



Brief report

The effect of familiarity on neural tracking of music stimuli is modulated by mind wandering

Joan Belo ^{1,2,*}, Maureen Clerc ¹ and Daniele Schön ^{2,3}

¹ Athena Project Team, INRIA, Université Côte d'Azur, Nice, France

² Aix Marseille University, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France

³ Institute for Language, Communication, and the Brain, Aix-en-Provence, France

* **Correspondence:** Email: joan.belo@inria.fr.

Abstract: One way to investigate the cortical tracking of continuous auditory stimuli is to use the stimulus reconstruction approach. However, the cognitive and behavioral factors impacting this cortical representation remain largely overlooked. Two possible candidates are familiarity with the stimulus and the ability to resist internal distractions. To explore the possible impacts of these two factors on the cortical representation of natural music stimuli, forty-one participants listened to monodic natural music stimuli while we recorded their neural activity. Using the stimulus reconstruction approach and linear mixed models, we found that familiarity positively impacted the reconstruction accuracy of music stimuli and that this effect of familiarity was modulated by mind wandering.

Keywords: auditory attention detection; stimulus reconstruction; EEG; familiarity; mind wandering; auditory attention

Abbreviations: AE: amplitude envelope; SR: stimulus reconstruction; RMS: root mean square; EEG: electroencephalography; ICA: independent component analysis; SBR: stepwise backward regression

1. Introduction

It is well established now that sounds are represented in the brain at multiple levels, including subcortical [1] and cortical areas [2]. Notably, it has been shown that the amplitude envelope (AE) of a dynamic acoustic stimulus, such as speech or music, is represented at a cortical level via a phase-coupling of neural rhythms onto the temporal structure of the auditory stimuli [3,4]. More precisely, this coupling takes place in the theta-delta range and also affects the amplitude of high frequency activity in the human auditory cortex [2,5–7]. One way to investigate this cortical tracking of AE of continuous auditory stimuli is to use the stimulus reconstruction approach (SR) which allows reconstructing the AE of an auditory source using neural activity recorded with magneto/electroencephalography [3]. Moreover, SR is interesting to gain insights about the factors that may influence the cortical representation of continuous auditory stimuli, such as speech or music. For instance, studies have investigated the effect of speech intelligibility on the cortical tracking of speech stimuli (e.g., [8]), the impact of musical expertise on speech and music representation (e.g., [9–11]) or even the effect of attention in more complex auditory environments on the cortical representation of attended and unattended speech streams (e.g., [12,13]). However, it is possible that other factors, cognitive as well as behavioral, may impact the cortical representation of sounds [11].

One possible candidate is the familiarity with the stimulus. This idea is supported by studies showing that familiar words [14] and voices [15] are recognized more promptly than unfamiliar ones. This is further supported by other works showing distinct processing for familiar and unfamiliar voices (e.g., [16]). Very recent studies directly investigated the effect of familiarity on the cortical representation of speech and found better cortical representation when listeners are more familiar with the stimuli [8]. While these results are rather consistent for speech stimuli, this is not the case when investigating the role of familiarity on the cortical representation of musical stimuli. Indeed, some studies show, as described with speech stimuli, an enhancement of cortical tracking for familiar compared to unfamiliar song utterances [17] and music [18,19]. However, other studies show the opposite pattern, meaning a better cortical tracking of unfamiliar music compared to familiar ones [20].

Beyond stimulus familiarity, another potential factor impacting cortical representations of auditory sources is the resistance to internal distractors. One way to address one's ability to resist internal distractors is to estimate the level of mind wandering during a task. While there is no clear definition of mind wandering, one can consider it as periods of inattention (toward the task at hand) during which self-generated thoughts arise (see [21]). Interestingly, it has been shown that when the mind wanders, attention is reduced toward external stimuli, which negatively impacts the processing of external information and the achievement of the external task [22,23]. Considering that cortical tracking is enhanced when attending to the stimuli (e.g., [12]), periods of inattention toward the auditory stimuli (i.e., mind wandering) may negatively impact its cortical representation.

In light of these considerations, the objective of this study was to investigate to what extent familiarity and mind wandering may predict the reconstruction accuracy of musical ecological stimuli.

2. Materials and methods

2.1. Participants

A total of forty-one subjects took part in the experiment (28.93 ± 10.94 years, min = 18 years, max = 54 years, 22 females). All the participants included in the analysis reported no history of hearing disorders, attention impairment or neurological disorder and were not under medication.

