

Musical training shapes neural responses to melodic and prosodic expectation



Ioanna Zioga^{a,*}, Caroline Di Bernardi Luft^{a,b}, Joydeep Bhattacharya^a

^a Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW, United Kingdom

^b School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Rd, London E1 4NS, United Kingdom

ARTICLE INFO

Article history:

Received 26 January 2016

Received in revised form

1 September 2016

Accepted 9 September 2016

Available online 10 September 2016

Keywords:

EEG

Expectation

Musical training

Language

Prosody

ABSTRACT

Current research on music processing and syntax or semantics in language suggests that music and language share partially overlapping neural resources. Pitch also constitutes a common denominator, forming melody in music and prosody in language. Further, pitch perception is modulated by musical training. The present study investigated how music and language interact on pitch dimension and whether musical training plays a role in this interaction. For this purpose, we used melodies ending on an expected or unexpected note (melodic expectancy being estimated by a computational model) paired with prosodic utterances which were either expected (statements with falling pitch) or relatively unexpected (questions with rising pitch). Participants' (22 musicians, 20 nonmusicians) ERPs and behavioural responses in a statement/question discrimination task were recorded. Participants were faster for simultaneous expectancy violations in the melodic and linguistic stimuli. Further, musicians performed better than nonmusicians, which may be related to their increased pitch tracking ability. At the neural level, prosodic violations elicited a front-central positive ERP around 150 ms after the onset of the last word/note, while musicians presented reduced P600 in response to strong incongruities (questions on low-probability notes). Critically, musicians' P800 amplitudes were proportional to their level of musical training, suggesting that expertise might shape the pitch processing of language. The beneficial aspect of expertise could be attributed to its strengthening effect of general executive functions. These findings offer novel contributions to our understanding of shared higher-order mechanisms between music and language processing on pitch dimension, and further demonstrate a potential modulation by musical expertise.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Music and language are two of the most characteristic human attributes, and there has been a surge of recent research interest in investigating the relationship between their cognitive and neural processing (e.g., Carrus et al., 2011; Koelsch and Jentschke, 2010; Maess et al., 2001; Patel et al., 1998a, 1998b). Music and language use different elements (i.e. tones and words, respectively) to form complex hierarchical structures (i.e. harmony and sentences, respectively), governed by a set of rules which determines their syntax (Patel, 1998, 2003, 2012; Slevc et al., 2009). However, analogies between the two domains should be done carefully, as grammatical categories (nouns, verbs) and functions (subject, object) have no parallels in music (Jackendoff, 2009; Patel, 2003). Further, musical elements can be played concurrently to form harmony, but this is not the case for language.

In this context, Patel (1998) hypothesised that what is common in music and language is that experienced listeners organise their elements in an hierarchical fashion based on learned rules (McMullen and Saffran, 2004; Slevc et al., 2009). Importantly, through everyday exposure to these rules, expectations for subsequent events are formed (Jonaitis and Saffran, 2009; Meyer, 2008). The fulfilment or violation of expectations constitutes a crucial component of the emotional and aesthetic experience of music (Huron, 2006). Expectations can also be disrupted in language, resulting in unexpected or incorrect sentences (Gibson, 2006).

Based upon their structural similarities, theoretical work has suggested that music and language use overlapping neural resources (Patel et al., 1998a, 1998b; Patel, 2003). In his 'Shared Syntactic Integration Resource Hypothesis' (SSIRH) Patel (2003) has suggested that shared domain-general neural areas in frontal regions would process syntactic information, which would be then integrated at posterior domain-specific representational sites (Fedorenko et al., 2009; Koelsch, 2012; Patel, 2008). Indeed, it has been shown that music influences simultaneous responses to

* Corresponding author.

E-mail address: iziog001@gold.ac.uk (I. Zioga).

language, due to competition for shared resources between harmony (chord sequences) and syntax (Carrus et al., 2013; Fedorenko et al., 2009; Jung et al., 2015; Koelsch et al., 2005; Kunert et al., 2015; Patel, 2003; Perruchet and Poulin-Charronnat, 2013; Slevc et al., 2009). For example, Carrus et al. (2013) found that harmonically unexpected notes influence neural responses to linguistic syntax, but not to semantics. Furthermore, studies have shown that sentence comprehension declines when an incongruent word co-occurs with an out-of-key note (Fedorenko et al., 2009; Slevc et al., 2009), providing behavioural evidence for interactions between the two domains. These interactions might provide evidence for shared, partially overlapping neural resources involved in the processing of syntax in language and music (e.g., Carrus et al., 2013; Koelsch et al., 2005). Critically, the present study did not investigate these aspects; rather we manipulated melodic and prosodic features to investigate potential shared processes in pitch dimension.

In fact, besides harmony and syntax, pitch is another important feature, forming the melody in music and the intonation or prosody (*‘the melody of speech’*) in language. Prosody is not only limited to fundamental frequency (F_0) fluctuations, but also refers to other properties of speech, such as fluctuations of loudness, stretching and shrinking of segments and syllable durations, speech rate and voice quality (Ashby and Maidment, 2005; Chun, 2002; Hirst and Cristo, 1998; Nolan, 2008; Nooteboom, 1997; Selkirk, 1995; Wells, 2006). Prosodic cues are used during on-line sentence comprehension to establish the syntactic structure and provide semantic information (Holzgrefe et al., 2013; Kotz et al., 2003; Steinhauer et al., 1999). Further, prosody serves communicative functions, as it allows to differentiate speech acts as questions or declaratives, and infers emotions (Pannekamp and Toepel, 2005; Paulmann et al., 2012).

In intonation languages, such as English and German, final tones are usually falling in pitch in statements (Meyer et al., 2002). The intonation contour of questions depends on the type of interrogation. Four main types of questions have been identified: alternative questions (e.g., *‘Is he alright or not?’*), yes-no questions (e.g., *‘Is he alright?’*), wh-questions (e.g., *‘Who is he?’*), and declarative questions (or ‘non-interrogative’) (e.g., *‘He is alright?’*) (Bartels, 1997; Bartels, 1999; Grabe, 2004). The rise of final intonation contour is significantly more common than low contour in all dialects, for yes-no questions, wh-questions, and declarative questions (Grabe, 2004). The exception is for wh-questions, where falling final pitch is more usual, as the more the lexical markers of interrogativity there are in an utterance, the less the final pitch rises (wh-questions have two markers: the initial “wh-” and the word inversion) (Haan, 2002). In our study, we used statements, and declarative questions with a final rise in the intonation contour.

In order to investigate potential effects of melodic expectancy on prosodic processing, our participants were asked to pay attention only to the speech while ignoring the music. Previous research has demonstrated qualitative differences between early vs. late ERP components, suggesting that early ERPs reflect sensory and perceptual processing modulated by attention (Hillyard et al., 1973; Näätänen et al., 1978), whereas late ERPs reflect integration and re-analysis processes (e.g., Astésano et al., 2004; Eckstein and Friederici, 2005). Studies using auditory stimuli have shown that N1 is enhanced when attention is directed to a stimulus (Alho et al., 1994; Schirmer and Kotz, 2006), as well as in response to unexpected events (Debener et al., 2005; Proverbio et al., 2004). In addition, musicians show larger P200 component, which is attributed to neuroplastic effects of musical expertise (Shahin et al., 2003; Trainor et al., 2003).

ERP responses to melodic and prosodic violation of expectations have been investigated separately in music and language

(e.g., Astésano et al., 2004; Besson and Faïta, 1995; Koelsch and Friederici, 2003; Koelsch and Jentschke, 2010; Lindsen et al., 2010; Patel et al., 1998a; Paulmann et al., 2012). It has been found that melodically unexpected notes elicit late positive components (LPCs) with a parietal scalp distribution around 300 ms post-stimulus onset (larger amplitude and shorter latency for non-diatonic compared to diatonic incongruities) (Besson and Faïta, 1995). Prosodic expectancy violations in speech utterances have been found to elicit a late positive component (*‘prosodic expectancy positivity’* or PEP) (Paulmann et al., 2012; Paulmann and Kotz, 2008), as well as a task-dependent, left temporo-parietal positivity (P800) (Astésano et al., 2004), associated with prosodic reanalysis processes. More specifically, when the prosody cannot be mapped onto the syntactic structure, the listener has to reanalyse the sentence in order to make sense of it (e.g., the utterance *Sarah is having lunch at the restaurant?* has unexpected prosody, as its syntax guides us to perceive it as a statement until the question is formed at the end). The P600 component has also been associated with violation of prosodic expectancy or prosody-syntax mismatch, reflecting structural revision processes (Eckstein and Friederici, 2005; Friederici and Alter, 2004; Schön et al., 2002; Steinhauer et al., 1999). Confirming the aforementioned findings, an fMRI study revealed increased BOLD activity in bilateral inferior frontal and temporal regions for unexpected compared to expected prosodic utterances (Doherty et al., 2004).

There is evidence that musical training alters the nervous system (Ragert et al., 2004), as well as the sensitivity to different aspects of musical sounds (Habib and Besson, 2009). Indeed, long-term musical training has been found to enhance brainstem responses to musical pitch (Skoe and Kraus, 2012), while musicians' auditory processing benefits have been positively correlated with the amount of musical expertise and the age that musical training started (Zendel and Alain, 2013). Supporting the view of shared processing, there is evidence for bidirectional influences between musical training and language skills (Asaridou and McQueen, 2013; Moreno, 2009; Schön et al., 2004; Wong et al., 2007; Zhao and Kuhl, 2015). For example, music experience has been shown to improve reading abilities in young children (Anvari et al., 2002). Musically trained listeners have shown better performance in detecting deviations of speech intonation (Schön et al., 2004), as well as interpreting affective prosody compared to untrained listeners (Thompson et al., 2004). Native English musicians have shown superior brainstem encoding of linguistic pitch patterns in a study using Mandarin Chinese syllables compared to non-musicians (Wong et al., 2007). Interestingly, there is also evidence for the opposite effect (e.g., Bidelman et al., 2011; Bradley, 2012). Specifically, speakers of Mandarin have been shown to perform better than English speakers in detection of pitch changes (Bradley, 2012). Thus, it seems that experience in one domain can be beneficial for the other one, suggesting that pitch tracking ability might be a potential shared mechanism (Asaridou and McQueen, 2013; Bidelman et al., 2011).

However, as both music performance and linguistic efficiency demand high levels of cognitive control, research has suggested that the aforementioned bidirectional influences could be attributed to enhanced executive functions due to musical training (Bialystok and Depape, 2009; Ho et al., 2003; Moreno et al., 2011; Schellenberg, 2004; Schellenberg, 2003, 2006) or bilingualism (e.g., Bialystok and Viswanathan, 2009; Festman et al., 2010; Krizman et al., 2012; Poarch and van Hell, 2012). For example, there is evidence that individuals who received music lessons show enhanced verbal (but not visual) memory (Ho et al., 2003) and intelligence (Moreno et al., 2011; Schellenberg, 2004; Schellenberg, 2006). Further, bilinguals possess a cognitive advantage towards attenuation of irrelevant information (inhibition), self-monitoring and intelligence, providing evidence for improved

general executive functions in non auditory tasks (Bialystok and Viswanathan, 2009; Festman et al., 2010; Poarch and van Hell, 2012).

Previous research on the interaction between music and language has focused on their syntactic and semantic elements, underestimating the role of pitch. The present study aims to fill this crucial gap by investigating the neural correlates (ERPs) of shared processing between melody and prosody and how they differ between musicians and nonmusicians, using a simultaneous auditory presentation paradigm. More specifically, the EEG and the reaction times of musicians and nonmusicians in a statement/question discrimination task were recorded, in order to reveal potential effects of melodic expectancy violations on prosodic processing.

Expectations for future events arise from learning the statistical properties of the auditory environment, in music (Dienes and Longuet-Higgins, 2004; Loui et al., 2010) and language (François and Schön, 2014; Maye et al., 2002), as also shown from studies demonstrating that song (merging of music and speech) facilitates language learning and stream segmentation (François et al., 2013; François and Schön, 2010; Schön et al., 2008). In line with the aforementioned research, in our study, we used a computational model of melodic pitch expectancy (Carrus et al., 2013; Pearce, 2005), which assumes that listeners' expectations are based on learning of statistical regularities in the musical environment. This model predicts upcoming events in an evolving melody based on the frequency with which each of these events has followed the context in a previously given corpus of music. It assumes that listeners' expectations are based on learning of statistical regularities in the musical environment, and are predicted successfully by the computational model (Wiggins and Pearce, 2006). Thus, high-probability notes are perceived as expected and low-probability notes as unexpected (Pearce et al., 2010). The degree of expectedness of the final notes can be expressed in units of information content, which is the negative logarithm, to the base 2, of the probability of an event occurring (MacKay, 2003). The final note of low-probability melodies had a higher information content ($M=11.75$, $SD=2.27$) than the information content of high-probability melodies ($M=1.98$, $SD=1.71$).

Linguistic stimuli consisted of speech utterances differing only in their final pitch: if falling they constituted a statement, if rising a question. As suggested by Ma et al. (2011), listeners' expectations are biased towards the perception of statements: first, because in English language statements are more frequently used than declarative questions (Bartels, 2014; Grabe, 2004), and, second, because declarative questions do not possess any word change (no inversion or wh- word) and are, thus, syntactically identical to statements (Bartels, 1997; Bartels, 2014; Grabe et al., 2003). Because it is less likely that a declarative syntax ends with interrogative intonation, we considered that statements are more expected compared to questions.

