Learning a Phonological Contrast Modulates the Auditory Grouping of Rhythm

H. Henny Yeung, a,b,c Anjali Bhatara, b,c Thierry Nazzi b,c

aDepartment of Linguistics, Simon Fraser University
bPsychology of Perception Laboratory (UMR 8242), Centre national de la recherche scientifique
[The French National Center for Scientific Research]
cPsychology of Perception Laboratory (UMR 8242), Université Paris Descartes [Paris Descartes University]

Received 3 March 2017; received in revised form 3 May 2018; accepted 9 May 2018

Abstract
Perceptual grouping is fundamental to many auditory processes. The Iambic–Trochaic Law (ITL) is a default grouping strategy, where rhythmic alternations of duration are perceived iambically (weak-strong), while alternations of intensity are perceived trochaically (strong-weak). Some argue that the ITL is experience dependent. For instance, French speakers follow the ITL, but not as consistently as German speakers. We hypothesized that learning about prosodic patterns, like word stress, modulates this rhythmic grouping. We tested this idea by training French adults on a German-like stress contrast. Individuals who showed better phonological learning had more ITL-like grouping, particularly over duration cues. In a non-phonological condition, French adults were trained using identical stimuli, but they learned to attend to acoustic variation that was not linguistic. Here, no learning effects were observed. Results thus suggest that phonological learning can modulate low-level auditory grouping phenomena, but it is constrained by the ability of individuals to learn from short-term training.

Keywords: Psycholinguistics; Language learning; Rhythm; Auditory perception; Speech; Iambic–trochaic law

1. Introduction
The parsing of higher-level structure from variable perceptual input, or perceptual grouping, is central to the study of perception. In the auditory domain, one of the most recognized examples of perceptual grouping comes from Bolton (1894), who described a tendency to perceive acoustically isochronous sounds (the beats of a metronome) as
rhythmic sequences of two-element units (a series of tick-tocks). Early studies of rhythmic grouping noted that louder sounds are more often perceived in the initial position of this two-element unit, while longer sounds are more often perceived in the final position, a pattern widely viewed as a basic, low-level property of the auditory system (Bolton, 1894; Vos, 1977; Woodrow, 1909). Pitch is also a grouping cue that patterns with intensity: Higher-pitched sounds are more often perceived in initial position (Nespor et al., 2008).

The idea of a universal pattern in rhythmic grouping is important for many fields of cognitive science, but it has gained special notice in linguistics, and in the broader study of language learning. Known as the Iambic–Trochaic Law or ITL, the above pattern offers a domain-general explanation of why linguistic units with initial-prominence (trochees) are marked more often by intensity or pitch in the world’s languages, while linguistic units with final-prominence (iambs) are more often marked by duration (Hayes, 1995; Nespor et al., 2008). The ITL is also relevant to research on first-language acquisition. As infants often use prosodic cues (the intensity, pitch, and duration of speech) to segment linguistic units from a continuous stream of syllables (Hirsh-Pasek et al., 1987; Jusczyk, Houston & Newsome, 1999), the ITL might explain how universal aspects of rhythmic perception can bootstrap infants’ learning of the native language (Abboub, Nazzi, & Gervain, 2016; Bion, Benavides-Varela, & Nespor, 2011; Hay & Saffran, 2012; Hayes, 1995; Molnar, Lallier, & Carreiras, 2014).

Nevertheless, even if the ITL is a universal rhythmic grouping, it also appears to be—at least in part—experience-dependent. First, human infants and non-human animals show inconsistent grouping across age and across species. Second, the strength of ITL-like groupings is modulated by one’s native language.

1.1. Rhythmic grouping in infants and non-human animals

Several studies have found evidence of ITL perception in infancy, but the robustness of rhythmic grouping across duration, intensity, and pitch cues is highly variable. For example, some studies report that ITL-like grouping of duration is less consistent and emerges later in development than ITL-like grouping of either intensity or pitch. In those studies, English- and Italian-learning infants from 5–7 months of age showed trochaic (strong-weak) groupings of intensity or pitch alternation, but they did not show iambic (weak-strong) groupings of duration alternation until around 2–4 months later (Bion et al., 2011; Hay & Saffran, 2012; Yoshida et al., 2010). At this older age, these duration-based groupings were still less consistent than intensity-based groupings when tested in bilingual infants (Molnar et al., 2014). Similar differences are seen in rats, who show trochaic grouping of pitch alternation, but fail at iambic grouping of duration alternation (de la Mora, Nespor & Toro, 2013), unless explicitly trained (Toro & Nespor, 2015).

In a second group of studies, the opposite pattern is found: ITL-like grouping of duration alternation is observed, but grouping for intensity and/or pitch alternation is not. Trainor and Adams (2000) found this pattern in both English-learning 8-month olds and English-speaking adults using parallel age-appropriate variants of the same behavioral
paradigm. A somewhat similar pattern of results was found in a series of behavioral experiments in 7.5-month-old infants exposed to French and German, where grouping by pitch and duration was observed, but not grouping by intensity (Abboub, Boll-Avetisyan, Bhatara, Höhle, & Nazzi, 2016). These findings were echoed in a study examining newborns from French-speaking mothers, where increased brain activation was found when hearing consistent ITL-like groupings over inconsistent ones. However, this effect held only for duration, and not intensity, and mixed performance with pitch-based groupings was observed depending on language exposure in the womb (Abboub, Boll-Avetisyan, et al., 2016).

