Musical Expertise Boosts Implicit Learning of Both Musical and Linguistic Structures

Clément François, D Schon

▶ To cite this version:

Clément François, D Schon. Musical Expertise Boosts Implicit Learning of Both Musical and Linguistic Structures. Cerebral Cortex, 2011, 21, pp.2357 - 2365. 10.1093/cercor/bhr022 . hal-03053499

HAL Id: hal-03053499

https://hal.science/hal-03053499

Submitted on 11 Dec 2020

Cerebral Cortex Advance Access published March 7, 2011

Cerebral Cortex
doi:10.1093/cercor/bhr022

Musical Expertise Boosts Implicit Learning of Both Musical and Linguistic Structures

Clément Francois and Daniele Schön

Institut de Neurosciences Cognitives de la Méditerranée CNRS and Aix-Marseille Universities, Marseille 13402, France

Address correspondence to Clement Francois, Institut de Neurosciences Cognitives de la Méditerranée CNRS and Aix-Marseille Universities, 31, Chemin Joseph Aiguier, 13402 Marseille, France. Email: clement.francois@incm.cnrs-mrs.fr.

Musical training is known to modify auditory perception and related cortical organization. Here, we show that these modifications may extend to higher cognitive functions and generalize to processing of speech. Previous studies have shown that adults and newborns can segment a continuous stream of linguistic and nonlinguistic stimuli based only on probabilities of occurrence between adjacent syllables or tones. In the present experiment, we used an artificial (sung) language learning design coupled with an electrophysiological approach. While behavioral results were not clear cut in showing an effect of expertise, Event-Related Potentials data showed that musicians learned better than did nonmusicians both musical and linguistic structures of the sung language. We discuss these findings in terms of practice-related changes in auditory processing, stream segmentation, and memory processes.

Keywords: brain plasticity, language, music, musical expertise, N400, statistical learning

Introduction

Musicians often undergo an intensive formal training period that can last 10–15 years and that can imply up to 10000-h practice by early adulthood (Krampe and Ericsson 1996). Moreover, they keep practicing several hours a day during their whole career. Thus, comparing musicians with nonmusicians allows studying the effects of extensive audiomotor training on the functional and structural organization of the brain. At a perceptual level, musicians have a lower frequency threshold than do nonmusicians (Kishon-Rabin et al. 2001). These differences might be related to functional differences in the auditory pathway. Musicians' brainstem responses show more robust encoding of sound pitch processing (Musacchia et al. 2007; Wong et al. 2007). Moreover, when listening to synthetic or instrumental sounds, musicians show larger N1 and P2 event-related potentials than nonmusicians do (Shahin et al. 2003, 2005). Moreover, they are more sensitive to sound spectral structure: For instance, musicians present a larger N1m to piano sounds than to pure tones, while nonmusicians are not sensitive to this contrast (Pantev et al. 1998). When deviant stimuli are introduced in a regular stream of identical sounds, participants, typically watching a silent movie (i.e., not paying attention to the sound stream), show a mismatch negativity (MMN) to deviant sounds. Interestingly, it has been shown that musicians have a larger MMN and MMNm than nonmusicians do when deviant chords are introduced in the stream (Koelsch et al. 1999; Brattico et al. 2009) as well as when a sound is omitted (Rüsseler et al. 2001). These differences can be interpreted as reflecting a greater efficiency of musicians' auditory system in processing sound features and can be accompanied by morphological differences, showing that musicians compared with nonmusicians have a different gray-matter concentration in the auditory cortex (Bermudez and Zatorre 2005) and a larger planum temporale (Schlaug et al. 1995; Keenan et al. 2001).

The fact that musicians perceive some sound features more accurately than nonmusicians do is not so surprising. After all, they spend hours and hours of their life focusing on sounds and the way they are generated, paying particular attention to pitch, timber, duration, and timing. However, what seems less evident to us is whether or not this intensive musical practice can affect nonmusical abilities. Several recent studies seem to confirm this possibility.

For instance, both adult and child musicians have a better performance than matched controls do when asked to detect fine contour modifications in the prosody of an utterance (Schön et al. 2004; Magne et al. 2006). These results are supported by other findings showing a possible correlation between musical and linguistic aptitudes in children (Anvari et al. 2002; Slevc and Miyake 2006; Milovanov et al. 2008, 2009). Indeed Slevc and Miyake (2006) have shown that musical aptitude is a rather good predictor of both receptive and productive phonological abilities in second language learning.

In this study, we took the challenge of focusing on a rather high cognitive function: word segmentation, namely, the ability to extract words from continuous speech. Because the speech stream does not systematically carry consistent acoustic cues such as pauses or accents at word boundaries, there must be other ways to learn how to segment words. According to several publications, speech might be segmented in a rather implicit manner based on one source of information: the statistical structure of the language (Saffran et al. 1996b; Kuhl 2004).

For instance, in English, the syllable pre can be followed by a small number of syllables (like ty, pare, clude); thus, the probability that pre is followed by ty is quite high in regular speech. By contrast, the syllable ty mostly occurs words finally and can thus be followed by any word initial syllables, making the probability that ty is followed by ba, as in "pretty baby," very low (Saffran 2003). In general, "syllables that are part of the same word tend to follow one another predictably, whereas syllables that span word boundaries do not" (Saffran et al. 2001).

The role of transitional probabilities in speech segmentation (viz., the probability of syllable X given syllable Y) has been elegantly shown by Saffran and collaborators with infants and adults and has also been extended to neonates (Saffran et al. 1996a, 1996b; Aslin et al. 1998; Kuhl 2004; Gervain et al. 2008; Teinonen et al. 2009). Throughout this series of studies, the authors showed that listening to an artificial language without acoustic cues at word boundaries yields correct word segmentation. Indeed, in a test following the listening phase, participants are able to discriminate words that are part of the language from similar words that are not part of the language. This learning paradigm has also been replicated using nonlinguistic stimuli such as sounds with different pitches (Saffran et al. 1999) and timber (Tillmann and McAdams 2004), as well as in the cotton top tamarin and in rats (Hauser et al. 2000, 2002; Ramus et al. 2000; Toro and Trobalón 2005).

We previously showed (Schön et al. 2008) that a sung language allowed a better word segmentation than a spoken language, and we interpreted this as a beneficial effect of the structuring and motivational properties of music on speech segmentation. Recently, we used the same sung material, and we investigated both melodic and word segmentation while recording electroencephalography (EEG). Nonmusician participants listened to a sung artificial language and were then tested using a 2-alternative forced-choice test (familiar/unfamiliar) (Francois and Schön 2010). While the linguistic dimension was learned better than the melodic dimension, at a behavioral level, a negative N400-like component was larger for unfamiliar than for familiar words in both linguistic and musical dimensions. This well fits the literature showing a larger N400 component to unfamiliar than familiar words (Young and Rugg 1992).

