Brain & Language 153-154 (2016) 13–19

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Short Communication

Speech rhythm sensitivity and musical aptitude: ERPs and individual differences

Cyrille Magne [a,*], Deanna K. Jordan [a], Reyna L. Gordon [b,c]

[a] Psychology Department, Middle Tennessee State University, United States
[b] Department of Otolaryngology, Vanderbilt University Medical Center, United States
[c] Vanderbilt Kennedy Center, United States

ARTICLE INFO

Article history:
Received 20 February 2015
Revised 13 December 2015
Accepted 21 January 2016

Keywords:
ERPs
Speech meter
Expectancy
Musical aptitude
Music

ABSTRACT

This study investigated the electrophysiological markers of rhythmic expectancy during speech perception. In addition, given the large literature showing overlaps between cognitive and neural resources recruited for language and music, we considered a relation between musical aptitude and individual differences in speech rhythm sensitivity. Twenty adults were administered a standardized assessment of musical aptitude, and EEG was recorded as participants listened to sequences of four bisyllabic words for which the stress pattern of the final word either matched or mismatched the stress pattern of the preceding words. Words with unexpected stress patterns elicited an increased fronto-central mid-latency negativity. In addition, rhythm aptitude significantly correlated with the size of the negative effect elicited by unexpected iambic words, the least common type of stress pattern in English. The present results suggest shared neurocognitive resources for speech rhythm and musical rhythm.

1. Introduction

Sensitivity to speech meter (i.e., recurring patterns of stressed and unstressed syllables) and rhythm (i.e., temporal organization of metrical structure) plays an important role in language acquisition (Jusczyk, 1999), speech segmentation (Mattys & Samuel, 1997), lexical access (Dilley & McAuley, 2008; Magne et al., 2007), and syntactic parsing (Schmidt-Kassow & Kotz, 2008). Listeners do not pay equal attention to all parts of the speech stream. Patterns of speech rhythm seem to influence how specific moments in the speech signal are attended at different hierarchical levels. Dynamic attending theory provides a framework in which auditory rhythms in music or speech are thought to create hierarchical expectancies for the signal as it unfolds over time (Jones & Boltz, 1989; Large & Jones, 1999). Prior work suggests that rhythmic expectations (Pitt & Samuel, 1990), acoustic cues (Kochanski & Orphanidou, 2008), and knowledge about predominant stress patterns (Jusczyk, 1999) all play a role in yielding the perception of some syllables as more prominent than others. These levels of prominence form hierarchies of attention and speech rhythm perception that give rise to meter (Kotz & Schwartze, 2010; Port, 2003; Rothermich, Schmidt-Kassow, & Kotz, 2012).

Listeners' entrainment to rhythmic regularities in the speech signal may allow these fluctuations in temporal attention to scaffold auditory input and create expectations for syllables and words (Port, 2003). There is mounting evidence in favor of rhythmic regularities in speech (Henrich, Alter, Wiese, & Domahs, 2014), despite lesser physical periodicity when compared to the temporal structure of music (e.g., Patel, 2008). To this point, recent investigations suggest that both temporal regularity of events (strong syllables) and metrical (i.e., stress pattern) regularity may contribute to guiding attention during speech perception. For instance, Quené and Port (2005) asked participants to detect phonemes in sequences of words that were either metrically regular or irregular and presented with either a constant or random inter-stress interval. Phoneme detection was better for sequences with a constant inter-stress interval regardless of the metrical regularity, suggesting that temporal expectancy was the primary factor guiding listeners' attention to specific portions of the speech signal. In contrast, recent event-related potential (ERP) studies showed that processing of syntactic incongruities (Schmidt-Kassow & Kotz, 2009a, 2009b) and syntactic ambiguities (Roncaglia-Denissen, Schmidt-Kassow, & Kotz, 2013), as well as semantic incongruities (Rothermich et al., 2012) is facilitated in sentences with regular metrical contexts, even when the inter-stress interval is not consistent (e.g., Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2009b). Thus, while it remains possible that temporal expectancies play a role in language, these later findings suggest that perceptual regu-

---

* Corresponding author at: Psychology Department, Middle Tennessee State University, 1301 E. Main Street, Murfreesboro, TN 37132, United States.
E-mail address: Cyrille.Magne@mtsu.edu (C. Magne).

http://dx.doi.org/10.1016/j.bandl.2016.01.001
0093-934X/

larities can arise from the abstract metrical structure of the signal even in absence of physical (i.e., temporal) regularities (Schmidt-Kassow & Kotz, 2009b).