Together, these developmental and animal studies show wide variation in ITL-like groupings of duration, pitch, and intensity cues across age, language exposure, and even across research sites. If the ITL is subject to these task- or experience-dependent factors, this casts doubt on the idea that it could be immune to linguistic influences. For the study of language acquisition, this suggests that the use of the ITL as a bootstrapping cue could be more language-specific than previously thought. For example, one possibility might be that the ITL exists only as a loose developmental constraint, with some early language learners drawing upon certain cues (pitch, intensity, duration) that happen to pattern with the ITL in their native language, while other language learners acquiring a different language might find the application of an ITL-based constraint to be less helpful.

1.2. Cross-linguistic rhythmic grouping

Besides anecdotal accounts of cross-linguistic differences (Jakobson, Fant, & Halle, 1952), the first studies on cross-linguistic variation showed that English speakers follow the ITL, while Japanese speakers show (a) even stronger trochaic grouping of intensity alternation, and (b) no consistent grouping of duration alternation at the population level (Iversen, Patel, & Ogushi, 2008; Kusumoto & Moreton, 1997). A related study showed that iambic grouping of duration alternation is found in English-learning infants around 7–8 months of age, but not in Japanese-learning infants (Yoshida et al., 2010). Another cross-linguistic study that compared Betaza Zapotec- and English-speaking adults also reported divergent rhythmic grouping (Crowhurst & Teodocio Olivares, 2014). Together these studies have argued that such group differences come from the way that pitch, intensity, and duration are used in the phonologies of each language (prosodic variation), but at the same time, it has been difficult to rule out cultural differences in exposure to Western versus non-Western music.

Other studies have compared American English, French, and German speakers (Bhatara, Boll-Avetisyan, Agus, Höhle, & Nazzi, 2016; Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Hay & Diehl, 2007), groups that have relatively similar exposure to Western musical forms, but whose languages are phonologically distinct when considering prosody at the phrasal and lexical levels. These phonological differences appear to affect auditory grouping: Although French- and German-learning infants appear to perceive the ITL similarly (Abboub, Boll-Avetisyan, et al., 2016), adults do not. That is, French speakers’ rhythmic grouping is seen to be less ITL-like than German adult
speakers, at least when stimuli are sufficiently complex (Bhatara et al., 2013, 2016; Hay & Diehl, 2007).

These cross-linguistic studies add support to the idea that rhythmic cues are not processed the same by every population, and that language learning is a possible source of variance. Together with the above line of developmental and cross-species work, there is emerging consensus that perceptual grouping according to the ITL is dependent on experience hearing a language.

1.3. The current study

There are two principal (but non-mutually exclusive) hypotheses about how acquiring a particular phonology—French versus, for example, German—affects auditory grouping of rhythms (i.e., the ITL). One is that prosodic prominence at the phrasal level differs according to the dominant word order within a particular language (e.g., head-final vs. head-initial languages), which may contribute to cross-linguistic differences in the ITL between languages like head-final Japanese and head-initial English (Christophe, Guast, & Nespor, 1997; Iversen et al., 2008). Prosodic cues can thus act as a bootstrapping cue to discover syntactic combinations if a learner, particularly an infant learner, uses the ITL to parse fluent speech (Bion et al., 2011).

Another idea, which is the focus of the current study, is that prosodic prominence at the lexical level also differs across languages, which may in turn drive differences in ITL perception. This hypothesis comes from the above observation that ITL perception differs even between languages with similar canonical word orders, like French and German. The general idea is that speakers of German-like languages must constantly pay attention to prosodic prominence within words, which is not the case for speakers of French-like languages (Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Peperkamp, Vendelin, & Dupoux, 2010), which do not have a lexically encoded pattern of word-level stress.

Here, we investigated learning mechanisms that may contribute to cross-linguistic differences in rhythmic grouping, focusing on the second idea above. Specifically, we designed a training study to attempt to make French speakers become more “German-like” in their perception of rhythmic alternations. Below, we describe two experimental conditions that tested distinct hypotheses about how auditory training may affect rhythmic grouping (Table 1). In one condition, participants learned a “phonological” contrast that was equivalent to word stress in German. In another condition, participants were trained with identical stimuli, and learned to categorize and discriminate the acoustic dimensions of duration and intensity.