In summary, we investigated behavioural and neural responses to expectancy violations of simultaneously presented melodies and speech utterances, and how these differed between musicians and nonmusicians. Following previous literature, we hypothesised that (a) reaction times in the behavioural statement/question discrimination task will be faster when expectancy violations in language and music are presented simultaneously (an unexpected event in music might facilitate the recognition of unexpected events in speech, i.e. questions); (b) musicians' performance will be faster and more accurate than nonmusicians', due to their increased pitch sensitivity; (c) prosodic violations will elicit the 'prosodic expectancy positivity' and the P600 component, reflecting prosodic reanalysis and integration processes; and (d) musicians will show enhanced early ERP components, as well as increased late positivities in response to unexpected events,

reflecting enhanced pitch sensitivity due to the neuroplastic effects of musical training. Based on the existing evidence for overlapping neural resources between music and language and bidirectional influences between music and language skills in the pitch dimension, we expected that melodic expectancy would interact with language processing. Finally, following previous studies showing enhanced pitch encoding in musicians, it could be suggested that interaction effects may differ between musicians and nonmusicians.

2. Results

2.1. Behavioural findings

We calculated mean reaction times for each participant across four conditions: SH (statements on high-probability notes), SL (statements on low-probability notes), QH (questions on high-probability notes), QL (questions on low-probability notes), and the results are shown in Fig. 1. A mixed factorial ANOVA on reaction times was performed (see Section 4). We found participants were significantly faster in response to statements than to questions (*prosody*: $F(1,38)=24.89$, $p<.001$, $\eta^2=.40$), and musicians were overall faster than nonmusicians (*musical training*: $F(1,38)=4.59$, $p=.039$, $\eta^2=.11$) (Fig. 1b). Further, there was a significant interaction between *prosody* and *note-probability* ($F(1,38)=6.53$, $p=.015$, $\eta^2=.15$), and this was primarily due to the difference between questions (QH-QL: $t(39)=2.93$, $p=.006$), but not between statements ($p>.05$). Specifically, reaction times were shorter when statements were paired with a high-probability note, compared to when statements were paired with a low-probability note. The opposite effect was observed for questions, namely reaction times were longer when paired with a high-probability note than a low-probability note. Therefore, reaction times across groups were longer when expectancy in music and language was incongruent, i.e. statements were paired with low-probability notes and questions with high-probability notes.

Mean d' scores for each of the four conditions are displayed in Fig. 2a. We found that the type of prosody made a significant difference between conditions (QH-SH, Wilcoxon $Z=-5.31$, $p<.001$, and QL-SL, $Z=-4.95$, $p<.001$). A marginal significance was found between SL-SH ($Z=-1.87$, $p=.062$) but no significant difference was found between QL-QH ($p>.05$). Fig. 2b shows the mean accuracies for two groups. There was a marginal significance in the statements on high-probability-notes condition (SH) (Mann-Whitney $U=-1.87$, $p=.062$), but no significance in the other conditions ($p>.05$).

2.2. ERPs

2.2.1. N1 time window (60–130 ms)

Within the N1 time window, ERP amplitudes were shown to be more negative in response to questions compared to statements (see Fig. 3). This was confirmed by the mixed ANOVAs, which yielded a significant main effect of *prosody* ($F(1,37)=5.23$, $p=.028$, $\eta^2=.12$). In addition, there was a main effect of *region* ($F(1, 37)=86.48$, $p<.000$, $\eta^2=.70$), and a marginal effect of *musical training* ($F(1, 37)=4.00$, $p=.053$, $\eta^2=.10$). There was no effect of *note-probability* or interaction between *prosody* and *note-probability* ($F<.29$, $p>.592$). These results showed an enhanced N1 component following violation of expectations in speech prosody.

Furthermore, there was a significant *prosody* \times *musical training* \times *hemisphere* interaction ($F(1, 37)=5.25$, $p=.028$, $\eta^2=.12$). In order to explore further this interaction, we carried out planned contrasts between Q-S in the left, and Q-S in the right hemisphere, for musicians and nonmusicians. The results revealed that mean

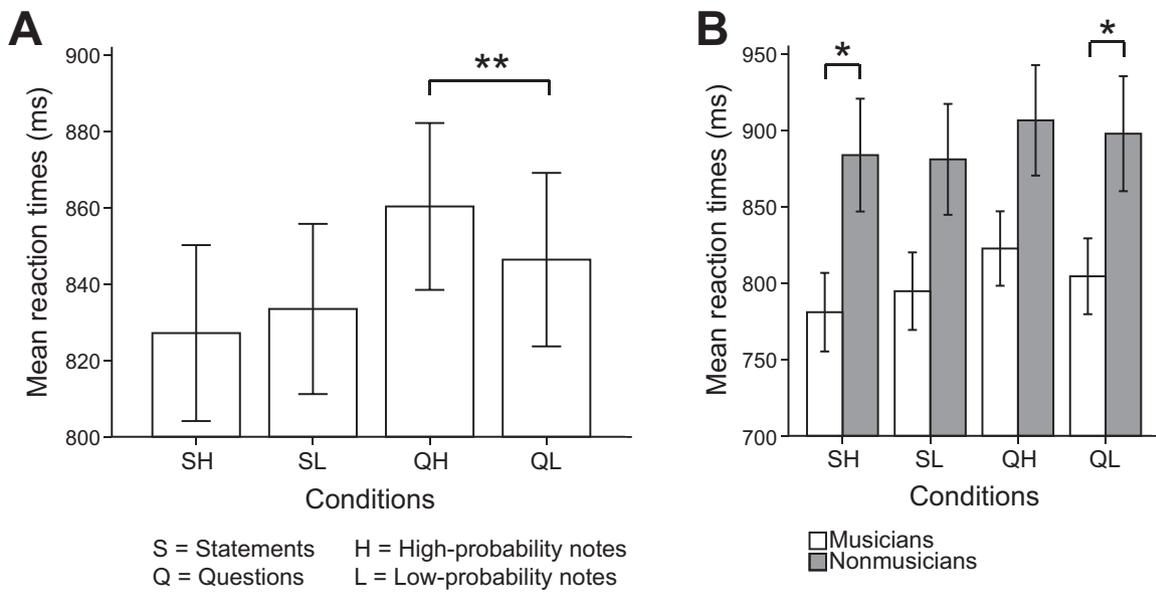


Fig. 1. (a) Bar chart of mean reaction times (ms) for each condition (SH=statements on high-probability notes, SL=statements on low-probability notes, QH=questions on high-probability notes, and QL=questions on low-probability notes). Double asterisks (**) denote statistical significance at $p < .01$; (b) Bar chart of mean reaction times (ms) for each condition for musicians (white) and nonmusicians (gray). Error bars represent ± 1 standard error mean (SEM).

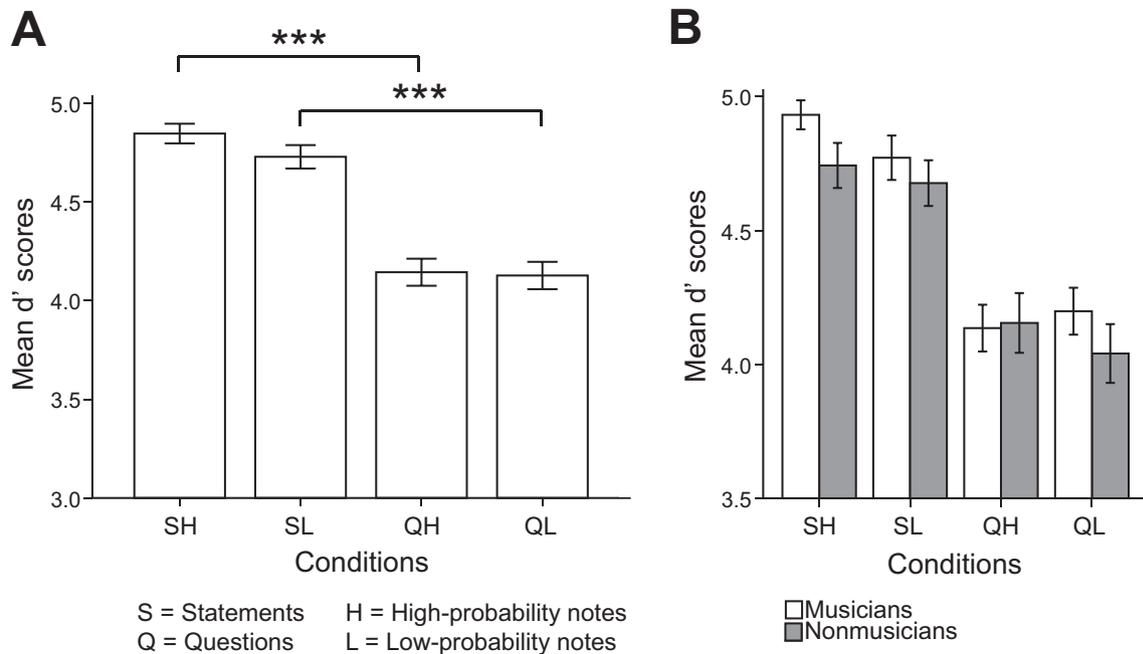


Fig. 2. (a) Bar chart of mean d' scores for each condition. Triple asterisks (***) denote statistical significance at $p < .001$; (b) Bar chart of mean d' scores for each condition for musicians (white) and nonmusicians (gray). Error bars represent ± 1 standard error mean (SEM).

amplitude difference between questions and statements was higher in the left hemisphere ($M = -.25$, $SE = .17$) ($p > .050$) than in the right hemisphere ($M = -.13$, $SE = .14$) ($p > .050$) for musicians, whereas it was lower in the left hemisphere ($M = -.18$, $SE = .17$) ($p > .050$) than in the right hemisphere ($M = -.40$, $SE = .14$) ($t(17) = -2.82$, $p = .012$) for nonmusicians. This interaction is due to the Q-S difference in the right hemisphere for nonmusicians.

2.2.2. P200 time window (100–200 ms)

Musicians showed lower ERPs in comparison to nonmusicians within the P200 time window, as confirmed by the significant main effect of *musical training* ($F(1,37) = 4.34$, $p = .044$, $\eta^2 = .11$).

There was no effect of *prosody* or *note-probability* or interaction between these factors ($F < 1.09$, $p > .304$). There was a main effect of *region* ($F(1,37) = 115.99$, $p < .001$, $\eta^2 = .76$), as well as a *region* \times *hemisphere* interaction ($F(1,37) = 7.59$, $p = .009$, $\eta^2 = .17$). In addition, the mixed ANOVA revealed a significant interaction between *prosody*, *note-probability*, and *hemisphere* ($F(1,37) = 5.24$, $p = .028$, $\eta^2 = .12$). This interaction was further analysed with planned contrasts between the left and right hemispheres in all conditions (SH left – SH right, SL left – SL right, QH left – QH right, QL left – QL right). The analysis revealed that the effect of hemisphere was important in conditions SL ($t(38) = 2.18$, $p = .035$) and

