

# Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening

Giovanni M. Di Liberto<sup>1,\*</sup>, Claire Pelofi<sup>2</sup>, Shihab Shamma<sup>1,3</sup> and Alain de Cheveigné<sup>1,4</sup>

<sup>1</sup>*Laboratoire des Systèmes Perceptifs, UMR 8248, CNRS, France. Ecole Normale Supérieure, PSL University, France*

<sup>2</sup>*Institut de Neurosciences des Système, UMR S 1106, INSERM, Aix Marseille Université, France*

<sup>3</sup>*Institute for Systems Research, Electrical and Computer Engineering, University of Maryland, College Park, USA*

<sup>4</sup>*UCL Ear Institute, London, United Kingdom*

(Received 30 September 2018, Accepted for publication 12 April 2019)

**Keywords:** Cortical tracking, EEG, Envelope, Acoustics, Brain signals

**PACS number:** 43.75.Cd [doi:10.1250/ast.41.361]

## 1. Introduction

Music is produced, listened, and enjoyed across all known cultures. Yet the question of what in the music makes it so ubiquitous remains largely unanswered. The answer to this fundamental question may be uncovered by a better understanding of the cortical underpinnings of music perception. Indeed, a large body of research has investigated the brain processes that relate to its fundamental attributes such as pitch, timbre, melodic contour, and rhythm [1]. However, most of these studies explored these attributes individually, often using short, tailored stimuli that largely differ from the music that we normally listen to.

Recent advances in the field of neuroscience have provided ways to investigate the cortical processing of continuous stimuli. Auditory research has indicated that brain signals *track* the energy fluctuations (i.e., envelope) of auditory inputs [2,3]. As a result, there are now frameworks to directly measure the brain responses to the acoustic envelope using continuous stimuli of arbitrary length and non-invasive brain recordings, such as scalp electroencephalography (EEG) [4,5].

In the present study, we use such an EEG framework [6] to investigate the cortical tracking of the acoustic envelope of continuous monophonic classical music as was previously applied to speech and other acoustic stimuli. In speech perception, this envelope tracking was shown to be significantly modulated by cognitive factors such as expectations [7] and attention [8], factors that are also central to music appreciation and perception [9]. Here we demonstrate the neural correlates of these influences, as well as the effects of musical expertise and of repetition on the cortical tracking of the music envelope.

## 2. Methods

### 2.1. Subjects and data acquisition

Fourteen healthy subjects (7 male, aged between 24 and 37,  $M = 29$ ) participated in this study. Seven of them were highly trained musicians with a degree in music and at least ten years of experience, while the other participants had no musical background. Each subject provided written

informed consent and was paid for their participation. Subjects reported no history of hearing impairment or neurological disorder. The study was undertaken in accordance with the Declaration of Helsinki and was approved by the CERES committee of Paris Descartes University (CERES 2013-11).

The experiment was carried out in a single session for each subject. EEG data were recorded from 64 electrode positions, digitised at 512 Hz using a BioSemi Active Two system. Audio stimuli were presented at a sampling rate of 44,100 Hz using Sennheiser HD650 headphones and Presentation software (<http://www.neurobs.com>). Testing was carried out in a dark room and subjects were instructed to maintain visual fixation on a crosshair centered on the screen, and to minimise motor activities while music was played.

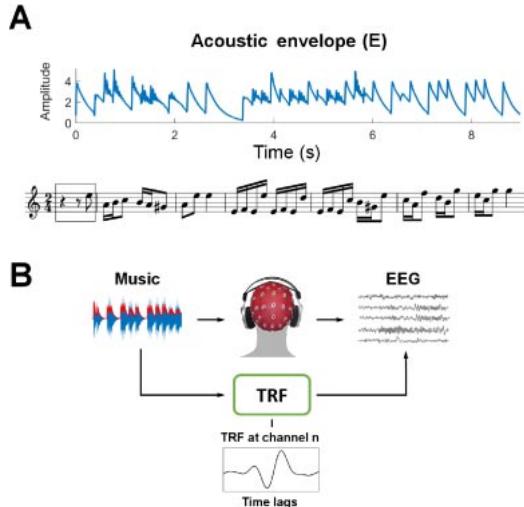
### 2.2. Stimuli and experimental procedure

