Children's early bilingualism and musical training influence prosodic discrimination of sentences in an unknown language

Arthur Stepanov,[a)] Matic Pavlič, and Penka Stateva
Center for Cognitive Science of Language, University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica, Slovenia
arthur.stepanov@ung.si, matic.pavlic@ung.si, penka.stateva@ung.si

Anne Reboul
Institute for Cognitive Sciences-Marc Jeannerod, CNRS UMR 5304, Bron, France
reboul@isc.cnrs.fr

Abstract: This study investigated whether early bilingualism and early musical training positively influence the ability to discriminate between prosodic patterns corresponding to different syntactic structures in otherwise phonetically identical sentences in an unknown language. In a same-different discrimination task, participants ($N = 108$) divided into four groups (monolingual non-musicians, monolingual musicians, bilingual non-musicians, and bilingual musicians) listened to pairs of short sentences in a language unknown to them (French). In discriminating phonetically identical but prosodically different sentences, musicians, bilinguals, and bilingual musicians outperformed the controls. However, there was no interaction between bilingualism and musical training to suggest an additive effect. These results underscore the significant role of both types of experience in enhancing the listeners' sensitivity to prosodic information.

[RS]
Date Received: July 10, 2017 **Date Accepted:** December 12, 2017

1. Introduction

Humans' ability to discriminate suprasegmental parts of the speech signal such as variations in pitch, amplitude, intensity, and duration of segments is a key to successful communication. This is so because prosodic information, easily recovered from the spoken input, serves as a powerful tool helping the speaker to encode or decode the syntactic structure of the sentence, especially when it concerns constituent groupings within sentences (Nespor and Vogel, 1986). A successful prosodic differentiation bears at least on the general hearing ability of the speaker/hearer, working memory (WM), and sensitivity to various segmental and suprasegmental features. Prosodic differentiation is instrumental in a child's acquisition of her native language. Evidence suggests that prenatal experience already shapes the perception of prosody in newborn infants (Abboub *et al.*, 2016). Monolingual children younger than 2 yrs of age may exploit prosodic information to predict the syntactic category of an upcoming word (de Carvalho *et al.*, 2017). They can also successfully use native as well as non-native prosodic information as a cue to constrain syntactic parsing and identify constituent-like units in an artificial language (Hawthorne *et al.*, 2015). In other studies, 4 to 6 yr old children were shown to reliably use prosodic cues to disambiguate syntactic structure in pairs of otherwise phonetically homonymous pre-ambles of sentences, such as [*The baby flies*] [*hide in the shadows*] vs [*The baby*] [*flies his kite all day long*] (de Carvalho *et al.*, 2016). These studies also suggest that children rely on prosody's general acoustic salience, rather than on correlations between prosody and syntax in their native language only. The importance of the prosodic differentiation abilities extends to learning a second or foreign language. For instance, learning of prosodic cues to word boundaries in segmenting foreign speech is negatively affected by similarity of one's native and second/foreign language prosodies (Tremblay *et al.*, 2016). In this context, of particular interest here are two subgroups of speakers that are often seen as better foreign language learners than others.

One subgroup includes bilingual population. Recent studies have shown that bilingualism improves young infants' perception of particular prosodic properties, such as stress, in discrimination tasks distinguishing stress-initial (trochaic), and stress-final

[a)]Author to whom correspondence should be addressed.