Recent studies have used ERPs to shed light on the neural basis of rhythmic and metric components of speech by studying the electrophysiological markers of rhythmic/metrical structure violations (e.g., Domahs, Wiese, Bornkessel-Schlesewsky, & Schlesewsky, 2008; Magne et al., 2007; Marie, Magne, & Besson, 2011; McCauley, Hestvik, & Vogel, 2012; Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2009a), words with correct but unexpected rhythmic/metrical patterns (Bohn, Knaus, Wiese, & Domahs, 2013; Böcker, Bastiaansen, Vroomen, Brunia, & de Gelder, 1999) or pseudowords with unexpected stress patterns (Rothermich, Schmidt-Kassow, Schwartze, & Kotz, 2010). An increased negativity, sometimes followed by a late positivity, is generally observed in response to rhythmically/metrically incongruous or unexpected words. The late positivity usually occurs between 500 and 900 ms over centro-parietal regions (Bohn et al., 2013; Domahs et al., 2008; Magne et al., 2007; Marie et al., 2011; McCauley et al., 2012; Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2009a). Because the effects are present only when the task explicitly directs participants' attention to the rhythmic/prosodic aspects of the stimuli (e.g., Magne et al., 2007; Rothermich et al., 2012), it has been proposed to reflect task-relevant processes (e.g., Domahs et al., 2008; Magne et al., 2007).

In contrast, the negative effect usually occurs within the first 400 ms post stimulus onset (Bohn et al., 2013; Böcker et al., 1999; Magne et al., 2007; Marie et al., 2011; McCauley et al., 2012; Rothermich et al., 2010, 2012; Schmidt-Kassow & Kotz, 2009a), though it has also been observed in later latency windows up to 1000 ms (Bohn et al., 2013; Domahs et al., 2008; McCauley et al., 2012). In addition, the scalp topography of this negative effect shows a bilateral distribution in most of the studies (Domahs et al., 2008; Magne et al., 2007, semantic task; Marie et al., 2011; Rothermich et al., 2010, 2012; Schmidt-Kassow & Kotz, 2009a), but is sometimes left-lateralized (Bohn et al., 2013; Böcker et al., 1999; McCauley et al., 2012) or right-lateralized (Magne et al., 2007, prosodic task; McCauley et al., 2012). Finally, the negativity occurs independently of the task demands in some studies (e.g., Magne et al., 2007; Marie et al., 2011; Rothermich et al., 2010; Schmidt-Kassow & Kotz, 2009a) while it was present only when participants are instructed to attend the rhythmic/metrical structure in others (Böcker et al., 1999; Rothermich et al., 2012).

It has been proposed that this negative effect represents a contingent negative variation (i.e., CNV) in response to an unstressed syllable for which stress was expected (Domahs et al., 2008; McCauley et al., 2012), an increased N400 component classically associated with lexico-semantic processing (Bohn et al., 2013; Domahs et al., 2008; Magne et al., 2007; McCauley et al., 2012), or a subcomponent of the left anterior negativity (i.e., LAN) reflecting a non-language-specific rule-based error-detection mechanism (Marie et al., 2011; Rothermich et al., 2010, 2012; Schmidt-Kassow & Kotz, 2009b). Additional support for this latter interpretation comes from ERP studies reporting early negativities in response to metric deviations in tone sequences (e.g., Brochard, Abecasis, Potter, Ragot, & Drake, 2003).

While the question remains open regarding exactly which cognitive processes are reflected in this negativity, it is important to note, however, that these interpretations are not necessarily mutually exclusive given that the aforementioned studies vary in terms of language, and task demands. For instance, most studies were conducted in languages with variable stress such as German (Bohn et al., 2013; Domahs et al., 2008; Rothermich et al., 2010, 2012; Schmidt-Kassow & Kotz, 2009b), Dutch (Böcker et al., 1999), and English (McCauley et al., 2012) whereas two used French (Magne et al., 2007; Marie et al., 2011), which has a fixed stress pattern. In addition, some studies only used an explicit task focused on the prosody (Bohn et al., 2013; Domahs et al., 2008) or pronunciation (McCauley et al., 2012) of the stimuli while others directly compared the effect of attentional task demand (explicit vs implicit) on the processing of rhythmically incongruous/unexpected words (Böcker et al., 1999; Magne et al., 2007; Marie et al., 2011; Rothermich et al., 2010, 2012; Schmidt-Kassow & Kotz, 2009b). Finally, the heterogeneity of the observed ERP effects could be due to difference in syllabic complexity of the stimuli used across the experiments. For instance, Domahs et al. (2008) directly examined the interplay between syllable structure and meter in German trisyllabic words with correct initial stress. In particular, metrical incongruities in words with a final closed syllable were compared to metrical incongruities in words with a final open syllable. Their results revealed that the ERP effects depended both on the location of the incorrect stress (second vs final syllable) and on the structure of the final syllable (open vs closed).

Musical expertise (acquired through formal music training) and musical aptitude (i.e., the "potential to achieve in music"; Gordon, 1989) have been associated with enhanced language skills (e.g., Besson, Schön, Moreno, Santos, & Magne, 2007; Milovanov & Tervaniemi, 2011; Schellenberg, 2005). For instance, individuals with more than four years of continuous formal musical training showed enhanced detection of sentence-final intonation contour violations (Magne, Schön, & Besson, 2006; Schön, Magne, & Besson, 2004) and words with incongruous stress patterns (e.g., Marie et al., 2011), as well as enhanced categorical perception of lexical tones in Chinese (Wu et al., 2015). Music training has also been linked to enhanced reading skills (Moreno et al., 2009; Rautenberg, 2013). In addition, there is evidence in favor of a causal influence of music training on language skills, as suggested by longitudinal studies with children randomly assigned to music instruction or control activity (e.g., art instruction) and using a pre-training vs post-training comparison procedure to examine the impact of music training on speech perception and language outcomes (Chobert, François, Velay, & Besson, 2012; Degé & Schwarzer, 2011; François, Chobert, Besson, & Schön, 2013; Kraus et al., 2014; Moreno et al., 2009).