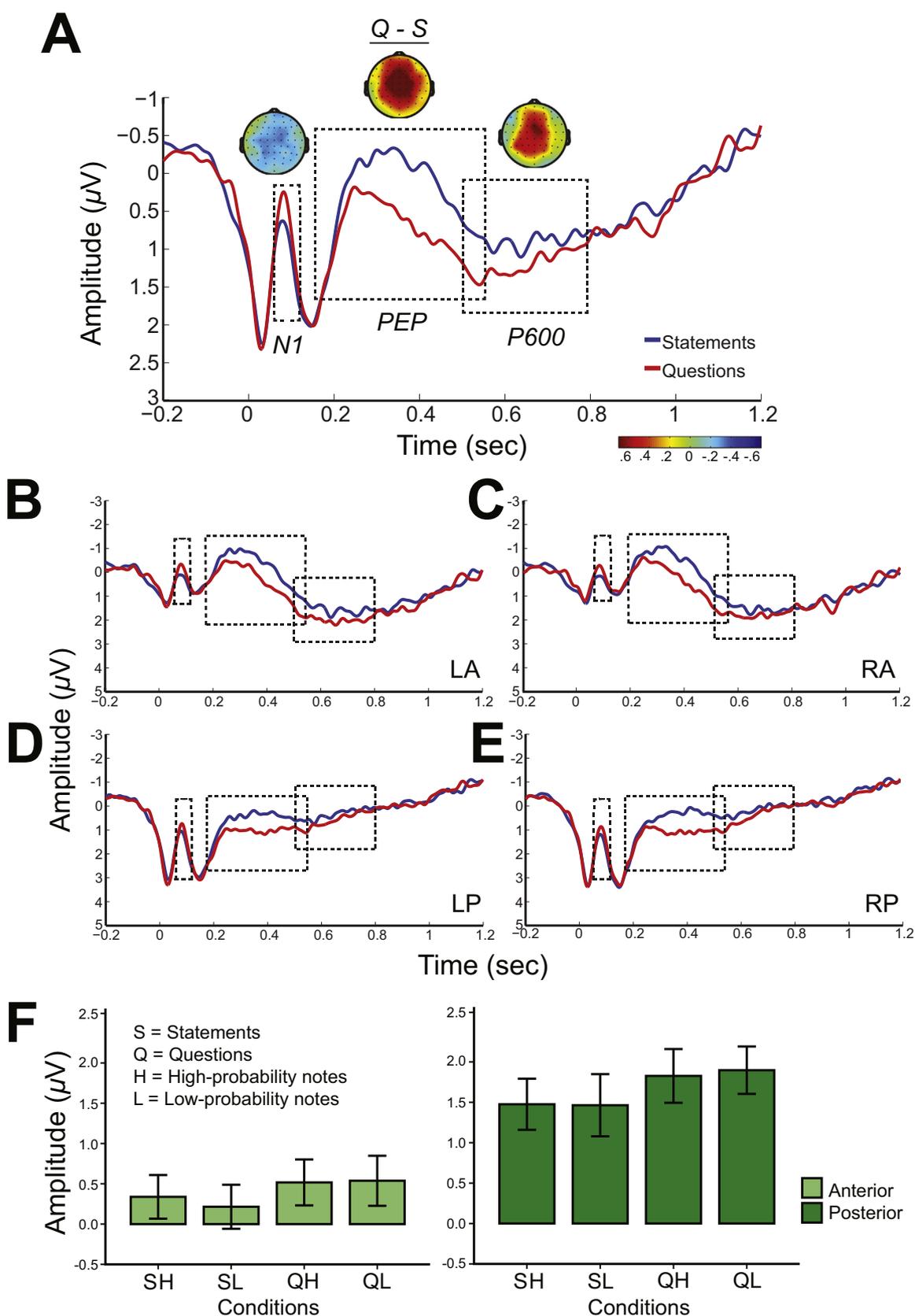


Fig. 3. (a) Grand average ERPs (all ROIs) following statements (blue) and questions (red). The time windows with a main effect of prosody are indicated by a rectangle (from left to right: N1, prosodic expectancy positivity ('PEP'), and P600). The difference scalp maps represent statements subtracted from questions, averaged across participants and electrodes. The grand average ERPs are displayed separately for the four ROIs in: (b) LA (left anterior), (c) RA (right anterior), (d) LP (left posterior) and (e) RP (right posterior); (f) Mean amplitudes averaged over the ROIs of anterior (light green) and posterior (dark green) electrode sites within the P600 time window in all four conditions. Error bars represent ± 1 standard error mean (SEM). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

QH ($t(38)=2.12, p=.041$), where the right was significantly more positive than the left hemisphere (SL (right-left): $M=.21, SE=.10$, QH (right-left): $M=.18, SE=.55$). Conditions SH and QL did not show significant differences between the two hemispheres ($p > .200$). In summary, in the P200 time window, the right hemisphere shows significantly higher ERP amplitudes when music and language expectancy is in a different direction, i.e. language is expected (statements) and music is unexpected (low-probability notes) (SL) or the opposite (questions on high-probability notes) (QH).

2.2.3. Prosodic expectancy positivity ('PEP') time window (150–550 ms)

As compared with statements, a larger ERP positivity between 150–550 ms, with fronto-central scalp distribution was observed for questions (see Fig. 4). The mixed ANOVA revealed a significant main effect of *prosody* ($F(1,37)=19.65, p < .001, \eta^2=.35$) and *region* ($F(1,37)=25.07, p < .001, \eta^2=.40$). Questions showed an enhanced PEP ($M=1.27, SE=.25$) in the anterior ROIs (LA, RA) compared to statements ($M=.75, SE=.22$), as well as in the posterior ROIs (LP, RP) (Q: $M=.40, SE=.20, S: M=-.48, SE=.16$). There was no effect of *note-probability* or *musical training* ($F < .26, p > .616$). The results confirmed the presence of an increased positivity in response to questions compared to statements in the PEP time window, which was more enhanced in the anterior sites.

Furthermore, the ANOVA revealed a significant interaction involving *prosody*, *note-probability* and *hemisphere* ($F(1,37)=4.67, p=.037, \eta^2=.011$) (Fig. 4d). Further analyses with planned contrasts between conditions QH–SH, and QL–SL, on the left and right hemisphere, revealed that mean amplitude difference between questions and statements was significantly higher on low-probability notes ($M=.66, SE=.20$) ($t(38)=3.25, p=.002$) than on high-probability notes ($M=.39, SE=1.05$) ($t(38)=2.32, p=.026$) at the left hemisphere, whereas this difference had the opposite direction (QL–SL: $M=.51, SE=.20, t(38)=2.61, p=.013$), QH–SH: $M=.58, SE=.17, t(38)=3.50, p=.001$) at the right hemisphere. In summary, low-probability notes in the left hemisphere elicit an enhanced PEP component compared to high-probability notes, whereas the opposite is true for the right hemisphere (high-probability notes are more positive than low-probability notes).

Finally, a significant interaction between *prosody*, *musical training* and *hemisphere* ($F(1,37)=8.24, p=.007, \eta^2=.18$) was found, which was further analysed with planned contrasts between Q–S in the left, and Q–S in the right hemisphere, for musicians and nonmusicians. The results revealed that mean amplitude difference between questions and statements was higher in the right hemisphere ($M=.51, SE=.18$) ($t(20)=2.90, p=.009$) than in the left hemisphere ($M=.28, SE=.16$) ($p=.087$, marg.) for musicians, whereas it was lower in the right hemisphere ($M=.58, SE=.19$) ($t(17)=3.06, p=.007$) than in the left hemisphere ($M=.81, SE=.21$) ($t(17)=3.80, p=.001$) for nonmusicians. In summary, for musicians the Q–S amplitude difference was higher in the right hemisphere, whereas for nonmusicians it was higher in the left hemisphere.

2.2.4. P600 time window (500–800 ms)

As with the PEP time window, ERP amplitudes in response to questions were shown to be more positive compared to statements, from 500 to 800 ms after the onset of the critical word/note (see Fig. 3). An ANOVA in this time window revealed significant main effects of both *prosody* ($F(1,37)=7.07, p=.012, \eta^2=.16$) and *region* ($F(1,37)=38.49, p < .001, \eta^2=.51$), but no effect of *note-probability* or *musical training* ($F < 1.76, p > .193$). As shown in Fig. 3f, positivity was enhanced in questions, and was also greater in the posterior ROIs.

Furthermore, the analysis revealed a significant interaction

between *prosody*, *note-probability* and *musical training* ($F(1,37)=5.62, p=.023, \eta^2=.13$). As shown from the difference topoplots for the P600 time window (Fig. 4b and c), the *prosody* \times *note-probability* interaction is mainly focused at mid-frontal sites. In order to compare differences between musicians and nonmusicians within this time window, we run a further 2 (*prosody*: statement vs. question) \times 2 (*note-probability*: high-probability vs. low-probability) \times 2 (*musical training*: musicians vs. nonmusicians) mixed ANOVA at the peak electrodes of the interaction (Fz, FCz, Cz, CPz, C1, C2) (see Fig. 5). Questions elicited significantly more positive responses compared to statements (main effect of *prosody*: $F(37)=14.93, p < .001, \eta^2=.29$). As expected, there was a significant interaction between *prosody*, *note-probability* and *musical training* ($F(37)=5.21, p=.028, \eta^2=.12$). In order to explore further this interaction, planned contrasts were carried out between conditions SH–SL, QH–QL, SH–QH and SL–QL, for musicians and nonmusicians. As shown in Fig. 5, in musicians the interaction was due to the difference between high-probability melodies (QH–SH: $t(20)=3.17, p=.005$), whereas in nonmusicians the difference was important between low-probability melodies (QL–SL: $t(17)=1.16, p=.004$), as well as questions (QL–QH: $t(17)=.82, p=.031$).

2.2.5. P800 time window (850–1200 ms)

A significant *note-probability* \times *musical training* interaction was observed ($F(1,37)=4.61, p=.039, \eta^2=.11$), which was further analysed with planned contrasts between high- and low-probability notes for the two groups. The results showed that mean ERP amplitudes for musicians were higher in response to high- compared to low-probability notes, whereas the opposite effect was observed for nonmusicians. However, neither of the contrasts was significant ($p > .535$). The ANOVA also revealed a significant interaction between *prosody* and *note-probability* ($F(1,37)=5.91, p=.02, \eta^2=.14$) (Fig. 4e). In order to explore further this interaction, planned contrasts were carried out between conditions SH–SL, QH–QL, SH–QH and SL–QL. Note-probability was found to be important only in statements (SH–SL: $t(38)=1.98, p=.055$), but not in questions (QH–QL: $p > .05$). Prosody made a marginally significant difference in the unexpected conditions (SL–QL: $t(38)=-1.80, p=.080$), but not in the expected conditions ($p > .05$). The time course of music and language interactions in the consecutive time windows mentioned above is illustrated in Fig. 4. There was no effect of *prosody* or *note-probability* ($F < .29, p > .594$). A main effect of *region* was found ($F(37)=40.00, p < .001, \eta^2=.52$).

Furthermore, we observed large variability in the P800 amplitudes across participants (as shown from the error bars in Fig. 4e). In order to further investigate this observation, we run a Pearson product-moment correlation between participants' (musicians and nonmusicians) Gold-MSI Musical Training scores and their mean ERP amplitudes. Interestingly, significant correlations were found between the ERP amplitudes of musicians and their level of musical training (SH: $r=.571, p=.007$; SL: $r=.525, p=.015$; QH: $r=.377, p=.092$ (marginal); QL: $r=.527, p=.014$): the higher the level of musical training, the more positive was the ERP response. In contrary, no significant correlations were found between nonmusicians' Gold-MSI scores and their ERP amplitudes ($r < .200$ and $r > -.200, p > .05$). The strongest positive correlation between musicians' amount of musical training and their ERP amplitude was observed for condition SH, and is illustrated in Fig. 4f.

2.2.6. ERPs and reaction times association

Potential associations between ERP components and reaction times (RTs) in the statement/question discrimination task were subsequently examined. More specifically, we wanted to investigate whether any specific ERP component could successfully predict faster RTs when questions were paired with low-

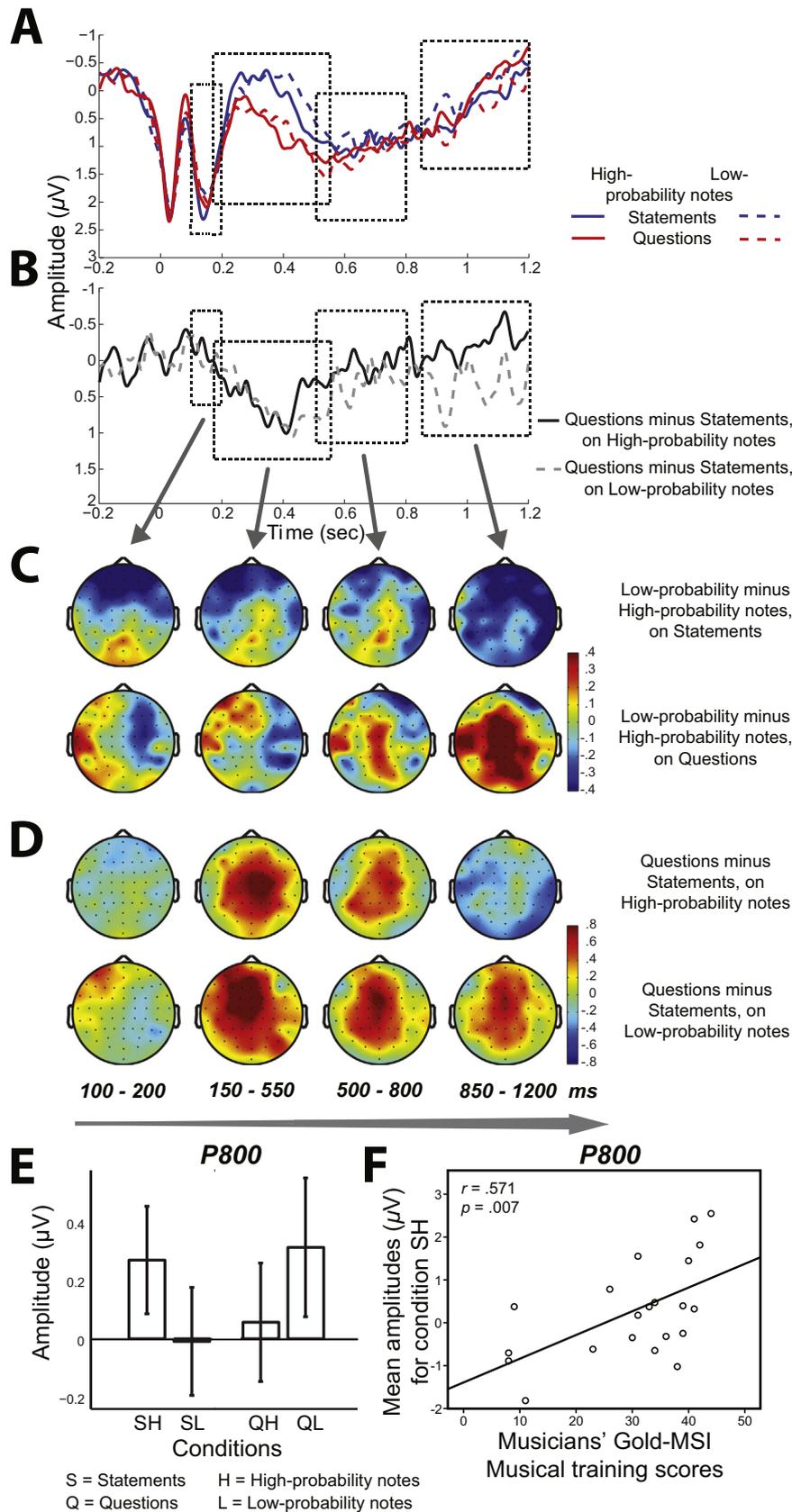


Fig. 4. (a) Grand average ERPs (all ROIs) for all conditions: statements on high-probability notes (solid blue line), questions on high-probability notes (dotted blue line), statements on low-probability notes (solid red line), and questions on low-probability notes (dotted red line). Music-language interaction time windows are indicated with a rectangle (P200, 'PEP', P600, and P800); (b) Difference grand average ERPs (all ROIs) for statements subtracted from questions on high-probability notes (QH-SH) (dotted black line), and statements subtracted from questions on low-probability notes (QL-SL) (dotted gray line). The scalp topographies corresponding to the differences between conditions are presented in: (c) Low-probability minus high-probability notes on statements (SL-SH), low-probability minus high-probability notes on questions (QL-QH), and (d) questions minus statements on high-probability notes (QH-SH), questions minus statements on low-probability notes (QL-SL). These represent averages across participants and electrodes; (e) Mean amplitudes averaged over ROIs within the P800 time window in all four conditions. Error bars represent ± 1 SEM; (f) Correlation between musicians' amount of musical training (Gold-MSI Musical Training score) and their mean ERP amplitudes (condition SH) within the P800 time window. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