In the present study, we used the same learning paradigm as in the study by Francois and Schön (2010) but comparing 2 groups. Participants listened to an artificial sung language (wherein music and language dimensions are highly intertwined) and were then tested with a 2-alternative forced-choice task on pairs of words and melodies (familiar vs. unfamiliar). The main goal of this study was to test whether musical expertise can facilitate word segmentation. With this aim, we compared 2 groups, one group with formal musical training and one without. At the behavioral level, we hypothesized that musicians may perform better than nonmusicians do on both musical and linguistic dimensions, because they would benefit the most from the musical information carried by the sung language. At the electrophysiological level, we hypothesized that the N400 familiarity effect with a frontocentral distribution would be larger for musicians than for nonmusicians for both linguistic and musical dimensions.

Materials and Methods

Participants

Two groups participated in the experiment. Sixteen professional musicians (mean age 27 years, 15 right-handed, 10 males, normal hearing, no known neurological problems, >12 years of formal musical learning and from 3 to 7 h of daily practice, 5 of them reported absolute pitch) and 20 nonmusicians (mean age 25 years, 20 right-handed, 11 males, normal hearing, no known neurological problems, not >2 years of formal musical training and no instrument practice). All participants were native French speakers and listened to 5.5 min of a continuous speech stream resulting from the concatenation of 5 3-syllable nonsense words (hereafter words) that were repeated in a pseudorandom order. All participants were paid 20 Euros. Informed consent was obtained from all participants, and the data were analyzed anonymously. This study was approved by the CNRS, Mediterranean Institute for Cognitive Neuroscience, and was conducted in accordance with national norms and guidelines for the protection of human subjects.

Materials

The language consisted of 4 consonants and 3 vowels, which were combined into a set of 11 syllables. The average syllable length was 230 ms, standard deviation was 16 ms. These syllables were then combined to give rise to 5 trisyllabic sung words (gimysy, mimosi, pogysi, pymiso, and sipygy). Each of the 11 syllables was associated with a distinct tone (C3, D3, F3, G3, A3, B3, C4, Db4, D4, E4, and F4). Therefore, each word was always sung on the same melodic contour (gimysy C3 D3 F3, mimosi E4 Db4 G3, pymiso B3 E4 F4, pogysi D4 C3 G3, and sipygy G3 B3 C4). The mean pitch interval within words was not significantly different from the mean interval between words ($P = 0.4$). Moreover, pitch-contour changes could not be used to segment the stream because they took place 50% of the time only at word boundaries.

Transitional probabilities within words ranged from 0.5 to 1.0. Transitional probabilities across word boundaries ranged from 0.1 to 0.5. While we used a 6-word language stream in our previous experiment (Francois and Schön 2010), in the present study, the language stream was built by a random concatenation of the 5 words (only constraint: no repetition of the same item twice in a row) and synthesized using Mbrola (http://tcts.fpms.ac.be/synthesis/mbrola.html). No acoustic cues were inserted at word boundaries. Each word was repeated 100 times in the stream.

Design and Procedure

During the learning phase, participants were told that they would listen to a continuous stream of sung syllables for several minutes, and they were asked to carefully listen to these sounds. During the test, participants had to choose, by pressing 1 of 2 buttons on a computer keyboard, which of 2 strings (first or second item) most closely resembled what they had just heard in the stream. Items had a flat contour ("spoken" version) in the linguistic test while they were played with a piano sound in the musical test (Fig. 1). In each test trial, one item was a "word" (linguistic test) or "melody" (musical test) from the artificial language (hereafter familiar word/melody); the other item was built by putting together the end of a word/melody with the beginning of another. Thus, these partial items were legal items of the language but had been heard 4 times less than the words/melodies (hereafter unfamiliar words/melodies). More precisely, these items contained either the last syllable of a word plus the first syllable pair of another word or the last syllable pair of a word plus the first syllable of another word. For instance, the last 2 syllables of the word pogysi were combined with the first syllable of mimosi to create the unfamiliar word gysimi (see Fig. 1). The 5 unfamiliar words used during the test were gysimi, mosigi, pygymi, sogimy, and sypogy. Each familiar word of the language was presented with each unfamiliar word, making up 25 pairs. During the musical test, the unfamiliar melodies were built with the same pitches defining the melodic contour of the unfamiliar words described above. Stimuli were presented via loudspeakers. Linguistic and musical tests lasted 5 min each, and their order was counterbalanced across participants.

Data Acquisition and Analysis

Participants were comfortably seated in a Faraday booth. The EEG was recorded from 32 scalp electrodes (Biosemi ActiveTwo system, Amsterdam University) located at standard left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas (International 10/20 system sites: Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F3, F4, C3, C4, P3, P4, P7, P8, Po3, Po4, O1, O2, F7, F8, T7, T8, Fc5, Fc1, Fc2, Fc6, Cp5, Cp1, Cp2, and Cp6). The bandpass was of 0–102.4 Hz, and sampling rate was 512 Hz. The data, acquired during the testing phase, were then rereferenced offline to the algebraic average of the left and right mastoids. Raw data containing movement artifacts or amplifier saturation were excluded. Signal containing ocular artifacts was corrected using Independant Component Analysis decomposition by removing the component containing the blink (Makeig et al. 1996). The EEG was then epoched for each item (familiar and unfamiliar) and the 200-ms baseline (before the stimulus onset) zero-mean normalized using Brain Vision Analyzer software (Brain Products, Munich, Germany).
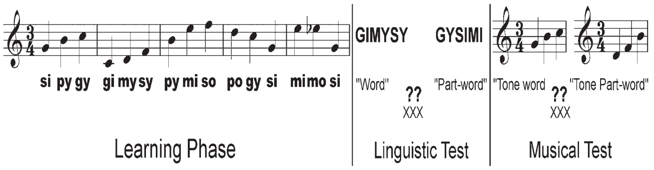
Figure 1. Illustration of the experimental design used in the present experiment. Stimuli were presented auditorily.

