



Humanwissenschaftliche Fakultät

Natalie Boll-Avetisyan | Anjali Bhatara | Annika Unger | Thierry Nazzi
| Barbara Höhle

Effects of Experience with L2 and Music on Rhythmic Grouping by French Listeners

Suggested citation referring to the original publication:
Bilingualism: Language and Cognition Vol. 19, No. 5 (2016)
DOI <http://dx.doi.org/10.1017/S1366728915000425>
ISSN (print) 1366-7289
ISSN (online) 1469-1841

Postprint archived at the Institutional Repository of the Potsdam University in:
Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 450
ISSN 1866-8364
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-413786>

Effects of experience with L2 and music on rhythmic grouping by French listeners*

NATALIE BOLL-AVETISYAN

Universität Potsdam

ANJALI BHATARA

Université Paris Descartes

CNRS

ANNIKA UNGER

Universität Potsdam

THIERRY NAZZI

Université Paris Descartes

CNRS

BARBARA HÖHLE

Universität Potsdam

(Received: February 28, 2014; final revision received: June 23, 2015; accepted: June 24, 2015; first published online 20 July 2015)

Rhythm perception is assumed to be guided by a domain-general auditory principle, the Iambic/Trochaic Law, stating that sounds varying in intensity are grouped as strong-weak, and sounds varying in duration are grouped as weak-strong. Recently, Bhatara et al. (2013) showed that rhythmic grouping is influenced by native language experience, French listeners having weaker grouping preferences than German listeners. This study explores whether L2 knowledge and musical experience also affect rhythmic grouping. In a grouping task, French late learners of German listened to sequences of coarticulated syllables varying in either intensity or duration. Data on their language and musical experience were obtained by a questionnaire. Mixed-effect model comparisons showed influences of musical experience as well as L2 input quality and quantity on grouping preferences. These results imply that adult French listeners' sensitivity to rhythm can be enhanced through L2 and musical experience.

Keywords: rhythmic grouping, second language acquisition, prosody, musicality, Iambic/Trochaic Law

Introduction

The perception of non-native or second language (L2) speech contrasts is initially restricted by the phonological (Best, McRoberts & Sithole, 1988) and the phonetic (Flege, 1995; Werker & Tees, 1984) properties of the native language. Because of this, L2 learners have difficulties perceiving segmental contrasts that do not occur in their L1, particularly if these contrasts assimilate to a single native category (Best, 1995) or are not acoustically salient (Burnham, 1986; Polka, 1991). Improvement in perceiving L2 contrasts can occur, but it depends on the relation between the phonological and

phonetic systems of the L2 and the L1 (Best, 1995; Flege, 1995) as well as the amount and quality of L2 experience and the learning conditions of the L2 (Flege & Liu, 2001).

Most L2 perception studies have investigated the perception of non-native phoneme contrasts. Much less is known on whether L2 acquisition can affect prosodic perception. The current study investigated an area in which effects of linguistic experience on perception have recently been demonstrated: the perception of rhythm and rhythmic grouping. Rhythm perception may be guided by universal biases (Hayes, 1995), which have been found to be affected by native language experience (Bhatara, Boll-Avetisyan, Agus, Höhle & Nazzi, *in press*; Bhatara, Boll-Avetisyan, Unger, Nazzi & Höhle, 2013; Crowhurst & Teodocio Olivares, 2014; Iversen, Patel & Ohgushi, 2008). The current study investigated whether knowledge of an L2 with lexical stress alters rhythmic grouping in French listeners. French is an interesting case because it does not have lexical stress: there are no minimal pairs that are distinguished on the basis of stress placement. French just has fixed phrasal stress, which is assigned to the phrase-final syllable (Delattre, 1938; Di Cristo, 1998), and is acoustically marked by pitch (rising

* Thanks to Wiebke Bruchmüller, Carina Hoppe, Silke Schunack, Alexandra Schmitterer for help with recruiting and testing participants, to Tanner Sorensen for proof-reading the manuscript, and to Tom Fritzsche, Reinhold Kliegl, Daniel Schad and Sarah Risse for help with “R” and useful suggestions for the statistical analysis, and to Adamantios Gafos and Elliott Moreton for comments on prior versions of the manuscript. This work was supported by the Agence Nationale de la Recherche - Deutsche Forschungsgemeinschaft grant # 09-FASHS-018 and HO 1960/14-1 to Thierry Nazzi and Barbara Höhle.

Supplementary material can be found online at <http://dx.doi.org/10.1017/S1366728915000425>

Address for correspondence:

Natalie Boll-Avetisyan, Universität Potsdam, Center of Cognitive Sciences, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany
nboll@uni-potsdam.de

sentence-internally, and falling sentence-finally) and lengthening (Féry, Hörnig & Pahaut, 2011; Jun & Fougeron, 2000; Jun & Fougeron, 2002; Welby, 2006). The L2 used in the present study is German, which has variable and contrastive lexical stress (as illustrated by the presence of minimal pairs such as *Party* ‘party’ / ‘pa:ti:/ vs. *Partie* ‘match’ / ,pa:’ti:/).

As yet, most evidence that prosodic properties are subject to perceptual reorganization comes from research on lexical stress, most of which, like the present study, focused on French and suggested that the lack of contrastive lexical stress has consequences for stress discrimination abilities: French monolinguals show lower accuracy than Spanish monolinguals (whose language uses contrastive lexical stress) in discriminating nonwords with different stress patterns (Dupoux, Pallier, Sebastián-Gallés & Mehler, 1997; Dupoux, Peperkamp & Sebastián-Gallés, 2001; see also Peperkamp, Vendelin & Dupoux, 2010 for similar results on other languages also lacking contrastive lexical stress). This is particularly the case in tasks that pose higher demands on processing capacities and go beyond low-level acoustic processing by presenting stimuli with phonetic variability (Dupoux et al., 1997; Dupoux et al., 2001). Recent studies with French-learning infants suggest that this relative stress ‘deafness’ emerges early in life, as their ability to discriminate word stress patterns decreases between 6 and 10 months (Abboub, Bijeljac-Babic, Serres & Nazzi, 2015; Bijeljac-Babic, Serres, Höhle & Nazzi, 2012; Höhle, Bijeljac-Babic, Herold, Weissenborn & Nazzi, 2009; Skoruppa, Pons, Bosch, Christophe, Cabrol & Peperkamp, 2013; Skoruppa, Pons, Christophe, Bosch, Dupoux, Sebastián-Gallés, Alves Limissuri & Peperkamp, 2009), and even simultaneous bilingual infants and adults are relatively stress ‘deaf’ if their dominant language was French during infancy (Bijeljac-Babic et al., 2012; Dupoux, Peperkamp & Sebastián-Gallés, 2010, but c.f. Abboub et al., 2015; Bijeljac-Babic, Serres, Höhle & Nazzi, 2013). This suggests that input factors during infancy are important in developing optimal prosodic processing abilities.

Rhythmic perception and the Iambic/Trochaic Law

The current study investigated how experience with L2 contrastive lexical stress and music affects rhythmic grouping. Rhythmic grouping studies have a long tradition in perception research. In these studies (e.g., Hay & Diehl, 2007; Woodrow, 1909; 1911), adults are required to listen to streams of sounds that vary in intensity, pitch, or duration and indicate whether they hear pairs of sounds that are strong-weak (i.e., trochees) or weak-strong (i.e., iambs). The general finding is that adults perceive a rhythmic structure when listening to such streams: sound stretches that alternate in intensity or pitch are grouped

into trochees, whereas sound stretches that alternate in duration are grouped into iambs.

Hayes (1995) proposed the Iambic/Trochaic Law (ITL), an innate domain-general perceptual primitive, to explain these biases. According to Hayes, effects of the ITL are also reflected in language and music typology: Cross-linguistically, initial word or phrase stress is usually marked by rising intensity and rising pitch, whereas final stress is usually marked by lengthening (e.g., Beckman & Pierrehumbert, 1986; Klatt, 1976; Nespor, Shukla, van de Vijver, Avesani, Schraudolf & Donati, 2008). The same distribution of rhythmic cues is found in music: initial beats are marked by higher intensity, and final notes are marked by longer duration (Lerdahl & Jackendoff, 1983; Narmour, 1990; Todd, 1985). Listeners perceive rhythmic structure even in invariant sequences of sounds such as the tick-tock of a clock, which are perceived as trochees (e.g., Bolton, 1894). This perceptual illusion is also captured by the ITL, as Hayes suggests that humans should perceive trochees unless durational cues (i.e., weight sensitivity, Hayes, 1985) trigger an iambic perception.