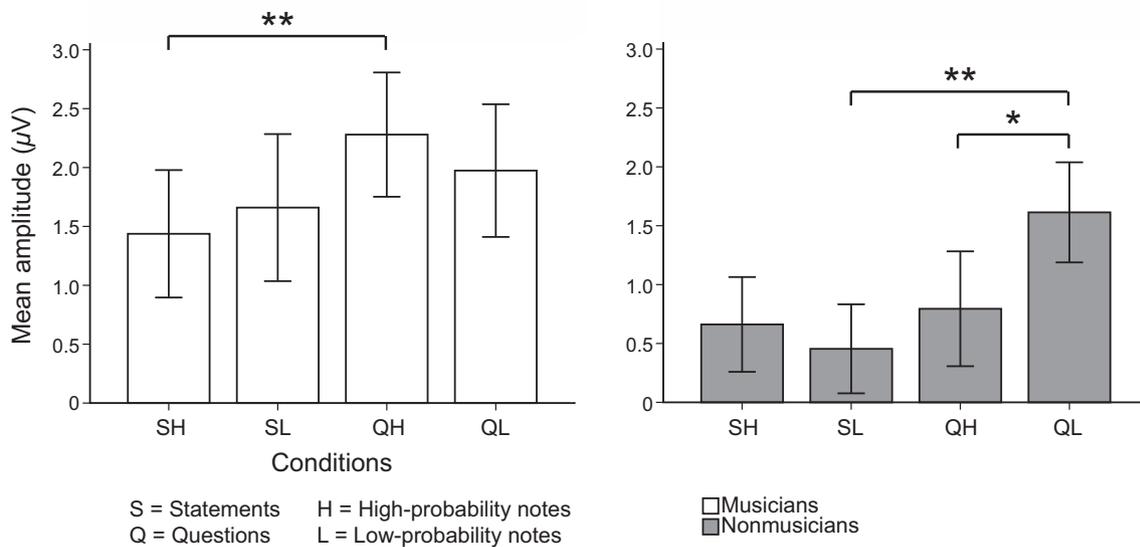


Fig. 5. Error bars with the mean ERP amplitude values for all conditions in the P600 time window (500–800 ms) (from left to right: statements on high-probability notes (SH), statements on low-probability notes (SL), questions on high-probability notes (QH), and questions on low-probability notes (QL)), for musicians (white) and non-musicians (gray). Error bars represent ± 1 SEM.

probability notes (QL) than when questions were paired with high-probability notes (QH) (QH–QL: $t(39)=2.93$, $p=.006$). We assumed that an ERP component which predicts RTs should precede in time the lowest RT observed (i.e. ERPs occurring after an RT cannot be used to predict it). The fastest RT across participants in the QH and QL conditions was 608.53 ms. Therefore, we examined potential associations between RTs and the ERPs preceding this minimum RT: N1 (60–130 ms), P200 (100–200 ms) and PEP (150–550 ms).

Linear backward regression was performed to predict RTs from each of the three ERP components mentioned above. The mean amplitude difference between conditions QL and QH across participants in each of the four ROIs (LA, LP, RA, RP) was the predictor variable. The mean difference of RTs between conditions QL and QH across participants was the dependent variable. Neither N1 nor PEP time-window QL–QH amplitudes significantly predicted the difference in RTs ($p > .220$). Interestingly, regression models using the P200 significantly predicted RTs. Specifically, a model with RA, RP and LA ROIs as predictors significantly predicted RTs ($R^2=.21$, $F(3,34)=3.02$, $p=.043$). Only the RP ROI was found to be significant for the prediction ($p=.009$). Another model including only the right ROIs (RA, RP) predicted more significantly the dependent variable ($R^2=.20$, $F(2,35)=4.47$, $p=.019$). Here both RA and RP ROIs added significantly to the prediction ($p=.017$, and $p=.009$, respectively). All four ROIs as predictors showed a marginally significant model ($R^2=.22$, $F(4,33)=2.30$, $p=.080$) (Fig. 6), in which none of the four ROIs was statistically significant ($p > .216$).

3. Discussion

The present study investigated the interactions between melodic expectancies and prosodic expectancies (statements, questions) using a statement/question discrimination task in musicians and nonmusicians, and demonstrated their behavioural and neural correlates. Behavioural results showed that musicians had superior performance to nonmusicians, providing evidence for transfer of pitch processing abilities from the music to the speech domain. At the neural level, questions were associated with increased N1 negativity compared to statements, as well as with a late fronto-central positivity (prosodic expectancy positivity), reflecting prosodic re-analysis processes. Furthermore, when violations

occurred in both music and language (double violation: questions on low-probability notes), a linear regression model significantly predicted faster reaction times from the corresponding ERPs within the P200 time window. Importantly, musicians showed lower P600 in the double violation condition, suggesting the usage of fewer neural resources to process strong pitch incongruities. Critically, musicians' P800 amplitudes were proportional to their level of musical training, suggesting that this expertise might shape the pitch processing of language. We speculate that the latter beneficial aspect of musical training could be explained by its intrinsic strengthening effect on general executive functions (e.g., attention and working memory).

In this section, we will first examine the effect of prosody on task performance and on neural responses to statements and questions. Then, we will focus on the interactions between prosody and note-probability, and finally, the role of musical expertise on pitch processing and cognition will be discussed.

3.1. Effect of prosody

Overall, statements were recognised faster and more accurately across groups, reflecting a bias towards expected prosody. Similar findings, in a statement/question discrimination task, have been previously reported in different languages (Ma et al., 2011; Peters and Pfitzinger, 2008). Moreover, higher recognition accuracy towards grammatically (Colé and Segui, 1994) or syntactically correct sentences (Hoch et al., 2011; Slevc et al., 2009) has been previously associated with the unexpectedness of the incorrect sentences. Similarly, higher accuracy has been observed for semantically neutral sentences as opposed to sentences with a violation of emotional or linguistic prosody (Paulmann et al., 2012). Noteworthy, in the aforementioned studies, expectancy was expressed by correct/incorrect sentences, whereas in our study for the same purpose we used statements/questions. This methodological difference might underlie not exactly comparable levels of expectancy, as incorrect sentences might be perceived as more unexpected than questions; however we assumed that our methodological choice would allow us to study the effect of language expectancy in a categorical fashion (expected vs. unexpected).

Although we did not have any specific hypothesis about early ERP responses, the results showed increased negativity within the

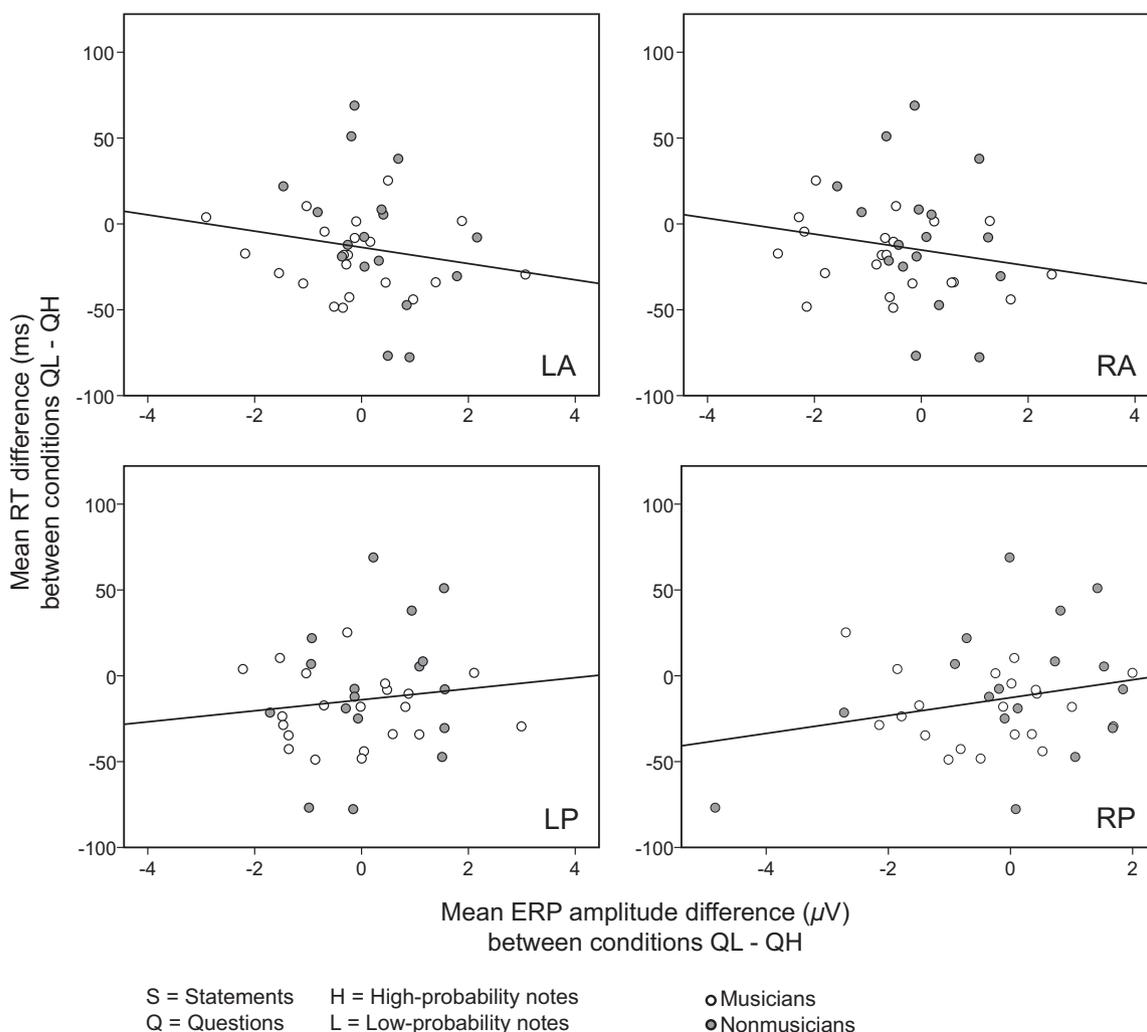


Fig. 6. Correlations between mean ERP amplitude differences (μV) between conditions QL (questions on low-probability notes) and QH (questions on high-probability notes), and mean RTs (reaction times) (ms) between the same conditions in the behavioural task, for musicians (white) and nonmusicians (gray). The scatterplots correspond to the four ROIs (LA=left anterior, LP=left posterior, RA=right anterior, RP=right posterior).

N1 component in response to questions, which is in line with previous findings on unexpected events (Debener et al., 2005; Horváth et al., 2008; Proverbio et al., 2004). As predicted, questions elicited a larger positivity at around 150–550 ms after the onset of the last word/note with a fronto-central scalp distribution. A similar ERP has been associated with violations of prosodic expectancy (prosodic expectancy positivity: PEP) (Kotz and Paulmann, 2007; Paulmann et al., 2012). Previous contradictory findings on the lateralisation of the PEP component (Chen et al., 2011; Paulmann and Kotz, 2008), could be attributed to the different type of prosodic manipulation (linguistic or emotional), as well as to the different task demands (Wildgruber et al., 2004). With regards to the essence of the PEP deflection, we argue that it is associated with re-analysis processes of prosodically unexpected utterances (Astésano et al., 2004; Paulmann et al., 2012; Paulmann and Kotz, 2008).

3.2. Interactions between prosody and note-probability

As predicted, reaction times across groups were shorter when expectancy violations in music and language occurred simultaneously. Specifically, responses to statements were faster on high- than low-probability notes, and the opposite effect was observed for questions. This interaction might be due to facilitation effects between music and language processing of expectation. Expected

musical chords (tonics) have been previously found to facilitate processing of concurrent language stimuli (phonemes), which constitutes the 'tonal function effect' (Bigand et al., 2001; Escoffier and Tillmann, 2008). Although the present study did not involve the use of harmony, the similar findings could be attributed to the fact that melodies can also engage harmonic processing by implying an underlying chord sequence (Giordano, 2011; Koelsch, 2012).