Analyses on the N1 and the P2 components were performed on peak amplitudes. Peak amplitudes were computed for each subject and defined as the mean value around the peak latency (i.e., the mean in a window of ±10 ms around the peak). Later components were analyzed by computing the mean amplitude using latency windows of 50 ms. Analyses were performed for correct trials only and for all trials (correct and incorrect trials). Repeated-measure analysis of variance (ANOVA) was used for statistical assessment, using expertise (musicians and nonmusicians) and familiarity (familiar and unfamiliar items) as factors. This differed from behavioral analyses wherein familiarity could not be taken into account (i.e., 1 response/trial in behavior vs. 2 separate items in EEG). Moreover, the topographical distribution of the effects was modeled by 2 additional factors (Hemisphere, left and right and Location, frontal, central, and parietal) defined as follows: left (AF3, F3, F7) and right (AF4, F4, F8) frontal, left (Fc1, C3, Cp1) and right (Fc2, C4, Cp2) central, and left (Po3, P3, P7) and right (Po4, P4, P8) parietal. All P values reported below were adjusted using the Greenhouse–Geisser correction for nonsphericity, when appropriate, and Sidak tests were used in post hoc comparisons. Because of the increased likelihood of type I errors associated with the large number of statistical tests, only effects that reached significance ($P < 0.05$) in at least 2 consecutive time windows were considered significant.

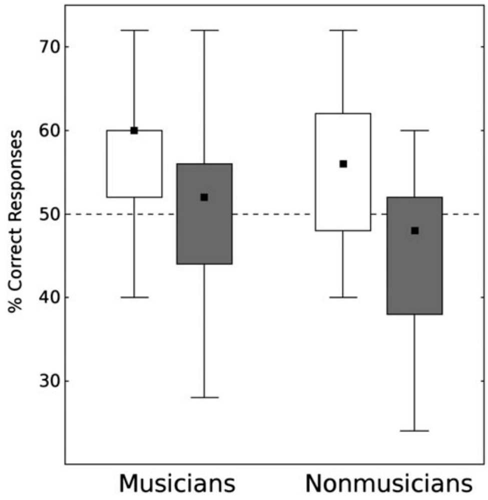

Figure 2. Percentage of correct responses: box plot of performances in the linguistic (white) and musical tests (gray) for musicians (left side) and nonmusicians (right side) (dashed line = chance level). Black squares indicate the medians.

Results

Behavioral Data

Results of a 2-way ANOVA (Expertise [as between factor with 2 levels] and Dimension [as within factor with 2 levels]) showed a main effect of dimension (F [1,35] = 12.52; P = 0.001): The linguistic dimension was learned better than the musical one for both groups. Although not significant, there was a trend for musicians to have a higher level of performance than nonmusicians in both tested dimensions (main effect of Expertise: (F [,35] = 2.7; P = 0.11) (see Fig. 2). This trend was equally distributed across the 2 dimensions (Expertise by Dimension interaction: [$F < 1$]). Comparison of performance with chance level (here 50%) showed that while the participants' level of performance in the linguistic test was above chance for both nonmusician and musician groups (56% and 60% of correct responses P = 0.03, P = 0.007, respectively, Wilcoxon test) this was not the case in the musical test (48% of correct responses for nonmusicians and 52% for musicians). Finally, an additional 3-way ANOVA including expertise (as between factor with 2 levels), dimension (as within factor with 2 levels) and "transitional probability (TP)" (as within factor with 2 levels) was conducted. With this aim, the average performances for the 3 words with the highest average TP (mean TP = 0.83) were compared with the average performances for the 2 words with the lowest average TP (mean TP = 0.5). The dimension by TP interaction was significant (F [1,35] = 8.6; P = 0.005). Post hoc analyses revealed that in the musical test, high TP items were better learned (52%) than low TP items (42%; P = 0.02) for the 2 groups. By contrast, in the linguistic test, no significant differences were found between performances for high TP items (54%) and performances for low TP items (60%; P = 0.10). Neither the main effect of Expertise was significant nor the interactions involving this factor (all $P > 0.2$).

Electrophysiological Data

In the following section, electrophysiological data are presented separately for each dimension for 2 reasons. First, at the behavioral level, musical dimension was not learned as well as the linguistic dimension, thus possibly leading to smeared ERP results because of confounding and/or overlapping cognitive processes engaged. Second, the acoustic characteristics of musical and linguistic stimuli are quite different; piano tones (as used for the musical dimension) have shorter attack time than syllables, thus leading to early differences in the ERPs (larger N1-P2 complex for piano tones).

Language Test

In order to compare ERPs to familiar items and to unfamiliar items, we performed a 4-way ANOVA including Expertise as between factor, Familiarity (Familiar and Unfamiliar), Location (Frontal, Central, and Parietal), and Hemispheres (Right and Left) as within factors on correct trials only.

Results of the ANOVA performed on the N1 peak amplitude showed a significant main effect of Expertise (F [1,35] = 5.81; P = 0.02): The N1 component was larger for musicians (−2.6 µV) than for nonmusicians (−1.3 µV). The familiarity effect was marginally significant (F [1,35] = 2.35; P = 0.13). No differences were significant on the P2 component (see Fig. 3).

In the 550- to 750-ms latency range, the Familiarity by Location interaction was significant (F [2,70] = 3.82; P = 0.04): ERPs were more negative for unfamiliar (−3.3 µV) than for familiar words (−2.1 µV; $P < 0.001$) over frontal regions. This familiarity effect spreads over the scalp in the 600- to 850-ms range, as revealed by the significant main effect of Familiarity (F [1,35] = 11.68; P = 0.004).

Most importantly, the Expertise by Familiarity interaction was significant in the 750- to 850-ms latency band (F [1,35] = 6.16; P = 0.01): Post hoc analyses showed a significant familiarity effect for musicians (2 µV of effect size; P = 0.02) but not for nonmusicians (0.5 µV of effect size; P = 0.5, see Figs. 4 and 5; see also the footnote for additional information concerning all trial analyses[1]).

---

[1] All trials analyses showed a significant main effect of Familiarity in the 600–850 ms and Expertise by Familiarity interaction in the 750- to 850-ms latency bands (P = 0.001 and 0.02, respectively). Only the P value for the Familiarity by Location interaction in the 550- to 750-ms latency band was no more significant (P = 0.23), pointing to a greater sensitivity of correct trials analyses.

Musical Test

The same analyses described for the linguistic test were performed but on all trials. Indeed, because in this test behavioral responses did not significantly differ from chance, an analysis on correct trials would not be appropriate. Moreover, participants' correct trials were not enough to yield an acceptable signal-to-noise ratio. This was also the reason why we decided not to pool together the data from the linguistic and musical tests.