Previous work has examined the influence of linguistic experience on rhythmic grouping by comparing monolingual speakers of different languages. Japanese and English listeners were presented with sequences of non-speech tones alternating in intensity or duration. Adult native listeners of Japanese and Japanese-learning 7–8-month-olds did not show a consistent pattern of grouping duration-varied sequences as iambs, whereas 7–8-month-old and adult listeners of English did (Iversen et al., 2008; Kusumoto & Moreton, 1997; Yoshida, Iversen, Patel, Mazuka, Nito, Gervain & Werker, 2010). One explanation for this difference provided by Iversen and colleagues is related to differences in word order across the two languages: whereas English function words (typically shorter and acoustically less salient than content words) mostly precede content words within a phrase, the reverse order holds for Japanese. Hence, Japanese listeners are more exposed to long-short patterns than English listeners, and this may interfere with the effects of the ITL. Compatible with this account, Spanish–Basque bilingual 10-month-olds behave like Japanese-learning infants if they are dominant in Basque, which has a similar word order to Japanese, whereas they behave like English-learning infants if they are dominant in Spanish, which has a similar word order to English (Molnar, Lallier & Carreiras, 2014).

Another recent line of research indicates that differences between languages in word-level stress also lead to differences in rhythmic grouping preferences. For example, Crowhurst and Teodocio Olivares (2014) found that native speakers of Zapotec group repetitions of the syllable pair /dege/ as long-short when syllables vary in duration, while native listeners of English group them as short-long. However, when syllables vary in intensity, both

Zapotec and English listeners group them as loud-soft. The authors explain this result by the fact that while stress is generally trochaic in both languages, the two languages differ with respect to weight-sensitivity. In English, heavy syllables usually attract stress even if they occur word-finally, and duration is a marker of prominence in such cases, which explains English listeners' preference for short-long groupings. In contrast, Zapotec is not weight-sensitive, and duration is an additional marker for trochaic stress, which explains Zapotec listeners' preference for long-short groupings (Crowhurst & Teodocio Olivares, 2014).

Other studies have examined grouping preferences by French listeners and compared them to native listeners of languages with contrastive lexical stress (Bhatara et al., *in press*; Bhatara et al., 2013; Hay & Diehl, 2007). Hay and Diehl (2007) tested rhythmic grouping in native listeners of French and English using both a speech and a non-speech condition, which consisted of sequences of repetitions of a tone or the syllable /ga/ separated by intervals of silence. In both conditions, the French and English listeners showed the same performance, grouping intensity-varied sequences as trochees and duration-varied sequences as iambs. However, differences in rhythmic perception were found in a study comparing native listeners of French versus German (Bhatara et al., 2013). In that study, the sequences were continuous streams of 16 segmentally varying coarticulated syllables (e.g., /...zu:le:bo:li:lo:zi:mu:be:.../) alternating in either intensity, duration or neither. Native listeners of both German and French showed grouping preferences as predicted by the ITL. However, the French were showing this effect less consistently and needed stronger acoustic contrasts (i.e., larger differences between syllables in intensity and duration) than the Germans. Moreover, only the Germans displayed a trochaic bias when listening to rhythmically invariant control sequences. These cross-linguistic differences were extended to complex non-speech sequences, made up of 16 different musical instrument sounds, while no differences between language groups were found when a single musical instrument sound was repeated (Bhatara et al., *in press*). Taken together, cross-linguistic differences between native listeners of French and native listeners of German or English were obtained when using streams of varying syllables or nonlinguistic sounds, but not when using a single repeated syllable or nonlinguistic sound.

To account for the above findings, Bhatara et al. (2013; *in press*) pointed out that they parallel the findings of studies on lexical stress 'deafness' (Dupoux et al., 1997; 2001). Hence, they proposed that these results indicate that whether or not languages have lexical stress affects not only the perception of word stress but also rhythmic grouping. Following up on this proposal, they argued that German listeners might try to build up a

foot structure when listening to rhythmic speech. While doing so, they would draw on abstract representations of lexical stress; especially when processing complex streams, which add a higher cognitive load on listeners compared to simple streams. However, French listeners do not establish these abstract representations of stress because they are not required for processing words in French. This could explain the presence or absence of cross-linguistic differences in ITL studies depending on the materials used.

Effects of L2 acquisition

The question explored in the present study is whether experience with an L2 with contrastive lexical stress in adulthood affects rhythmic grouping, i.e., the manifestation of the ITL, in native listeners of French. Previous studies have found no effect of L2 knowledge on French listeners' lexical stress perception: European French L2 learners of Spanish were as stress 'deaf' as monolingual French listeners with no effect due to their L2 proficiency level (Dupoux, Sebastian-Gallés, Navarrete & Peperkamp, 2008). Similarly, in cross-linguistic studies with L2 learners of English (Altmann, 2006) and Polish (Kijak, 2009) with various language backgrounds, French participants were among those with the poorest stress perception abilities. Furthermore, French L2 learners of German had greater difficulties in detecting rhythmic violations than syntactic violations in German sentences (Schmidt-Kassow, Rothermich, Schwartze & Kotz, 2011). Hence, these studies suggest that stress 'deafness' persists with knowledge of an L2 with lexical stress and irrespective of L2 proficiency.

Three recent studies, however, indicate that L2 experience may have an effect on lexical stress perception. Lin, Wang, Idsardi and Xu (2014) found that native listeners of Korean, which also has no lexical stress, showed increased stress perception ability with increasing L2 proficiency in English. Furthermore, Canadian French listeners' self-reported amount of daily use of L2 English (Tremblay, 2009) and their time immersed in an English-speaking environment (Tremblay, 2008) were significant predictors of their ability to perceive stress, while their general level of L2 proficiency was not.

The differences in outcome in the above studies may be accounted for by prosodic differences between the L1 languages (Korean, European and Canadian French). More importantly for our study, they might also relate to how L2 proficiency is assessed. First, whereas Lin et al. (2014) used continuous measures, the other studies used a categorical strategy, subdividing L2 learners into different proficiency groups (Tremblay, 2009; Dupoux et al., 2008). This categorical division could obscure subtle effects that might otherwise be visible in a continuous analysis. Second, there were differences in the way

L2 proficiency was measured (Lin et al., 2014: cloze test; Tremblay, 2009: cloze test and read aloud task; Dupoux et al., 2008: language history questionnaire). Maybe measures reflecting an impact of L2 knowledge need to be more specific. Studies have shown effects of quantity of input (Altenberg, 2005; Boll-Avetisyan, 2012; Flege & Liu, 2001), quality of input (such as interactions with friends/family, e.g., Moyer, 2011, or passive exposure to TV/radio, e.g., Flege, Yeni-Komshian & Liu, 1999) and diversity of exposure (for a review, see Munoz & Singleton, 2011) on different aspects of L2 phonological acquisition. Therefore, some aspects of L2 proficiency or exposure may be better predictors of speech perception than others. Hence, in the current study, we took continuous quantitative and qualitative indicators of L2 input into account.

Effects of music experience

A number of studies (e.g., Wong, Skoe, Russo, Dees & Kraus, 2007; Zuk, Ozernov-Palchik, Kim, Lakshminarayanan, Gabrieli, Tallal & Gaab, 2013) have provided evidence for a link between phonological processing and musical experience (as measured by the number of years of playing an instrument and/or the age of instrument acquisition): musically experienced native listeners of English are, for example, better at discriminating syllables contrasting in non-native lexical tone (Wong et al., 2007) and at discriminating consonants that vary in their voice onset time (Zuk et al., 2013) than listeners with less musical experience. Furthermore, Slevc and Miyake (2006) reported correlations between musical ability and L2 phonological perception and production skills (but not lexical and syntactic knowledge) in Japanese L2 learners of English. Hence, musical experience might generally facilitate L2 phonological acquisition and/or processing.

Only a few studies have explored whether musical experience may influence sensitivity to suprasegmental information. Kolinsky, Cuvelier, Goftry, Peretz and Morais (2009) found that French trained musicians outperformed non-musicians in the perception of lexical stress if only weak stress cues were provided. Bhatara, Yeung and Nazzi (2015) showed a correlation between musical rhythm perception and years of foreign language experience among French speakers. This parallels findings on tone perception with native listeners of English, in which musicians identified and discriminated Mandarin tone contrasts more accurately than non-musicians (e.g., Gottfried, Staby & Ziemer, 2004). These results indicate cross-domain transfer, which is not unexpected as music and speech use the same acoustic parameters to signal prominence.