The size of the final interval in our melodic stimuli might constitute a limitation, as the majority of the high-probability melodies had small intervals, whereas low-probability melodies had an equal amount of small and large intervals (see Section 4.2). For example, small intervals are more frequent than large intervals in Western tonal music (Krumhansl, 1995; Vos and Pasveer, 2002; Vos and Troost, 1989). Therefore, small intervals might constitute a feature *sine qua non* melodies are usually built, thus constituting an intrinsic characteristic of high-probability melodies. Faster reaction times in response to high-probability intervals may thus be related to the high frequency of small intervals. However, besides the actual pitch difference, perceived interval size has been found to depend on the level of musical training, the melodic direction, as well as whether the interval is larger than an octave (Russo and Thompson, 2005).

The faster reaction times in the double violation condition (questions on low-probability notes) were successfully predicted

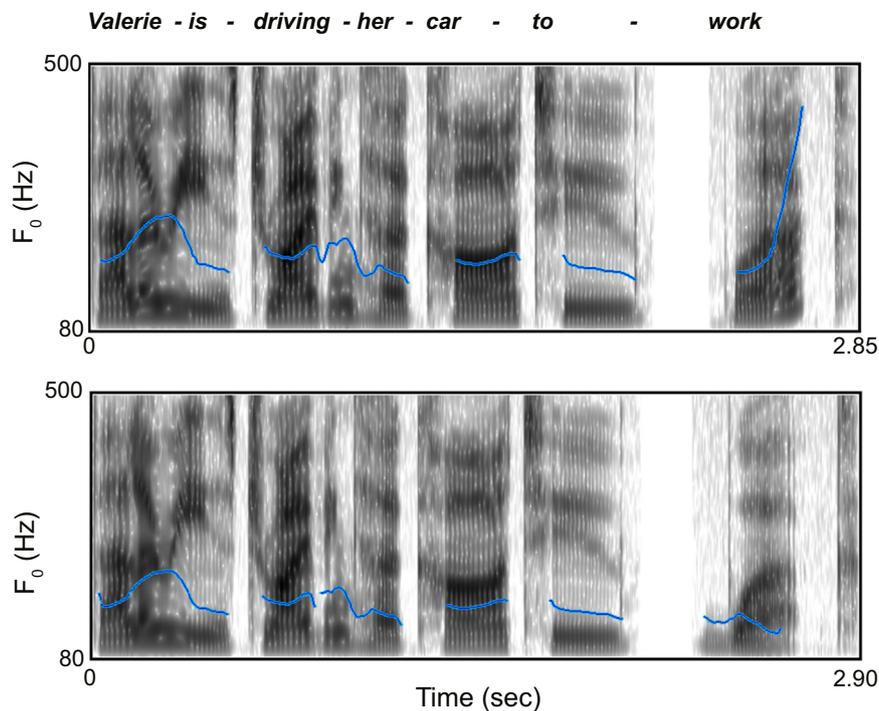


Fig. 7. Two paired stimuli illustrating the different intonation conditions in the experiment. These are the voice spectrograms with their fundamental frequency contours (blue line). Top: the fundamental frequency F_0 (Hz) of the final word rises, thus forming a question (*Valerie is driving her car to work?*); bottom: the fundamental frequency F_0 (Hz) of the final word falls, thus forming a statement (*Valerie is driving her car to work.*). Figures were created using the PRAAT software (Boersma, 2001). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from the corresponding P200 component in the right ROIs, as shown from a linear regression model we implemented. Therefore, we suggest that the P200 might facilitate pitch expectancy processing in music and language. Future research could assess this hypothesis in a similar simultaneous presentation paradigm, investigating violation of grammar or syntax, instead of linguistic prosody.

Our main hypothesis was related to the investigation of the neural effect of melodic violations on prosodic processing. Interestingly, we observed neural interactions in consecutive time windows, from the P200 until the P800, suggesting interdependencies at an early stage (linked to sensory and perceptual processing (Hillyard et al., 1973; Näätänen et al., 1978)), as well as at a later stage (integration and re-analysis processes (Eckstein and Friederici, 2005)). The P800 showed the largest positivity when music and language expectancy were in the same direction, i.e. both expected (statements on high-probability notes) and both unexpected (questions on low-probability notes). The latter condition (double violation) showed the highest amplitude overall (not significant), which could be linked to fundamental frequency re-analysis processes of unexpected intonation (Astésano et al., 2004). Although Astésano et al. (2004) linked the P800 to prosodic manipulations, considering the music-language interaction observed, we argue for its amodal rather than language-specific role in detecting expectancy violations.

Previous studies have interpreted music-language interactions as a competition of neural resources used for the simultaneous processing of their syntactic elements (Carrus et al., 2013; Koelsch and Gunter, 2005). Therefore, interactions in the pitch dimension in a simultaneous presentation paradigm would suggest that melodic and prosodic expectancy are interdependent. Our study is the first to find this effect, favouring the possibility that music and language share access to a limited pool of resources for pitch processing. In particular, recent research suggested that these shared resources between music and language for pitch processing

are attributed to more general higher-order cognitive mechanisms (e.g., working memory (WM)), rather than specialised lower-level pitch perception processes (Marie et al., 2011; Moreno and Bidelman, 2014; Smayda et al., 2015). In support of this hypothesis, fundamental structures of WM, such as Broca's area, premotor cortex, pre-SMA/SMA, left insular cortex and inferior parietal lobe, are involved in both tonal and verbal WM (Schulze et al., 2011). Accordingly, in an fMRI study Janata et al. (2002) have shown that attentive listening to music recruits neural circuits serving general functions, such as attention, WM, and semantic processing. Strong evidence favouring shared higher-order, executive functions have reported enhanced attention and WM associated with musical expertise, suggesting that this strengthening of executive functions is responsible for the speech processing benefits (Besson et al., 2011; Bialystok and Depape, 2009; Carey et al., 2015; George and Coch, 2011; Parbery-Clark et al., 2012; Patel, 2014; Rigoulot et al., 2015; Smayda et al., 2015; Strait et al., 2010; Strait and Parbery-Clark, 2013). In the next Subsection 3.4 we will discuss the effects of musical training on melodic and speech pitch processing considering the potential role of general functions.

3.3. Effects of musical training

Confirming our hypothesis on musical expertise, musicians showed overall better performance in the statement/question discrimination task. They were significantly faster and showed a trend for higher accuracy compared to nonmusicians, which is in line with studies on increased pitch sensitivity for musicians in linguistic tasks (Alexander et al., 2005; Deguchi et al., 2012; Magne et al., 2003, 2006; Schön et al., 2004). For example, it has been shown that musicians are more accurate in detecting pitch violations not only in music, but also in language (Magne et al., 2003), and show better discrimination abilities between weak incongruities and congruities in both domains (Besson and Faïta, 1995; Magne et al., 2006; Schön et al., 2004). It is therefore likely that

musicians are able to transfer their musical pitch processing abilities to speech pitch tasks, due to common underlying pitch processing mechanisms.

At the neural level, musicians showed overall lower P200 amplitude (linked to attention processes (Shahin et al., 2003)) compared to nonmusicians. Although this contradicts previous findings on training effects reporting enhanced P200 in musicians (Atienza et al., 2002; Tremblay et al., 2001), evidence from other cognitive domains demonstrated lower amplitude in the early ERP components explained as less attentional effort needed (Berry et al., 2009, 2010). That is, musicians might require reduced attentional demands (lower P200) to out-perform nonmusicians in the behavioural task (higher reaction times and accuracy), suggesting greater efficiency at prosodic pitch processing. Another possible explanation of this effect could be related to the allocation of attention, as participants were asked to focus on the speech and ignore the music. Future research could investigate this hypothesis by instructing the participants to rate the melodic endings while ignoring the language.

Importantly, we observed a neural interaction between prosodic and melodic violation and musical training, in the P600 time window. Specifically, musicians showed an overall larger positivity compared to nonmusicians. This is in line with previous literature related to harmonic violations (Besson and Faïta, 1995; Featherstone et al., 2013; Regnault et al., 2001), and prosodic or melodic violations (Schön et al., 2002). For example, Regnault et al. (2001) found enhanced late positivities elicited by dissonant chords in musicians compared to nonmusicians. Critically, we found that strong incongruities (questions on low-probability notes) elicited smaller P600 than weaker incongruities (questions on high-probability notes) in musicians, whereas nonmusicians showed the opposite pattern (non significant). This trend confirms previous studies which demonstrated that strong music and language incongruities elicit lower P600 amplitudes after auditory pitch training in musicians, but not in nonmusicians (Besson et al., 2007; Moreno and Besson, 2006). Considering that P600 reflects working memory demands (Gibson, 1998), we suggest that musicians need less neural resources to process and integrate strong pitch incongruities (Featherstone et al., 2013; Tatsuno and Sakai, 2005). In contrary, nonmusicians find simultaneous violations of expectations more demanding and difficult to integrate, due to lower working memory capacity. Therefore, we speculate that pitch processing might become automatic in musically trained people.

Further analysis revealed that Gold-MSI musical training scores of musicians (but not of nonmusicians) correlated positively with their P800 amplitudes: the higher the level of musical training, the more positive was the ERP response. This finding might provide evidence for neuroplastic changes in the pitch domain due to musical expertise (Moreno and Bidelman, 2014; Pantev et al., 2001; Ridding et al., 2000; Schlaug et al., 1995; Steele et al., 2013; Wan and Schlaug, 2010).

To sum up, we propose that expertise-related effects might result in lower-level perceptual benefits, as well as higher-order cognitive enhancement. In particular, one possibility is that musical training enhances pitch perception, and such improvement may be mediated by tuning of neurons in auditory cortical regions (Schneider et al., 2002) and the brainstem (Bidelman et al., 2011; Wong et al., 2007). Another possible explanation is that expertise results in more efficient suppression of task-irrelevant auditory stimuli. Thus, musicians could better inhibit the musical stimuli while focusing on the speech (as they were instructed), which had an impact on their allocation of attention (lower P200). This is in line with evidence about musicians demonstrating benefits in sound segregation, such as speech-in-noise conditions (Parbery-Clark et al., 2009) and the “cocktail party problem” (Swaminathan et al., 2015; Zendel and Alain, 2009). Therefore, successful

inhibition of task-irrelevant stimuli might constitute one of the mechanisms of improved cognitive control following expertise.

3.4. Conclusion

Our findings suggest that melodic expectancy influences the processing of language pitch expectancy. We reveal that musical expertise modulates the nature of such influence, by facilitating the processing of unexpected events, and by providing a more refined response to pitch not only in music, but also in language. Critically, musicians' neural responses were found to be proportional to their level of musical expertise, suggesting that expertise shapes prosodic processing. Therefore, these results provide evidence for the beneficial effects of musical training on general cognitive functions (e.g., allocation of attention, working memory), during the simultaneous processing of expectancy violations in music and language. We suggest that these findings have implications for investigating potential shared higher-order mechanisms between music and language.

4. Methods and materials

4.1. Participants

Forty-two neurologically healthy adult human volunteers (25 female) aged between 18 and 37 years old with mean \pm s.d. age of 23.79 ± 4.25 participated in a behavioural and an EEG experiment. All participants were native speakers of English (L1), not tested for potential second language proficiency, with normal hearing and normal or corrected-to-normal vision (self-reported). Participants were divided into two groups according to their self-reported level of musical training: musicians (22 subjects, mean age of 22.59 years, 15 female) which had a mean \pm s.d. Gold-MSI score of 35.91 ± 6.80 , and nonmusicians (20 subjects, mean age of 24.89 years, 10 female) which had a mean \pm s.d. Gold-MSI score of 17.80 ± 10.12 . The ‘Goldsmith's Musical Sophistication Index’ (Gold-MSI) questionnaire was administered to validate participants' self-reported musicality (Müllensiefen et al., 2014). Participants gave written informed consent in accordance with procedures approved by the local ethics committee of the Department of Psychology at Goldsmiths, and received a monetary compensation for their participation.

4.2. Materials

The Gold-MSI assesses musical engagement and behaviour. For our group validation, we used the ‘Musical Training’ factor, the Gold-MSI Dimension 3, which includes seven statements regarding formal musical training experience and musical skill, and has a possible score of 7–49 points. Each statement (e.g., ‘I have never been complimented for my talents as a musical performer’) requires a response on a 7-point scale ranging from 1 (*Completely Disagree*) to 7 (*Complete Agree*).

