orthogonality). We coded the contrast so that duration was compared to intensity (Dur-Int), and control to duration (Cont-Dur), while control and intensity (Cont-Int) were not compared. This coding (which was also used by Boll-Avetisyan et al., 2016) allows us to see if an effect of the ITL is enhanced by musicality (that is, fewer trochaic responses to duration compared to more trochaic responses to intensity and control sequences respectively with increasing musicality). As we did not expect an effect of musicality on Cont-Int, and as such an effect would anyway be uninformative about the ITL, we disregarded this comparison.

In order to assess whether the three different musicality factors were valuable predictors of the participants' rhythmic grouping preferences, we ran three separate models, each one including the interaction between Condition and one of these three continuous variables. These three models were compared by means of the Akaike Information criterion (Akaike, 1998). The model including Condition*METrhythm yielded the lowest Akaike values (AIC = 2030, BIC = 2074), and, hence, accounted best for the variance in the data. We could not test any models including more than two fixed factors or random slopes, as these did not converge, so they are not interpretable.

For our analysis, we used the statistics program *R* (R Core Team, 2012). Models were fitted using the package *lme4* (Bates et al., 2015); graphs were generated using the package *ggplot2* (Wickham, 2009). The successive difference contrast was coded using the `contr.sdif()` function available from the *MASS* package (Venables & Ripley, 2002).

3. Results

3.1. Mixed-effects model results

The model coefficients are provided in **Table 2**. The estimates (β) indicate the logit-transformed proportion of trochaic responses. Results indicate a significant effect of Dur-Int, the negative β suggesting that participants gave fewer trochaic responses in the duration condition than in the intensity condition. Moreover, there was a significant effect of Cont-Dur, the positive β suggesting that participants gave more trochaic responses in the control condition than in the duration condition. There was no main effect of METrhythm. However, METrhythm interacted significantly with both Dur-Int and Cont-Dur. As illustrated by linear regression lines in **Figure 1**, the higher participants scored in the rhythm part of the MET, the more consistent they were in their rhythmic grouping performance.

3.2. Correlations

METmelody and METrhythm scores were marginally correlated ($r = .43, p = .06$). METmelody and musical experience were significantly correlated ($r = .44, p = .05$). However, METrhythm and musical experience were not correlated ($r = .12, p = .61$).

4. Discussion

The aim of the present study was to investigate the link between speech rhythm processing and musicality. We tested whether rhythmic grouping of speech by native listeners of German was affected by receptive musical abilities. We predicted that rhythm discrimination abilities would be a better predictor for speech rhythm grouping preferences than

Fixed effects	β	SE	z	P
Grand mean (Intercept)	0.65	0.13	5.01	< .001
Dur-Int	-2.12	0.15	-14.26	< .001
Cont-Dur	1.79	0.22	8.31	< .001
METrhythm	1.30	1.39	0.94	.35, n.s.
Dur-Int*METrhythm	-7.40	1.36	-5.45	< .001
Cont-Dur*METrhythm	13.30	2.16	6.14	< .001

Random effects	Variance	SD
item (Intercept)	0.29	0.54
id (Intercept)	0.21	0.46

Table 2: Parameters of the linear mixed-effects logit regression.

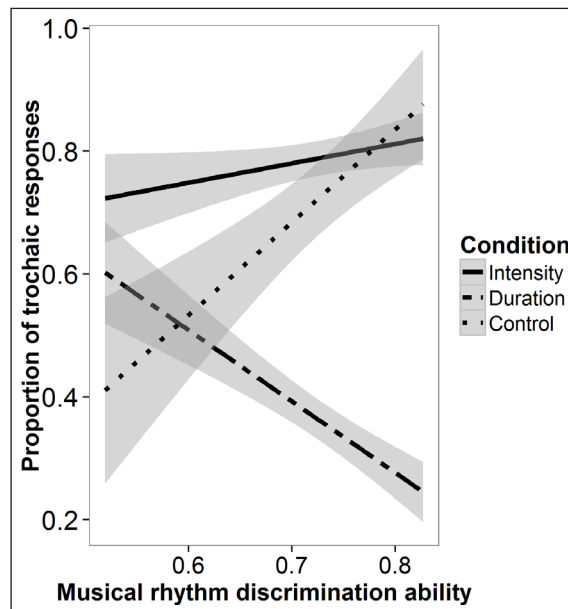


Figure 1: Linear regression lines reflecting the effect of musical rhythm discrimination ability (measured in % correct in Musical Ear Test) on the mean proportion of trochaic responses (0 = iambic, 1 = trochaic) broken down by condition. Shaded areas indicate the standard deviations.