(iambic) patterns in the stimuli, with and without segmental variability in these patterns (Bijeljac-Babic *et al.*, 2016). In these studies, monolingual infants acquiring a language lacking lexical stress, such as French, were compared with bilingual infants exposed to French and a language that has a lexical stress, e.g., German. Furthermore, bilingual infants appear to show a perceptual advantage in the discrimination of non-native tones, or prosodic cues over syllables in languages like Mandarin (Liu and Kager, 2013). Bilingual infants also showed better performance in discriminating musical pitch variations than monolingual infants, which has been attributed to the bilinguals' higher sensitivity to subtle acoustic differences in the stimuli due to the fact that bilinguals face a more complicated learning environment (Liu and Kager, 2017). Studies on older subjects reached somewhat controversial results: while some suggest that bilingual adults may lose the extra perceptual sensitivity to distinguish non-native phonetic contrasts in comparison to infants (Werker, 1986), other studies found that bi/multilinguals may outperform monolinguals on non-native contrasts after receiving a training on these contrasts, and be able to transfer their learned knowledge to other non-native contrasts (Tremblay and Sabourin, 2012). Bilinguals also outperformed monolinguals in artificial language learning tasks, and their success was correlated with factors such as phonetic similarity to the native language, and the complexity of the target phonetic contrasts (Antoniou *et al.*, 2015).

The second subgroup involves musicians. The idea that musicians are better than non-musicians in learning foreign languages finds increasing empirical support in recent psycho- and neurolinguistic literature which suggests that music and speech processing share a great deal of properties at the behavioral and neuro-anatomical levels. For instance, Broca's area, a key region in linguistic processing, is also operative in music processing, as well as a number of other brain regions (Patel, 2008). The relevance of studying musicians for the purposes of prosodic processing is especially highlighted by the fact that music and speech share fundamental prosodic markers such as pitch, duration, and rhythm (the Ancient Greek word *prosōidíā* literally means "song sung to music"). Music perception and prosody perception, specifically, word stress, was found to correlate in tasks discriminating compound words from homonymous phrasal utterances (Hausen *et al.*, 2013). Musical experience improves general auditory skills, including the brain's sensitivity to linguistic pitch patterns (Kraus and Chandrasekaran, 2010). Marques *et al.* (2007) investigated sensitivity of French-speaking musicians and non-musicians with respect to violations of language-specific pitch patterns in sentences in an unknown language (Portuguese) using behavioral markers and event-related potentials. Musicians in this study outperformed non-musicians in reliably detecting pitch incongruity in the speech signal. These and other results point to a strong connection between musical training and linguistic abilities that might also participate in second language learning.

The above considerations suggest that both bilingualism and early music training may enhance speakers' sensitivity to particular aspects of natural language prosody. However, it remains unclear whether bilinguals and musicians might also be more efficient in simultaneously exploiting various types of prosodic information to discriminate between different syntactic structures in an unknown language, in the absence of other obvious cues. We may therefore ask whether bilingualism and music training may be factors boosting sensitivity to prosodic cues signaling different constituent structure, compared to monolinguals and non-musicians. If this is indeed the case, we will have converging evidence of an important role that these two different types of experience play in making use of sentential prosody.

2. Experiment

We set out to investigate if the bilingual and musician groups' better general predisposition toward linguistic learning, however limited, might positively affect their performance in a prosodic discrimination task that taps on the prosody's general acoustic salience at the sentence level. Specifically, we used a same-different task to test whether bilingualism and early musical training influence the ability to discriminate between prosodic patterns corresponding to different syntactic structures in otherwise phonetically identical sentences in French, a language unknown to participants. We were also interested in knowing whether both factors induce a cumulative (additive) effect on this ability.