Linguistic stimuli consisted of 200 seven-word utterances equally divided into two conditions: (i) statements, if the last word had a falling pitch, and (ii) questions, if it had a rising pitch. Therefore, the two conditions differed only in the final pitch (see Fig. 7 for an illustration of the intonation contours produced by the statement and question version of a typical utterance). One prominent global rule governing statement/question discrimination is that listeners tend to perceive utterances with a final pitch rise as questions, whereas utterances with a final fall as statements (Studdert-Kennedy and Hadding, 1973). Questions that are syntactically declarative sentences are commonly used (besides *wh*- and *yes/no* questions) in different English dialects, as well as

American English (Bartels, 2014; Grabe et al., 2003; Grabe, 2004). The absence of word-order change in declarative questions results in statement/question discrimination judgments based mainly on the final pitch contour, which is typically rising (Bartels, 1997; Bartels, 2014). As suggested by Ma et al. (2011), listeners' expectations are biased towards the perception of statements: first, because in English language statements are more frequently used than declarative questions (Bartels, 2014; Grabe, 2004), and, second, because declarative questions do not possess any word change (no inversion or *wh-* word) and are, thus, syntactically identical to statements (Bartels, 1997; Bartels, 2014; Grabe et al., 2003). Because it is less likely that a declarative syntax ends with interrogative intonation, we therefore considered that statements are more expected compared to questions.

The language stimuli were recorded by a female opera singer experienced in musical theatre and also a native speaker of British English. She was asked to pronounce the utterances in a natural way. The recording took place in a soundproof booth and a Zoom H4n recorder was used for this purpose (mono, 44.1 kHz, 24-bit recording). The 200 spoken sentences were recorded in pair, so that they were lexically identical but differed in prosody. Using WaveLab software, sentences were normalised to the same amplitude (the lower-amplitude sentence of a statement-question pair was scaled up in amplitude to match the higher). In order to control for differences between statements and questions other than the pitch of the final word, half of the statements were considered 'stems' for the paired questions, the last words of which were attached to them. Half of the questions were 'stems' for their paired statements (following the method of Patel et al., 1998a). PRAAT software was used for this purpose.

Musical stimuli consisted of 200 five-note isochronous melodies ending either with a high-probability or a low-probability note (Carrus et al., 2013), created using Pearce's (2005) computational model of melodic expectation. Specifically, the model created low-probability final notes for high-probability melodies, by choosing a new pitch with a probability lower than that of the actual continuation. Half of the final notes of the low-probability melodies produced were preceded by large intervals of six or more semitones, whereas the rest were preceded by small intervals of less than six semitones (Carrus et al., 2013). In order to investigate potential differences between note-probability (high vs. low) and interval size (small vs. large), we performed a chi-square test which revealed a significant relationship between the two variables ($\chi^2(1)=55.64, p < .001$). Specifically, only 7% of the high-probability melodies had large final intervals, in contrast to low-probability melodies of which 56% had large intervals. This intrinsic characteristic of high-probability melodies to use small final intervals is confirmed by studies revealing that the majority of closures of melodic sequences in Western tonal music consist of one or two semitones (Krumhansl, 1995; Vos and Pasveer, 2002).

The linguistic and melodic stimuli were played binaurally, both left and right, and had the same volume (Fig. 8). The presentation time was 600 ms for each of the first four musical notes and 1200 ms for the final note. As the temporal distribution of syllables is not isochronous in speech, an isochronous presentation of the sentences' words would result in utterances sounding unnatural (Geiser et al., 2008). Care was taken so that the onset of the final note coincided with the onset of the utterances' final word by inserting a 200 ms pause between the penultimate and the final word. Overall, there were 400 trials, as each of the 200 linguistic stimuli was presented twice: once combined with the 100 high-probability melodies and once with the 100 low-probability melodies. The pairing of a specific linguistic stimulus with a melodic stimulus, as well as the presentation order, was randomised across participants. During the simultaneous presentation of the auditory language and melodic stimuli, a fixation cross was presented at the

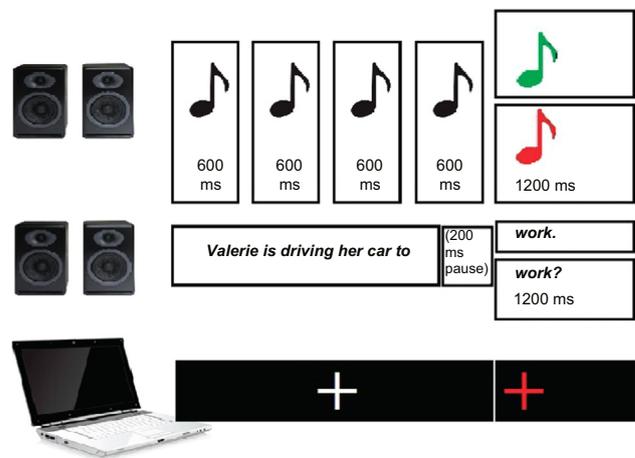


Fig. 8. An illustration of the experimental design. Speech and melodies are presented simultaneously via speakers. Participants listened simultaneously to seven-word sentences and five-note melodies. The linguistic stimuli ended with a falling pitch (statements) or with a rising pitch (questions). Melodies ended with either a high- or a low-probability note. The onset of the final word was the event of interest in the analysis. A fixation cross in the centre of the screen was shown during stimuli presentation, and turned red at the onset of the last word/note.

centre of the screen. Two speakers (Creative Gigaworks, Creative Technology Ltd.) were used for the stimuli reproduction, and the volume was kept constant across participants and for the duration of the experiment. Cogent 2000 (www.vislab.ucl.ac.uk/cogent.php), a MATLAB (MATLAB 7.12.0, The MathWorks, Inc., Natick, MA, 2011) Toolbox was used to present the stimuli.

4.3. Procedure

All participants took part in two separate experiments: one EEG and one behavioural experiment. Participants completed first the EEG and then the behavioural task in order to avoid neural habituation to the stimuli; a 15-min break was provided between the two sets of experiments.

In order to investigate whether participants' neural responses changed depending on the type of melody (expected or unexpected) the utterances were paired with, their EEG was recorded. At the beginning all participants completed the Gold-MSI questionnaire. Then they were seated in front of a computer in a dimly lit room. Through written instructions, they were informed that they would listen simultaneously to speech and melodies. They were instructed to attend only to the speech, ignoring the music. They were informed about the different sentence types, but not about the different melody types. For the EEG experiment, they were prompted only for 10% of trials to indicate whether the spoken sentence they heard was a statement or a question by pressing two buttons of a response box. The inter-trial interval was randomised between 1.5 and 2 s. Two practice trials (one statement, and one question) familiarised them with the task. Breaks were provided after each of the four blocks of 100 trials (about 12 min). Across participants the presentation order of the trials was randomised, and each sentence was randomly paired with a melody. At the end of the EEG, participants performed in the behavioural version of the experiment in which almost identical procedures were followed, except participants had to indicate at every trial as fast and accurate as possible the type of sentence (statement or question), by pressing two keys in the keyboard. The overall procedure lasted for approximately 2 h.

4.4. EEG recording and pre-processing

The EEG signals were recorded with sixty-four Ag-AgCl

electrodes placed according to the extended 10–20 electrode system (Jasper, 1958) and amplified by a BioSemi ActiveTwo amplifier (www.biosemi.com). The vertical and horizontal EOGs were recorded in bipolar fashion, in order to monitor eye-blinks and horizontal eye-movements. The EEG recording used a sampling frequency of 512 Hz, and the signals were band-pass filtered between .16 and 100 Hz. MATLAB Toolbox EEGLAB (Delorme and Makeig, 2004) was used for data preprocessing, and FieldTrip (Oostenveld et al., 2011) for data analysis. EEG data were re-referenced to the algebraic mean of the right and left earlobe electrodes (Essl and Rappelsberger, 1998). Continuous data were high-pass filtered at .5 Hz and then epoched from –1000 ms to 2000 ms time-locked to the onset of the last word/note. Artefact rejection was done in a semi-automatic fashion. Specifically, independent component analysis was run to correct for eye-blink related artefacts. Data from electrodes with consistently poor signal quality, as observed by visual inspection and by studying the topographical maps of their power spectra, was then removed and reconstructed by interpolation from neighbouring electrodes (5.65% of the data). Subsequently, epochs containing amplitude exceeding $\pm 85 \mu\text{V}$ were removed after visual inspection. Three participants were removed due to poor EEG data quality (more than 25% of the trials rejected). Additional preprocessing included low-pass filtering the epoched data at 30 Hz, and baseline correcting to 200 ms prior to last word/note onset.

4.5. Statistical analysis

Behavioural data: As the percentage of correct responses would constitute a biased measure of accuracy, signal detection theory was used to score discriminability (d' prime scores, or d') at the statement/question task (Macmillan and Creelman, 2005). Hits (correctly recognised statements) and false alarms (falsely recognised questions) were calculated for each experimental condition across participants (statements on high-probability notes ('SH'), statements on low-probability notes ('SL'), questions on high-probability notes ('QH'), and questions on low-probability notes ('QL')). All 100% and 0% scores were altered to 99.50% and .50%, in order to correct for ceiling and floor effects, respectively. Mean reaction times and d' accuracy scores were calculated for each condition across participants. Reaction times were analysed for correct trials only; trials with reaction times above and below two standard deviations were considered as outliers and removed from subsequent analysis (Ratcliff, 1993). First, a $2 \times 2 \times 2$ mixed ANOVA was performed with *prosody* (statement, question) and *note-probability* (high, low) as within-subjects factors, and *musical training* (musicians, nonmusicians) as the between-subjects factor. In order to further explore the *prosody* \times *note-probability* interaction, planned contrasts were run between conditions. As the accuracy scores were non-normally distributed ($p < .05$, Shapiro-Wilk test), non-parametric test (Wilcoxon signed-rank) test was used for planned contrasts in d' scores. All follow-up tests were Bonferroni corrected. Two outliers were identified by inspection of boxplots. One-sample t tests confirmed that the accuracy of these two participants was significantly different compared to the accuracy of the rest of the participants in the QE condition ($t(39) = 22.02, p < .001$) and the QU condition ($t(39) = -16.10, p < .001$), respectively.

ERP data: After Carrus et al. (2013), we had four regions of interest (ROIs): right anterior (RA) (F6, FC4, F4, F2, FC2, FC6), left anterior (LA) (F3, F5, F1, FC3, FC5, FC1), right posterior (RP) (P6, PC4, P4, P2, PC2, PC6), and left posterior (LP) (P5, PC5, PC1, P3, P1, PC3). The following time windows were used for the analysis, based on previous literature (Astésano et al., 2004; Carrus et al., 2013; Eckstein and Friederici, 2005; Paulmann et al., 2012; Pinheiro, Vasconcelos, Dias, Arrais, & Gonçalves, 2015) and visual

inspection of the ERPs: N1 (60–130 ms), P200 (100–200 ms), prosodic expectancy positivity ('PEP') (150–550 ms), P600 (500–800 ms), and P800 (850–1200 ms). Mean ERP amplitudes were calculated at each of the ROIs and individual time windows. As the ANOVA assumptions were met, a mixed factorial ANOVA was performed separately for each time window with the following five factors: *prosody* (statement, question), *note-probability* (high, low), *hemisphere* (left, right), *region* (anterior, posterior), and *musical training* (musicians, nonmusicians).

All statistical analyses were carried out using the IBM Statistical Package for the Social Sciences (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.).

Acknowledgements

The authors are supported by the CREAM project that has been funded by the European Commission under Grant Agreement no 612022. This publication reflects the views only of the authors, and the European Commission cannot be held responsible for any use which may be made of the information contained therein. We thank Dr Marcus T. Pearce for providing the melodic stimuli and Mara Golemme for her careful reading and suggestions on this manuscript. We also thank the anonymous reviewers for their valuable comments toward the improvement of the paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.brainres.2016.09.015>.

References

- Alexander, J.A., Wong, P.C., Bradlow, A.R., 2005. Lexical tone perception in musicians and non-musicians. *Interspeech*, 397–400.
- Alho, K., Teder, W., Lavikainen, J., Näätänen, R., 1994. Strongly focused attention and auditory event-related potentials. *Biol. Psychol.* 38 (1), 73–90.
- 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. *J. Exp. Child Psychol.* 83 (2), 111–130. [http://dx.doi.org/10.1016/S0022-0965\(02\)00124-8](http://dx.doi.org/10.1016/S0022-0965(02)00124-8).
- Asaridou, S.S., McQueen, J.M., 2013. Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Front. Psychol.* 4, 1–14. <http://dx.doi.org/10.3389/fpsyg.2013.00321>.
- Ashby, M., Maimment, J., 2005. *Introducing Phonetic Science*. Cambridge University Press, Cambridge.
- Astésano, C., Besson, M., Alter, K., 2004. Brain potentials during semantic and prosodic processing in French. *Cognit. Brain Res.* 18 (2), 172–184. <http://dx.doi.org/10.1016/j.cogbrainres.2003.10.002>.
- Atienza, M., Cantero, J.L., Dominguez-Marin, E., 2002. The time course of neural changes underlying auditory perceptual learning. *Learn. Mem.* 9 (3), 138–150.
- Bartels, C., 1997. *Towards a Compositional Interpretation of English Statement and Question Intonation*. University of Massachusetts, Amherst, Doctoral dissertation.
- Bartels, C., 1999. *The Intonation of English Statements and Questions: A Compositional Interpretation*. Garland Publishing, New York/London.
- Bartels, C., 2014. *The intonation of English statements and questions: a compositional interpretation*. New York, Routledge.
- Berry, A.S., Zanto, T.P., Rutman, A.M., Clapp, W.C., Gazzaley, A., 2009. Practice-related improvement in working memory is modulated by changes in processing external interference. *J. Neurophysiol.* 102 (3), 1779–1789. <http://dx.doi.org/10.1152/jn.00179.2009>.
- Berry, A.S., Zanto, T.P., Clapp, W.C., Hardy, J.L., Delahunt, P.B., Mahncke, H.W., Gazzaley, A., 2010. The influence of perceptual training on working memory in older adults. *PLoS One* 5 (7), e11537. <http://dx.doi.org/10.1371/journal.pone.0011537>.
- Besson, M., Faïta, F., 1995. An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *J. Exp. Psychol.: Hum. Percept. Perform.* 21 (6), 1278–1296. <http://dx.doi.org/10.1037/0096-1523.21.6.1278>.
- Besson, M., Chobert, J., Marie, C., 2011. Transfer of training between music and speech: common processing, attention, and memory. *Frontiers: research topics. The relationship between music and language*. *Front. Media SA*, 147–158.