In sum, some factors have been identified in prior studies that facilitate perception of non-native prosody:

experience with music and quantity and quality of L2 experience. Consequently, we tested native listeners of French who had begun learning German after childhood to determine if their responsiveness to intensity and duration as cues for rhythmic grouping may have been enhanced by their knowledge of an L2 with lexical stress, using the same task as Bhatara et al. (2013) and comparing their performance to French and German monolinguals. Based on the finding that L1 experience with contrastive lexical stress leads to stronger rhythmic grouping preferences (Bhatara et al., 2013; *in press*), we hypothesized that L2 experience with contrastive lexical stress would have similar (though smaller) effects. Furthermore, as the same acoustic cues are used in speech and in music, and since the ITL has been proposed to be relevant for the perception of both (Hayes, 1995), we hypothesized that L2 learners with more musical experience may be more sensitive to rhythm in speech than L2 learners with less musical experience. To thoroughly explore which factors may contribute to the achievement of more German-like rhythmic perception in French L2 learners, we provided our participants with an extensive questionnaire concerning their past and present L2 and musical exposure and use.

Method

Participants

Participants were 40 French late L2 learners of German (28 F, 12 M, $M_{\text{age}} = 30.1$, age range: 22–52) tested in the Berlin area. Recruitment criteria included having been raised monolingually with European French, having normal hearing and no language disorders, and being able to understand a German task instruction. They were compensated for their participation.

Measures for German and French proficiency were obtained by using German and French versions of a c-test, a standardized way of assessing second-language proficiency of learners similar to a cloze test: participants have to fill in missing letters to complete words in a text. Performance on this test correlates with multiple domains of linguistic knowledge (e.g., Eckes, 2010). Although a written measure of proficiency may not ideally reflect spoken language comprehension skills, it has the advantage of not being redundant with our auditory rhythmic perception task. Furthermore, we obtained data about their language history using a questionnaire based on the LEAP-Q (Marian, Blumenfeld & Kaushanskaya, 2007). A summary of the results of the proficiency test and the questionnaire is provided in Table 1.

To assess the impact of L2 acquisition, the L2 learners were compared to the French and German monolinguals tested by Bhatara et al. (2013). Furthermore, to assess the potential influence of musical experience, we

Table 1. Participant language background information.

Source of information	Factor	Mean	Range
Proficiency test	German proficiency (c-test score/100)	69.4	28–95
Language history questionnaire	Education (in years)	18.3	8–23
	Length of residence in Germany (in years)	5.6	0.3–28
	Age of arrival in Germany (in years)	22.3	12–33
	Age of acquisition (in years)	12.3	7–32
	Self-estimated amount of daily exposure (in %)	43	5–83
	L2 self-rating (from 0 = “none” to 10 = “perfect”)		
	• Comprehension	7.3	3–10
	• Reading	6.9	2–10
	• Oral	7.1	3–9
	• Writing	6.2	2–9
	• Vocabulary	6.5	2–10
	• Grammar	6.3	2–9
	• Pronunciation	6.2	1–9

Table 2. Participant music background information. Note: Information on the age of acquisition does not include 16 French, 7 German monolinguals and 9 L2 learners who never learned an instrument.

Factor	French monolinguals (40 N)		French L2 learners (40 N)		German monolinguals (33 N)	
	Mean	(Range)	Mean	(Range)	Mean	(Range)
Age (in years)	27	19–44	30	22–52	24	18–50
Length of formal musical training (in years)	4.2	(0–29)	5.8	(0–19)	6.1	(0–21)
Number of acquired musical instruments	1.0	(0–3)	1.4	(0–4)	1.5	(0–4)
Age of acquisition of the first instrument (in years)	10.1	(5–27)	10.9	(4–28)	9.2	(3–38)

collected three measures for each participant: number of instruments played, years of musical experience (if the years of two instruments overlapped, these were counted only once), and age of first musical experience (see Table 2).

Material

We used the same stimuli as Bhatara et al. (2013). These were ninety speech streams containing sixteen different CV syllables made up of four vowels (/e: i: o: u:/) and four consonants (/b z m l/; e.g., /...zu:le:bo:li:lo:zi:mu:be:.../). The syllable sequences were synthesized using the text-to-speech software MBROLA (Dutoit, Pagel, Pierret, Bataille & van der Vreken, 1996) in two different female voices – the German voice De5 and the French voice Fr4. The selection of the phonemes was based on the consideration that they exist in both languages and are discriminable for both language groups in both the French and German synthesis in spite of differences between their French and German acoustic

realizations. Baseline syllable intensity as measured in Praat (Boersma & Weenink, 2010) was set at 70 dB, baseline syllable duration at 260 ms, and baseline F0 at 200 Hz. For two of the experimental conditions (Duration and Intensity), these sequences were further manipulated. For the duration condition, every second syllable in a stream was lengthened, and for the intensity condition, every second syllable had increased intensity. The third experimental condition was the control condition, for which the sequences contained no variation. There were four levels of intensity variation (2, 4, 6, or 8 dB above baseline) and four levels of duration variation (50, 100, 150 or 200 ms above baseline).

In order to prevent the beginnings of the stimuli from giving grouping cues, the first three seconds of the streams were masked by white noise fading out and intensity fading in. In addition, half of the sequences started with a weak syllable, and the other half started with a strong syllable. Each participant heard 10 repetitions of each level of intensity or duration variation and 10 repetitions of the control sequences, resulting in a total of 90 trials.

Procedure

The procedure was the same as that of Bhatara et al. (2013). Participants were tested in a quiet room. Participants were instructed to listen carefully to each syllable sequence and to report whether they heard the syllables alternating as “strong-weak” pairs, or as “weak-strong” pairs. They were told not to wait until the end of a sequence and to respond as fast as possible.

Half of the participants received instructions in French, and the other half in German, using the same instructions as in Bhatara et al. (2013). As French does not use lexical stress, the examples in the French instructions were trochaic and iambic words from Spanish as well as examples of syllable emphasis in response to a word misunderstanding such as: “You say, ‘*J’aime le bateau*’ [I like the boat] and your friend says ‘*Le gâteau?*’ [the cake]. You respond, ‘*Non, le BAteau*’ placing the emphasis on the first syllable of the word.” The German instructions gave a couple of examples of trochaic (e.g., *Apfel, Auto*) and iambic words in German (e.g., *Salat, Proband*). By counterbalancing the language of instruction, we could control for language mode effects and test whether it would matter that different examples were used to explain what “weak-strong” and “strong-weak” meant.

Participants responded by pressing one of two labeled buttons (either a tall bar to the left of a short bar, symbolizing trochaic, or a short bar to the left of a tall bar, symbolizing iambic). Their responses were recorded. The testing procedure began with four practice trials. They were followed by ninety test trials presented in random order. After half of the trials, participants could take a short break. The experiment was run on a MacBook laptop, controlled by the software PsyScope X (available at <http://psy.ck.sissa.it/>). The stimuli were presented over AKG K 55 headphones with the volume set at a comfortable level. Left-right position of the response keys and MBROLA voice (German De5 or French Fr4) were counterbalanced between participants. After the rhythmic perception task, they completed the two c-tests and filled out the language and music background questionnaire.

Data processing, analysis, and results

Three separate analyses were carried out. Part 1 was a simple model testing whether the L2 learners’ grouping preferences followed the ITL by testing each condition against chance. Part 2 assessed whether the rhythmic grouping preferences of the L2 learners differed from those of the German and French monolinguals. Part 3 tested the potential influence of multiple continuous and categorical predictors on the L2 learners’ rhythm perception.

In all three models, the dependent variable was response type (1 = “trochaic” versus 0 = “iambic”). As

Table 3. Parameters of the linear mixed-effects logit regression, Part 1. For fixed effects, regression coefficients (β), their standard errors (SE), z-scores and the respective p-values are given. For random effects, variance and standard deviations are given.