Monophonic MIDI versions of ten musical pieces from Bach's monodic instrumental corpus were partitioned into short snippets of approximately 120 s. Specifically, violin and flute tracks were used, replacing the original instrument with MIDI piano sounds. This was done in order to reduce familiarity for the expert pianist participants while enhancing their neural response by using their preferred instrument timbre [10]. Each 120 s piece, corresponding to an EEG trial, was presented three times throughout the experiment, adding up to 30 trials in total that were presented in a random order. At the end of each trial, participants were asked to report on their familiarity to the piece (from 1: unknown; to 7: know the piece very well). This rating could take into account both their familiarity with the piece at its first occurrence in the experiment, as well as the build-up of familiarity across repetitions.

### 2.3. EEG preprocessing and data analysis

EEG data were analysed offline using MATLAB (*The Mathworks Inc.*), digitally filtered between 0.5 and 45 Hz using a Butterworth zero-phase filter (low- and high-pass filters with order 2 and implemented with the function *filtfilt*), and down-sampled to 128 Hz. EEG channels with a variance exceeding three times that of the surrounding ones were replaced by an estimate calculated using spherical spline interpolation. All channels were re-referenced to the average

\*e-mail: diliberg@tcd.ie



**Fig. 1** EEG data were recorded while subjects listened to monophonic music. (A) The acoustic envelope was calculated for each musical piece. (B) Linear regression was used to fit multivariate temporal response functions (TRFs) between the music envelope and the corresponding brain responses measured with EEG.

of the two mastoid channels with the goal of maximising the auditory EEG responses [11].

The broadband amplitude envelope was extracted from the acoustic waveform using the Hilbert transform (Fig. 1A). Then, a system identification technique was used to compute the channel-specific mapping between the music envelope and the recorded EEG data. This method, here referred to as the Temporal Response Function (TRF; [4,5]), uses a regularized linear regression [6] to estimate a filter that optimally describes how the brain transforms the acoustic envelope into the corresponding neural response (forward model; Fig. 1B). The TRF takes into consideration multiple time-lags between stimulus and neural signal, providing us with patterns of model weights interpretable in both space (scalp topographies) and time (music-EEG latencies). Additionally, leave-one-out cross-validation was used to assess how well the model can predict unseen data. This was quantified by calculating Pearson's correlation between the recorded signals and the corresponding predictions for each scalp electrode. A time-lag window of  $-50$ – $350$  ms was used to fit the TRF models in the first instance. The window was then restricted to the significant peaks (0– $200$  ms) for the EEG prediction analysis.

While forward models offer a view on the spatio-temporal dynamics of the cortical responses to music, backward approaches combine multivariate EEG data to reconstruct the univariate sound envelope [6]. This produces a single correlation score that is generally larger and more reliable than what can be obtained with forward models [12]. Here, we use this approach to further investigate the effect of musical expertise and trial repetition. Specifically, data within each subject-group were combined using multiway canonical correlation analysis [13,14] which, given temporally aligned brain responses, extracts common brain components that are most shared (i.e., largest correlation) across subjects. This

allows for more accurate modelling and larger reconstruction correlations.

### 3. Results

The primary focus of this experiment was to investigate the effect of musical expertise on the cortical tracking of the envelopes of musical signals. We used forward- and backward-models to describe how the music envelope is transformed into EEG signals and vice versa. The analysis below details the contrast among groups of participants and the effects of stimulus repetition.