Results of the ANOVA on the N1 peak amplitude showed a main effect of Expertise (F [1,35] = 4.24; P = 0.04): The N1 amplitude was larger for musicians (−4.8 μV) than for nonmusicians (−3.2 μV, see Fig. 6). The familiarity effect was not significant (F [1,35] = 1.1; P = 0.3). Analyses of the P2 component also showed a significant main effect of Expertise (F [1,35] = 9.73; P = 0.003): the P2 amplitude was larger for musicians (4.4 μV) than for nonmusicians (2.2 μV). Most interestingly, this analysis also revealed a significant Expertise by Familiarity by Hemisphere interaction (F [1,35] = 4.13; P = 0.04). Post hoc analyses showed a significant Familiarity effect size for musicians but not for nonmusicians (1 μV vs. 0.1 μV). Musicians' Familiarity effect was larger over the left hemisphere (1.2 μV) than over the right (0.7 μV).

Between 350 and 500 ms, the Expertise by Familiarity by Location by Hemisphere interaction was significant (F [2,70] = 4.98; P = 0.02). Post hoc analyses revealed a significant familiarity effect for musicians but not for nonmusicians (0.75 vs. 0.15 μV). Musicians Familiarity effect was largest over central and left frontal regions (0.9-μV effect size, P < 0.001, see Figs. 7 and 8).[2]

Between 600 and 700 ms, results of the ANOVA showed a significant main effect of Familiarity (F [1,35] = 3.99; P = 0.05). In this latency range, unfamiliar melodies (−1.2 μV) elicited more negative ERPs than familiar melodies (−0.7 μV).

Between 700 and 800 ms, results of the ANOVA showed a significant Expertise by Familiarity by Location by Hemisphere interaction (F [2,70] = 3.42; P = 0.05). Post hoc analyses revealed a significant familiarity effect for musicians but not for nonmusicians (0.4 vs. 0.1 μV). Musicians' Familiarity effect was significant over right frontal regions only (0.4 μV effect size, P < 0.001; see Fig. 8).

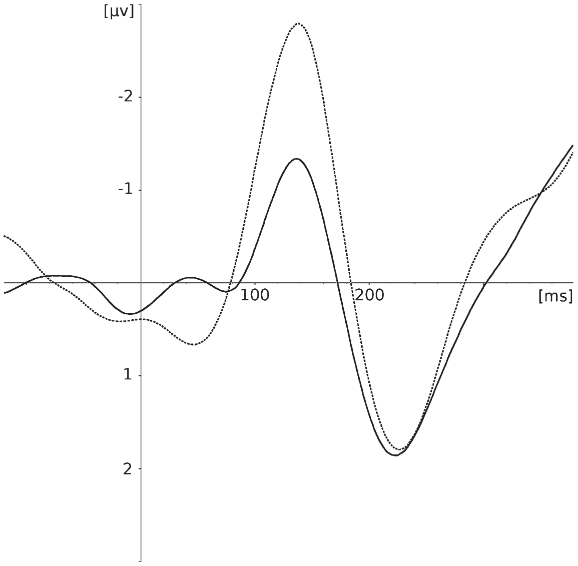

Figure 3. Linguistic test: modulation of the N1-P2 complex as a function of expertise at Cz electrode (dashed line = musicians, solid line = nonmusicians).

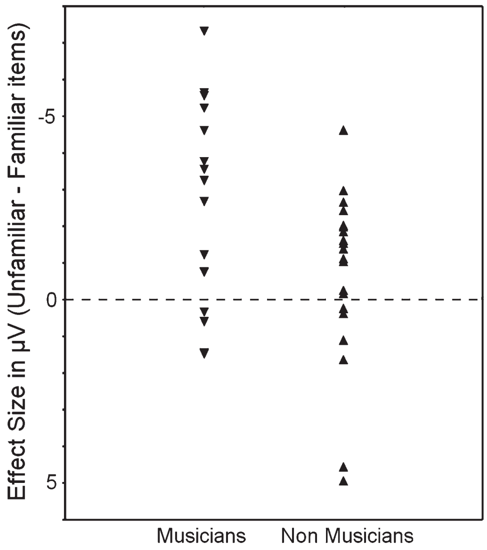

Figure 4. Familiarity effect size in the linguistic dimension for musicians and nonmusicians in the 750–850 latency band at Cz electrode (negativity is up).

Discussion

The main goal of this study was to test whether artificial language learning (segmentation) may benefit from formal music training. With this aim, we recorded both behavioral and electrophysiological measures during the testing sessions in an artificial language learning paradigm and we compared musician and nonmusician participants.

Music Training Shapes the Auditory System

In both tested dimensions, electrophysiological data revealed an effect of expertise on the N1 component with larger amplitude for musicians than for nonmusicians. Previous studies described a similar effect of expertise on the N1 in response to musical sounds (Pantev et al. 1998; Shahin et al. 2003). This effect is even stronger for timbers of the played instrument (e.g., trumpet sound for trumpeters, Pantev et al. 2001). These functional changes may rely on anatomical changes induced by musical training in the auditory regions

---

[2] We interpret this negativity as an MMN-like component to the second tone. One way to test this is to analyze separately the 2 types of unfamiliar items. Type 312 are built using the last syllable of a word plus the first syllable pair of another word whereas type 231 are built using the last syllable pair of a word plus the first syllable of another word). Because in type 312 items the "illegal" transition takes place at the second note, whereas for type 231 it does not, we expected a larger negativity to type 312 than to type 231. We conducted additional analyses in the 350- to 500-ms latency band aiming at comparing mean amplitudes elicited by these 2 different types of unfamiliar items. The Expertise by Item by Hemisphere was significant (F [1,35] = 6.41; P = 0.01). Post hoc analyses revealed that, for musicians only, ERPs for 312 unfamiliar items were significantly more negative than for 231 items over the right (P = 0.04) but not over the left Hemisphere (P = 0.99). No significant differences were found for nonmusicians.

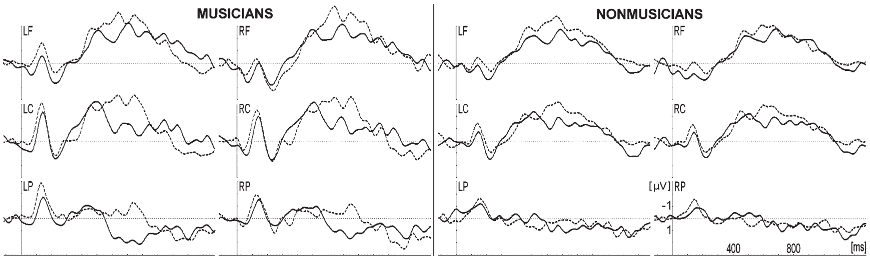

Figure 5. Musicians (Left side) and nonmusicians (right side), ERPs to familiar (solid) and unfamiliar words (dashed) in the linguistic session. Each site represents the mean of the 3 electrodes included in the ROI (LF: left frontal, RF: right frontal, LC: left central, RC: right central, LP: left parietal, RP: right parietal).