Fixed effects	β	SE	z	p
Intensity	0.66	0.06	10.23	<.001
Duration	-0.23	0.06	-3.73	<.001
Control	0.31	0.11	2.82	<.001
Random effects	Variance	Std.Dev.		
Stimulus(Intercept)	0.03	0.16		
Participant(Intercept)	0.05	0.23		

the data had a binomial distribution, a logit generalized linear mixed model (e.g., Jaeger, 2008) was applied to the unaggregated data. Coefficients (β) indicate the logit-transformed proportion of trochaic responses. To account for the variability between individual participants and individual items, each model included random factors for participants and items, but random slopes were not specified as their inclusion led to false convergence.

To reduce collinearity and the number of highly correlated variables in our models, we performed principal component analyses (PCA). We always used the first principal component (PC) obtained by a PCA as a predictor, as this is the one that accounts for most of the variance of the combined variables. Models were calculated in “R” (R Core Team, 2012) using the package “lme4” (Bates, Maechler & Bolker, 2012); graphs were generated using the package “ggplot2” (Wickham, 2009).

Part 1: Testing L2 learners’ grouping preferences against chance

Data processing and analysis

The following model included the data of all 40 L2 learners. The intercept was set at zero, so that each level (Duration, Intensity and Control) of the fixed factor condition was tested against chance. A negative β indicates performance below chance, and a positive β indicates performance above chance.

Results

The mixed model results (see Table 3) revealed that for the L2 learners, as predicted by the ITL, the proportion of trochaic responses was significantly above chance level in both the intensity and the control condition, while it was below chance level in the duration condition. Results are presented in Figure 1.

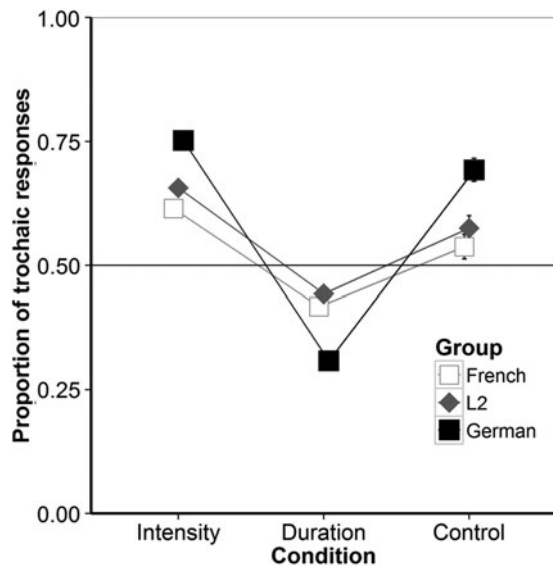


Figure 1. Mean proportion of trochaic responses (0 = “iambic”, 1 = “trochaic”) and their standard errors broken down by condition and group.

Part 2: Comparing Groups (L2 learners versus monolinguals)

Data processing and Analysis

The data of the 40 L2 learners were compared to the data from 40 monolingual French and 40 monolingual German listeners obtained by Bhatara et al. (2013). The fixed factors were group, condition and musical experience. The three levels of both the condition and the group factors were compared in sliding contrasts, using the R function `contr.sdif()`, which serially compare different groups (i.e., 1 to 2, and 2 to 3). Sliding contrasts have the statistical benefits of being orthogonal and using the grand mean as the intercept, for which the baseline is much more stable than in contrasts that only use a subset of the data as a baseline. For group, we compared the French monolinguals to the L2 learners (henceforth L2–F) and the L2 learners to the German monolinguals (henceforth G–L2), thus not comparing the two monolingual groups already compared in Bhatara et al. (2013).

For condition, comparisons were made between Duration and Intensity (henceforth D–I) and Control and Duration (henceforth C–D). This contrast was chosen because in Bhatara et al.’s (2013) study, both intensity-varied and control sequences were grouped more consistently as trochaic and duration-varied sequences were grouped more consistently as iambic by the German than by the French monolinguals. Hence, to ascertain whether the L2 learners develop German-like grouping preferences, we should examine whether they show increased differences relative to French monolinguals

when comparing duration-varied sequences to both intensity-varied and control sequences.

The effect of musical experience was assessed by means of a PC, which equally represented the participants’ years of experience, number of instruments, and age of first musical experience (loadings of 0.58, 0.57 and 0.58; the three variables being correlated) and captured 80% of the variance contained in these three variables (see Supplementary Material, Table S1, for more details on this PCA). As the D–I contrast is more relevant for testing the predictions of the ITL than the C–D contrast, we will report the results of the D–I comparison before the results of the C–D comparison.

Results

The model coefficients (β) and estimated standard errors (*SE*) are provided in Table 4. Coefficients indicate the difference scores of the logit-transformed proportion of trochaic responses. Only significant effects are discussed.

Intensity–Duration comparison

Note that for the D–I contrast, the negative β means that there were more trochaic responses for Intensity than for Duration since Intensity is subtracted from Duration. Interactions including the D–I contrast that have negative β s indicate an increase in the magnitude of the D–I difference, which corresponds to more German-like responses.

Across all groups, there were significantly more trochaic responses to intensity-varied than to duration-varied sequences. Moreover, differences were observed between L2 learners and German monolinguals (significant G–L2*D–I). The L2 learners showed less of a difference than the German monolinguals in the number of trochaic responses between the duration and intensity conditions. However, the L2 learners did not differ from the French monolinguals (nonsignificant L2–F*D–I).

The only difference between the L2 learners and the French monolinguals was found in interactions with musical experience (see Figure 2). D–I**Musical experience* was significant: the difference in grouping preferences between Intensity and Duration increased with musical experience. However, the L2 learners’ grouping preferences were more enhanced by musical experience than the French monolinguals’ (significant L2–F*D–I**Musical experience*).

Control–Duration comparison

Note that for the C–D contrast, the positive β means that there were more trochaic responses for Control than for Duration since Duration is subtracted from Control. Interactions including the C–D contrast that have positive β s indicate an increase in the magnitude of the C–D difference, which corresponds to more German-like responses.

Table 4. Parameters of the linear mixed-effects logit regression, Part 2. For fixed effects, regression coefficients (β), their standard errors (SE), z-scores and the respective p-values are given. For random effects, variance and standard deviations are given.

Fixed effects	β	SE	z	p
Intercept (= grand mean)	0.24	0.04	6.41	<.001
D-I	-1.20	0.05	-25.74	<.001
C-D	0.88	0.07	12.13	<.001
L2-F	0.14	0.09	1.55	.12, n.s.
G-L2	0.13	0.09	1.50	.13, n.s.
Musical experience	0.03	0.02	1.07	.28, n.s.
L2-F*D-I	-0.08	0.11	-0.70	.48, n.s.
G-L2*D-I	-1.08	0.11	-9.79	<.001
L2-F*C-D	0.06	0.18	0.35	.73, n.s.
G-L2*C-D	1.16	0.17	6.78	<.001
D-I*Musical experience	-0.07	0.03	-2.42	.015
C-D*Musical experience	0.03	0.05	0.69	.49, n.s.
L2-F*Musical experience	0.04	0.05	0.74	.46, n.s.
G-L2*Musical experience	-0.01	0.06	-0.16	.88, n.s.
L2-F*D-I*Musical experience	-0.19	0.07	-2.94	<.01
G-L2*D-I*Musical experience	0.12	0.08	1.52	.13, n.s.
L2-F*C-D*Musical experience	0.21	0.10	2.06	.04
G-L2*C-D*Musical experience	-0.23	0.12	-1.97	<.05
Random effects	Variance	Std.Dev.		
Stimulus(Intercept)	0.03	0.17		
Participant(Intercept)	0.08	0.28		

Across all groups, there were significantly more trochaic responses to control than to duration-varied sequences. The contrast between the control and the duration-varied sequences was larger for the German monolinguals than for the L2 learners (significant G-L2*C-D), while there was no difference between the L2 learners and French monolinguals (nonsignificant L2-F*C-D).

However, differences between the L2 learners and the French monolinguals as well as the L2 learners and the German monolinguals became evident in the interaction with musical experience: the more musically experienced the L2 learners were, the more German-like their grouping preferences were, while the monolinguals' grouping preferences did not change with musical experience (significant L2-F*C-D*Musical experience and G-L2*C-D*Musical experience).¹

¹ We separately investigated whether there were group differences with respect to the different manipulation steps (200 ms, 150 ms...), excluding musicality because of non-convergence. This model generally did not add much information beyond the models presented in the current study, but a difference was found between French monolinguals and French L2 learners between the 50 and 100 ms duration steps, L2 learners changing their grouping

Part 3: Exploring predictors of L2 learners' rhythmic perception

Data processing and analysis

In order to reveal which factors are associated with the L2 group's performance on the grouping task, a third model was applied to the L2 learners' data. The fixed factor Condition was specified as a sliding contrast as described above. Furthermore, several potential (continuous and categorical) predictors of L2 rhythmic perception were added as fixed factors.