melody discrimination abilities or musical (production) experience. The results confirmed our predictions. Musical rhythm receptivity was the best predictor of rhythmic grouping preferences. Higher scores in the rhythm test of the MET were related to more iambic responses in the duration condition and more trochaic responses in both the intensity and the control condition. Previous studies had found evidence for an effect of musical *experience* on rhythmic speech grouping preferences by native listeners of French but not German (Bhatara et al., 2016; Boll-Avetisyan et al., 2016). A potential explanation of this finding was that the effect of musical experience among the German listeners was not strong enough, perhaps because they already have a substantial amount of rhythmic experience in speech due to their linguistic experience with lexical stress. The French

listeners must use skills gained in other domains for rhythmic grouping, whereas the German speakers have ample knowledge gained from the linguistic domain upon which they can call. However, the current results rule out the hypothesis that German listeners' processing of rhythmic speech is only influenced by their linguistic knowledge. Rather, they suggest a role for cross-domain transfer between music and language in speech rhythm perception.

A first finding of the current study was that musical aptitude as measured by the MET compared to a measure of (productive) musical experience was a better predictor of rhythmic speech grouping. The distinction between *perception* and *production* of music and the abilities associated with each of these types of musical experience may explain the differences between prior and current findings. In the prior studies, Musical Experience measured only the time spent learning an instrument, age of first musical experience, and number of musical activities learned; that is, experience producing music with an instrument, the body, or the voice. Hence, the musical experience measure would correlate most strongly with musical production skills rather than perception skills. However, musical production and perception skills do not necessarily need to be correlated (Levitin, 2012), and, in fact, musical rhythm discrimination ability and general musical experience were not correlated in the presented sample. Hence, it is not surprising that a measure of musical perception abilities might be more informative for investigating the link between musicality and speech rhythm perception. Together, the results suggest a perceptual measure of musicality may be more sensitive than measures of musical experience when accounting for inter-listener variability in the processing of speech rhythm. Whereas the musical experience measure in the present and previous study included only production experience, the rhythm perception ability test may more directly measure the skill that is used when rhythmically grouping speech.

A second finding in this study was that a comparison between musical rhythm and melody perception abilities showed that rhythm abilities were a better predictor of speech rhythm grouping preferences. This provides support for the view that music is processed in different components (Levitin, 2012), and suggests that cross-domain transfer between rhythm related components is more likely than an involvement of other components of music (e.g., melody) in the processing of speech rhythm. In the present study, both the rhythmic grouping task and the rhythm perception test relied on processing of timing information. Because of this, it is likely that these two tasks relied on similar auditory mechanisms. In contrast, melodic perception would not have overlapped as strongly with the rhythmic grouping task, resulting in a weaker association between the two. This result is also in line with those of previous studies: There is evidence that speech rhythm experience and musical rhythm discrimination abilities are linked (Roncaglia-Denissen et al., 2016; Roncaglia-Denissen et al., 2013a), and experience with lexical tone has been linked with receptive skills for musical melody (Bidelman et al., 2013; Deutsch et al., 2006; Wong et al., 2012).

A third finding was that musical rhythm perception abilities predicted grouping preferences in all three conditions. In the introduction, we raised the issue of whether musicality would affect lower-level acoustic or higher-level abstract processing of rhythmic speech. We predicted that if musicality was linked to a lower-level acuity for acoustic cues in rhythmic speech, then participants with higher MET scores would show more iambic groupings in the duration condition and more trochaic groupings in the intensity condition, but there should have been no effect on the acoustically invariant control condition. If musicality was (furthermore) linked to higher-level abstract prosodic processing, then participants with higher MET scores should not only have enhanced grouping preferences

in the acoustically variant duration and intensity condition, but also in the invariant control condition.