- 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. *Restor. Neurol. Neurosci.* 25 (3–4), 399–410.
- Bialystok, E., Depape, A.-M., 2009. Musical expertise, bilingualism, and executive functioning. *J. Exp. Psychol. Hum. Percept. Perform.* 35 (2), 565–574. <http://dx.doi.org/10.1037/a0012735>.
- Bialystok, E., Viswanathan, M., 2009. Components of executive control with advantages for bilingual children in two cultures. *Cognition* 112 (3), 494–500. <http://dx.doi.org/10.1016/j.cognition.2009.06.014>.
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cognit. Neurosci.* 23 (2), 425–434. <http://dx.doi.org/10.1162/jocn.2009.21362>.
- Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D.A., Madurell, F., 2001. The effect of harmonic context on phoneme monitoring in vocal music. *Cognition* 81 (1), 11–20. [http://dx.doi.org/10.1016/S0010-0277\(01\)00117-2](http://dx.doi.org/10.1016/S0010-0277(01)00117-2).
- Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott international* 5 (9/10), 341–345.
- Bradley, E.D., 2012. Tone language experience enhances sensitivity to melodic contour. In: *Proceedings of the LSA Annual Meeting Extended Abstracts*, vol. 3, pp. 40–41.
- Carey, D., Rosen, S., Krishnan, S., Pearce, M.T., Shepherd, A., Aydelott, J., Dick, F., 2015. Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition* 137, 81–105. <http://dx.doi.org/10.1016/j.cognition.2014.12.005>.
- Carrus, E., Koelsch, S., Bhattacharya, J., 2011. Shadows of music-language interaction on low frequency brain oscillatory patterns. *Brain Lang.* 119 (1), 50–57. <http://dx.doi.org/10.1016/j.bandl.2011.05.009>.
- Carrus, E., Pearce, M.T., Bhattacharya, J., 2013. Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex* 49 (8), 2186–2200. <http://dx.doi.org/10.1016/j.cortex.2012.08.024>.
- Chen, X., Zhao, L., Jiang, A., Yang, Y., 2011. Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biol. Psychol.* 86 (3), 158–167. <http://dx.doi.org/10.1016/j.biopsycho.2010.11.004>.
- Chun, D.M., 2002. *Discourse Intonation in L2: From Theory and Research to Practice*. John Benjamins Publishing, Amsterdam, Philadelphia.
- Colé, P., Segui, J., 1994. Grammatical incongruity and vocabulary types. *Mem. Cogn.* 22 (4), 387–394. <http://dx.doi.org/10.3758/BF03200865>.
- Debener, S., Makeig, S., Delorme, A., Engel, A.K., 2005. What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognit. Brain Res.* 22 (3), 309–321. <http://dx.doi.org/10.1016/j.cogbrainres.2004.09.006>.
- Deguchi, C., Boureau, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., Colombo, L., 2012. Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: behavioral, electrophysiological and psychoacoustic study. *Brain Res.* 1455, 75–89. <http://dx.doi.org/10.1016/j.brainres.2012.03.034>.
- Delorme, A., Makeig, S., 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* 134 (1), 9–21.
- Dienes, Z., Longuet-Higgins, C., 2004. Can musical transformations be implicitly learned? *Cognit. Sci.* 28 (4), 531–558.
- Doherty, C.P., West, W.C., Dilley, L.C., Shattuck-Hufnagel, S., Caplan, D., 2004. Question/statement judgments: an fMRI study of intonation processing. *Hum. Brain Mapp.* 23 (2), 85–98. <http://dx.doi.org/10.1002/hbm.20042>.
- Eckstein, K., Friederici, A.D., 2005. Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognit. Brain Res.* 25 (1), 130–143. <http://dx.doi.org/10.1016/j.cogbrainres.2005.05.003>.
- Escoffier, N., Tillmann, B., 2008. The tonal function of a task-irrelevant chord modulates speed of visual processing. *Cognition* 107 (3), 1070–1083. <http://dx.doi.org/10.1016/j.cognition.2007.10.007>.
- Essl, M., Rappelsberger, P., 1998. EEG coherence and reference signals: experimental results and mathematical explanations. *Med. Biol. Eng. Comput.* 36 (4), 399–406. <http://dx.doi.org/10.1007/BF02523206>.
- Featherstone, C.R., Morrison, C.M., Waterman, M.G., MacGregor, L.J., 2013. Semantics, syntax or neither? A case for resolution in the interpretation of N500 and P600 responses to harmonic incongruities. *PLoS One* 8 (11). <http://dx.doi.org/10.1371/journal.pone.0076600>.
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., Gibson, E., 2009. Structural integration in language and music: evidence for a shared system. *Mem. Cogn.* 37 (1), 1–9. <http://dx.doi.org/10.3758/MC.37.1.1>.
- Festman, J., Rodriguez-Fornells, A., Münte, T.F., 2010. Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behav. Brain Funct.* 6 (1), 5. <http://dx.doi.org/10.1186/1744-9081-6-5>.
- Francois, C., Schön, D., 2010. Learning of musical and linguistic structures: comparing event-related potentials and behavior. *Neuroreport*, 928–932. <http://dx.doi.org/10.1097/WNR.0b013e32833ddd5e>.
- François, C., Schön, D., 2014. Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: the role of musical practice. *Hear. Res.* 308, 122–128. <http://dx.doi.org/10.1016/j.heares.2013.08.018>.
- François, C., Chobert, J., Besson, M., Schön, D., 2013. Music training for the development of speech segmentation. *Cereb. Cortex* 23 (9), 2038–2043. <http://dx.doi.org/10.1093/cercor/bhs180>.
- Friederici, A.D., Alter, K., 2004. Lateralization of auditory language functions: a dynamic dual pathway model. *Brain and language* 89 (2), 267–276.
- Geiser, E., Zaehle, T., Jancke, L., Meyer, M., 2008. The neural correlate of speech rhythm as evidenced by metrical speech processing. *J. Cognit. Neurosci.* 20 (3), 541–552. <http://dx.doi.org/10.1162/jocn.2008.20029>.
- George, E.M., Coch, D., 2011. Music training and working memory: an ERP study. *Neuropsychologia* 49 (5), 1083–1094. <http://dx.doi.org/10.1016/j.neuropsychologia.2011.02.001>.
- Gibson, E., 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68 (1), 1–76. [http://dx.doi.org/10.1016/S0010-0277\(98\)00034-1](http://dx.doi.org/10.1016/S0010-0277(98)00034-1).
- Gibson, E., 2006. The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *J. Mem. Lang.* 54 (3), 363–388. <http://dx.doi.org/10.1016/j.jml.2005.12.005>.
- Giordano, B.L., 2011. Music perception (springer handbook of auditory research): invited book review. *J. Acoust. Soc. Am.*, vol. 129, p.4086.
- Grabe, E., 2004. Intonational variation in urban dialects of English spoken in the British Isles. In: Gilles, P., Peters, J. (Eds.), *Regional Variation in Intonation*. Linguistische Arbeiten, Tübingen, Niemeyer, pp. 9–31.
- Grabe, E., Kochanski, G., Coleman, J., 2003. Quantitative modelling of intonational variation. *Proc. SAS RTLM*, 45–57.
- Haan, J., 2002. Speaking of questions. *Neth. Grad. Sch. Linguist.*
- Habib, M., Besson, M., 2009. What do music training and musical experience teach us about brain plasticity? *Music Percept.: Interdiscip. J.* 26 (3), 279–285.
- Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. *Science* 182, 177–180.
- Hirst, D., Cristo, A., 1998. *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, Cambridge.
- Ho, Y.-C., Cheung, M.-C., Chan, A.S., 2003. Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology* 17 (3), 439–450. <http://dx.doi.org/10.1037/0894-4105.17.3.439>.
- Hoch, L., Poulin-Charronnat, B., Tillmann, B., 2011. The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Front. Psychol.* 2, 1–10. <http://dx.doi.org/10.3389/fpsyg.2011.00112>.
- Holzgrefe, J., Wellmann, C., Petrone, C., Truckenbrodt, H., Höhle, B., Wartenburger, I., 2013. Brain response to prosodic boundary cues depends on boundary position. *Front. Psychol.* 4, 1–14. <http://dx.doi.org/10.3389/fpsyg.2013.00421>.
- Horváth, J., Winkler, I., Bendixen, A., 2008. Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* 79 (2), 139–147. <http://dx.doi.org/10.1016/j.biopsycho.2008.04.001>.
- Huron, D.B., 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, Cambridge, MA.
- Jackendoff, R., 2009. Parallels and nonparallels between language and music. *Music Percept.: Interdiscip. J.* 26 (3), 195–204.
- Janata, P., Tillmann, B., Bharucha, J.J., 2002. Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognit. Affect. Behav. Neurosci.* 2 (2), 121–140. <http://dx.doi.org/10.3758/CABN.2.2.121>.
- Jonaitis, E.M., Saffran, J.R., 2009. Learning harmony: the role of serial statistics. *Cognit. Sci.* 33 (5), 951–968. <http://dx.doi.org/10.1111/j.1551-6709.2009.01036.x>.
- Jung, H., Sontag, S., Park, Y.S., Loui, P., 2015. Rhythmic effects of syntax processing in music and language. *Front. Psychol.* 6, 1762. <http://dx.doi.org/10.3389/fpsyg.2015.01762>.
- Koelsch, S., 2012. *Brain and Music*. John Wiley & Sons, New York.
- Koelsch, S., Friederici, A.D., 2003. Toward the neural basis of processing structure in music. *Ann. N. Y. Acad. Sci.* 999 (1), 15–28. <http://dx.doi.org/10.1196/annals.1284.002>.
- Koelsch, S., Gunter, T., 2005. Interaction between syntax processing in language and in music: an ERP study. *J. Cognit. Neurosci.* 17 (10), 1565–1577.
- Koelsch, S., Jentschke, S., 2010. Differences in electric brain responses to melodies and chords. *J. Cognit. Neurosci.* 22 (10), 2251–2262. <http://dx.doi.org/10.1162/jocn.2009.21338>.
- Koelsch, S., Gunter, T.C., Wittfoth, M., Sammler, D., 2005. Interaction between syntax processing in language and in music: an ERP study. *J. Cognit. Neurosci.* 17 (10), 1565–1577. <http://dx.doi.org/10.1162/089892905774597290>.
- Koelsch, S., Fritz, T., Schulze, K., Alsup, D., Schlaug, G., 2005. Adults and children processing music: An fMRI study. *NeuroImage* 25 (4), 1068–1076. <http://dx.doi.org/10.1016/j.neuroimage.2004.12.050>.
- Kotz, S.A., Paulmann, S., 2007. When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Res.* 1151 (1), 107–118. <http://dx.doi.org/10.1016/j.brainres.2007.03.015>.
- Kotz, S.A., Meyer, M., Alter, K., Besson, M., Von Cramon, D.Y., Friederici, A.D., 2003. On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang.* 86 (3), 366–376. [http://dx.doi.org/10.1016/S0093-934X\(02\)00532-1](http://dx.doi.org/10.1016/S0093-934X(02)00532-1).
- Krizman, J., Marian, V., Shook, A., Skoe, E., Kraus, N., 2012. Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proc. Natl. Acad. Sci. USA* 109 (20), 7877–7881. <http://dx.doi.org/10.1073/pnas.1201575109>.
- Krumhansl, C.L., 1995. Music psychology and music theory: problems and prospects. *Music Theory Spectr.* 17 (1), 53–80. <http://dx.doi.org/10.1525/mts.1995.17.1.02a00030>.
- Kunert, R., Willems, R.M., Casasanto, D., Patel, A.D., Hagoort, P., 2015. Music and language syntax interact in Broca's area: an fMRI study. *PLoS One* 10 (11), e0141069. <http://dx.doi.org/10.1371/journal.pone.0141069>.
- Lindsen, J.P., Pearce, M.T., Doyné, M., Wiggins, G.A., Bhattacharya, J., 2010. Implicit brain responses during fulfillment of melodic expectations. In: *Proceedings of the 12th International Conference on Music Perception and Cognition*, pp. 606–607.