#### 3.1. Behavioural results

As expected, behavioural results in Fig. 2A illustrate that participants perceived repeated pieces as more familiar (paired *t*-test on the average familiarity ratings for all participants across repetitions:  $rep_2 > rep_1$ ,  $p = 2.4 \times 10^{-5}$ ,  $rep_3 > rep_2$ ,  $p = 0.005$ ). No significant differences emerged between musicians and non-musicians on this account (two-sample *t*-test,  $p = 0.4, 0.5, 0.6$  for repetitions 1, 2, and 3 respectively), indicating that musical expertise does not play a significant role in shaping the ability to recognise the monophonic pieces chosen for this study.

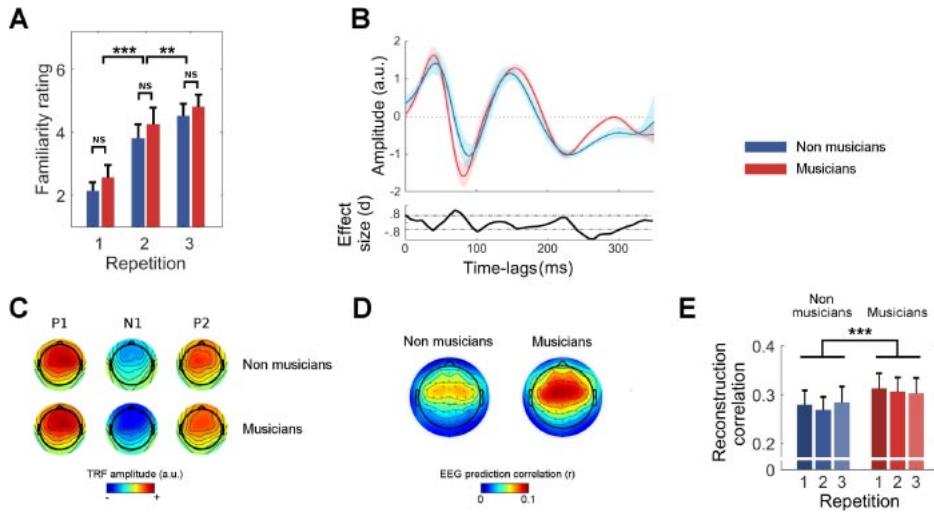
#### 3.2. Musical expertise enhances envelope tracking

64-channel EEG measurements were made from musicians and non-musician participants as they listened to the musical excerpts. First, forward TRF models were fitted for individual subjects and scalp channels. Data from one non-musician subject was excluded from further analysis as it did not show meaningful TRF responses. After averaging the model weights across subjects for each group, normalized TRFs for the nine best predicted channels were averaged to investigate the temporal dynamics of the system (Fig. 2B). TRFs for both groups have magnitude significantly larger than zero for components that are typical of envelope responses (P1, N1, and P2; [4]). The effect-size of the group amplitude difference was calculated at each time-lag, showing a strong effect of musical expertise corresponding to the N1 peak (Cohen's  $d > 0.8$ ).

The data show trends that suggest a small latency shift for the N1 and P2 peaks, which occurred at earlier and longer latencies for musicians than non-musicians respectively (permutation test, 10,000 repetitions,  $p_{P1} = 0.316$ ,  $p_{N1} = 0.053$ ,  $p_{P2} = 0.063$ ). Figure 2C provides another view of the TRF result by showing the topographical maps for each group and peak of interest, which are consistent with what was previously seen in other studies with TRFs and auditory stimulation (e.g., [4,15]).

EEG data were then predicted using these TRF models (Fig. 2D). The best predicted electrodes emerged in one broad centro-parietal area of the scalp and, importantly predictions were far more accurate or better correlated with the acoustic envelopes for musicians than for non-musician participants (two-sample *t*-test,  $p \lesssim 0.05$  for a cluster of 8 electrodes over central scalp areas).