To reduce collinearity, we applied a predictor reduction via PCA whenever at least two questions of the questionnaires assessed the same kind of information and were correlated. Otherwise, we included the variable in isolation. To reduce collinearity further, all continuous variables were centered. Although some of the variable combinations included in the model comparisons were correlated, we did not further control for collinearity by means of residualization (see Wurm & Fiscaro, 2014; York, 2012, who show that an inclusion of

preferences to "iambic" sooner than the French monolinguals. This might suggest that L2 experience led to an increase in sensitivity to duration (see details in the Supplementary Material, Figure S1 and Table S2).

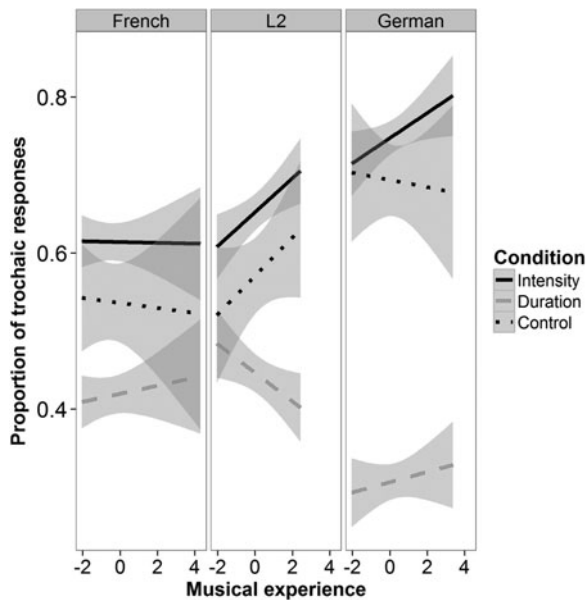


Figure 2. Linear regression lines reflecting the mean proportion of trochaic responses (0 = “iambic”, 1 = “trochaic”) for the effect of musical experience broken down by condition and group. The values for musical experience were the output of a Principal Component Analysis including information on participants’ years of musical experience, age of acquiring an instrument, and the number of learned instruments. Lower values indicate less musical experience. Shaded areas indicate the standard deviations.

residualized factors has stronger detrimental impacts on the interpretability of effects than an inclusion of correlated variables).

The method of testing various potential variables is an exploratory approach. For that reason, we used the Akaike information criterion (AIC) as a measure for comparing the relative goodness of fit between models. Our model comparisons included 12 predictors, described in more detail below. To test the effects of quality and quantity of exposure to German, we included six predictors from the questionnaire. The first one (Exposure-G, amount of daily German input in %) corresponded to the amount of L2 exposure, while the other five provided information regarding the type of exposure. These five variables were based on participants’ ratings on an 11-point scale ranging from not important/never (0) to most important/all the time (10). Of these five predictors, three were related to the type of input considered by the participants to be relevant for their ACQUISITION (Acq) of German: a) Acq-Reading, hence non-auditory input, b) Acq-Listening, a PC for non-interactive auditory exposure to TV/radio input, and c) Acq-Interacting, a PC combining the relevance of interacting with family and friends for acquisition, that is interactive auditory exposure. The other two predictors

were related to the amount of the same types of input IN CURRENT LIFE (Cur): a) Cur-Reading, and b) Cur-Auditory, a PC combining four correlated factors relating to TV/radio/friends/family input. The remaining five variables were tested because they were identified as predictors of L2 perception in prior studies. Two of these reflected general L2 proficiency: a) the c-test score obtained from the proficiency test, and b) a PC for the self-estimated oral L2 skills in responses to five questions regarding the L2 learners’ self-estimated German skills (oral language, comprehension, pronunciation, strength and detectability of foreign accent). Two variables were inherent to the experimental design and tapped into potential language mode effects: a) the language in which the instructions were provided (French or German), and b) the voice (French or German) with which the stimuli were presented.² The fifth variable was age of arrival in Germany. Details on the PCs are provided in the Supplementary Materials (Supplementary Materials, Table S3).³

Our model comparisons started out with a model including all covariates, which was then reduced in a step-wise fashion. The final model, which yielded the lowest AIC score (AIC: 4771) and hence can be considered to provide an optimal account for the variance in the L2 learners’ data, included the predictors Musical experience, Exposure-G, Acq-Reading, Cur-Reading, Acq-Listening, and Cur-Auditory. The first three predictors were tested in interaction with Condition, while the latter three predictors were not tested in this interaction, as this would not have helped to capture more variance.

Results

Results of the covariate analysis of the L2 group are given in Table 5 and displayed in Figure 2. As in the previous model, there were effects of Condition: duration-varied sequences were judged less often to be trochaic than intensity-varied sequences (negative D–I) and control sequences (positive C–D).

² No effects of language mode were found, and this factor will not be further discussed.

³ Length of residence was not included because it was confounded with age (L2ers who had spent more time in Germany were generally older [$r = .61$]). Hence, we cannot interpret the “negative” effect of length of residence as it might be due to related factors such as hearing decrease. Furthermore, we did not include age of acquisition because all but two participants had acquired German in school, long before moving to Germany – a lack of variability reducing its potential informativeness. Instead, we used participants’ age at arrival in Germany, which should be informative for L2 phonological acquisition as it marks the beginning of consequent L2 exposure. Finally, we did not include factors relating to L3/L4/L5 knowledge. First, had we done so, we would have needed to test monolingual control groups in other frequent L2s. Second, the number of data points obtained in the current study may not have sufficed to increase the number of tested predictors much further.

Table 5. Parameters of the linear mixed-effects logit regression, Part 3. For fixed effects, logit-transformed regression coefficients (β), their standard errors (SE), z-scores and the respective p-values are given. For random effects, variance and standard deviations are given.

Fixed effects	β	SE	z	p
Intercept (= grand mean)	0.25	0.05	4.96	<.001
D-I	-0.91	0.07	-12.17	<.001
C-D	0.55	0.12	4.69	<.001
Musicality	0.03	0.03	0.88	.38, n.s.
Exposure-G	-0.0003	0.003	-0.12	.90, n.s.
Acq-Reading	0.001	0.02	0.04	.97, n.s.
Acq-Listening	-0.09	0.04	-2.16	.04
Cur-Reading	-0.04	0.02	-2.05	.04
Cur-Auditory	0.13	0.04	3.17	<.01
D-I*Musical experience	-0.11	0.05	-2.16	.03
C-D*Musical experience	0.14	0.08	1.76	.07, n.s.
D-I*Exposure-G	-0.01	0.004	-2.45	<.05
C-D*Exposure-G	0.01	0.01	1.32	.19, n.s.
D-I*Acq-Reading	0.11	0.02	4.39	<.001
C-D*Acq-Reading	-0.07	0.04	-1.75	.08, n.s.
Random effects	Variance	Std.Dev.		
Stimulus(Intercept)	0.05	0.21		
Participant(Intercept)	0.03	0.17		

Effects of the covariates are illustrated in Figure 3. Three variables, Musical experience, Exposure-G and Acq-Reading, interacted with D-I. Again, given the direction of these effects, interactions with the D-I contrast that have negative coefficients correspond to more German-like responses. Musical experience (as already described in Part 1) and Exposure-G had “positive” effects on the D-I difference: L2 learners with more musical experience and L2 learners who are currently generally more exposed to German input gave more German-like responses with more trochaic responses in the intensity condition, and more iambic responses in the duration condition. In contrast, Acq-Reading had a “negative” influence on rhythm perception: the less importance was attributed to reading (which is non-interactive and non-auditory) input for the acquisition of German, the more German-like the responses in the task were (positive D-I*Acq-Reading).

Three further predictors had a significant main effect: first, higher ratings of TV, radio, friends and family as a current source of German input go along with more trochaic responses across conditions, including the duration condition (positive Cur-Auditory). Moreover, both Acq-listening and Cur-Reading had effects in the opposite direction (reflected by the negative coefficients): L2 learners who attributed LESS importance to listening to radio and TV for their acquisition of German and L2 learners who are currently LESS exposed to written

German input gave more trochaic responses across conditions, again including duration.

