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
Available online 29 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).