1.3.1. The Iam-Tro condition

The first training condition was called the iambic–trochaic (Iam–Tro) condition, which tested an argument made by Bhatara et al. (2013). Those authors attributed cross-linguistic differences in ITL perception to lexical stress—the German distinction between iambic
“Café (“coffee shop”) and trochaic *Kaffee* (“coffee”)—and the lack of such a distinction in French (who just drink their *café* in a *café*). French speakers exhibit relatively poor lexical stress perception (Abboub, Bijeljac-Babic, Serres, & Nazzi, 2015; Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012; Dupoux et al., 2001; Skoruppa et al., 2009) but—at least in limited cases—can be trained in the laboratory (Carpenter, 2015; Yeung & Nazzi, 2014).

This account suggests that the phonological use of lexical stress in German enhances the distinctiveness of the perceptual units or templates in a kind of *imprinting* or *unitization* process (Goldstone, 1998, 2000; Welham & Wills, 2011). Thus, German speakers might have more clearly differentiated iambic and trochaic templates in perception, which could result in a stronger perception of ITL-like groupings relative to French speakers, who might switch more easily between trochaic and iambic templates without a similar contrast in their native phonology. French speakers in the Iam–Tro learning condition were therefore trained on a “German-like” stress contrast over sets of disyllables, which was hypothesized to enhance the distinctiveness of two-element auditory templates that French listeners might not otherwise employ when listening to rhythmic alternations.

### 1.3.2. The Dur-Int condition

The *duration-intensity* (Dur-Int) learning condition used identical stimuli, but instead trained participants to differentiate duration from intensity variation. This tested a second possible way in which experience with a German-like phonology might result in increased ITL performance: German speakers might be relatively more sensitive to acoustic cues in rhythmic grouping tasks (a kind of increased *attentional weighting* of those cues; Goldstone, 1998), because pitch, intensity, and duration are more phonologically relevant in German versus French (see also Deutsch, Henthorn, Marvin, & Xu, 2006 for a similar argument about pitch in Mandarin vs. English). This account suggests that German speakers may be better at distinguishing acoustic variation in duration and intensity (and pitch) than French speakers, because of the respective linguistic structure of these languages. This might then carry over to ITL perception, giving an advantage to German speakers in a task that relies upon the same acoustic cues for rhythmic grouping.

Overall, the logic of the Dur-Int learning condition was thus to train French adults to better hear duration and intensity variation. In doing so, they might become more “German-like” in their attention to acoustic variation. This might then have enhanced rhythmic grouping using a different, non-phonological mechanism than in the Iam-Tro condition.

<table>
<thead>
<tr>
<th>Word Type</th>
<th>Category A</th>
<th>Category B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iam-Tro condition</td>
<td>Duration words</td>
<td>Short-Long</td>
</tr>
<tr>
<td></td>
<td>Intensity words</td>
<td>Soft-Loud</td>
</tr>
<tr>
<td>Dur-Int condition</td>
<td>Iambic words</td>
<td>Short-Long</td>
</tr>
<tr>
<td></td>
<td>Trochaic words</td>
<td>Long-Short</td>
</tr>
</tbody>
</table>

*Note.* Category assignment (bird or fish to Category A or Category B) was counterbalanced across participants.
1.3.3. Classification of training stimuli

Because the above literature had suggested that duration and intensity alternations are not processed equivalently in rhythmic grouping, we separately analyzed the two categories of disyllabic words used in phonological training. As shown in Table 1, we asked whether learning was different for *Duration words* (short-long, long-short) and *Intensity words* (soft-loud, loud-soft) in the Iam-Tro condition. In a complementary fashion, we examined sets of *Iambic words* (short-long, soft-loud) and *Trochaic words* (long-short, loud-soft) in the Dur-Int condition.

2. Method

In both conditions, the method was identical, except for the applied training manipulation. The experiment had three phases: First, there was a pre-test assessing rhythmic grouping, which was actually a shortened version of Experiment 1 from Bhatara et al. (2013). Only intensity and duration alternations were tested because previous cross-linguistic work had also focused on these cues (Hay & Diehl, 2007; Iversen et al., 2008), and because French speakers show more consistent performance with these cues than with pitch (Bhatara et al., 2013). In the second phase, we trained participants in either Iam-Tro or Dur-Int conditions. In the third and final phase, we administered a post-test identical to the pre-test.

2.1. Participants

In the Iam-Tro condition, 36 French adults were tested (8 males; $M_{\text{age}} = 22.4$ years, *Range* = 18–30 years). Two additional participants interrupted or refused to complete the experiment, finding training to be too frustrating. Fifteen participants from the final sample had no musical experience, and the remaining participants had a mean of 3.6 years of instrument and/or voice lessons (*Range* = 0–17 years). Foreign language education is compulsory in France, but all participants identified themselves as monolingual European French speakers (*Mean number of years learning a foreign language* = 10.6, *Range* = 5–18 years). In the Dur-Int learning condition, 36 participants were tested from the same population (4 males; $M_{\text{age}} = 22.9$ years, *Range* = 19–27 years; *Mean number of years learning a foreign language* = 11.6, *Range* = 6–21 years). Thirteen of the participants had no musical experience, and the remaining had an average of 4.1 years of music lessons (*Range* = 0–15 years). Welch two-sample *t* tests showed no differences between the two conditions in years of musical or foreign language experience [Music: $t(69) = -0.54$, *p* = .59; Language: $t(69) = -1.41$, *p* = .16].