Recent findings suggest that such relations between musical aptitude and language skills exist even in non-musicians (i.e., with less than two years of formal music education). Higher levels of musical aptitude were associated with superior phonological awareness (Moritz, Yampolsky, Papadelis, Thomson, & Wolf, 2013; Peynircioğlu, Durgunoğlu, & Öney-Küsefoğlu, 2002) and reading skills (Anvari, Trainor, Woodside, & Levy, 2002; Strait, Hornickel, & Kraus, 2011) in children. Musical rhythm perception abilities were also associated with expressive grammar skills in children (Gordon et al., 2015). In addition, musical aptitude correlates with second language learning proficiency. Slevc and Miyake (2006) found that musical aptitude was strongly correlated with both productive and receptive phonology in Japanese immigrants. Similarly, Finnish children and adults with higher musical aptitude exhibited more accurate reproductions of English phonemes for which there are no direct Finnish equivalents (Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008; Milovanov, Pietilä, Tervaniemi, & Esquef, 2010).

In sum, musical skills and aptitude appear to be important sources of variance, and should arguably be taken into account when studying language skills. Moreover, the potential benefit of musical aptitude for language skills may come from the many shared anatomical and functional bases between the two domains (e.g., Patel, 2008). In particular, several findings favor the domain-generality of rhythm processing (Gordon, Magne, & Large, 2011; Hausen, Torppa, Salmela, Vainio, & Sarkamo, 2013; Peter, McArthur, & Thompson, 2012).

In the present study, speech rhythm sensitivity was examined by recording EEG while participants were presented with sequences of four bisyllabic words. The first three words were either all stressed on the first syllable (i.e., trochaic) or all stressed on the second syllable (i.e., iambic). In addition, the fourth word either had the same (i.e., expected) or opposite (i.e., unexpected) stress pattern as the previous three words. Thus, four types of word sequences were created by manipulating both the stress pattern and metrical expectancy (see Table 1 for examples in each experimental condition). Based on previous work on speech rhythm, critical words with an unexpected stress pattern were predicted to elicit an increased negativity between 200 and 600 ms (e.g., Domahs et al., 2008; Magne et al., 2007; Rothermich et al., 2010). We also aimed to study how individual differences in musical aptitude predict variance in speech rhythm sensitivity. Music perception skills were tested with the Advanced Measures of Music Audiation (AMMA; Gordon, 1989), which yields a rhythm aptitude score and a tonal aptitude score. The size of the negative ERP effect was expected to positively correlate with the degree of music aptitude, such that individuals who performed better on the musical aptitude test would tend to have larger negative ERP responses to words with unexpected stress patterns. Finally, because trochaic words are more common than iambic words in English (85–90% of spoken English words are trochaic, according to Cutler & Carter, 1987), data were analyzed separately for the two types of stress patterns.

2. Results

2.1. Metrical expectancy

Words with an unexpected trochaic stress pattern elicited an increased negativity that was significant from 288 to 576 ms over a centro-frontal cluster of electrodes ($p < 0.001$, see Fig. 1). Words with an unexpected iambic stress pattern also elicited an increased negativity over a centro-frontal cluster of electrodes, but in a later time window from 398 to 594 ms ($p < 0.001$, see Fig. 1).

To compare the onset latency of the negative effects for unexpected trochaic and iambic words, difference waves were computed separately for trochaic and iambic words by subtracting metrically expected critical words from metrically unexpected words. The resulting difference waves were then analyzed using the same cluster-based permutation procedure as described in the Methods section. Results revealed significant differences between difference waves corresponding to the expectancy effect on trochaic words and the effect on iambic words, between 298 and 338 ms post-word onset over centro-parietal regions of the scalp ($p < 0.028$), thus suggesting that the negative effect to unexpected trochaic words started 40 ms earlier.

2.2. Musical aptitude

Participants had a mean AMMA tonal score of 25.6 (SD = 4.3), and a mean AMMA rhythm score of 27.9 (4.5). The mean percentile score was 54.7th percentile (SD = 20.1), putting these participants in line with average American students not majoring in Music (the AMMA provides separate norms for Music majors and non-Music-majors). A significant moderate correlation was found between the size of the negative effect elicited by unexpected iambic words and the rhythm score on the musical aptitude test ($r = -0.51$, $p = 0.022$), suggesting that the higher the musical rhythm aptitude, the larger the negativity elicited in response to unexpected iambic words (see Fig. 2). The correlation between the negativity and the tonal score trended toward significance ($r = -0.43$, $p = 0.056$). In contrast, the negative effect elicited by unexpected trochaic words did not significantly correlate with either the rhythm score ($r = -0.12$, $p = 0.607$) or the tonal score ($r = 0.19$, $p = 0.418$). The maximum Cook's distance for the reported correlations indicated no undue influence (i.e. max Cook's $d < 1.00$).