(Schlaug et al. 1995; Keenan et al. 2001; Schneider et al. 2002). Moreover, recent studies using musical or linguistic stimuli and focusing on an earlier stage of auditory processing (Musacchia et al. 2007; Wong et al. 2007; Lee et al. 2009; Parbery-Clark et al. 2009; Bidelman and Krishnan 2010), found a larger frequency-following response in the brainstem of musicians compared with that of nonmusicians. Our results show that these differences can also be found at later stages of processing, such as the N1, probably originating in the belt and parabelt auditory regions (Liégeois-Chauvel et al. 1994; Hackett et al. 2001). However, care must be taken in the interpretation of the present results, as one cannot exclude a top-down effect such as an increase of selective attention to sounds for musicians. Nonetheless, recent results indicate that the effect of music expertise on the N1-P2 complex is not primarily caused by selective attention. Rather, they support the view that increased auditory-evoked potentials in musicians reflect an enlarged neuronal representation for specific sound features of these tones (Baumann et al. 2008).

Enhanced Categorization of Tonal and Interval Information in Musicians

In the musical test, only musicians showed a significant familiarity effect on the P2 component. The P2 is known to be sensitive to the level of attention and to the level of musical expertise (Hillyard et al. 1987; Shahin et al. 2003; Kuriki et al. 2006). P2 amplitude modulations have also been reported in an auditory stream segregation paradigm: The P2 amplitude is greater when a stream of 2 sounds can be perceptually segregated in 2 independent streams (Snyder et al. 2006). Interestingly, a P2 modulation has been reported using artificial language learning paradigms, using linguistic stimuli presented in the auditory modality: The P2 amplitude increases along the listening phase for good learners but not for poor learners (De Diego Balaguer et al. 2007). In our case, the P2 observed in musicians was larger for unfamiliar than for familiar melodies. This is probably due to an increase in short-term memory load (Conley et al. 1999) or to a more difficult categorization of the unfamiliar than familiar items (Liebenthal et al. 2010). Interestingly, this familiarity effect took place 200 ms after the first tone onset that is before the beginning of the second tone. Because the musical structure of the sung language contained a rather strong tonal center (10 notes of 11 were in C major), it is possible that musicians kept in memory a representation of this tonal center that influenced the processing of the first tone of each item (Krumhansl and Kessler 1982). In fact, once a tonal center is established, each individual pitch can be processed and categorized relative to the tonal center (with no need of absolute pitch). Also note that, because the pitches of the first tone of familiar and unfamiliar items all differed (except for one), a good performance was theoretically possible on the sole basis of the information carried by the first pitch. Taken together, these results suggest that musicians might be able to discriminate familiar and unfamiliar musical items based on the tonal function of the first pitch.

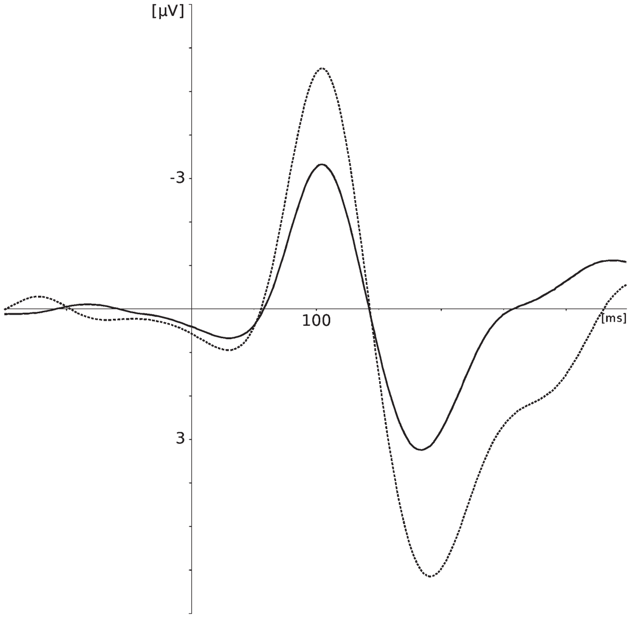

Figure 6. Musical test: modulation of the N1-P2 complex as a function of expertise at Cz electrode (dashed line = musicians, solid line = nonmusicians).

The familiarity effect on the P2 was followed by another familiarity effect on a negative component, only significant for musicians in the music test. This component, larger for unfamiliar than familiar items had its peak around 450 ms. Due to its narrow width, rather than interpreting it as a late negative component, we consider it as an MMN in response to the second tone of the melody (see Fig. 7). This interpretation is supported by the fact that this effect was stronger for the unfamiliar items, wherein the second tone was an illegal transition (i.e., "312" type: last syllable of a word plus the first syllable pair of another word) compared with items, wherein the second tone was a legal transition (i.e., "231" type: last syllable pair of a word plus the first syllable of another word).

Although the MMN is typically elicited by simple changes in a regular sequence (Näätänen et al. 1978), it has also been shown for more abstract rules (Tervaniemi et al. 1994). Interestingly, Pantev and collaborators (Herholz et al. 2008) compared musicians and nonmusicians in a musical imagery

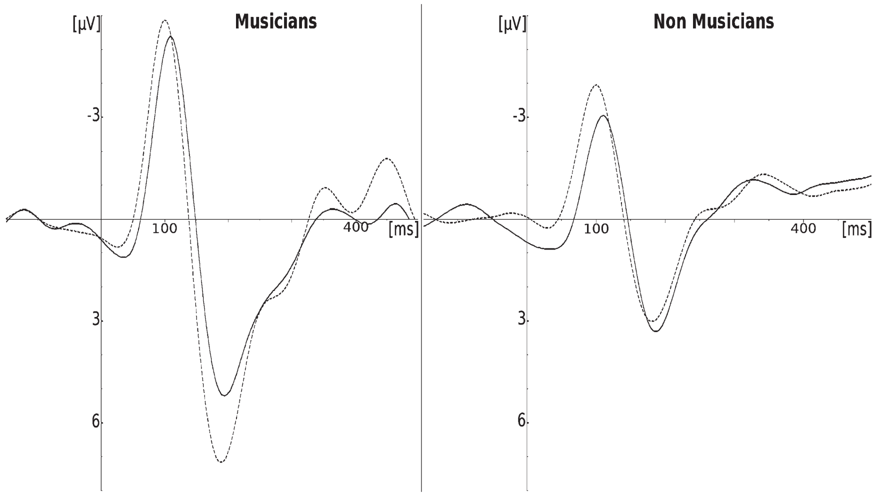

Figure 7. Musicians (left side) and nonmusicians (right side), ERPs to familiar (solid) and unfamiliar melodies (dashed) in the musical session, at Cz electrode.