3. Method

3.1 Participants

Children (108) from elementary schools in the area of Nova Gorica (Slovenia) and Gorizia (Italy) were tested. On the basis of questionnaire data, the participants were assigned to the following four groups: monolingual non-musicians [$N = 27$, $M_{age} = 11.8$ yrs, standard deviation (SD) = 0.86], bilingual non-musicians ($N = 28$, $M_{age} = 11.9$, SD = 0.96), monolingual musicians ($N = 31$, $M_{age} = 12.1$, SD = 1.12), and bilingual musicians ($N = 22$, $M_{age} = 12.1$, SD = 0.78). There were no significant differences in age across groups [$F(3, 107) = 0.80$, ns]. Thirty-four (68%) of the bilingual children came from bilingual families and thus spoke two languages from birth. The remaining 16 bilinguals came from monolingual families and were systematically exposed to two languages within the first 3 years after birth. One of the languages spoken by the bilinguals was always Slovenian; the other was Albanian, Bosnian/Croatian/Serbian, Bulgarian, Friulian, German, Italian, Macedonian, Russian, Spanish, or Tamil. The musicians had an uninterrupted training in at least one musical instrument and musical theory in state-accredited musical schools for an average of 5.23 yrs (SD = 0.95) for monolinguals, and 5.35 yr (SD = 1.09) for bilinguals, at the time of experiment. The non-musician groups did not have a regular exposure to musical training. The four groups thus differed in the mean number of years of musical experience, $F(3, 107) = 499.7$, $p < 0.0001$, the two musician groups differing from the non-musicians (Tukey HSD, $p < 0.0001$), and there was no difference within each of these pairs (Tukey HSD, ns). None of the participants had a previous exposure to French. An additional six children were tested, but their results were not included in the final analysis: two children had a hearing and an attention disorder, respectively, and four others performed at chance on identical controls, identical fillers, or lexically different controls (a 50% accuracy threshold was assumed). One child erroneously terminated the experimental program at a pause stage; the collected subset of data, however, was retained.

3.2 Materials

The stimuli comprised pairs of short sentences in French. We exploited the fact that French is a case-less language, in which words that differ in syntactic category (nouns, adjectives, or verbs) can have the same morpho-lexical makeup, e.g., *ferme* for "farm" or "closes." Clitics and definite determiners are also phonologically identical: *le*, *la*, *les*. This makes it possible to construct phonologically identical sequences of words that correspond to sentences with different syntactic construction and different meaning, as shown in Example (1),

$[_{NP}$Le jeune garde$]$ $[_{VP}$la voit$]$.
The young guard her sees
"The young guard sees her" (1a)

$[_{NP}$Le jeune$]$ $[_{VP}$garde la voie$]$.
The young guards the road
"The young man guards the road" (1b)

While there are no phonological differences between Examples (1a) and (1b), there are prosodic differences, with a different prosodic contour corresponding to each sentence. There was a prosodic boundary between noun and verb phrases in the sentences. There were three conditions, each composed of 30 stimuli. The target condition involved a pair of phonologically identical but prosodically different sentences. Two other conditions were controls, one where the sentences (with the same two different syntactic organizations as the target condition) differed both in phonology and prosody and one where the sentences were both prosodically and phonologically identical inside a pair (half of them had one syntactic structure, the other half the other). The order of presentation between two syntactic structures (long NP-short VP, hereafter referred to as "noun condition" vs short NP-long VP, hereafter referred to as "verb condition") was counterbalanced within a block, with 15 stimuli beginning with one structure, and 15 with the other. To ensure that "different" answers should not predominate creating a habituation among subjects, we also had 30 "identical" fillers, using twice the same sentence. Half of those fillers were taken from the noun condition sentences in the target condition, the other half from the verb condition sentences. Thus overall, there were 120 stimuli, balanced by length which amounted to 7 ± 2 syllables.

We assessed prosodic differences between the noun and verb conditions by conducting acoustic analyses of duration and pitch on the phonetic segments in the vicinity of the prosodic boundary between the noun and verb phrases in each condition, using Praat. A comparison of acoustic features in a sample pair is illustrated in Fig. 1. The total durations of sentences in the noun and verb condition did not differ significantly [$M_{noun} = 2235$ ms, $SD_{verb} = 33$ vs $M_{verb} = 2315$ ms, $SD_{verb} = 25$; $t(58) = 1.04$, ns].