Table 1
Examples of stimuli in each experimental condition.

Condition	Word 1	Word 2	Word 3	Target word
Expected trochaic	Zebra	Bacon	Easter	Pedal
Expected iambic	Morale	Embrace	Delight	Caffeine
Unexpected trochaic	Morale	Embrace	Delight	Pedal
Unexpected iambic	Zebra	Bacon	Easter	Caffeine

3. Discussion

The present study investigated the electrophysiological correlates of the detection of metrical expectancy violations in spoken English and examined how individual differences in musical aptitude account for speech rhythm sensitivity. In line with previous studies in German (Bohn et al., 2013; Domahs et al., 2008; Rothermich et al., 2010, 2012; Schmidt-Kassow & Kotz, 2009b), French (Magne et al., 2007; Marie et al., 2011) and Dutch (Böcker et al., 1999), English words with an unexpected stress pattern elicited an increased negativity over the centro-frontal region of the scalp. Given that the task requirements did not explicitly direct attention toward the rhythmic aspects of the stimuli, this finding clearly suggests that metrical expectancy can be automatically generated during speech perception. In addition, this effect occurred despite using a variable ISI between successive words, thus supporting the idea that metrical expectancies can be generated in speech even in absence of strict temporal regularity (Schmidt-Kassow & Kotz, 2009b). These expectancies are continually built up during speech perception: even distal (non-local) prosodic patterns can influence word segmentation of subsequently occurring lexically ambiguous syllable sequences (Dilley & McAuley, 2008).

Previous studies have interpreted similar negative ERP effects as reflecting either a general error detection response to unexpected occurrences within sequences (e.g., Rothermich et al., 2010), or an N400 associated with increased semantic processing difficulty (Magne et al., 2007; Marie et al., 2011). In the present study, it is unlikely that the observed negativity was a semantic N400 component, mainly for two reasons. First, the semantic relatedness between successive words and their lexical frequency was controlled for with each word sequence. Second, the latency range of this negative effect was modulated by the type of unexpected stress pattern, starting earlier for trochaic words than for iambic words (298 ms vs 338 ms). Since trochaic words are stressed on the first syllable, these findings suggest the acoustic salience of the syllable directly affected the latency range of this negativity. Note that these latency differences could also be due to differences in syllabic complexity between trochaic and iambic words in English. The influence of phonological factors has indeed been largely documented in the ERP literature. For instance, Praamstra, Meyer, and Levelt (1994) found that unrelated word pairs elicited an increased negativity that was early (between 250 and 450 ms) when compared to alliterating word pairs, but later (between 450 and 700 ms) when compared to rhyming word pairs. In addition, Domahs et al. (2008) found that the ERP effects elicited by violations of stress patterns varied in function of both the location of the incorrect stress and the syllabic structure. While a similar design has not yet been carried out in English, several standard descriptions of the English language indeed propose that stress