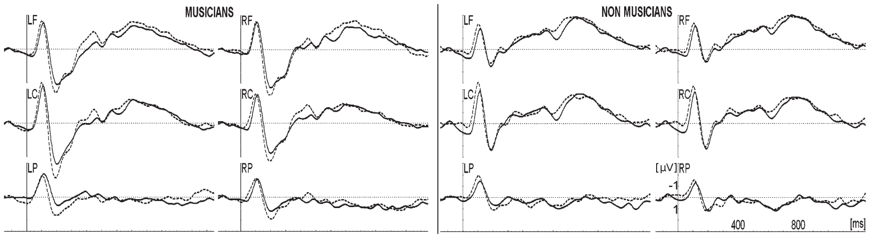

Figure 8. Musicians (left side) and nonmusicians (right side): ERPs to familiar (solid) and unfamiliar melodies (dashed) in the musical session. Each site represents the mean of the 3 electrodes included in the ROI (LF: left frontal; RF: right frontal, LC: left central, RC: right central, LP: left parietal, RP: right parietal).

task with familiar melodies. Participants listened to the incipit of the melodies, continued them in their mind and then heard a tone that fitted in or not with the melody. Only musicians showed an MMN to incorrect continuations. These results suggest that a long-term memory trace (familiar melody) can induce, in musicians, a sufficiently strong representation to yield an MMN to an incorrect item. This is in line with our results. First, participants stored in memory the representations of melodies segmented during learning; then they heard, during the test, familiar (correct) or unfamiliar (incorrect) items. The larger MMN to the second tone of unfamiliar items can thus be interpreted as a mismatch of the first interval with respect to a memory trace stored during the learning session. Indeed, the second tone allows the building of the first interval of the melodies, also giving, in theory, all the necessary information to discriminate familiar from unfamiliar items.

Overall, both familiarity effects on the P2 and MMN-like components point to the fact that musicians did learn the musical structure better than nonmusicians, and this in turn affected the way they processed the musical items presented during the test. This confirms and extends previous findings showing that implicit learning of 12-tone serialist music is facilitated by experience with this music style. Indeed, while participants without a specialized training in atonal music cannot (implicitly) perceive the transforms of serialist music, participants with routine exposure to atonal music do (implicitly) perceive the distinction between different types of transforms (Dienes and Longuet-Higgins 2004). Again, as stated for the results on the N1, one cannot exclude that musicians pay more attention to sounds in general and to the musical material in particular, even when the experimental tasks do not require them to do so. Indeed, also in our experiment, attentional processes during the passive listening may differ between musicians and nonmusicians in terms of selective attention or in terms of switching attention from one dimension to the other, and this may in turn affect behavioral and/or electrophysiological data beyond direct influences related to expertise.

Familiarity Effect and Memory Retrieval

In both tested dimensions, we found a larger frontal negativity for unfamiliar words than for familiar words (600- to 850-ms interval for the linguistic dimension and 600- to 800-ms interval for the musical one). This result is interesting as it confirms our previous results with nonmusicians showing a similar fronto-central negativity modulated by the degree of familiarity in both dimensions (Francois and Schön 2010). Due to the topography and the sensitivity to word familiarity, we previously interpreted this component as an N400-like component, possibly reflecting the memory trace stored during the listening phase. Although the latency is delayed with respect to the typical N400, one should consider that the items used here were sung words without any meaning. Interestingly, previous findings showed that the latency of the N400 to pseudowords is around 600 ms (O'Rourke and Holcomb 2002). Moreover, the involvement of N400-like components has been reported in various studies focusing on artificial language learning (Sanders et al. 2002; Cunillera et al. 2006, 2009; De Diego Balaguer et al. 2007). Its modulation is often reported during the listening phase and is interpreted as an index of word prelexicalization due to speech segmentation processes. To our knowledge, only one study briefly analyzed and discussed the ERP data obtained during the testing phase, reporting a larger N400 for words than for nonwords (De Diego Balaguer et al. 2007). However, 2 important differences with the present design might explain this discordance. First, in the De Diego Balaguer experiment, participants underwent 2 successive learning phases, one with a speech stream similar to ours, and one with a stream including violations. Second, the testing phase contrasted words to illegal words (i.e., not present in the speech stream), while we contrasted familiar and unfamiliar words. While De Diego Balaguer and collaborators interpreted the N400 reduction to nonwords as reflecting the detection of impossible linguistic items, we interpret the N400 increase to unfamiliar words in terms of a classic N400 familiarity effect. Moreover, an alternative interpretation of the De Diego Balaguer results could also be that illegal words may lead to a quickly abandoned search resulting in smaller N400 amplitude for illegal compared with legal words. Indeed, Young and Rugg (1992) showed that unfamiliar words elicit a larger N400 than familiar words do, probably due to a longer lexical search.

Overall, the presence of this familiarity effect points to an ability of both musicians and nonmusicians to segment both linguistic and musical structures. As we previously found in this type of paradigm for nonmusicians (Francois and Schön 2010), the electrophysiological measures (on a N400-like component) seem to be more sensitive than behavioral measures. While accuracy measures pointed to a lack of learning of the musical dimension by nonmusicians and musicians, ERP results are clear cut in showing that both groups process differently familiar and unfamiliar items.

Musical Training Facilitates Implicit Learning of Both Linguistic and Musical Structures

The most important finding of this experiment is that such a familiarity N400-like effect was significantly greater in musicians compared to nonmusicians in both dimensions. This greater N400-like effect can be interpreted in terms of more distinguished memory representations for familiar and unfamiliar items in musicians. Indeed, the N400 amplitude seems to be sensitive to the ease of accessing information from long-term memory (Neville et al. 1986) and predicts very well immediate recall performance after learning (Elger et al. 1997).

These results reveal that musicians do have more “robust” representations of musical and linguistic structures shaped during the listening phase than nonmusicians. Even though the behavioral data are not clear-cut concerning an advantage of musicians over nonmusicians, several electrophysiological indices confirm the trend observed in the behavioral results. First, the familiarity effects present on the P2 to the first tone and on the MMN to the second tone show that musicians process differently familiar and unfamiliar musical items quite early on. Second, the familiarity effect present on a later negative component is larger for musicians than nonmusicians in both tested dimensions, pointing to a greater distinctiveness between familiar and unfamiliar items in musicians.