The time until the onset of the prosodic boundary demarcating the noun and verb phrases differed among the conditions [$M_{\text{noun}} = 1141$ ms, $\text{SD}_{\text{verb}} = 180$ vs $M_{\text{verb}} = 705$ ms, $\text{SD}_{\text{verb}} = 149$; $t(58) = 10.17$, $p < 0.001$]. The analysis of duration revealed a significant phrase-final lengthening, consistently with the previous literature (Jun and Fougeron, 2002). The syllable immediately preceding the prosodic phrase boundary in the noun condition (e.g., *garde* in the first graph) was lengthened by 25% compared to the verb condition [$M_{\text{noun}} = 375$ ms, $\text{SD}_{\text{verb}} = 94$ vs $M_{\text{verb}} = 300$ ms, $\text{SD}_{\text{verb}} = 88$; $t(58) = 2.974$, $p = 0.004$]. The syllable immediately preceding the prosodic phrase boundary in the verb condition (e.g., *jeune* in the second graph) was lengthened by 49% compared to the noun condition [$M_{\text{verb}} = 400$ ms, $\text{SD}_{\text{verb}} = 84$ vs $M_{\text{noun}} = 269$ ms, $\text{SD}_{\text{verb}} = 66$; $t(58) = 6.62$, $p < 0.0001$]. We also analyzed phrase-initial articulatory strengthening (Fougeron and Keating, 1997): the onset of the target syllable in our stimuli (e.g., /g/ in *garde*) was lengthened by 33% in the verb condition (phrase-initial) compared to the noun condition [phrase-medial; $M_{\text{verb}} = 130.8$ ms, $\text{SD}_{\text{verb}} = 53.4$ vs $M_{\text{noun}} = 98.2$ ms, $\text{SD}_{\text{noun}} = 47.3$; $t(58) = 2.50$, $p = 0.015$]. Note that each pair of stimulus sentences always contains two words that are on different sides of a prosodic boundary in the verb condition, but on the same side in the noun condition, e.g., *jeune* and *garde* in Fig. 1, henceforth referred to as Word 1 and Word 2. The prosodic boundaries in our stimuli are also marked by a silent pause between the noun and verb phrases. There was no significant difference in the duration of the pause separating prosodic units across the conditions [$M_{\text{noun}} = 490$ ms, $\text{SD}_{\text{noun}} = 174$ vs $M_{\text{verb}} = 534$ ms, $\text{SD}_{\text{verb}} = 116$; $t(58) = 1.14$, ns]. In contrast, there was a significant difference in the duration of the pause between Word 1 and Word 2 across the conditions [$M_{\text{noun}} = 0.021$ ms, $\text{SD}_{\text{noun}} = 0.04$ vs $M_{\text{verb}} = 534$ ms, $\text{SD}_{\text{verb}} = 116$; $t(58) = 22.36$, $p < 0.001$].

In the analysis of pitch, we compared maximum *F*0s on the last syllables at Word 1 and Word 2. This analysis revealed a significant rise of pitch in the respective condition when each of these words immediately precedes a prosodic boundary, compared to the condition in which it does not [Word 1: $F0^{\max}$ (noun condition) $= 139.31$ Hz, $F0^{\max}$ (verb condition) $= 155.49$ Hz, $t(58) = 3.86$, $p < 0.001$; Word 2: $F0^{\max}$ (verb condition) $= 128.56$ Hz, $F0^{\max}$ (noun condition) $= 150.73$ Hz, $t(58) = 2.29$, $p = 0.025$]. Furthermore, Word 2's mean $F0^{\max}$ is higher than that of Word 1 by 11.42 Hz in the noun condition [$t(58) = 1.91$, $p = 0.03$], but lower than that of Word 1 by 26.93 Hz in the verb condition [$t(58) = 3.92$, $p = 0.0002$]. These patterns are consistent with the general tendency of French for a rising pitch contour toward the end of the prosodic units. Overall, participants in the current experiment had to exploit (at least) the pause, phrase-final lengthening, and pitch variation as prosodic cues to disambiguate similarly sounding sentences in French. Notably, these cues are common in the world's languages (Nespor and Vogel, 1986; Shattuck-Hufnagel and Turk, 1996), and were shown to be effectively exploited in music perception as well (Knosche *et al.*, 2005).

The stimuli were read by a male native French speaker with an interval of 5 s between the sentences within each pair, and recorded with the stereo recorder H2next by Zoom. Speech was digitized to an Apple Macbook Air computer with a sampling rate of 44.1 KHz and a 16-bit sampling depth.