Total sample size was determined by a power analysis (Faul, Erdfelder, Lang, & Buchner, 2007), where input parameters came from previous comparisons between French and German adults on a similar grouping task in Bhatara et al. (2013), including the average effect size from both experiments, and the overall correlation between dependent variables. An analysis with $\alpha = 0.05$ and power = 0.99 predicted a maximum total sample of
\( n = 76 \) (given conservative assumptions about sphericity). In consideration of this analysis, we ran a total sample size of \( n = 72 \), which was a slightly smaller number due to other constraints in our study’s counterbalancing design.

2.2. Pre- and post-test stimuli

The same rhythmically alternating 32-syllable sequences from Bhatara et al. (2013) were used. Each sequence was composed of 16 different CV syllables constructed from four vowels (/e:/, /i:/, /o:/, /u:/) and four consonants (/b/, /z/, /m/, /l/), using the German voice (De5) in MBROLA, a text-to-speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). All sequences were randomly created from these syllables with the following limitations: Each syllable occurred twice, once in the strong form (louder or longer) and once in the weak form (softer or shorter); no syllable duplications occurred; no consonant or vowel occurred in consecutive syllables more than twice; and no two-syllable sequence could be a word in French. All consecutive syllables were co-articulated, and no pauses were included between syllables.

A total of 54 sequences were created (6 control sequences with no alternation, 24 sequences with intensity alternation, and 24 sequences with duration alternation). The baseline pitch of each syllable was 200 Hz, the intensity was 70 dB, and the duration was 260 ms, of which the vowel was 160 ms. As in Bhatara et al. (2013), four levels of intensity-alternated sequences were used: While weak syllables were at baseline, strong syllables were either 2, 4, 6, or 8 dB above baseline. This intensity manipulation was applied over the whole syllable, and these intensity-varied sequences had a total duration of 8.32 s. Similarly, four levels of duration-alternated sequences were used such that strong syllables were either 50, 100, 150, or 200 ms longer than baseline in any given stream. This duration manipulation was applied only to the vowel of the strong syllables, as the vowel is the primary carrier of prosodic information in French. These sequences thus had a total duration of 9.12–11.52 s, depending on the level of duration alternation.

For all intensity- and duration-varied sequences, half began with a weak syllable, and half began with a strong syllable. The control sequences had no intensity or duration alternation, and pitch never varied in any of the 54 sequences. The first 3 s of all sequences were masked with white noise, which faded out following a raised-cosine intensity function, while the syllable sequence simultaneously faded in following the same function. Together, these measures prevented participants from using a grouping strategy that was based on the initial pair of sounds (Hay & Diehl, 2007).

2.3. Training stimuli

Participants in both conditions categorized a set of 32 disyllabic non-words. These non-words were generated from eight CVCV forms, created using the same constraints as the pre- and post-test stimuli: /bime/, /lezul/, /mibo/, /zeltu/, /boli/, /lumo/, /muzi/, and /zobe/. Four versions of each CVCV sequence were created (short-long, long-short, soft-
loud, loud-soft), using the largest level of intensity variation (8 dB) or the largest level of duration variation (200 ms).

2.4. Procedure

The experiment lasted approximately 40 min in total.

2.4.1. Pre-test

Participants listened through headphones (Sennheiser HG 558) on a MacBook Pro laptop running PsyScope X (http://psy.ck.sissa.it/). For each trial, participants were instructed to listen to one of the 54 sequences and report whether they heard a weak-strong or strong-weak grouping by pressing one of two buttons on the keyboard (see Bhatara et al., 2013, for further details about the instructions). After four practice trials (two with the maximal intensity variation and two with the maximal duration variation), participants were presented with the 54 sequences in random order. No feedback was given. This phase of the experiment lasted about 10–12 min.

2.4.2. Training task

Participants were told they would hear words from an alien language, and they had to decide if each word referred to a type of bird or a type of fish. For example, one participant in the Iam–Tro condition might have been taught that iambic words (soft-loud and short-long) referred to fish, whereas trochaic words (loud-soft and long-short) referred to birds. We hypothesized that this Iam-Tro training might strengthen the distinction between weak-strong and strong-weak perceptual templates analogous to learning a phonological word-stress distinction and, in doing so, perhaps boost ITL-like perception in the post-test.

In a complementary way, a participant in the Dur-Int condition might have been taught that duration-varying words (short-long and long-short) referred to fish, whereas intensity-varying words (loud-soft and soft-loud) referred to birds. We hypothesized that this training might strengthen attention to the acoustic dimensions of duration and intensity in a non-phonological way, possibly boosting sensitivity to ITL-like groupings in the post-test as well.