Proficiency

A correlation analysis including the c-test scores in German, the self-estimated rates of L2 proficiency and the factors found to be relevant predictors of L2 rhythmic perception in the generalized mixed model (Part 3) revealed that the c-test scores significantly correlate with a PC of the self-assessed rates of oral proficiency (PC-Proficiency; $r = .52, p < .001$), and with reading ($r = .80$), writing ($r = .84$), vocabulary ($r = .69$) and grammar ($r = .80$; all p 's $< .05$) skills. Hence, the c-test scores are likely to give a good indication of the participants' general L2 proficiency. Nonetheless, neither the c-test scores nor the PC-proficiency were included into the described model because neither were significant predictors of L2 rhythmic perception, nor could they account for variability in the data. Furthermore, the c-test scores were uncorrelated with Musical experience ($r = -.08, p = .61$), Exposure-G ($r = .28, p = .08$), Acq-Reading ($r = .29, p = .07$), Acq-Listening ($r = .21, p = .21$), but significantly correlated with Cur-Reading ($r = .46, p < .01$) and Cur-Auditory ($r = .39, p < .05$).

Discussion

French advanced L2 learners of German were tested on their rhythmic grouping abilities. We predicted that L2

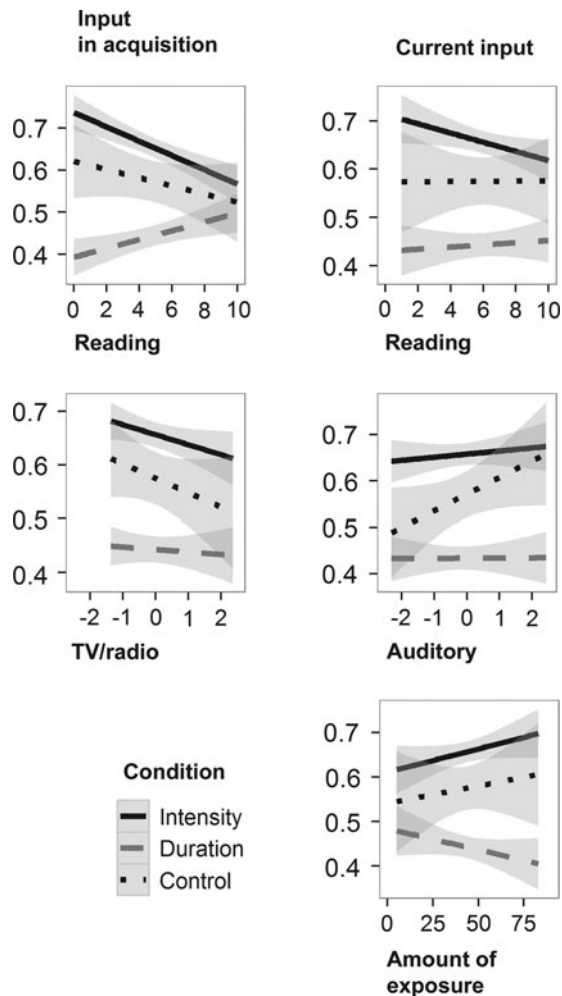


Figure 3. Linear regression lines reflecting the mean proportion of trochaic responses (0 = “iambic”, 1 = “trochaic”) for the effect of specific types of L2 input on condition. Lower values indicate less relevance of a specific input type during acquisition (left column) and less amount of a specific input type in current life (right column). Units of measurement on the x-axis: Likert scale (0-10) for reading, principal component for auditory and TV/radio, and % for exposure. Shaded areas indicate the standard deviations.

experience with contrastive lexical stress should lead to more consistent rhythmic grouping preferences compared to those of French monolinguals. This prediction was motivated by the finding that German monolinguals are more consistent in their rhythmic grouping preferences than French monolinguals (Bhatara et al., 2013; *in press*), and that L2 experience can affect L2 prosodic processing in other languages (Lin et al., 2014; Tremblay, 2009). However, given some studies suggesting that stress perception in European French listeners is not improved by L2 experience with lexical stress (Altmann, 2006; Dupoux et al. 2008; Kijak, 2009; Schmidt-Kassow et al.,

2011), we explored various predictors that are known to affect L2 phonological acquisition, such as the amount and type of exposure. In so doing, our exploratory data analysis indicated that several of these language-related factors, together with musical experience, modulate sensitivity to prosodic information among French L2 learners of German.

First, we established that rhythmic grouping by the L2 learners was in line with the predictions of the ITL: like the German and French monolinguals (Bhatara et al., 2013), the L2 learners perceived a trochaic structure when listening to syllable sequence alternating in intensity and an iambic structure when listening to sequences with syllables alternating in duration. Like the German but not the French monolinguals, the L2 listeners perceived trochees when listening to control sequences without rhythmic alternation.

Second, we tested whether L2 learners differed in their grouping preferences from German and French monolinguals. We found that they did not differ significantly from French monolinguals as a group, but differences were revealed if musical experience was taken into account: L2 learners who had more musical experience performed more similarly to the German monolinguals. Still, although the grouping abilities of the L2 learners were affected by their L2 experience, they continued to differ as a group from the German monolinguals, suggesting limits in the way prosodic processing can change as a result of L2 learning.

Third, we investigated the influence of several potential predictors of rhythm perception related to L2 experience. We found that neither general measures of language proficiency, age of arrival, nor the experimental variables “language of instruction” and “stimulus voice” accounted for any variance in the data. Instead, we identified five input quality and quantity factors that were related to rhythmic grouping preferences in the French L2 learners of German. Three of these language-related factors concern current L2 input while the other two concern L2 acquisition. Moreover, these factors were associated with different perceptual outcomes: two were associated with a more German-like perception, while the other three were associated with a more trochaic perception across conditions. A more German-like perception was found when L2 learners were currently exposed to more German, and when they had NOT considered written input (Acq-Reading) to be relevant for their acquisition of German. A general trochaic bias was found in L2 learners who indicated that their current contact with German was mainly through auditory input. Furthermore, this general trochaic bias was negatively correlated with reliance on TV/radio during acquisition and with current reading of German – i.e., the more the participants read or the more they reported that TV/radio exposure had been important during acquisition, the less they demonstrated a trochaic bias.

In the following, we discuss the effects of the factors that improved the statistical model with respect to the AIC, after a brief discussion of the factors that did not improve the model. Finally, we discuss potential explanations for why some predictors were associated with an increase of the effects of the ITL whereas others were associated with a general trochaic bias.

Factors not affecting rhythmic grouping

L2 proficiency

General L2 proficiency had no effect on rhythmic perception in the present study. Neither the German c-test scores nor self-estimated oral proficiency had significant effects on grouping, and neither improved the statistical model. While L2 studies frequently use c-tests as a measure for individual differences among L2 learners (Tremblay, 2011) because they often correlate with different aspects of L2 proficiency (Eckes & Grotjahn, 2006), they may not relate closely to prosodic proficiency. As indicated by the present results, the acquisition of L2 stress appears to depend on exposure to spoken language, just like the acquisition of other phonological aspects of an L2 (e.g., Altenberg, 2005; Flege & Liu, 2001). L2 proficiency, however, can be influenced by many other factors that do not relate to exposure at all, such as motivation (Bongaerts, 1999), language aptitude (Abrahamsson & Hyltenstam, 2008), and general cognitive skills such as working memory (Service & Kohonen, 1995), just to name a few. This indicates that measures of general L2 proficiency might be too global to predict L2 prosodic development (though see Lin et al., 2014). This could explain why Tremblay (2009) did not find differences across groups with varying proficiency levels but found an impact of daily L2 exposure on the processing of word stress in native listeners of Canadian French. It might also explain why in Dupoux et al.'s study (2008), a global measure of L2 proficiency was not a significant predictor of L2 contrastive lexical stress acquisition in European French learners of Spanish.

Factors affecting rhythmic grouping

Musical experience

Musical experience throughout the lifespan was a positive predictor of German-like rhythmic perception in L2 learners. This finding is in line with the proposal of the ITL as an extra-linguistic principle accounting for rhythmic patterns in music and language (Hayes, 1995). However, musical experience did not generally affect speech rhythm perception. Rather, it appeared to interact with L2 learning. Indeed, with increased musical experience, the difference between the duration and both the intensity and control conditions became larger in the L2 learners but

not in the monolinguals. Hence, the present results rather suggest an effect of musical experience on L2 learning.