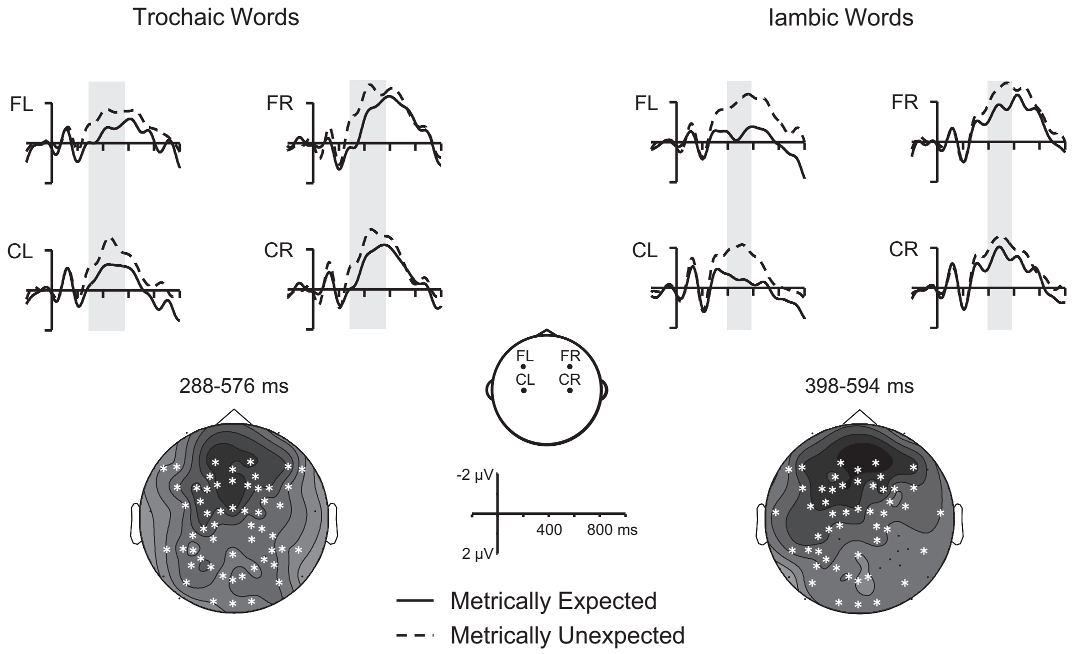

Fig. 1. Grand-average ERPs recorded for trochaic (Left) and iambic (Right) critical words. (Top panel) Averaged waveforms for metrically expected words (solid line) and metrically unexpected words (dashed line) at 4 selected electrodes (FL = Frontal Left, FR = Frontal Right, CL = Central Left, CR = Central Right). The latency range of the significant clusters are indicated in gray. Negative amplitude values are plotted upward. (Bottom panel) Topographic maps showing mean differences in scalp amplitudes in the latency range of the significant clusters. Electrodes belonging to the cluster are indicated with an asterisk.

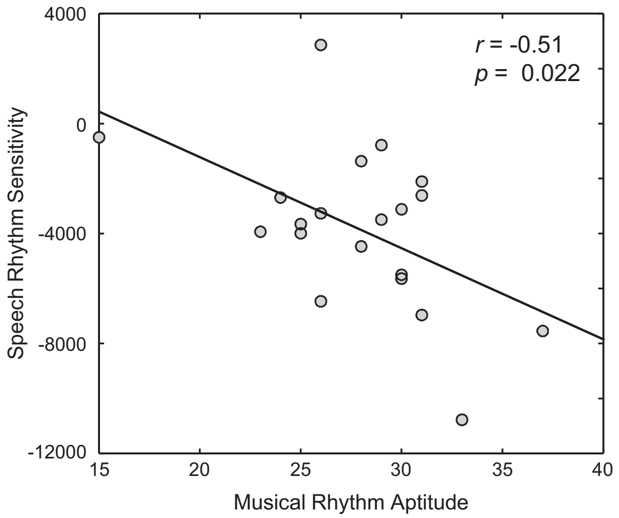

Fig. 2. Correlation between Musical rhythm aptitude (i.e., AMMA rhythm raw scores) and speech rhythm sensitivity (as indexed by the negative cluster sum for unexpected iambic words; lower values thus indicate larger ERPs and better speech rhythm sensitivity). The solid line represents a linear fit.

assignment is dependent on the word syllabic structure (Halle & Vergnaud, 1987; Hayes, 1995). For instance, a word will often be stressed on the final syllable if it has a long vowel, but will be stressed on the penultimate syllable if the latter is heavy (Duanmu, Kim, & Stiennon, 2005). Taken together, we thus favor the interpretation that the observed negativity reflected an error detection mechanism when the stress pattern of the critical word did not meet the expectation automatically set by the regular context of the preceding word sequence.

Participants' scores on the rhythm subtest of the musical aptitude test significantly correlated with the size of the brain response elicited by unexpected iambic stress pattern, converging with behavioral findings of robust associations between musical rhythm skills and speech prosody perception studied in adults (Hausen et al., 2013). In a previous study in French, Marie et al. (2011) also found enhanced neural sensitivity to speech rhythm violations in professional musicians. The present results extend this line of research by showing a direct relationship between the levels of rhythmic aptitude in music and speech rhythm sensitivity, in individuals who do not have professional musical experience. Other work on non-musician children has shown that individual differences in musical rhythm are also associated with phonological awareness (Moritz et al., 2013), reading (Strait et al., 2011), and grammatical competence (Gordon et al., 2015). In light of the literature showing sensitivity to the rhythmic aspect of native speech in infants (e.g., Nazzi & Ramus, 2003), and that reduced stress pattern discrimination in infants is associated with later language impairment (Weber, Hahne, Friedrich, & Friederici, 2005), early domain-general rhythm abilities may play a key role in language acquisition (Brandt, Gebrian, & Slevc, 2012; Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014). It is important to note that the literature on training-driven plasticity of musical abilities suggests that many musical skills are a subset of a larger category of auditory skills (e.g., Hyde et al., 2009; Shahin, Bosnyak, Trainor, & Roberts, 2003; Tallal & Gaab, 2006). Thus, the effects observed in the present study could reflect either a general auditory advantage during language acquisition, or a more specific advantage of enhanced rhythm processing.

The finding that an association between musical aptitude and speech rhythm sensitivity was found for unexpected iambic words, but not trochaic words, may result from the high frequency of the trochaic pattern in English. Interestingly, Jusczyk (1999) reported that at about 7.5 months of age, infants are only able to segment trochaic words, and are unable to segment the less common iambic words until 10.5 months. Thus this difference in sensitivity between the two stress patterns appears very early during language acquisition. As the present study used adult participants, in future studies it would be interesting to examine whether correlations between musical aptitude and sensitivity to the trochaic pattern exist earlier in life. Segmentation of iambic words appears to rely more on phonotactic and allophonic rules (Jusczyk, 1999), and individuals with music aptitude may possibly be able to more efficiently recruit neural resources during the perception of rhythmic acoustic cues in order to bootstrap segmentation.