3.3 Procedure

Children were tested in a quiet room on school premises. They heard pairs of stimulus sentences played binaurally through circumaural headphones at a comfortable listening level in pseudo-randomized order (different for each child) and were asked to indicate, for each pair, whether the sentences sound the same or different, by clicking on one of the two designated buttons that appeared on the computer screen after presentation of

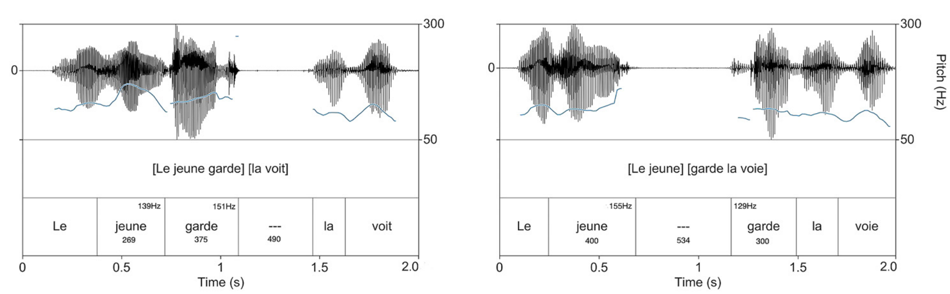

Fig. 1. (Color online) Comparison of prosodic differences between the noun and verb condition, using a sample stimulus.

each stimulus (to control for handedness, for half of the participants the "Same" button appeared on the left, and for the other half, on the right, of the "Different" button). Only when the two response buttons appeared on the screen, and not earlier, could the participant's response on a particular trial be registered (the response buttons were not available during actual playback of the stimulus). Clicking on the response button after hearing the entire stimulus would then advance the next trial. There was a self-timed break after every 40 stimulus pairs. There was no prior familiarization stage in our experiment. The main experiment was preceded by a practice session during which participants listened to four to six exemplars of each experimental condition through computer-internal loudspeakers and were given feedback by the experimenter after each response. For their contribution, the participants were rewarded a pen or a sticker.

4. Results

We checked for response biases by calculating hit and false alarm rates and the d' value for each participant over all conditions. D' values were significantly different from 0, indicating that the performance is not random [$t(107) = 30.81$, $p < 0.0001$], and they were also significantly larger than 1.35 [$t(107) = 14.78$, $p < 0.0001$], indicating an overall accuracy of more than 75%. This demonstrates that our participants were in general sufficiently attentive to the content of the task. Children were on average 98.4% accurate on lexically different pairs and mean response time (RT) was 1437 msec (SD = 300). Within the lexically identical pairs, mean accuracy was 88.2% on prosodically identical control pairs, and 70.9% on prosodically different target pairs. RTs were 1751 msec (SD = 273) and 1737 msec (SD = 278), respectively.

The distribution of accuracy means per participant on lexically identical-prosodically different trials in each tested subgroup is shown in Fig. 2.

We modeled our results concerning accuracy on target items using mixed effects logistic regression (logit) models for accuracy, and linear mixed models for RTs, entering bilingualism and musical training as fixed factors, and participants and stimuli as random factors with an intercept. The models revealed a main effect of bilingualism [$\chi^2(1) = 8.65$, $p = 0.003$], corresponding to an increase in probability of correct discrimination in bilinguals by 0.73 compared to monolinguals, as well as a main effect of music training [$\chi^2(1) = 16.68$, $p < 0.0001$] corresponding to an increase in probability of correct discrimination in musicians by 0.80 compared to non-musicians.[1] However, there was no significant interaction between the two factors [$\chi^2(1) = 0.11$, ns]. With respect to RTs (only RTs on the correct responses were analyzed), we found a marginal effect of musical training [$\chi^2(1) = 3.435$, $p = 0.06$], with musically trained children being slightly faster (by about 427 ms) in discrimination than their non-musically educated peers. There was no main effect of bilingualism [$\chi^2(1) = 0.50$, ns] and no interaction between the two factors ($p > 0.1$).