- Loui, P., Wessel, D., Kam, C., 2010. Humans rapidly learn grammatical structure in a new musical scale. *Music Percept.: Interdiscip. J.* 27 (5), 377–388.
- Ma, J.K.-Y., Ciocca, V., Whitehill, T.L., 2011. The perception of intonation questions and statements in Cantonese. *J. Acoust. Soc. Am.* 129 (2), 1012–1023. <http://dx.doi.org/10.1121/1.3531840>.
- MacKay, D., 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge.
- Macmillan, N., Creelman, C., 2005. *Detection Theory: A User's Guide*. Psychology Press, New York.
- Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4 (5), 540–545. <http://dx.doi.org/10.1038/87502>.
- Magne, C., Schön, D., Besson, M., 2003. Prosodic and melodic processing in adults and children: behavioral and electrophysiological approaches. *Ann. N. Y. Acad. Sci.* 999, 461–476. <http://dx.doi.org/10.1196/annals.1284.056>.
- 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. *J. Cognit. Neurosci.* 18 (2), 199–211. <http://dx.doi.org/10.1162/jocn.2006.18.2.199>.
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M.O., Besson, M., 2011. Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *J. Cognit. Neurosci.* 23 (10), 2701–2715.
- Maye, J., Werker, J., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82 (3), B101–B111.
- McMullen, E., Saffran, J.R., 2004. Music and language: a developmental comparison. *Music Percept.* 21 (3), 289–311. <http://dx.doi.org/10.1525/mp.2004.21.3.289>.
- Meyer, L.B., 2008. *Emotion and meaning in music*. University of Chicago Press.
- Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., Von Cramon, D.Y., 2002. fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* 17 (2), 73–88. <http://dx.doi.org/10.1002/hbm.10042>.
- Moreno, S., 2009. Can music influence language and cognition? *Contemp. Music Res.* 28 (3), 329–345. <http://dx.doi.org/10.1080/07494460903404410>.
- Moreno, S., Besson, M., 2006. Musical training and language-related brain electrical activity in children. *Psychophysiology* 43 (3), 287–291. <http://dx.doi.org/10.1111/j.1469-8986.2006.00401.x>.
- Moreno, S., Bidelman, G.M., 2014. Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* 308, 84–97. <http://dx.doi.org/10.1016/j.heares.2013.09.012>.
- Moreno, S., Bialystok, E., Barac, R., Schellenberg, E.G., Cepeda, N.J., Chau, T., 2011. Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* 22 (11), 1425–1433. <http://dx.doi.org/10.1177/0956797611416999>.
- Müllensiefen, D., Gingras, B., Musil, J., Stewart, L., 2014. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS ONE* 9 (2). <http://dx.doi.org/10.1371/journal.pone.0089642>.
- Näätänen, R., Gaillard, A.W., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42 (4), 313–329.
- Nolan, F., 2008. *Intonation. The Handbook of English Linguistics*, pp. 433–457. <http://dx.doi.org/10.1002/9780470753002.ch19>.
- Nooteboom, S., 1997. *The prosody of speech: melody and rhythm*. *Handb. Phon. Sci.* 5, 640–673.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M. *FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data*. *Computational Intelligence and Neuroscience*, 2011.
- Pannekamp, A., Toepel, U., 2005. *Prosody-driven sentence processing: an event-related brain potential study*. *J. Cognit. Neurosci.* 17 (3), 407–421.
- Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12 (1), 169–174.
- Parbery-Clark, A., Skoe, E., Kraus, N., 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.: Off. J. Soc. Neurosci.* 29 (45), 14100–14107. <http://dx.doi.org/10.1523/JNEUROSCI.3256-09.2009>.
- Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N., 2012. Musical experience offsets age-related delays in neural timing. *Neurobiol. Aging* 33, 1–4. <http://dx.doi.org/10.1016/j.neurobiolaging.2011.12.015>.
- Patel, A.D., 1998. *Syntactic processing in language and music: different cognitive operations, similar neural resources?* *Music Percept.* 16 (1), 27–42.
- Patel, A.D., 2003. Language, music, syntax and the brain. *Nat. Neurosci.* 6 (7), 674–681. <http://dx.doi.org/10.1038/nn1082>.
- Patel, A.D., 2014. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. <http://dx.doi.org/10.1016/j.heares.2013.08.011>.
- Patel, A.D., Peretz, I., Tramo, M., Labrègue, R., 1998a. Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* 61 (1), 123–144. <http://dx.doi.org/10.1006/brln.1997.1862>.
- Patel, A.D., Gibson, E., Ratner, J., Besson, M., Holcomb, P.J., 1998b. Processing syntactical relations in language and music: an event-related potential study. *J. Cognit. Neurosci.* 10 (6), 717–733. <http://dx.doi.org/10.1162/089892998563121>.
- Paulmann, S., Kotz, S., 2008. An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain Lang.* 105 (1), 59–69. <http://dx.doi.org/10.1016/j.bandl.2007.11.005>.
- Paulmann, S., Jessen, S., Kotz, S., 2012. It's special the way you say it: an ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia* 50 (7), 1609–1620. <http://dx.doi.org/10.1016/j.neuropsychologia.2012.03.014>.
- Pearce, M.T., 2005. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. City University, London.
- Pearce, M.T., Ruiz, M.H., Kapasi, S., Wiggins, G.A., Bhattacharya, J., 2010. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage* 50 (1), 302–313. <http://dx.doi.org/10.1016/j.neuroimage.2009.12.019>.
- Perruchet, P., Poulin-Charronnat, B., 2013. Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.* 20 (2), 310–317. <http://dx.doi.org/10.3758/s13423-012-0344-5>.
- Peters, B., Pfitzinger, H., 2008. Duration and F0 interval of utterance-final intonation contours in the perception of German sentence modality. In: *Proceedings of the 35th International Symposium on Computer Architecture*. Beijing, China, interspeech, 2008, pp. 65–68.
- Poarch, G.J., van Hell, J.G., 2012. Executive functions and inhibitory control in multilingual children: evidence from second-language learners, bilinguals, and trilinguals. *J. Exp. Child Psychol.* 113 (4), 535–551. <http://dx.doi.org/10.1016/j.jecp.2012.06.013>.
- Proverbio, A.M., Leoni, G., Zani, A., 2004. Language switching mechanisms in simultaneous interpreters: an ERP study. *Neuropsychologia* 42 (12), 1636–1656. <http://dx.doi.org/10.1016/j.neuropsychologia.2004.04.013>.
- Ragert, P., Schmidt, A., Altenmüller, E., 2004. Superior tactile performance and learning in professional pianists: evidence for meta-plasticity in musicians. *Eur. J. Neurosci.* 19 (2), 473–478. <http://dx.doi.org/10.1111/j.1460-9568.2003.03142.x>.
- Ratcliff, R., 1993. *Methods for dealing with reaction time outliers*. *Psychol. Bull.* 114 (3), 510.
- Regnault, P., Bigand, E., Besson, M., 2001. Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: Evidence from auditory event-related brain potentials. *J. Cognit. Neurosci.* 13 (2), 241–255. <http://dx.doi.org/10.1162/089892901564298>.
- Ridding, M.C., Brouwer, B., Nordstrom, M.A., 2000. Reduced interhemispheric inhibition in musicians. *Exp. Brain Res.* 133 (2), 249–253.
- Rigoulot, S., Pell, M.D., Armony, J.L., 2015. Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience* 290, 175–184. <http://dx.doi.org/10.1016/j.neuroscience.2015.01.033>.
- Russo, F.A., Thompson, W.F., 2005. The subjective size of melodic intervals over a two-octave range. *Psychon. Bull. Rev.* 12 (6), 1068–1075.
- Schellenberg, E., 2003. Does exposure to music have beneficial side effects? In: Peretz, I., Zatorre, R.J. (Eds.), *The Cognitive Neuroscience of Music*. Oxford University Press, Oxford, UK, pp. 430–448.
- Schellenberg, E., 2006. Long-term positive associations between music lessons and IQ. *J. Educ. Psychol.* 98 (2), 457.
- Schellenberg, E.G., 2004. Music lessons enhance IQ. *Psychol. Sci.* 15 (8), 511–514. <http://dx.doi.org/10.1111/j.0956-7976.2004.00711.x>.
- Schirmer, A., Kotz, S., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cognit. Sci.* 10 (1), 24–30.
- Schlaug, G., Jäncke, L., Huang, Y., Staiger, J.F., Steinmetz, H., 1995. Increased corpus callosum size in musicians. *Neuropsychologia* 33 (8), 1047–1055.
- 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. *Nat. Neurosci.* 5 (7), 688–694. <http://dx.doi.org/10.1038/nn871>.
- 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. <http://dx.doi.org/10.1111/1469-8986.00172.x>.
- Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., Kolinsky, R., 2008. Songs as an aid for language acquisition. *Cognition* 106 (2), 975–983. <http://dx.doi.org/10.1016/j.cognition.2007.03.005>.
- Schön, D., Magne, C., Schrooten, M., Besson, M., 2002. The music of speech: electrophysiological approach. In: *Proceedings of the Speech Prosody 2002, International Conference*.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A.D., Koelsch, S., 2011. Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* 32, 771–783. <http://dx.doi.org/10.1002/hbm.21060>.
- Selkirk, E., 1995. *Sentence prosody: Intonation, stress, and phrasing*. In: Goldsmith, J.A. (Ed.), *The Handbook of Phonological Theory*. Blackwell, Oxford, pp. S550–S569.
- Shahin, A., Bosnyak, D.J., Trainor, L.J., Roberts, L.E., 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23 (13), 5545–5552.
- Skoe, E., Kraus, N., 2012. A little goes a long way: how the adult brain is shaped by musical training in childhood. *J. Neurosci.* 32 (34), 11507–11510. <http://dx.doi.org/10.1523/JNEUROSCI.1949-12.2012>.
- Slevc, L.R., Rosenberg, J.C., Patel, A.D., 2009. Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16 (2), 374–381. <http://dx.doi.org/10.3758/16.2.374>.
- Smayda, K.E., Chandrasekaran, B., Maddox, W.T., 2015. Enhanced cognitive and perceptual processing: a computational basis for the musician advantage in speech learning. *Front. Psychol.* 6, 682. <http://dx.doi.org/10.3389/fpsyg.2015.00682>.
- Steele, C.J., Bailey, J.A., Zatorre, R.J., Penhune, V.B., 2013. Early musical training and white-matter plasticity in the corpus callosum: evidence for a sensitive period. *J. Neurosci.* 33 (3), 1282–1290.
- Steinhilber, K., Alter, K., Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2 (2), 191–196. <http://dx.doi.org/10.1038/5757>.
- Strait, D., Parbery-Clark, A., 2013. *Biological impact of preschool music classes on*

- processing speech in noise. *Dev. Cognit. Neurosci.* 6, 51–60.
- Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261 (1–2), 22–29. <http://dx.doi.org/10.1016/j.heares.2009.12.021>.
- Studdert-Kennedy, M., Hadding, K., 1973. Auditory and linguistic processes in the perception of intonation contours. *Lang. Speech* 16 (4), 293–313. <http://dx.doi.org/10.1177/002383097301600401>.
- Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V., Kidd, G., Patel, A.D., 2015. Musical training, individual differences and the cocktail party problem. *Sci. Rep.* 5, 11628. <http://dx.doi.org/10.1038/srep11628>.
- Tatsuno, Y., Sakai, K.L., 2005. Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *J. Neurosci.: Off. J. Soc. Neurosci.* 25 (7), 1637–1644. <http://dx.doi.org/10.1523/JNEUROSCI.3978-04.2005>.
- Thompson, W.F., Schellenberg, E.G., Husain, G., 2004. Decoding speech prosody: do music lessons help? *Emotion* 4 (1), 46–64. <http://dx.doi.org/10.1037/1528-3542.4.1.46>.
- Trainor, L.J., Shahin, A., Roberts, L.E., 2003. Effects of musical training on the auditory cortex in children. *Ann. N. Y. Acad. Sci.* 999 (1), 506–513. <http://dx.doi.org/10.1196/annals.1284.061>.
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., Otis, B., 2001. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear.* 22 (2), 79–90.
- Vos, P., Troost, J., 1989. Ascending and descending melodic intervals: statistical findings and their perceptual relevance. *Music Percept.: Interdiscip. J.* 6 (4), 383–396.
- Vos, P.G., Pasveer, D., 2002. Goodness ratings of melodic openings and closures. *Percept. Psychophys.* 64 (4), 631–639. <http://dx.doi.org/10.3758/BF03194731>.
- Wan, C.Y., Schlaug, G., 2010. Music making as a tool for promoting brain plasticity across the life span. *Neuroscientist* 16 (5), 566–577.
- Wells, J.C., 2006. *English Intonation* PB and Audio CD: An Introduction. Cambridge University Press, Cambridge.
- Wiggins, G., Pearce, M., 2006. Expectation in melody: the influence of context and learning. *Music Percept.: Interdiscip. J.* 23 (5), 377–405. <http://dx.doi.org/10.1525/mp.2006.23.5.377>.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H., 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cereb. Cortex* 14 (12), 1384–1389. <http://dx.doi.org/10.1093/cercor/bhh099>.
- Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10 (4), 420–422. <http://dx.doi.org/10.1038/nn1872>.
- Zendel, B.R., Alain, C., 2009. Concurrent sound segregation is enhanced in musicians. *J. Cognit. Neurosci.* 21 (8), 1488–1498.
- Zendel, B.R., Alain, C., 2013. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *J. Cognit. Neurosci.* 25 (4), 503–516. http://dx.doi.org/10.1162/jocn_a_00329.
- Zhao, T.C., Kuhl, P.K., 2015. Higher-level linguistic categories dominate lower-level acoustics in lexical tone processing. *J. Acoust. Soc. Am.* 138 (2), EL133–EL137. <http://dx.doi.org/10.1121/1.4927632>.