To summarize, the present study extends the previous literature by showing that rhythmic expectancies can be automatically generated in spoken English, and that individual differences in music rhythm abilities can account for levels of speech rhythm sensitivity. More broadly, these findings are also consistent with the literature showing overlapping cognitive and neural resources between language and music (Patel, 2008) as well as the positive influence of musical training on language abilities (Kraus et al., 2014; Milovanov & Tervaniemi, 2011). The overlap between musical rhythm and speech rhythm skills may result from the existence of a shared subcortico-cortical network that is crucial for timing and beat perception, not only in music but also in language (Kotz & Schwartze, 2010). Interestingly, recent ERP studies showed similar negativities in response to unexpected stress patterns during silent reading as well (e.g., Luo & Zhou, 2010; Magne, Gordon, & Midha, 2010). Thus the present study provides further support in favor of the hypothesis that music training may enhance speech processing skills that are the foundations of the acquisition of good literacy skills (e.g., Moritz et al., 2013).

4. Methods

4.1. Participants

Twenty-two college students (8 females, *mean age* = 23.7, age range: 19–36) were recruited for the study. All were right-handed, native English speakers with less than two years of formal musical training. None of the participants were enrolled in a Music major. Data from two participants were discarded for excessive eye artefacts in the EEG data. The study was approved by the Institutional Review Board at Middle Tennessee State University and written consent was obtained from the participants prior to the start of the experiment.

4.2. Stimuli

A total of 480 bisyllabic English words (240 trochaic and 240 iambic) were selected from the English Lexicon Project database (Balota et al., 2007) to build 240 metrically priming sequences of three bisyllabic words followed by a consistent (120) or inconsistent (120) final trochee or iamb (see Supplementary Content 1 for the full set of stimuli). To increase the number of trials per condition, nouns (360) and adjectives (120) were used. However, word sequences only contained nouns or adjectives. Metrically unexpected versions of each word sequence were created by switching the final words (i.e., the critical word) between sequences with opposite stress patterns. In both metrically expected and unexpected word sequences, the lexical frequency of all the words within each sequence was controlled using the log HAL frequency (Lund & Burgess, 1996). The mean log HAL frequency for each set of stress patterns was 8.904 (SD = 1.529) for trochaic sequences and 8.896 (1.527) for iambic sequences (Mean = 8.941, SD = 1.587 and Mean = 8.914, SD = 1.583 for critical words). The semantic relatedness of the words within each sequence was evaluated by two independent judges with no knowledge of the purpose of the experiment, as well as using a web-based Latent Semantic Analysis (LSA) tool to measure similarity between words within each sequence (LSA@CU, http://lsa.colorado.edu). LSA gives a similarity score comprised between 0 (not related) to 1 (highly related). The final stimulus set showed very low mean LSA similarity scores for the first three words composing the metrical priming sequences (*trochaic:* Mean = 0.07, SD = 0.08; *iambic:* Mean = 0.1, SD = 0.1). LSA values for the final critical words were also very low and there was no difference between the two counterbalanced lists of stimuli (List A: Mean = 0.09, SD = 0.10; List B: Mean = 0.08, SD = 0.09). In addition, LSA scores for the critical words were low for both metrically expected (Mean = 0.1, SD = 0.1) and metrically unexpected conditions (Mean = 0.08, SD = 0.09). Finally, critical words were controlled so that they did not share either their initial or final syllable with any of the three first words in the sequence.

Participants were presented with 30 different sequences of words in each of the 4 experimental conditions (i.e., Expected Trochaic, Expected Iambic, Unexpected Trochaic, and Unexpected Iambic). While the same critical words were used in the metrically expected and unexpected conditions, participants only saw one of the two versions of each word sequence during the experiment. However, two different lists of 120 word sequences were used so that each critical word was presented in both metrically expected and unexpected conditions between participants, without being repeated within participants.

Words were produced by a male voice at a sampling rate of 44 kHz using the Neospeech Text-to-Speech software (Neospeech, Inc., Santa Clara, CA) in order to have a consistent intonation and speech rate across all stimuli (Chen, Wong, & Hu, 2014). Words had a mean duration of 536 ms (SD = 78 ms).