Finally, we sought to support this result by conducting a more powerful analysis that combines all data for each group (subjects, EEG channels, time-lags) to reconstruct the univariate and clean (as opposed to the noisy EEG signal) music envelope. This analysis produced an envelope reconstruction



**Fig. 2** (A) Familiarity ratings for each subject group and repetition ( $M \pm SE$  across subjects) showed a significant effect of repetition ( $^{**}p < 0.01$ ,  $^{***}p < 0.001$ ) but not of group ( $p > 0.05$ ). (B) Average TRF weights at the best predicted electrodes. The shaded areas indicate the standard error of the mean across subjects. Time-lag zero indicates the instant the sound is heard, while subsequent lags relate to brain responses after the specific duration of time. (C) Topographical maps for the TRF weights for the peaks P1, N1, and P2. (D) EEG prediction correlations for musicians and non-musicians. (E) Envelope reconstruction correlations using EEG and a backward modelling approach ( $M \pm SE$  across trials). This analysis shows significant differences between musicians and non-musicians ( $^{***}p < 0.001$ ) and no effect of repetition.

score (Pearson's  $r$ ) for each subject group and trial (Fig. 2D). Again, envelope reconstructions were significantly larger from the musicians' data (paired  $t$ -test across trials,  $p = 1.03 \times 10^{-6}$ ) than the non-musicians.

#### 4. Discussion

Music perception engages a wide range of complex processing mechanisms of many attributes. In auditory perception at large, one indicator of these processes is the fidelity of the cortical tracking of the sound envelope, which has been shown to fluctuate substantially reflecting the cognitive engagement compared to passive listening. Recent research has highlighted this phenomena in the tracking of speech signals, especially with attentional engagement [8], prior knowledge and expectations [7], and has even shown different response patterns dependent on particular language deficits [16]. Musical signals evoke similar cortical responses that can be measured with EEG, shedding light on the processes underlying its perception. Here, we demonstrated that musical expertise increases the accuracy of cortical tracking—note that this effect was not due to differences in familiarity between groups. This is in line with previous research on event-related potentials [17,18] and cortical oscillations [19], and complements data showing that expertise modulates subcortical tracking of pitch information [20].

This finding could be the result of the positive effects of musical expertise on brain plasticity and development [21]. Other factors may also contribute such as attention or the stronger ability to identify patterns of music that comes with expertise [22]. Future studies should test how exactly musical expertise leads to an increased tracking, which could be related to changes in the precision and the strength of the responses to music and various time-scales. Additional

analysis showed that envelope tracking is not affected by repetition of a musical piece. Indeed, one issue is our focus on long musical pieces (2 minutes), thus further work is required to assess possible effects of repetition at different time-scales (e.g., [23]). Future research should also assess whether stimulus repetition affects higher-order cortical processes that are not captured by envelope tracking measures, as previously seen in the context of speech perception [7].

Finally, while this framework may allow us to address some of these questions in the future, we are fully aware that there are many enhancements that can be applied to this analysis. For example, one issue concerns the stimulus representation, where we realise that the amplitude envelope is a drastic simplification, which mainly encodes rhythm and intensity while ignoring other features that are central to music, such as pitch and timbre. Another improvement might be the use of stimuli from real musical performances, which would allow us to investigate more fully a multitude of musical performance factors whose integration produces emotions and musical enjoyment.

#### Acknowledgments

This study was supported by the EU H2020-ICT grant 644732 (COCOHA). The authors would like to thank Gaëlle Rouvier for contributing to the data collection process. Part of this work was started at the Neuromorphic Engineering Workshop 2018.

#### References

- [1] S. Koelsch and W. A. Siebel, "Towards a neural basis of music perception," *Trends Cogn. Sci.*, **9**, 578–584 (2005).
- [2] N. Ding and J. Z. Simon, "Cortical entrainment to continuous speech: Functional roles and interpretations," *Front. Hum.*

*Neurosci.*, **8** (2014).