5. Discussion and conclusions

Our results indicate that both bilingual as well as musically trained children detect prosodic patterns disambiguating different syntactic structures in an unknown language more accurately than their respective monolingual and non-musically trained peers. Musicians also showed a tendency for faster overall recognition than non-musicians. However, there was no cumulative effect of musicianship and bilingualism. Given the nature of our stimuli with a pause at the prosodic boundary in different places in the

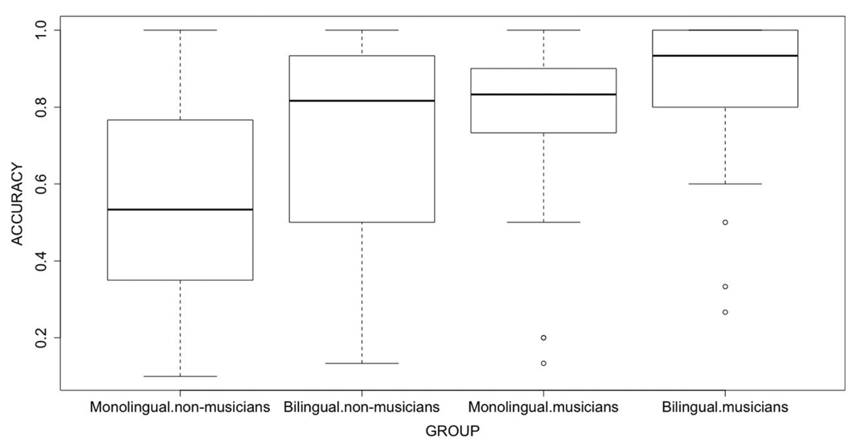

Fig. 2. Accuracy score means per participant on target trials, across the bilingualism and music training factors.

noun and verb conditions, one may wonder whether children could have responded to the task simply by noting the respective discrepancy midway through the second sentence in the target pair (e.g., [*le jeune garde*] + PAUSE vs [*le jeune*] + PAUSE) in Eqs. (1a) and 1(b) and discarding any material afterwards. If that were the case, it could be expected that participants in all groups would respond faster on the target stimuli than on the identical controls or fillers (where one ought to attend to the whole material to report identity). In fact, RTs on targets and identical controls or fillers did not differ significantly [targets vs controls: $t(4506) = 1.06$, ns; targets vs fillers: $t(4600) = 0.3004$, ns]. One could also expect, under this scenario, that target RTs may not be greater than those on lexically different controls. However, the opposite was the case, with the former RTs (mean 1862 ms) being greater than the latter [mean 1455 ms; $t(4763) = 3.2963$, $p = 0.0009$]. This suggests that participants did not base their decision on pause-related discrepancies only, but considered other prosodic cues available in the input as well, as described in Sec. 3.2.

The monolingual non-musician group showed above chance (53.6%) mean accuracy. This result reinforces earlier findings concerning younger participants who could successfully exploit non-native prosody cues to detect constituent-like boundaries (Hawthorne *et al.*, 2015). Our results also point to a perceptual advantage in the domain of prosodic processing due to bilingualism and music training, respectively. Concerning the latter, the results are consistent with previous studies that established a cognitive advantage of musicians in linguistic pitch perception tasks (Marques *et al.*, 2007), but suggest a further dimension to this line of inquiry by establishing that this advantage holds in a discrimination task, across a range of prosodic cues, and, furthermore, that it can be successfully exploited in helping differentiate between two syntactic structures in the absence of other obviously perceptible cues.