The most straightforward explanation of such an effect of Expertise is that musicians may be better in using the musical structure contained in song language than nonmusicians and that this might be due to a better perception and/or memory for pitch. Another explanation might be that musical expertise facilitates regularity extractions and sequence learning in general. It is known that musicians can organize a sound sequence according to a number regularity, for instance, implicitly distinguishing segments containing 4 tones from the segments containing 5 tones, while such a perceptual organization of sound according to number is less relevant for nonmusicians (Van Zuijen et al. 2005). Moreover, recent findings showed that deaf children with cochlear implants are impaired in visual sequence learning, suggesting that a period of auditory deprivation may have a major impact on cognitive processes that are not specific to the auditory modality (Conway et al. 2011). Therefore, sound seems to provide a cognitive scaffolding for the development of serial-order behavior: whether sound processing is impaired or whether it is extensively practiced have opposite effects. Interestingly, Sluming et al. (2002) reported for musicians, an increased gray matter density and volume in the left-inferior frontal gyrus, Broca’s area. Moreover, neuroplastic development throughout musicians’ life seems to promote the retention of cortical tissue. This region is known to be involved in on-line speech stream segmentation as well as in music perception (Tillmann et al. 2003; McNealy et al. 2006). Overall, our results support an “auditory scaffolding hypothesis” (Conway et al. 2009), as we presently show that increased exposure to sounds leads to a benefit for implicit learning, putatively via anatomical and/or functional modifications going beyond the auditory regions. This second explanation makes the prediction that musicians should also have an advantage in learning an artificial spoken language, that is, without any specific musical cue. Further work will be necessary to test this possibility.

Finally, it is important to note that the differences due to expertise were smaller in the musical (familiarity effect musicians − familiarity effect nonmusicians = 0.4 μV) than in the linguistic dimension (1.5 μV). This is not surprising insofar as musicians, as well as nonmusicians, had troubles in learning the musical structure. While one may want to interpret such an advantage in speech compared with music processing as a general advantage of language learning over musical learning or as a greater resistance to interference of verbal items than musical items (Schendel and Palmer 2007), care must be taken in comparing results across dimensions because it is very difficult to control and balance the perceptual saliency of the linguistic and musical structures.

To conclude, these results contribute new evidence to the positive effects of musical practice on language segmentation. Although more research is needed to disentangle whether this effect is a general effect or more specific to the sung material we used, the results described above show that musicians are better able to segment both linguistic and musical structures. Because we focused here on the data acquired during tests taking place straight after learning, future studies will need to address whether these differences are due to an advantage in stream segmentation or in lexical storage or both.

Funding

French National Research Agency (ANR 05-BLAN-0214 “Music and Memory” to D.S.).

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Notes

We thank Julie Chobert and Jérome Daltrozzo for assistance in EEG recordings. We thank Deirdre Bolger, Nia Cason, Mireille Besson, and 2 anonymous reviewers for constructive comments on a previous version. *Conflict of Interest* : None declared.

References

Anvari SH, Trainor LJ, Woodside J, Levy BA. 2002. Relations among musical skills, phonological processing, and early reading ability in preschool children. J Exp Child Psychol. 83(2):111-130.

Aslin RN, Saffran JR, Newport EL. 1998. Computation of conditional probability statistics by 8-month-old infants. Psychol Sci. 9:321-324.

Baumann S, Meyer M, Jäncke L. 2008. Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. J Cogn Neurosci. 20(12):2238-2249.

Bermudez P, Zatorre RJ. 2005. Differences in gray matter between musicians and nonmusicians. Ann New York Acad Sci. 1060: 395-399.

Bidelman GM, Krishnan A. 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 1355:112-125.

Brattico E, Pallesen KJ, Varyagina O, Bailey C, Anourova I, Järvenpää M, Eerola T, Tervaniemi M. 2009. Neural discrimination of non-prototypical chords in music experts and laymen: an MEG study. J Cogn Neurosci. 21(11):2230-2244.

Conley EM, Michalewski HJ, Starr A. 1999. The N100 auditory cortical evoked potential indexes scanning of auditory short-term memory. Clin Neurophysiol. 110(12):2086-2093.

Conway CM, Pisoni DB, Anaya EM, Karpicke J, Henning SC. 2011. Implicit sequence learning in deaf children with cochlear implants. Dev Sci. 14(1):69-82.

Conway CM, Pisoni DB, Kronenberg WG. 2009. the importance of sound for cognitive sequencing abilities: the auditory scaffolding hypothesis. Curr Dir Psychol Sci. 18(5):275-279.

Cunillera T, Càmara E, Toro JM, Marco-Pallares J, Sebastián-Galles N, Ortiz H, Pujol J, Rodríguez-Fornells A. 2009. Time course and functional neuroanatomy of speech segmentation in adults. Neuroimage. 48(3):541-553.

Cunillera T, Toro JM, Sebastián-Gallés N, Rodríguez-Fornells A. 2006. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. Brain Res. 1123(1):168-178.

De Diego Balaguer R, Toro JM, Rodriguez-Fornells A, Bachoud-Lévi AC. 2007. Different neurophysiological mechanisms underlying word and rule extraction from speech.. PLoS ONE. 2(11):e1175.

Dienes Z, Longuet-Higgins C. 2004. Can musical transformations be implicitly learned? Cogn Sci. 28:531-558.

Elger CE, Grunwald T, Lehnertz K, Kutas M, Helmstaedter C, Brockhaus A, Van Roost D, Heinze HJ. 1997. Human temporal lobe potentials in verbal learning and memory processes. Neuropsychologia. 35(5):657-667.

Francois C, Schön D. 2010. Learning of musical and linguistic structures: comparing event-related potentials and behavior. Neuroreport. 21(14):928-932.

Gervain J, Macagno F, Cogoi S, Peña M, Mehler J. 2008. The neonate brain detects speech structure. Proc Natl Acad Sci U S A. 105: 14222-14227.

Hackett TA, Preuss TM, Kaas JH. 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J Comp Neurol. 441(3):197-222.

Hauser MD, Newport EL, Aslin RN. 2000. Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. Cognition. 75:1-12.

Hauser MD, Weiss D, Marcus G. 2002. Rule learning by cotton-top tamarins. Cognition. 86(1):B15-B22.

Herholz SC, Lappe C, Knief A, Pantev C. 2008. Neural basis of music imagery and the effect of musical expertise. Eur J Neurosci. 28(11):2352-2360.

Hillyard SA, Woldorff M, Mangun GR, Hansen JC. 1987. Mechanisms of early selective attention in auditory and visual modalities. Electroencephalogr Clin Neurophysiol. 39(Suppl):317-324.