Participants were told they would be given feedback about their response accuracy with brief audio and visual feedback, and then would be shown the correct answer (a bird or a fish). Without any further instruction, participants heard one of the disyllabic training stimuli and pressed one of two buttons to indicate category membership. Each response button was covered with an image of a bird or a fish (the same images as in Fig. 1). If the response was correct, a chime was heard and a green checkmark appeared on the screen for 300 ms; if incorrect, a buzz was heard and a red “x” appeared on the screen for the same duration. Regardless of response accuracy on that trial, an image of the correct response was then displayed for 700 ms followed by a blank screen for 300 ms (Fig. 1).

All participants completed 8 blocks of 32 trials, for a total of 256 trials. Within a single block, trials were pseudo-randomly presented, sampled without replacement from the
full set of 32 disyllables. Stimulus presentation was identical across learning conditions, except for the type of categories to learn (see Table 1). The assignment of bird and fish categories was counterbalanced across participants. This phase of the experiment lasted about 15–20 min.

2.4.3. Post-test
A protocol identical to the pre-test was administered, lasting about 10–12 min.

3. Results

3.1. Pre- and post-tests

The ITL predicted more trochaic groupings for intensity alternation \((N = 24\) trials in each test) than for duration alternation \((N = 24\) trials in each test). An ANOVA on the proportion of trochaic responses was conducted with within-subjects factors of Test (pre or post) and Trial
Type (duration or intensity alternation), and a between-subjects factor of Condition (Iam-Tro or Dur-Int conditions). Results showed a main effect of Trial Type, $F(1, 70) = 9.17, p = .003, \eta^2_p = 0.12$, which reflected overall ITL-like grouping: more trochaic responses for intensity ($M = 0.58; SD = 0.22$) than for duration alternation ($M = 0.45; SD = 0.18$). There was also a marginal effect of Test, $F(1, 70) = 3.16, p = .080, \eta^2_p = 0.04$, reflecting a slight decline in trochaic responding from pre- ($M = 0.52; SD = 0.087$) to post-tests ($M = 0.50; SD = 0.11$). Means indicated that this was due to a decline specifically in duration alternation trials (Fig. 2), but the interaction between Test and Trial Type did not reach significance, $F(1, 70) = 1.62, p = .207$, nor did any other main effects or interactions (all $p$’s $>.648$). As shown in Fig. 2, a planned comparison within condition indicated that the global pattern of ITL-like grouping held true at pre- and post-tests within both Iam-Tro, $t(35) = 2.20, p = .034, d = 0.37$, and Dur-Int conditions, $t(35) = 2.09, p = .044, d = 0.35$.

3.1.1. No alternation (control) sequences

The ITL does not make clear predictions for control trials, and we had fewer trials using control sequences than Bhatara et al. (2013; $N = 6$ in each test), but these results were still reported and analysed for completeness (Fig. 2). An ANOVA on the proportion of trochaic responses in control sequences was conducted with a within-subjects factor of Test and a between-subjects factor of Condition. There were only marginal effects, including an interaction, $F(1, 70) = 3.11, p = .082, \eta^2_p = 0.043$, as well as a main effect of Test, $F(1, 70) = 3.11, p = .082, \eta^2_p = 0.043$. As shown in Fig. 2, a planned comparison confirmed that there was stable trochaic responding before and after Iam-Tro training (means were identical), $t(35) > .01, p > .99$, but that there was a significant drop in trochaic responding after Dur-Int training, $t(35) = 2.33, p = .026, d = 0.39$. This latter effect likely reflects subjects’ increased attention to the acoustic dimensions relevant to the grouping task. That is, learning to categorize duration variation separately from intensity variation likely led to less of a grouping bias for control sequences in the post-test, reflected in the chance (50%) responding seen when there was no rhythmic variation.

3.1.2. Interim summary

Results replicate previous reports of ITL-like grouping of speech rhythms in French adults, showing greater trochaic responses for intensity than duration alternation (Bhatara et al., 2013; Hay & Diehl, 2007). However, there was only a marginal effect of training, showing a slight decrease in overall trochaic groupings from pre- to post-test. In order to determine which aspects of our training manipulations were linked to ITL-like responding, and to account for learning effects during the training, we investigated the training portions of both conditions in the following section.

3.2. Learning in Iam-Tro and Dur-Int conditions

3.2.1. The Iam-Tro condition

An ANOVA was conducted on the proportion of correct categorizations with within-subjects factors of Block (8 blocks of 32 trials) and Word Type (Duration or Intensity words,
see Table 1). Results revealed significant main effects of Block, $F(7, 245) = 22.18$, $p < .001$, $\eta^2_p = 0.39$, and of Word Type, $F(1, 35) = 38.18$, $p < .001$, $\eta^2_p = 0.52$, but their interaction did not reach significance, $F(7, 245) = 1.55$, $p = .151$. As seen in Fig. 3, the main effect of Block reflected the fact that correct responses increased over time as participants learned the correct categorizations, while the main effect of Word Type suggested that participants had more difficulty with Iam-Tro learning for Intensity words, $M = 0.58$, $SD = 0.12$, compared to Duration words, $M = 0.72$, $SD = 0.14$. This is likely due to the relatively more important prosodic role for duration versus intensity in French (Delattre, 1966), which may have made it easier for participants to identify iambic–trophic differences using this cue.