Furthermore, the results support the converging view in the literature that linguistic and musical training tap on the same or similar cognitive resources (Fedorenko *et al.*, 2009). Both population groups described above could be characterized by a common cognitive denominator in the form of enhanced executive control mechanisms, trained by the relevant type of experience, viz., bilingual or musical (Bialystok, 2010; Chan *et al.*, 1998). When examined together, both groups perform better at some executive control measures such as the Simon arrows test, than monolingual and non-musical groups (Bialystok and DePape, 2009). Similar results were reported when taking bilingual musicians into the picture, although no additive effects of bilingual and musical proficiency were found (Schroeder *et al.*, 2016). Both groups were also tested in task-switching and dual task paradigms (Moradzadeh *et al.*, 2015). Musicians outperformed non-musicians in this study, but there was no obvious bilingualism effect, and, again, no interaction with the music factor. In terms of phonetic processing, early musical or second language training was shown to have similar short-term as well as long-lasting effects on the brain's ability to discriminate relevant and irrelevant sounds, both in music and language (Moreno *et al.*, 2015). It thus seems reasonable to conjecture that bilinguals and musicians are similarly attuned to variations of the relevant sound properties; the former because of their greater sensitivity to acoustic properties of a linguistic sound, facilitating access to the phonetic vocabularies of their two or more languages, and the latter because of their acuter sensitivity to the prosodic structure of music that comes with regular music training. We further conjecture that this greater sensitivity can potentially be seen in the context of the view that bilingual experience enhances attentional processes in WM (see Calvo *et al.*, 2016 for a review). Similar claims have independently been made concerning the impact of music training on WM (Vandervert, 2015). Common models of WM assume the existence of a phonological loop responsible for articulatory rehearsal of information for immediate recall and/or manipulation (Baddeley, 2012). In a same-different task involving a pair of audio stimuli in the form of two sentences, a participant needs to temporarily store and/or rehearse verbal representations being compared. A greater capacity for rehearsal, including maintaining finer-grained details of the stimulus, along with other relevant aspects of WM, may thus be advantageous for this task.

In addition, our results support the hypothesis that bilinguals and musicians might have an advantage in foreign language learning, at least with respect to prosody as well as syntax (to the extent that prosody bootstraps syntactic learning; see Sec. 1). This does not, however, preclude the possibility that bilinguals, in particular, might have a disadvantage or display greater limitations than monolinguals in other linguistic domains, such as lexical retrieval (Bialystok, 2010). Nevertheless, theories of bilingualism and of music cognition might need to take the respective groups' increased sensitivity to prosodic patterns into account.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Program for research, technological development, and demonstration under Grant Agreement No. 613465. We thank Jean-Baptiste Van der Henst for recording the stimuli, Tatjana Nazir for advice on the experimental design (both from the Institute for Cognitive Sciences-Marc Jeannerod), and the administrative staff of all collaborating schools and school centers, where testing was done.

References and links

[1]Logit model estimates are in log-odds (logarithms of odds of giving a targeted answer), which can be converted to probabilities by taking an inverse logit [$\text{logit}^{-1}(\alpha) = \exp(\alpha)/(1 + \exp(\alpha))$].

Abboub, N., Nazzi, T., and Gervain, J. (**2016**). "Prosodic grouping at birth," Brain Lang. **162**, 46–59.

Antoniou, M., Liang, E., Ettlinger, M., and Wong, P. C. M. (**2015**). "The bilingual advantage in phonetic learning," Bilingualism: Lang. Cognit. **18**, 683–695.

Baddeley, A. (**2012**). "Working memory: Theories, models, and controversies," Ann. Rev. Psychol. **63**, 1–29.

Bialystok, E. (**2010**). "Bilingualism," Wiley Interdiscipl. Rev. Cognit. Sci. **1**, 559–572.

Bialystok, E., and DePape, A.-M. (**2009**). "Musical expertise, bilingualism, and executive functioning," J. Exp. Psychol. **35**(2), 565–574.

Bijeljac-Babic, R., Höhle, B., and Nazzi, T. (**2016**). "Early prosodic acquisition in bilingual infants: The case of the perceptual trochaic bias," Front. Psychol. **7**, 210.