Keenan JP, Thangaraj V, Halpern AR, Schlaug G. 2001. Absolute pitch and planum temporale. Neuroimage. 14(6):1402-1408.

Kishon-Rabin L, Amir O, Vexler Y, Zaltz Y. 2001. Pitch discrimination: are professional musicians better than non-musicians? J Basic Clin Physiol Pharmacol. 12(2 Suppl):125-143.

Koelsch S, Schröger E, Tervaniemi M. 1999. Superior pre-attentive auditory processing in musicians. Neuroreport. 10(6):1309-1313.

Krampe RT, Ericsson KA. 1996. Maintaining excellence: deliberate practice and elite performance in young and older pianists. J Exp Psychol. 125:331-359.

Krumhansl CL, Kessler EJ. 1982. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychol Rev. 89:334-368.

Kuhl PK. 2004. Early language acquisition: cracking the speech code. Nat Rev Neurosci. 5:831-843.

Kuriki S, Kanda S, Hirata Y. 2006. Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. J Neurosci. 26(15):4046-4053.

Lee KM, Skoe E, Kraus N, Ashley R. 2009. Selective subcortical enhancement of musical intervals in musicians. J Neurosci. 29(18): 5832-5840.

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR. 2010. Specialization along the left superior temporal sulcus for auditory categorization. cerebral cortex. 20(12):2958–2970.

Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P. 1994. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol. 92(3):204-214.

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 Cogn Neurosci. 18(2):199-11.

Makeig S, Bell AJ, Jung TP, Sejnowski TJ. 1996. Independent component analysis of electroencephalographic data. Adv Neural Inform Proc Syst. 8:145-151.

McNealy K, Mazziota JC, Dapretto M. 2006. Cracking the language code: neural mechanisms underlying speech parsing. J Neurosci. 26: 7629-7639.

Milovanov R, Huotilainen M, Esquef PA, Alku P, Välimäki V, Tervaniemi M. 2009. The role of musical aptitude and language skills in preattentive duration processing in school-aged children. Neurosci Lett. 460(2):161-165.

Milovanov R, Huotilainen M, Välimäki V, Esquef PA, Tervaniemi M. 2008. Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. Brain Res. 1194:81-89.

Musacchia G, Sams M, Skoe E, Kraus N. 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci U S A. 104(40):15894-15898.

Näätänen R, Gaillard AW, Mäntysalo S. 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. 42(4): 313-329.

Neville HJ, Kutas M, Chesney G, Schmidt AL. 1986. Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. J Mem Lang. 25(1):75-92.

O'Rourke TB, Holcomb PJ. 2002. Electrophysiological evidence for the efficiency of spoken word processing. Biol Psychol. 60:121-150.

Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. 1998. Increased auditory cortical representation in musicians. Nature. 392(6678):811-814.

Pantev C, Roberts LE, 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. 29(45):14100–14107.

Ramus F, Hauser MD, Miller C, Morris D, Mehler J. 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. Science. 288:349–351.

Rüsseler J, Altenmüller E, Nager W, Kohlmetz C, Münte TF. 2001. Event-related brain potentials to sound omissions differ in musicians and non-musicians. Neurosci Lett. 308(1):33–36.

Saffran JR. 2003. Statistical language learning: mechanisms and constraints. Curr Direc Psychol Sci. 12(4):110–114.

Saffran JR, Aslin RN, Newport EL. 1996a. Statistical learning by 8-month old infants. Science. 274:1926–1928.

Saffran JR, Johnson EK, Aslin RN, Newport EL. 1999. Statistical learning of tones sequences by human infants and adults. Cognition. 70:27–52.

Saffran JR, Newport EL, Aslin RN. 1996b. Word segmentation: the role of distributional cues. J Mem Lang. 35:606–621.

Saffran JR, Senghas A, Trueswell JC. 2001. The acquisition of language by children. Proc Natl Acad Sci U S A. 98:12874–12875.

Sanders LD, Newport EL, Neville JH. 2002. Segmenting nonsense: an event-related-potential index of perceived onsets in continuous speech. Nat Neurosci. 5(7):700–703.

Schendel ZA, Palmer C. 2007. Suppression effects on musical and verbal memory. Mem Cogn. 35(4):640–650.

Schlaug G, Jäncke L, Huang Y, Staiger JF, Steinmetz H. 1995. Increased corpus callosum size in musicians. Neuropsychologia. 33(8): 1047–1055.

Schneider P, Scherg M, Dosch HG, Specht HJ, Gutschalk A, Rupp A. 2002. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nat Neurosci. 5:688–694.

Schön D, Boyer M, Moreno S, Besson M, Peretz I, Kolinsky R. 2008. Song as an aid for language acquisition. Cognition. 106(2):975–983.

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.

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci. 23(13):5545–5552.

Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B. 2005. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. Neuroreport. 16(16):1781–1785.

Slevc LR, Miyake A. 2006. Individual differences in second-language proficiency: does musical ability matter? Psychol Sci. 17(8):675–681.

Sluming V, Barrick T, Howard M, Cezayirli E, Mayes A, Roberts N. 2002. Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. Neuroimage. 17(3):1613–1622.

Snyder JS, Alain C, Picton TW. 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci. 18(1):1–13.

Teinonen T, Fellman V, Näätänen R, Alku P, Huotilainen M. 2009. Statistical language learning in neonates revealed by event-related brain potentials. BMC Neurosci. 13:10–21.

Tervaniemi M, Maury S, Näätänen R. 1994. Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. Neuroreport. 5(7):844–846.

Tillmann B, Janata P, Bharucha JJ. 2003. Activation of the inferior frontal cortex in musical priming. Cogn Brain Res. 16(2):145–161.

Tillmann B, McAdams S. 2004. Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)-similarities.. J Exp Psychol Learn Mem Cogn. 30:1131–1142.

Toro JM, Trobalón JB. 2005. Statistical computations over a speech stream in a rodent. Percept Psychophys. 67(5):867–875.

Van Zuijen TL, Sussman E, Winkler I, Näätänen R, Tervaniemi M. 2005. Auditory organization of sound sequences by a temporal or numerical regularity: a mismatch negativity study comparing musicians and non-musicians. Brain Res Cogn Brain Res. 23(2–3):270–276.

Wong PC, Skoe E, Russo NM, Dees T, Kraus N. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci. 10(4):420–422.

Young MP, Rugg MD. 1992. Word frequency and multiple repetitions as determinants of the modulation of event-related-potentials in a semantic classification task. Psychophysiology. 29(6):664–676.