[3] E. Ahissar, S. Nagarajan, M. Ahissar, A. Protopapas, H. Mahncke and M. M. Merzenich, "Speech comprehension is correlated with temporal response patterns recorded from auditory cortex," *Proc. Natl. Acad. Sci. USA*, **98**, 13367–13372 (2001).

[4] E. C. Lalor, A. J. Power, R. B. Reilly and J. J. Foxe, "Resolving precise temporal processing properties of the auditory system using continuous stimuli," *J. Neurophysiol.*, **102**, 349–359 (2009).

[5] N. Ding, M. Chatterjee and J. Z. Simon, "Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure," *Neuroimage*, **88**, 41–46 (2014).

[6] M. J. Crosse, G. M. Di Liberto, A. Bednar and E. C. Lalor, "The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli," *Front. Hum. Neurosci.*, **10** (2016).

[7] G. M. Di Liberto, M. J. Crosse and E. C. Lalor, "Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech," *Eneuro*, **5**, 0084-18 (2018).

[8] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma and E. C. Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cereb. Cortex*, p. bht355 (2014).

[9] D. B. Huron, *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, 2006).

[10] C. Pantev, L. E. Roberts, M. Schulz, A. Engelien and B. Ross, "Timbre-specific enhancement of auditory cortical representations in musicians," *Neuroreport*, **12**, 169–174 (2001).

[11] S. J. Luck, *An Introduction to the Event-related Potential Technique* (MIT Press, 2005).

[12] D. D. E. Wong, S. A. Fuglsang, J. Hjortkjaer, E. Ceolini, M. Slaney and A. De Cheveigne, "A comparison of regularization methods in forward and backward models for auditory attention decoding," *Front. Neurosci.*, **12**, 531 (2018).

[13] A. de Cheveigne, G. M. Di Liberto, D. Arzounian, D. D. Wong, J. Hjortkjaer, S. Fuglsang and L. C. Parra, "Multiway canonical correlation analysis of brain data," *Neuroimage*, **186**, 728–740 (2019).

[14] I. Sturm, S. Dähne, B. Blankertz and G. Curio, "Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli," *PLoS One*, **10**, e0141281 (2015).

[15] G. M. Di Liberto, J. A. O'Sullivan and E. C. Lalor, "Low-frequency cortical entrainment to speech reflects phoneme-level processing," *Curr. Biol.*, **25**, 2457–2465 (2015).

[16] G. M. Di Liberto, V. Peter, M. Kalashnikova, U. Goswami, D. Burnham and E. C. Lalor, "Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia," *Neuroimage*, **NIMG-17-29**, 70–79 (2018).

[17] H. K. Sanju and P. Kumar, "Enhanced auditory evoked potentials in musicians: A review of recent findings," *J. Otol.*, **11**, 63–72 (2016).

[18] A. Shahin, D. J. Bosnyak, L. J. Trainor and L. E. Roberts, "Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians," *J. Neurosci.*, **23**, 5545–5552 (2003).

[19] K. B. Doelling and D. Poeppel, "Cortical entrainment to music and its modulation by expertise," *Proc. Natl. Acad. Sci. USA*, **112**, E6233–E6242 (2015).

[20] P. C. M. Wong, E. Skoe, N. M. Russo, T. Dees and N. Kraus, "Musical experience shapes human brainstem encoding of linguistic pitch patterns," *Nat. Neurosci.*, **10**, 420–422 (2007).

[21] E. A. Miendlarzewska and W. J. Trost, "How musical training affects cognitive development: Rhythm, reward and other modulating variables," *Front. Neurosci.*, **7**, 279 (2014).

[22] D. Schön and C. François, "Musical expertise and statistical learning of musical and linguistic structures," *Front. Psychol.*, **2**, 167 (2011).

[23] C. Neuhaus, T. Knösche and A. Friederici, "Similarity and repetition," *Ann. N.Y. Acad. Sci.*, **1169**, 485–489 (2009).