Calvo, N., Ibáñez, A., and García, A. M. (**2016**). "The impact of bilingualism on working memory: A null effect on the whole may not be so on the parts," Front. Psychol. **7**, 265.

Chan, A. S., Ho, Y. C., and Cheung, M. C. (**1998**). "Music training improves verbal memory," Nature **396**, 128.

de Carvalho, A., Dautriche, I., Lin, I., and Christophe, A. (**2017**). "Phrasal prosody constrains syntactic analysis in toddlers," Cognition **163**, 67–79.

de Carvalho, A., Lidz, J., Tieu, L., Bleam, T., and Christophe, A. (**2016**). "English-speaking preschoolers can use phrasal prosody for syntactic parsing," J. Acoust. Soc. Am. **139**, EL216–EL222.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (**2009**). "Structural integration in language and music: Evidence for a shared system," Memory Cognit. **37**, 1–9.

Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., and Särkämö, T. (**2013**). "Music and speech prosody: A common rhythm," Front. Psychol. **4**, 566.

Hawthorne, K., Mazuka, R., and Gerken, L. (**2015**). "The acoustic salience of prosody trumps infants' acquired knowledge of language-specific prosodic patterns," J. Mem. Lang. **82**, 105–117.

Jun, S., and Fougeron, C. (**2002**). "Realizations of accentual phrase in French intonation," Probus **14**(1), 147–172.

Knosche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., and Friederici, A. D. (**2005**). "Perception of phrase structure in music," Hum. Brain Mapp. **24**, 259–273.

Kraus, N., and Chandrasekaran, B. (**2010**). "Music training for the development of auditory skills," Nature Rev. Neurosci. **11**, 599–605.

Liu, L., and Kager, R. (**2013**). "How bilingualism alters nontone-learning infants' tone perception in the first year of life," in *Proceedings of the 37th Annual Boston University Conference on Child Language Development*, edited by S. Baiz, N. Goldman, and R. Hawkes (Cascadilla Press, Somerville, MA).

Liu, L., and Kager, R. (**2017**). "Enhanced music sensitivity in 9-month-old bilingual infants," Cognitive Process. **18**, 55–65.

Marques, C., Moreno, S., Luís Castro, S., and Besson, M. (**2007**). "Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence," J. Cognit. Neurosci. **19**, 1453–1463.

Moradzadeh, L., Blumenthal, G., and Wiseheart, M. (**2015**). "Musical training, bilingualism, and executive function: A closer look at task switching and dual-task performance," Cognitive Sci. **39**, 992–1020.

Moreno, S., Lee, Y., Janus, M., and Bialystok, E. (**2015**). "Short-term second language and music training induces lasting functional brain changes in early childhood," Child Develop. **86**, 394–406.

Nespor, M., and Vogel, I. (**1986**). *Prosodic Phonology* (Foris, Dordrecht).

Patel, A. D. (**2008**). *Music, Language, and the Brain* (Oxford University Press, New York).

Schroeder, S. R., Marian, V., Shook, A., and Bartolotti, J. (**2016**). "Bilingualism and musicianship enhance cognitive control," Neural Plasticity **2016**, 1–11.

Shattuck-Hufnagel, S., and Turk, A. (**1996**). "A prosody tutorial for investigators of auditory sentence processing," J. Psycholing. Res. **25**(2), 193–247.

Tremblay, A., Broersma, M., Coughlin, C. E., and Choi, J. (**2016**). "Effects of the native language on the learning of fundamental frequency in second-language speech segmentation," Front. Psychol. **7**, 985.

Tremblay, M.-C., and Sabourin, L. (**2012**). "Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals," J. Acoust. Soc. Am. **132**, 3465–3474.

Vandervert, L. (**2015**). "How music training enhances working memory: A cerebrocerebellar blending mechanism that can lead equally to scientific discovery and therapeutic efficacy in neurological disorders," Cerebellum Ataxias **2**, 11.

Werker, J. F. (**1986**). "The effect of multilingualism on phonetic perceptual flexibility," Appl. Psycholing. **7**, 141–155.