The individual variability seen in Fig. 3—some participants showed no evidence at all of learning—suggested that learning performance should be taken into account in our analysis. In order to examine this factor, we used learning performance as a predictor of increased ITL-like rhythmic groupings on those trials with either duration or intensity alternation (excluding control trials). Thus, we coded the proportion increase in ITL-like grouping responses from pre- to post-test and measured this separately for duration
alternation (increases in iambic responding), intensity alternation (increases in trochaic responding), and their average (i.e., overall increase in ITL-like grouping). With these measures, we were able to generate Table 2, which shows the bivariate correlations between overall learning performance in each condition, and those three measures of increase in ITL-like grouping.

Fig. 3. The proportion of correct responses for the eight blocks of training in the Iam-Tro (top) and Dur-Int conditions (bottom). Light gray lines indicate overall performance for individual participants, and the black lines indicate the group averages for overall performance (black solid lines), as well as by each of the two word types in both Iam-Tro and Dur-Int conditions.
As seen in Table 2, participants had a significant positive correlation between learning in the Iam-Tro condition and increased ITL-like grouping of duration alternation, but no relation between Iam-Tro learning and increased ITL-like grouping of intensity alternation. When averaging these grouping scores, a marginal correlation was found. The data in Fig. 4 illustrate these individual correlations graphically and confirm that those participants who succeeded at learning the Iam-Tro contrast showed greater increases in ITL-like iambic groupings of duration alternation in the post-test.

We also examined Iam-Tro learning performance over the two word types used in the training task: Duration words (short-long vs. long-short) and Intensity words (soft-loud vs. loud-soft). Learning performance for each word type was positively correlated, $r (36) = .42, p = .011$, and so Table 3 displays two analyses. First, bivariate correlations are shown, measuring learning performance for each word type and the three measures of pre- to post-test increase in ITL-like groupings. Second, partial correlations are shown, which control for the variance related to learning of the other word type. This analysis reveals that it was primarily Iam-Tro learning over Duration words (distinguishing short-long vs. long-short words) that drove the association between training and increases in ITL-like iambic groupings of duration alternation in the post-test.

### Table 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>ITL-like Grouping</th>
<th>Correlation</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iam-Tro condition</td>
<td>Overall</td>
<td>0.300*</td>
<td>.076</td>
</tr>
<tr>
<td></td>
<td>Duration alternation</td>
<td><strong>0.406</strong></td>
<td>.014</td>
</tr>
<tr>
<td></td>
<td>Intensity alternation</td>
<td>0.053</td>
<td>.760</td>
</tr>
<tr>
<td>Dur-Int condition</td>
<td>Overall</td>
<td>−0.018</td>
<td>.916</td>
</tr>
<tr>
<td></td>
<td>Duration alternation</td>
<td>−0.012</td>
<td>.945</td>
</tr>
<tr>
<td></td>
<td>Intensity alternation</td>
<td>−0.020</td>
<td>.909</td>
</tr>
</tbody>
</table>

*Note. *$p < .10.$  
**$p < .05.$

As seen in Table 2, participants had a significant positive correlation between learning in the Iam-Tro condition and increased ITL-like grouping of duration alternation, but no relation between Iam-Tro learning and increased ITL-like grouping of intensity alternation. When averaging these grouping scores, a marginal correlation was found. The data in Fig. 4 illustrate these individual correlations graphically and confirm that those participants who succeeded at learning the Iam-Tro contrast showed greater increases in ITL-like iambic groupings of duration alternation in the post-test.

We also examined Iam-Tro learning performance over the two word types used in the training task: Duration words (short-long vs. long-short) and Intensity words (soft-loud vs. loud-soft). Learning performance for each word type was positively correlated, $r (36) = .42, p = .011$, and so Table 3 displays two analyses. First, bivariate correlations are shown, measuring learning performance for each word type and the three measures of pre- to post-test increase in ITL-like groupings. Second, partial correlations are shown, which control for the variance related to learning of the other word type. This analysis reveals that it was primarily Iam-Tro learning over Duration words (distinguishing short-long vs. long-short words) that drove the association between training and increases in ITL-like iambic groupings of duration alternation in the post-test.