4.3. Procedure

Participants were first administered the Advanced Measures of Music Audiation (AMMA; Gordon, 1989) in order to assess their musical aptitude. The AMMA has been used previously to measure correlations between musical aptitude and indices of brain activity (e.g., Schneider et al., 2002; Seppänen, Brattico, & Tervaniemi, 2007; Vuust, Brattico, Seppänen, Näätänen, & Tervaniemi, 2012). In addition, this measure was nationally standardized with a normed sample of 5336 U.S. students and offers percentile rank norms for both music and non-music majors. The AMMA lasts 15 min and consists of 30 pairs of melodies. For each pair, participants are asked to determine whether the two melodies are the same, tonally different or rhythmically different. Thus, this standardized test gives an objective measure of their ability to differentiate rhythmic and tonal variations. In particular, for non-Music majors, reliability scores are 0.80 for tonal score and 0.81 for rhythm score (Gordon, 1990). Following the AMMA test, participants were seated in a soundproofed and electrically shielded room at approximately 3 feet in front of a computer screen. Speech stimuli were presented through headphones using a Toshiba Portege Tablet PC and the software E-prime 2.0 Professional with Network Timing Protocol (Psychology software tools, Inc., Pittsburgh, PA). Participants were presented with 3 blocks of 40 word lists each. The word lists were randomized within each block and the order of the blocks was counterbalanced across participants. Each sequence of words was introduced by a fixation cross displayed at the center of a computer screen and remaining until 2 s after the onset of the fourth word. Because we were interested in the perception of metrical regularity regardless of any apparent physical temporal regularity, successive words were presented with a random inter-stimulus interval varying between 300 and 500 ms, in order to minimize any potential effects of temporal expectancy. The perception of a regular meter is further enhanced when the stressed syllables fall at regular time intervals (e.g., Gordon et al., 2011; Quené & Port, 2005). However, previous studies have found that listeners can still perceive a regular metrical structure in speech, even when the inter-stress time interval is not kept constant (e.g., Schmidt-Kassow & Kotz, 2009b).

Finally, to ensure that participants attentively listened to each word of the sequences without having explicit knowledge of the metrical manipulation, they were required to perform a short memory task. To that end, an additional word was visually presented on a computer screen 2 s after the onset of the fourth word of each sequence. For half of the word sequences, the visual target word was new, while for the other half it was a repetition of one of the four previous spoken words. Participants were asked to pay attention to each word in the sequence and to press a button if they thought the visual target word was new or another button if they thought it was a repetition of one of the previous aurally presented words (see Supplementary Content 2 for analysis of the behavioral data). The entire experimental session lasted 1.5 h.

4.4. EEG acquisition and preprocessing

EEG was recorded continuously from 64 Ag/AgCl electrodes embedded in sponges in a Hydrocel Geodesic Sensor Net (EGI, Eugene, OR) placed on the scalp, connected to a NetAmps 300 amplifier, and using a MacBook Pro computer. Electrode impedances were kept below 50 kOhm. Data was referenced online to Cz and re-referenced offline to the averaged mastoids. In order to detect the blinks and vertical eye movements, the vertical and horizontal electrooculograms (EOG) were also recorded. The EEG and EOG were digitized at a sampling rate of 500 Hz.

EEG preprocessing was carried out with NetStation Viewer and Waveform tools. The EEG was first filtered with a bandpass of 0.1–100 Hz. Data time-locked to the onset of the fourth word (i.e., critical word) of each list was then segmented into epochs of 1100 ms, starting with a 100 ms prior to the onset of the fourth words and continuing 1000 ms post-word-onset. Trials containing movements, ocular artifacts or amplifier saturation were discarded. ERPs were computed separately for each participant and each condition by averaging together the artefact-free EEG segments relative to a 100 ms pre-baseline.

4.5. Data analysis

Statistical analyses were performed using MATLAB and the FieldTrip open source toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). The cluster-based permutation method implemented in the Fieldtrip toolbox is optimally designed for pairwise comparisons only. Thus, we conducted planned pairwise comparisons between critical words with same stress pattern to avoid any potential confounds relative to the acoustical differences in the realization of the iambic and trochaic stress patterns (Unexpected Trochaic vs Expected Trochaic and Unexpected Iambic vs Expected Iambic). The advantage of this non-parametric data-driven approach is that it does not require the user to specify any latency range or region of interest a priori, while also offering a solution to the problem of multiple comparisons (see Maris & Oostenveld, 2007).

To relate the ERP results to the musical aptitude measure, cluster sums were calculated as in Lense, Gordon, Key, and Dykens (2014). Correlations were then tested between the ERP cluster sum difference (i.e., difference between the cluster sums of each condition) and the participants' scores on the AMMA.

Acknowledgments

This study was partially funded by NSF Grant # BCS-1261460 awarded to Cyrille Magne and by the MTSU Faculty Research and Creative Activity Grant Program. The funding sources had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. We wish to thank two anonymous reviewers for their helpful comments and suggestions on previous versions of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandl.2016.01.001.

References

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83*(2), 111–130.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience, 25*(3–4), 399–410.

Böcker, K. B., Bastiaansen, M. C., Vroomen, J., Brunia, C. H., & de Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology, 36*(6), 706–720.

Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir) regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia, 51*(4), 760–771.

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology, 3*, 327.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The tick-tock of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science, 14*, 362–366.

Chen, F., Wong, L. L. N., & Hu, Y. (2014). Effects of lexical tone contour on mandarin sentence intelligibility. *Journal of Speech, Language, and Hearing Research, 57*(1), 338–345.

Chobert, J., François, C., Velay, J. L., & Besson, M. (2012). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex, 24*(4), 956–967.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*(3–4), 133–142.

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology, 2*, 124.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*(3), 294–311.

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). The processing of German word stress: Evidence for the prosodic hierarchy. *Phonology, 25*(1), 1–36.

Duanmu, S., Kim, H. Y., & Stiennon, N. (2005). Stress and syllable structure in English: Approaches to phonological variations. *Taiwan Journal of Linguistics, 3*(2), 45–57.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex, 23*(9), 2038–2043.