To account for this effect, we speculate that musical experience does not directly influence rhythmic grouping preferences when perceiving speech (which explains why the French monolinguals were unaffected by musical experience). However, musical experience may have influenced the L2 learners' acquisition of word stress in German and, hence, an establishment of abstract representations of stress which was proposed to have a relevant role in rhythmic grouping (Bhatara et al., 2013). As pointed out in the introduction, musical experience can enhance listeners' sensitivity to suprasegmental features of a non-native language (Gottfried et al., 2004; Kolinsky et al., 2009). Hence, we suggest that musical experience may facilitate the detection of the acoustic correlates of word stress in French L2 learners of German, which is a necessary condition to build up a representation of stress. This process may be supported by the acoustic overlap of signaling rhythm in music and language. These explanations could also account for why musical experience was not associated with rhythmic grouping in the monolingual French listeners, as they have no abstract representation of lexical stress. Taken together, the present results could be interpreted as indicative of an effect of abstract knowledge of either L1 or L2 lexical stress rather than a direct effect of musical experience on rhythmic grouping.

Note that musical experience was not correlated with general L2 proficiency. This is in line with findings by Slevc and Miyake (2006) reporting effects of musical ability on L2 segmental phonology production and perception (discrimination of non-native phoneme contrasts), but not syntax and lexical knowledge. In the current study, we may have found a link between music and L2 regarding suprasegmental L2 knowledge; a connection that is even more logical since music and language share rhythmic properties.

While our PC combining the variables of years of experience, age of acquisition and number of instruments provides us with a global measure of the participants' musical experience, the three questions we used did not allow us to measure the quality of this musical experience. Therefore, future studies should endeavor to use more fine-grained or precise measures for assessing effects of musicality or musical experience (as we did in the present study for the language measures), possibly by manipulating these factors experimentally. For example, in addition to the questions used in the present study, future studies could collect information on whether a person had individual or group lessons, studied music in a conservatory or in basic music classes, and whether they practiced daily or not, both in the past and currently.

Furthermore, it would also be interesting for future studies to use musicality tests for assessing listeners'

musical aptitude. It is likely that musical aptitude and musical experience are correlated given that people with higher musical aptitude are more likely to pursue musical training. However, future studies should directly address the question of whether individuals who are inherently more musical are better at picking up prosodic properties of an L2, and how or even if this relates to their musical experience.

Amount of L2 exposure

The current quantity of daily exposure to German significantly predicted rhythmic grouping. With more daily exposure to German, the L2 learners in our study also developed more German-like grouping preferences. This is in line with previous findings of effects of daily exposure on other aspects of L2 perception and production such as phonemes and phonotactics (Altenberg, 2005) and word stress (Tremblay, 2009).

Type of input

In this study, we examined the effects of the type of input during acquisition and in current life. With respect to current input, we found that more L2 auditory input and, paralleling this, less L2 written input was associated with more trochaic responses in all conditions, including the duration condition. Hence, certain types of current input were associated with a general trochaic bias. Though duration cues are not processed as they should be, this general trochaic grouping strategy is German-like in the other two conditions (see further discussion below). The finding that current quality of input is related to L2 phonology is not new. Previous studies have found that the interactive use of the L2 with native speakers has a significant impact on phonological abilities in perception (e.g., phoneme identification, Díaz-Campos, 2004; Flege & Liu, 2001). It has also been found previously that current media, written and oral L2 input are predictors of L2 prosody production (Huang & Jun, 2011).

To our knowledge, we found for the first time indications that the type of input during the initial stages of acquisition may have a sustained impact on rhythm perception in L2 learners (although previous studies have assessed, but failed to reveal, its potential influence, e.g., Huang & Jun, 2011). The less relevance L2 learners attributed to reading during acquisition of German, the more German-like grouping preferences were found. Moreover, the less relevance L2 learners attributed to TV/radio exposure during acquisition, the more they demonstrated a general trochaic bias. Our findings also show that the same type of input can have different effects during the initial stages of L2 learning versus in current life: while less importance attributed to reading at the acquisition stage was associated with a more German-like perception in accordance with the ITL, less reading in current life was associated with a general trochaic bias.

ITL or trochaic bias: a different status of intensity/(pitch) versus duration changes?

Our study has identified some predictors related to L2 German and musical experience that affect rhythmic grouping preferences in French listeners. However, not all predictors were associated with similar perceptual outcomes. More musical experience, more daily exposure to German and less importance attributed to reading exposure at the initial stages of L2 acquisition were associated with enhanced grouping preferences in accordance with the ITL. This relation is logical, as this would mirror the grouping preferences of the native speakers of the target language. More puzzling, however, is that less reading and more auditory exposure in current life as well as more importance attributed to passive auditory exposure at the initial stages of L2 acquisition were associated with relatively more trochaic groupings across all conditions, including the grouping of duration-varied syllable sequences.

One possible explanation for this effect is that L2 learners may have overgeneralized the trochaic pattern. Work by Cutler and colleagues suggests that postulating word beginnings at strong syllables is an extremely successful segmentation strategy for English listeners, as English words are predominantly trochaic (for a summary, see Cutler, 1994). German words are also predominantly trochaic. Hence, even if L2 experience does not *per se* enhance sensitivity to all stress cues, an overgeneralization of a trochaic listening strategy while ignoring duration as a cue for an iambic stress pattern is still prone to be a successful strategy for processing German speech. Hence, the emergence of an overall trochaic bias in the French L2 learners who were frequently exposed to auditory German and infrequently exposed to written German may indicate the development of a partial German-like rhythmic perception.

A second possible explanation for these results is that there is a qualitative difference in the way listeners perceive intensity and duration variation in speech. Perhaps duration is not an early, robust cue to iambic structure. Indeed, while pitch and/or intensity variation consistently leads to trochaic groupings in adults (Bhatara et al., 2013; *in press*; Crowhurst & Teodocio Olivares, 2014; Iversen et al., 2008; Kusumoto & Moreton, 1997) and infants (Bion, Benavides-Varela & Nespor, 2011; Molnar et al., 2014; Yoshida et al., 2010) with various native languages, grouping by duration variation depends on the particular native language (adults: Crowhurst & Teodocio Olivares, 2014; Iversen et al., 2008; Kusumoto & Moreton, 1997; infants: Hay & Saffran, 2012; Molnar et al., 2014; Yoshida et al., 2010), and duration-based grouping preferences seem to emerge later in infancy (Bion et al., 2011; Hay & Saffran, 2012; Yoshida et al., 2010). Our results that some factors are associated with an overall increase in trochaic responses even in

duration-varied sequences might constitute additional evidence that duration is not as robust a grouping cue as intensity and pitch. Moreover, the results suggest that this is true for both L1 and L2 acquisition.

Conclusion

The present results on French L2 learners of German show that the effects of the Iambic/Trochaic Law, extended here to this population of late bilinguals, are not only modulated by L1 experience (Bhatara et al., 2013, *in press*; Crowhurst & Teodocio Olivares, 2014; Iversen et al., 2008; Kusumoto & Moreton, 1997; Yoshida et al., 2010), but also by specific factors involved with adult L2 experience. This first establishes that even after childhood, learning an L2 can affect rhythm perception. Second, some of the factors identified relate to quantitative and qualitative (e.g., written versus auditory input) indicators of L2 experience. Lastly, we found that musical experience influences the perception of rhythmically structured speech, if listeners have sufficient knowledge of an L2 for which stress is relevant. On the contrary, general measures of L2 proficiency did not capture variance in L2 learners' rhythmic grouping. The current study did not directly investigate the acquisition of word stress in an L2, but rather the modulation of rhythmic perception of speech streams. It would be interesting for future studies to directly assess whether factors relating to L2 and musical experience influence the acquisition of L2 lexical stress. However, given the proposal that rhythmic grouping preferences relate to experience with word-level prosodic structure (Bhatara et al., 2013; *in press*), our results have implications for the study of stress 'deafness'. Many studies have reported that French adults cannot overcome their stress 'deafness' even if they are advanced learners of an L2 with lexical stress (Altmann, 2006; Dupoux et al., 2008; Kijak, 2009, Schmidt-Kassow et al., 2011). We also failed to find an overall difference between French monolinguals and French L2 learners of German. However, by making use of more fine-grained information related to L2 and musical experience, we found that French L2 learners become more sensitive to rhythmic structure under specific circumstances. Such an approach should be used in future studies on stress 'deafness'.