### 3.2.2. The Dur-Int condition

An ANOVA was similarly conducted on the proportion of correct categorizations with within-subjects factors of Block and Word Type (Iambic or Trochaic words, see Table 1). Results revealed significant main effects of Block, $F(7, 245) = 20.00, p < .001, \eta_p^2 = 0.36$, and of Word Type, $F(1, 35) = 12.50, p < .001, \eta_p^2 = 0.26$, and their interaction did not reach significance, $F(7, 245) = 0.99, p = .442$. The main effect of Block confirmed that the proportion of correct responses increased over time, while the main effect of Word Type suggested that participants had more difficulty discriminating between duration and intensity variation for iambic items, $M = 0.76, SD = 0.18$, than for trochaic items, $M = 0.80, SD = 0.17$. This latter effect may be related to the dominance of iambic prosody in French (Delattre, 1966), which may make the more typical iambic form less salient to French listeners.
Correlational analyses in Table 2 (and Fig. 4) revealed no relation between Dur-Int learning and increases in ITL-like grouping (for any type of rhythmic alternation). Training performance for Iambic and Trochaic word types was significantly correlated, $r (36) = .90$, $p < .001$. However, as shown in Table 3, there was no relation between learning for either word type in this condition and any increases in ITL-like grouping (all $p$’s $> .197$).
4. Discussion

This study examines how perceptual grouping, specifically the ITL, changes with linguistic experience. Here, we built upon previous work (Bhatara et al., 2013, 2016), which had suggested that French speakers have fewer ITL-like grouping strategies for auditory rhythms, compared to German speakers. We trained French native speakers in two conditions, which were aimed at pinpointing ways in which linguistic experience may influence perceptual grouping. One hypothesis (the Iam-Tro condition) was that training French speakers on a “German-like” stress contrast would lead to more contrastive perceptual templates for iambs and trochees (see also Goldstone, 2000; Welham & Wills, 2011), which could then affect rhythmic grouping. Another hypothesis was that French speakers, compared to German speakers, might not be as accustomed to attending to the acoustic variation used in rhythmic tasks. We therefore exposed French speakers to the same stimuli, but we trained them to process acoustic intensity versus duration variation (in the Dur-Int condition).

Group results suggested that rhythmic grouping at the population level was not greatly affected by either type of training, except in one case in the Dur-Int condition, where the identification of no variation (control) trials was more accurately rated as having no consistent grouping pattern. However, participants also showed wide inter-individual differences in learning during our training manipulation. Our main result comes from an analyses of this individual variability, which showed that successful learning in the Iam-Tro condition at the individual level was strongly associated with individual increases in ITL-like rhythmic groupings, particularly of duration alternation. Further analyses revealed that it was specifically the individual learning of duration words—the

<table>
<thead>
<tr>
<th>Condition: Word Type</th>
<th>ITL-like Groupings</th>
<th>Bivariate</th>
<th>Partial</th>
<th>p</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iam-Tro condition: Duration words</td>
<td>Overall</td>
<td>0.434**</td>
<td>.008</td>
<td>0.465**</td>
<td>.005</td>
</tr>
<tr>
<td>Iam-Tro condition: Intensity words</td>
<td>Duration alternation</td>
<td>0.475**</td>
<td>.003</td>
<td>0.448**</td>
<td>.007</td>
</tr>
<tr>
<td></td>
<td>Intensity alternation</td>
<td>0.202</td>
<td>.238</td>
<td>0.294*</td>
<td>.086</td>
</tr>
<tr>
<td>Dur-Int condition: Iambic words</td>
<td>Overall</td>
<td>−0.009</td>
<td>.958</td>
<td>0.034</td>
<td>.844</td>
</tr>
<tr>
<td></td>
<td>Duration alternation</td>
<td>−.052</td>
<td>.764</td>
<td>−0.190</td>
<td>.275</td>
</tr>
<tr>
<td></td>
<td>Intensity alternation</td>
<td>.029</td>
<td>.869</td>
<td>0.215</td>
<td>.216</td>
</tr>
<tr>
<td>Dur-Int condition: Trochaic words</td>
<td>Overall</td>
<td>−.026</td>
<td>.878</td>
<td>−0.042</td>
<td>.809</td>
</tr>
<tr>
<td></td>
<td>Duration alternation</td>
<td>.032</td>
<td>.852</td>
<td>0.185</td>
<td>.286</td>
</tr>
<tr>
<td></td>
<td>Intensity alternation</td>
<td>−.070</td>
<td>.687</td>
<td>−0.223</td>
<td>.197</td>
</tr>
</tbody>
</table>

Note. *p < .10. **p < .01.
categorization of short-long pairs as iambic, and long-short pairs as trochaic—that drove this effect. Those French speakers who became better at distinguishing trochees and iambs defined by duration variation also had a tendency to perceive more ITL-like rhythmic grouping of duration alternation. In contrast, no relation between individual performance in learning and grouping was observed for intensity cues.

In addition to further analyses in the Supporting Information for this article, these findings globally suggest that duration-based learning in the Iam-Tro condition (at least for those participants who seemed to learn in our training paradigm) is selectively associated with a subsequent increase in ITL-like groupings of duration alternation. This supports the first hypothesis outlined in the Introduction, which was that learning a phonological iambic-trochaic contrast (at least over duration cues) leads to an increase in the stability of contrastive templates in rhythmic perception (at least of duration), and thus modulates the perceptual grouping of rhythmic alternation (of duration). Two further points relevant to this conclusion merit further discussion: First, to what degree are our results explained by phonological learning versus natural, pre-existing variance in perceptual ability? Second, why was learning only observed for duration-based contrasts, rather than both duration- and intensity-based ones?