The present results suggest effects of musical aptitude on higher-order processing (potentially in addition to effects on acoustic sensitivity): The most unmusical participants actually failed to perceive a trochaic structure in the control sequences. This result may suggest that in fact some level of receptive rhythm ability is needed in order to establish higher level abstract representations. However, as only two subjects showed this pattern and we do not know if they were affected by any perceptual disabilities (e.g., undiscovered amusia or language delay) we cannot draw any strong conclusions. The current findings have implications for studies on prosodic processing. It is generally assumed that rhythmic grouping preferences facilitate prosodically-cued speech segmentation in adulthood (e.g., Bion et al., 2011; Roncaglia-Denissen et al., 2013b; Tyler & Cutler, 2009) and infancy (Abboub et al., 2016; Bion et al., 2011; Hay & Saffran, 2012). Hence, it is possible that rhythmic perception ability is one of the factors that can account for the variability we see in adults' speech processing performance. Moreover, in light of the prosodic bootstrapping hypothesis (Gleitman & Wanner, 1982; Weissenborn & Höhle, 2001) by which infants use prosody to bootstrap into higher-order linguistic domains like lexical and syntax acquisition, musicality may even play a role in language acquisition. In support of the bootstrapping hypothesis, a longitudinal study with children has found that prosodic perception abilities at four months of age predict language abilities at five years (Höhle et al., 2014). If musicality affected early prosodic acquisition, then, potentially, it would also account for some of the variability seen in language development. It would be interesting to follow up on these issues in future studies.

In sum, the results of the presented study support the view of the ITL as a domain-general bias with links to music and language (Bolton, 1894; Hayes, 1985, 1995) while providing further evidence that some of the variability between listeners can be captured by factors relating to their individual musicality.

Acknowledgements

We would like to thank Sophie Gruhn, Franz Hildebrandt-Harangozó, Olivia Malotka, and Isabelle Mackuth for help with testing participants. This work was supported by two Agence Nationale de la Recherche – Deutsche Forschungsgemeinschaft grants (# 09-FASHS-018 and HO 1960/14-1 to Barbara Höhle and Thierry Nazzi, and HO 1960/15-1 and ANR-13-FRAL-0010 to Ranka Bijeljic-Babic and Barbara Höhle).

Competing Interests

The authors have no competing interests to declare.

References

- Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B. and Nazzi, T. 2016. An exploration of rhythmic grouping of speech sequences by French- and German-learning infants. *Frontiers in Neurosciences*, 10, 292. DOI: <https://doi.org/10.3389/fnhum.2016.00292>
- Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. *Selected Papers of Hirotugu Akaike* (pp. 199–213). New York: Springer. DOI: https://doi.org/10.1007/978-1-4612-1694-0_15
- Bates, D., Maechler, M., Bolker, B. and Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>

- Bhatara, A., Boll-Avetisyan, N., Agus, T., Höhle, B. and Nazzi, T. 2016. Language experience affects grouping of musical instrument sounds. *Cognitive Science*, 40(7), 1816–1830. DOI: <https://doi.org/10.1111/cogs.12300>
- 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(5), 3828–3843. DOI: <https://doi.org/10.1121/1.4823848>
- Bidelman, G. M., Hutka, S. and Moreno, S. 2013. Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music. *PLoS ONE*, 8(4), 1–11. DOI: <https://doi.org/10.1371/journal.pone.0060676>
- Bion, R. A. H., Benavides-Varela, S. and Nespore, M. 2011. Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, 54(1), 123–140. DOI: <https://doi.org/10.1177/0023830910388018>
- Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T. and Höhle, B. 2016. Effects of experience with L2 and music on rhythmic grouping by French listeners. *Bilingualism: Language and Cognition*, 19(5), 971–986. DOI: <https://doi.org/10.1017/S1366728915000425>
- Bolton, T. L. 1894. Rhythm. *American Journal of Psychology*, 6(2), 145–238. DOI: <https://doi.org/10.1177/0023830910388018>
- Crowhurst, M. J. 2016. Iambic-Trochaic Law Effects among Native Speakers of Spanish and English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 12. DOI: <https://doi.org/10.5334/labphon.42>
- Crowhurst, M. J. and Teodocio Olivares, A. 2014. Beyond the Iambic-Trochaic Law: The joint influence of duration and intensity on the perception of rhythmic speech. *Phonology*, 31(1), 51–94. DOI: <https://doi.org/10.1017/S0952675714000037>
- Deutsch, D., Henthorn, T., Marvin, E. and Xu, H. 2006. Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, 119(2), 719–722. DOI: <https://doi.org/10.1121/1.2151799>
- Dupoux, E., Pallier, C., Sebastián-Gallés, N. and Mehler, J. 1997. A distressing 'deafness' in French? *Journal of Memory and Language*, 36(3), 406–421. DOI: <https://doi.org/10.1006/jmla.1996.2500>
- Dupoux, E., Peperkamp, S. and Sebastián-Gallés, N. 2001. A robust method to study stress "deafness". *Journal of the Acoustical Society of America*, 110(3), 1606–1618. DOI: <https://doi.org/10.1121/1.1380437>
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F. and van der Vreken, O. 1996. *The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes*. Proceedings of The Fourth International Conference on Spoken Language Processing (ICSLP'96), Philadelphia, PA, pp. 1393–1396. DOI: <https://doi.org/10.1109/ICSLP.1996.607874>
- Fletcher, J. 2010. The prosody of speech: Timing and rhythm. In: Hardcastle, W. J., Laver, J. and Gibbon, F. E. (eds.), *The Handbook of Phonetic Sciences, Second Edition* (pp. 523–602). Oxford: Wiley-Blackwell & Sons. DOI: <https://doi.org/10.1002/9781444317251.ch15>
- Frost, R. L. A., Monaghan, P. and Tatsumi, T. 2016. Domain-general mechanisms for speech segmentation: The role of duration information in language learning. *Journal of Experimental Psychology: Human Perception and Performance*, In press. Online First View: November 28, 2016. DOI: <https://doi.org/10.1037/xhp0000325>
- Gleitman, L. R. and Wanner, E. 1982. Language acquisition: The state of the state of the art. In: Wanner, E. and Gleitman, L. R. (eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, MA: Cambridge University Press.