Gordon, E. E. (1989). *Manual for the advanced measures of music audiation*. Chicago, IL: G.I.A. Publications.

Gordon, E. E. (1990). *Predictive validity study of AMMA: A one-year longitudinal predictive validity study of the Advanced Measures of Music Audiation*. Chicago, IL: G.I.A. Publications.

Gordon, R. L., Magne, C. L., & Large, E. W. (2011). EEG correlates of song prosody: A new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology, 2*, 352.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & McAuley, J. D. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science, 18*(4), 635–644.

Halle, M., & Vergnaud, J. R. (1987). Stress and the cycle. *Linguistic Inquiry, 18*, 45–84.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Sarkamo, T. (2013). Music and speech prosody: A common rhythm. *Frontiers in Psychology, 4*, 566.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Henrich, K., Alter, K., Wiese, R., & Domahs, U. (2014). The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain and Language, 136*, 19–30.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience, 29*(10), 3019–3025.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459–491.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*(9), 323–328.

Kochanski, G., & Orphanidou, C. (2008). What marks the beat of speech? *The Journal of the Acoustical Society of America, 123*(5), 2780–2791.

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences, 14*(9), 392–399.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., & White-Schwoch, T. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *The Journal of Neuroscience, 34*(36), 11913–11918.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119–159.

Lense, M., Gordon, R., Key, A. P. F., & Dykens, E. (2014). Neural correlates of cross-modal affective priming by music in Williams syndrome. *Social Cognitive and Affective Neuroscience, 9*(4), 529–537.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28*(2), 203–208.

Luo, Y. Y., & Zhou, X. L. (2010). ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading. *NeuroImage, 49*, 2836–2849.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex, 17*(11), 2659–2668.

Magne, C., Gordon, R. L., & Midha, S. (2010). Influence of metrical expectancy on reading words: An ERP study. *Proceedings of the Fifth International Conference on Speech Prosody, 100432*, 1–4. <http://speechprosody2010.illinois.edu/papers/100432.pdf>.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18*(2), 199–211.

Marie, M., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience, 23*(2), 294–305.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.

Mattys, S., & Samuel, A. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language, 36*(1), 87–116.

McCauley, S. M., Hestvik, A., & Vogel, I. (2012). Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials. *Language and Speech, 56*(1), 23–44.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research, 1194*, 81–89.

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P. A. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences, 20*(1), 56–60.

Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: A review. *Frontiers in Psychology, 2*, 321.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex, 19*(3), 712–723.

Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., & Wolf, M. (2013). Links between early rhythm skills, musical training, and phonological awareness. *Reading and Writing, 26*(5), 739–769.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41*, 233–243.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869.

Patel, A. D. (2008). *Music, language, and the brain*. NY: Oxford University Press.

Peter, V., McArthur, G., & Thompson, W. F. (2012). Discrimination of stress in speech and music: A mismatch negativity (MMN) study. *Psychophysiology, 49*(12), 1590–1600.

Peynircioğlu, Z. F., Durgunoğlu, A. Y., & Öney-Küsefoğlu, B. (2002). Phonological awareness and musical aptitude. *Journal of Research in Reading, 25*(1), 68–80.

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 564–573.

Port, R. (2003). Meter and speech. *Journal of Phonetics, 31*, 599–611.

Praamstra, P., Meyer, A. S., & Levelt, W. J. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component timelocked to phonological mismatch. *Journal of Cognitive Neuroscience, 6*(3), 204–219.

Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62*(1), 1–13.

Rautenberg, I. (2013). The effects of musical training on the decoding skills of German-speaking primary school children. *Journal of Research in Reading*. http://dx.doi.org/10.1111/jrir.12010.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS ONE, 8*(2), e56000.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia, 50*(2), 232–244.

Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: Rules versus meaning. *NeuroReport, 21*(8), 580–584.

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science, 14*(6), 317–320.

Schmidt-Kassow, M., & Kotz, S. A. (2008). Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Research, 1226*, 144–155.

Schmidt-Kassow, M., & Kotz, S. A. (2009a). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience, 21*(9), 1693–1708.

Schmidt-Kassow, M., & Kotz, S. A. (2009b). Attention and perceptual regularity in speech. *NeuroReport, 20*(18), 1643–1647.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5*, 688–694.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341–349.

Seppänen, M., Brattico, E., & Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiology of Learning and Memory, 87*, 236–247.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience, 23*(13), 5545–5552.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Association for Psychological Science, 17*(8), 678–681.

Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions: BBF, 7*, 44.

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences, 29*(7), 382–390.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia, 50*, 1432–1443.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neurophysiological evidence. *Cognitive Brain Research, 25*(1), 180–187.

Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences of the United States of America, 111*(40), 14559–14564.

Wu, H., Ma, X., Zhang, L., Liu, Y., Zhang, Y., & Shu, H. (2015). Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Frontiers in Psychology*, *6*, 436. http://dx.doi.org/10.3389/fpsyg.2015.00436.