4.1. Phonological learning and population variance in perceptual ability

Our results illustrate an interaction between phonological learning and population variance in perceptual learning ability, as grouping in the Iam-Tro condition (at least for duration words) was possible for only a subset of individuals. It thus cannot be said that phonological learning was possible for everyone tested in our sample. Only those who were already sensitive auditory perceivers (likely before entering our study) were able to learn in this training condition. Nevertheless, it should be clear from a comparison of the two training conditions—as participants were randomly assigned—that the type of training received was a crucial factor in showing greater ITL-like perception from pre- to post-test. In other words, we likely captured substantial population variance in auditory perception by randomly assigning participants to the Dur-Int training condition, yet no correlation with greater ITL-responding was observed there.

Together, these results suggest that our current version of perceptual training, which was limited to only 15–20 min—the same length as a previously reported stress-training paradigm with adult French speakers (Carpenter, 2015)—is just a proof-of-concept that illustrates the potential of phonological learning to change auditory perception of the ITL. Future work may consider more comprehensive, longer, and in-depth training regimens, which may reduce the effects of population variance, as has been done in other prosodic training studies, like ones that train non-tone language speakers to perceive pitch contours. These studies involved 2–3 times more training per day, and often over several days (Wang, Spence, Jongman, & Sereno, 1999; Wayland & Li, 2008). Future stress training studies may thus consider manipulating the degree of training to better understand the interactive effects of pre-existing perceptual sensitivity and phonological learning.
4.2. Duration-based phonological learning, but not intensity-based learning

There are two general possibilities about why the effect in the Iam-Tro condition was found for duration groupings, but not for intensity groupings. The first is that the grouping of duration alternation may be inherently more learnable than the grouping of intensity alternation. As discussed in the Introduction, some have suggested that ITL-like groupings emerge earlier for intensity than duration variation in infancy (Bion et al., 2011; Hay & Saffran, 2012; Yoshida et al., 2010), while adult studies have also shown greater cross-linguistic differences for rhythmic grouping of duration alternation than intensity or pitch alternation (Crowhurst & Teodocio Olivares, 2014; Iversen et al., 2008; Molnar et al., 2014). In addition, perceptual learning of ITL-like groupings is possible in rats only for duration alternation (Toro & Nesp, 2015). These studies support the idea that grouping by duration could be more subject to experience-driven change than grouping by either intensity or pitch.

Other work casts doubt upon this idea, however, showing that the coupling between duration-based ITL perception and experience is not as clear as the above literature may claim (Abboub, Boll-Avetisyan, et al., 2016; Trainor & Adams, 2000). An alternative possibility is that observed learnability of duration-based patterns could be explained by the relative prominence of those cues in French (Delattre, 1966). Future studies will need to tease apart these hypotheses, first by explaining the diverse results seen throughout the literature in the grouping of duration, intensity, and pitch alternation across age and linguistic background, and second, by training rhythmic grouping in speakers of other languages.

This latter point is also related to the question of bilingualism: Consider French speakers whose second language has lexical stress (e.g., English, German, Spanish). Previous work has shown that bilinguals, even the subset of highly proficient ones, are unable to robustly encode contrastive stress in their phonological representations (Dupoux, Peperkamp, & Sebastián-Gallés, 2010; Dupoux et al., 2008). Our results demonstrate a case where limited encoding of iambic and trochaic units is possible under certain conditions: for example, when categories vary in only a single acoustic cue or when a training task is used. A recent report also shows that certain aspects of foreign language learning (specifically, French speakers’ learning of oral German) and musical training are correlated with increased ITL-like grouping (Boll-Avetisyan et al., 2016). Together with our study, this underlines the potential applications for second-language learning that may follow from understanding the perceptual consequences of phonological learning.

5. Conclusions

The current work contributes to an emerging literature that is asking how experience with specific language input shapes perceptual processes. Recent studies have suggested that various non-linguistic auditory processes are shaped by learning a particular language
(Bhatara, Yeung & Nazzi, 2015; Bhatara et al., 2013; Iversen et al., 2008; Peng et al., 2010; Salmelin et al., 1999), but it remained unclear how language experience interacts with more fundamental properties of the auditory system to produce certain patterns in perception. Here, we detail a specific case of phonologically driven changes in perception.

The particular case that we describe here shows how individual variability in laboratory training of an iambic–trochaic contrast (using duration cues) appears to increase individual tendencies to perceive ITL-like groupings (of duration alternation). These findings are important because the ITL has long been assumed to play an important role in guiding phonological acquisition. Our results show the reverse pattern: Brief exposure to a new phonological pattern does induce learning in some individuals, and for these people, this training results in modifications to the perception of the ITL.

Acknowledgments

The first two authors contributed equally to this work. This research was supported by a grant from an Agence Nationale de la Recherche (ANR) grant no. 09-FASHS-018 to T. Nazzi. We thank Angelica Buerkin-Salgado and Sylvie Margules for help running participants.

References


Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Data S1. Further analyses confirming that individual performance in training was the source of variation in ITL-like perception.