- Hay, J. F. and Diehl, R. L. 2007. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, 69, 113–122. DOI: <https://doi.org/10.3758/BF03194458>
- Hay, J. F. and Saffran, J. R. 2012. Rhythmical grouping biases constrain infant statistical learning. *Infancy*, 17(6), 610–641. DOI: <https://doi.org/10.1111/j.1532-7078.2011.00110.x>
- Hayes, B. 1985. *Iambic and trochaic rhythm in stress rules*. Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society, pp. 429–446.
- Hayes, B. 1995. *Metrical stress theory: principles and case studies*. Chicago & London: The University of Chicago Press.
- Höhle, B., Pauen, S., Hesse, V. and Weissenborn, J. 2014. Discrimination of rhythmic pattern at 4 months and language performance at 5 years: A longitudinal analysis of data from German-learning children. *Language Learning*, 64, 141–164. DOI: <https://doi.org/10.1111/lang.12075>
- Hyde, B. 2011. The Iambic-Trochaic Law. In: van Oostendorp, M., Ewen, C. J., Hume, E. and Rice, K. (eds.), *The Blackwell Companion to Phonology: Suprasegmental and Prosodic Phonology* (Vol. II, pp. 1052–1077). Oxford: Blackwell-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(4), 2263–2271. DOI: <https://doi.org/10.1121/1.2973189>
- Kusumoto, K. and Moreton, E. 1997. Native language determines parsing of nonlinguistic rhythmic stimuli. *Journal of the Acoustical Society of America*, 102(5), 3204. DOI: <https://doi.org/10.1121/1.420936>
- Langus, A., Mehler, J. and Nespors, M. 2016a. Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews*, In press. DOI: <https://doi.org/10.1016/j.neubiorev.2016.12.012>
- Langus, A., Seyed-Allaei, S., Uysal, E., Pirmoradian, S., Marino, C., Asaadi, S. and Nespors, M. 2016b. Listening natively across perceptual domains? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(7), 1127–1139. DOI: <https://doi.org/10.1037/xlm0000226>
- Lerdahl, F. and Jackendoff, R. 1983. *Generative Theory of Tonal Music*. Cambridge: MIT Press.
- Levitin, D. J. 2012. What Does It Mean to Be Musical? *Neuron*, 73(4), 633–637. DOI: <https://doi.org/10.1016/j.neuron.2012.01.017>
- McMullen, E. and Saffran, J. R. 2004. Music and language: A developmental comparison. *Music Perception*, 21(3), 289–311. DOI: <https://doi.org/10.1525/mp.2004.21.3.289>
- Molnar, M., Carreiras, M. and Gervain, J. 2016. Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition*, 152, 150–159. DOI: <https://doi.org/10.1016/j.cognition.2016.03.023>
- Molnar, M., Lallier, M. and Carreiras, M. 2014. The Amount of Language Exposure Determines Nonlinguistic Tone Grouping Biases in Infants From a Bilingual Environment. *Language Learning*, 64(2), 45–64. DOI: <https://doi.org/10.1111/lang.12069>
- Narmour, E. 1990. *The analysis and cognition of basic melodic structures*. Chicago, IL: University of Chicago Press.
- Nespors, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H. and Donati, C. 2008. Different phrasal prominence realizations in VO and OV languages. *Lingue e linguaggio*, 7(2), 1–28. DOI: <https://doi.org/10.1418/28093>
- Patel, A. D. 2003. Rhythm in language and music. *Annals of the New York Academy of Sciences*, 999(1), 140–143. DOI: <https://doi.org/10.1196/annals.1284.015>