To conclude, our results imply that the processes underlying the perception of speech rhythm are more dynamic and flexible than previously thought. Future research should further investigate L2 experience as well as both musical experience and ability in more detail to better understand the contributions of linguistic and non-linguistic factors to rhythmic perception.

Supplementary Material

For supplementary material accompanying this paper, visit <http://dx.doi.org/10.1017/S1366728915000425>

References

- Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30, 481–509.
- Abboub, N., Bijeljac-Babic, R., Serres, J., & Nazzi, T. (2015). On the importance of being bilingual: Word stress processing in a context of segmental variability. *Journal of Experimental Child Psychology*, 132, 111–120.
- Altenberg, E. P. (2005). The judgment, perception, and production of consonant clusters in a second language. *International Review of Applied Linguistics in Language Teaching*, 43, 53–80.
- Altmann, H. (2006). *The Perception and Production of Second-Language Stress: A Cross-linguistic Experimental Study*. Ph.D. Dissertation, University of Delaware.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using Eigen and syntax. R package version 0.999999-0.
- Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pp. 171–204. Timonium, MD: York Press.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345–360.
- Bhatara, A., Boll-Avetisyan, N., Agus, T., Höhle, B., & Nazzi, T. (in press). Language experience affects grouping of musical instrument sounds. *Cognitive Science*.
- Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language and stimulus complexity affect rhythmic grouping of speech. *Journal of the Acoustical Society of America*, 134, 3828–3843.
- Bhatara, A., Yeung, H. H., & Nazzi, T. (2015). Foreign language learning in French speakers is associated with rhythm perception, but not with melody perception. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 277–282.
- Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLOS ONE*, 7, 1–8.
- Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2013). Effect of bilingualism on the perception of lexical stress in 6-month-old French-learning infants. In S. Baiz, N. Goldman & R. Hawkes (eds.), *Proceedings of the 37th annual Boston University Conference on Language Development*.
- Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, 54, 123–140.

- Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.2.05, retrieved 10 October 2010 from <http://www.praat.org/>.
- Boll-Avetisyan, N. (2012). *Phonotactics and its Acquisition, Representation, and Use: An Experimental-Phonological Study* (Vol. 298), Utrecht: LOT Dissertation series.
- Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, 6, 145–238.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, pp. 133–159. Mahwah, NJ: Erlbaum.
- Burnham, D. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, 7, 207–240.
- Crowhurst, M. J., & Teodocio Olivares, A. (2014). Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech. *Phonology*, 31, 51–94.
- Cutler, A. (1994). The perception of rhythm in language. *Cognition*, 50, 79–81.
- Delattre, P. (1938). L'accent final en français: accent d'intensité, accent de hauteur, accent de durée ('The final accent in French: Intensity accent, pitch accent, duration accent'). *The French Review*, 12, 141–145.
- Di Cristo, A. (1998). Intonation in French. In D. Hirst & A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*, pp. 195–218. New York: Cambridge University Press.
- Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26, 249–273.
- Dupoux, E., Pallier, C., Sebastian-Gallés, N., & Mehler, J. (1997). A destressing 'deafness' in French? *Journal of Memory and Language*, 36, 406–421.
- Dupoux, E., Peperkamp, S., & Sebastian-Gallés, N. (2010). Limits on bilingualism revisited: stress 'deafness' in simultaneous French-Spanish bilinguals. *Cognition*, 114, 266–275.
- Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". *Journal of the Acoustical Society of America*, 110, 1606–1618.
- Dupoux, E., Sebastian-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, 106, 682–706.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vreken, O. (1996). The MBROLA project: Towards a Sst of high-quality speech synthesizers free of Uue for non-commercial purposes. *Proceedings of the Proceedings of The Fourth International Conference on Spoken Language Processing (ICSLP'96)* 3.
- Eckes, T. (2010). Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In R. Grotjahn (ed.), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-test: Contributions from current research*, pp. 125–192. Frankfurt am Main: Lang.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23, 290–325.
- Féry, C., Hörnig, R., & Pahaut, S. (2011). Correlates of phrasing in French and German from an experiment with semi-spontaneous speech. In C. Gabriel & C. Lleó (eds.), *Intonational Phrasing in Romance and Germanic: Cross-linguistic and Bilingual Studies*. Vol. 10, pp. 11–42. Amsterdam: John Benjamins Publishing Company.
- Flège, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in cross-language research*, pp. 233–272. Timonium, MD: York Press.
- Flège, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a Second language. *Studies in Second Language Acquisition*, 23, 527–552.
- Flège, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78–104.
- Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination. *Journal of the Acoustical Society of America*, 115, 2545–2545.
- Hay, J. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, 69, 113–122.
- Hay, J. F., & Saffran, J. R. (2012). Rhythmical grouping biases constrain infant statistical learning. *Infancy*, 17, 610–641.
- Hayes, B. (1985). Iambic and trochaic rhythm in stress rules. In M. Niepokuj et al. (eds.), *Proceedings of the Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*.
- Hayes, B. (1995). *Metrical stress theory: principles and case studies*, Chicago & London: The University of Chicago Press.
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, 32, 262–274.
- Huang, B. H., & Jun, S.-A. (2011). The Effect of Age on the Acquisition of Second Language Prosody. *Language and Speech*, 54, 387–414.
- Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, 124, 2263–2271.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (ed.), *Intonation: Analysis, Modeling and Technology*, pp. 209–242. Dordrecht: Kluwer Academic.
- Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, 14, 147–172.
- Kijak, A. M. (2009). *How stressful is L2 stress? A cross-linguistic study of L2 perception and production of metrical systems* (Vol. 214), Utrecht: LOT.
- Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1221.

- Kolinsky, R., Cuvelier, H., Goftry, V., Peretz, I., & Morais, J. (2009). Music training facilitates lexical stress processing. *Music Perception, 22*, 235–246.
- Kusumoto, K., & Moreton, E. (1997). Native language determines parsing of nonlinguistic rhythmic stimuli. *Journal of the Acoustical Society of America, 105*, 3204.
- Lerdahl, F., & Jackendoff, R. (1983). *Generative Theory of Tonal Music*. Cambridge: MIT Press.
- Lin, C. Y., Wang, M., Idsardi, W. J., & Xu, Y. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition, 17*, 316–346.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940–967.
- Molnar, M., Lallier, M., & Carreiras, M. (2014). The Amount of Language Exposure Determines Nonlinguistic Tone Grouping Biases in Infants From a Bilingual Environment. *Language Learning, 64*, 45–64.
- Moyer, A. (2011). An investigation of experience in L2 phonology: Does quality matter more than quantity? *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes, 67*, 191–216.
- Munoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching, 44*, 1–35.
- Narmour, E. (1990). *The analysis and cognition of basic melodic structures*. Chicago, IL: University of Chicago Press.
- Nespor, M., Shukla, M., Vijver, R. v. d., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e linguaggio, 7*, 1–28.
- Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics, 38*, 422–430.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961–2977.
- Schmidt-Kassow, M., Rothermich, K., Schwartze, M., & Kotz, S. A. (2011). Did you get the beat? Late proficient French-German learners extract strong-weak patterns in tonal but not in linguistic sequences. *Neuroimage, 54*, 568–576.
- Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155–172.
- Skoruppa, K., Pons, F., Bosch, L., Christophe, A., Cabrol, D., & Peperkamp, S. (2013). The Development of Word Stress Processing in French and Spanish Infants. *Language Learning and Development, 9*, 88–104.
- Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Alves Limissuri, R., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science, 12*, 914–919.
- Slevc, L. R., & Miyake, A. (2006). Individual differences in Second-language proficiency: does musical ability matter? *Psychological Science, 17*, 675–681.
- R Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33–58.
- Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29*, 553–584.
- Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by French Canadian listeners. *International Journal of Bilingualism, 13*, 35–62.
- Tremblay, A. (2011). Proficiency assessment standards in Second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition, 33*, 339–372.
- Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics, 34*, 343–371.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.
- 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, 54*–77.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*, 420–422.
- Wurm, L. H., & Fisičaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language, 72*, 37–48.
- York, R. (2012). Residualization is not the answer: rethinking how to address multicollinearity. *Social Science Research, 41*, 1379–1386.
- Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115*, 356–361.
- Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D. E., Tallal, P., & Gaab, N. (2013). Enhanced syllable discrimination thresholds in musicians. *PLOS ONE, 8*(12).