- Patel, A. D. 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2, 1–14. DOI: <https://doi.org/10.3389/fpsyg.2011.00142>
- Patel, A. D. and Iversen, J. R. 2007. The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11(9), 369–372. DOI: <https://doi.org/10.1016/j.tics.2007.08.003>
- Peperkamp, S., Vendelin, I. and Dupoux, E. 2010. Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, 38(3), 422–430. DOI: <https://doi.org/10.1016/j.wocn.2010.04.001>
- Rauscher, F. H. and Hinton, S. C. 2003. Type of music training selectively influences perceptual processing. In: Kopiez, R., Lehmann, A., Wolther, I. and Wolf, C. (eds.), *Proceedings of the 5th Triennial Conference of the European Society for the Cognitive sciences of Music*. Hannover: University of Music and Drama.
- R Core Team 2012. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rice, C. C. 1992. *Binarity and ternarity in metrical theory: parametric extensions*: Doctoral dissertation, University of Texas at Austin.
- Roncaglia-Denissen, M. P., Roor, D. A., Chen, A. and Sadakata, M. 2016. The enhanced musical rhythmic perception in Second language learners. *Frontiers in Neurosciences*, 10, 288. DOI: <https://doi.org/10.3389/fnhum.2016.00288>
- Roncaglia-Denissen, M. P., Schmidt-Kassow, M., Heine, A., Vuust, P. and Kotz, S. A. 2013a. Enhanced musical rhythm perception in Turkish early and late learners of German. *Frontiers in Psychology*, 4(645). DOI: <https://doi.org/10.3389/fpsyg.2013.00645>
- Roncaglia-Denissen, M. P., Schmidt-Kassow, M. and Kotz, S. A. 2013b. Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS ONE*, 8(2), e56000. DOI: <https://doi.org/10.1371/journal.pone.0056000>
- Todd, N. 1985. A model of expressive timing in tonal music. *Music Perception*, 3(1), 33–58. DOI: <https://doi.org/10.2307/40285321>
- Tyler, M. D. and Cutler, A. 2009. Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*, 126(1), 367–376. DOI: <https://doi.org/10.1121/1.3129127>
- Venables, W. N. and Ripley, B. D. 2002. *Modern Applied Statistics with S. Fourth Edition*. New York: Springer. DOI: <https://doi.org/10.1007/978-0-387-21706-2>
- Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C. and Vuust, P. 2010. The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20(3), 188–196. DOI: <https://doi.org/10.1016/j.lindif.2010.02.004>
- Weissenborn, J. and Höhle, B. eds. 2001. *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. Philadelphia, PA: John Benjamins.
- Wickham, H. 2009. *ggplot2. Elegant Graphics for Data Analysis*. New York: Springer. DOI: <https://doi.org/10.1007/978-0-387-98141-3>
- Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L.-H. and Peretz, I. 2012. Effects of culture on musical pitch perception. *PLoS ONE*, 7(4). DOI: <https://doi.org/10.1371/journal.pone.0033424>
- Woodrow, H. 1909. A quantitative study of rhythm: The effect of variations in intensity, rate, and duration. *Archives of Psychology*, 14, 1–66.
- Woodrow, H. 1911. The role of pitch in rhythm. *Psychological Review*, 18(1), 54–77. DOI: <https://doi.org/10.1037/h0075201>
- Woodrow, H. 1951. Time perception. In: Stevens, S. (ed.), *Handbook of experimental psychology* (pp. 1224–1236). Oxford: Wiley.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J. and Werker, J. F. 2010. The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*, 115(2), 356–361. DOI: <https://doi.org/10.1016/j.cognition.2010.01.005>

How to cite this article: 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, pp. 1–16, DOI: <https://doi.org/10.5334/labphon.91>

Published: 27 April 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[

Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 