The experiment was approved by the Operational Committee for the Evaluation of Legal and Ethical Risks (COERLE) of INRIA and was undertaken in accordance with the Declaration of Helsinki. Each participant provided written informed consent and was financially compensated.

2.2. Main experimental task

2.2.1. Stimuli

Monodic stimuli consisted of 60-second recordings of four well-known and four less-known classical piano pieces. The well-known musical excerpts were taken from Beethoven's "Für Elise," Bach's "Prelude in C Major," Mozart's "Sonata in C Major," K545, and Beethoven's "Pathétique" Sonata. The less-known musical excerpts were taken from Beethoven's "Lustig und Traurig," Scarlatti's "Sonata in A Minor," K54, Beethoven's "Sonata in F Presto" and Bach's "Prelude No. 8." All musical excerpts were RMS normalized. Stimuli were presented via Sennheiser HD-25 supra-aural headphones at a comfortable level.

2.2.2. Procedure

Participants were instructed to concentrate their attention on the musical excerpt as if they were trying to memorize it. While the musical excerpt was played for 60 seconds, a fixation cross was displayed at the screen center. Each of the musical excerpts was presented twice for a total of sixteen 60-second trials. The pseudo-random order of presentation avoided direct repetition of a given excerpt. The whole experiment was monitored using a custom version of OpenSesame developed by Oticon Medical [24].

2.2.3. Behavioral indicators

At the end of each trial, the participants were asked several questions (see table 1).

Table 1. Questions asked to the participants after each excerpt and the corresponding behavioral indicator.

Questions	Possible answers	Behavioral indicator
<i>Do you know this music?</i>	<i>Yes / No / Not sure</i>	Familiarity
<i>How much did your mind wander?</i>	<i>0-25% / 25-50% / 50-75% / 75-100% (of the time)</i>	Mind Wandering

For *Familiarity* we computed two categories across stimuli: *Low* and *High familiarity*. The *Low familiarity* category combines the answers *No* and *Not sure* while the *High familiarity* category is composed of the answer *Yes*. Importantly, because the musical excerpts were presented twice during the experiment, we only used the answer to the first presentation of each stimulus to avoid a repetition bias.

For *Mind Wandering* we computed two categories across participants: *Low* and *High Mind Wandering*, corresponding to participants for which the average mind wandering was lower or higher than the grand mean, respectively.

In addition to these behavioral indicators created from the questions, because each musical excerpt was presented twice, we also extracted a *Repetition* indicator, composed of two categories (i.e., *First Presentation* and *Second Presentation* of the stimulus). Importantly, this indicator allowed us to explore the effect of the acquired (short-term) knowledge on neural tracking.

2.2.4. EEG data

Electroencephalography data were recorded using an ANT Refa8 amplifier and a 21-electrode cap (arranged in accordance with the International 10–20 system) with a sampling rate of 256 Hz and an average reference. They were preprocessed using the *MNE-python* package [25] according to the procedure described in [26]. First, raw EEG data were cut into 60 s epochs, each epoch representing a trial, and the first 500 ms of each epoch was discarded to avoid modeling the response to the stimulus onset. Next, EEG data were digitally filtered between 1–40 Hz, and bad channels were interpolated via spherical spline interpolation when necessary. Then, ICA was performed to remove eye-blinks and saccade artifacts. Afterward, EEG data were digitally filtered between 1–9 Hz [27,28] using a 4th order Butterworth zero-phase-shift filter and downsampled to 64 Hz. Finally, 60 s epochs were cut into 30 s epochs, resulting in 32 epochs with the aim of having enough data to train and test the stimulus-reconstruction model.

2.2.5. Audio feature extraction

Amplitude envelopes of the audio monodic stimuli were obtained using the function *human_cochleagram* from the Python *Pycochleogram* package [29]. This function allows 1) computing an equivalent rectangular bandwidth (ERB) filter bank and 2) using this filter bank to decompose the signal into subband envelopes. Afterwards, subband envelopes were averaged to obtain a unique envelope. Each envelope was then digitally filtered between 1–9 Hz with a 4th order Butterworth zero-phase-shift filter, downsampled to 64 Hz and cut into 30 s long epochs in order to match EEG preprocessing.

2.2.6. Stimulus reconstruction

In this study, we used the same stimulus reconstruction approach as [30] (this approach is also described in several other articles; see [26] and [31] for a comprehensive description and see Figure 1 for an illustration)

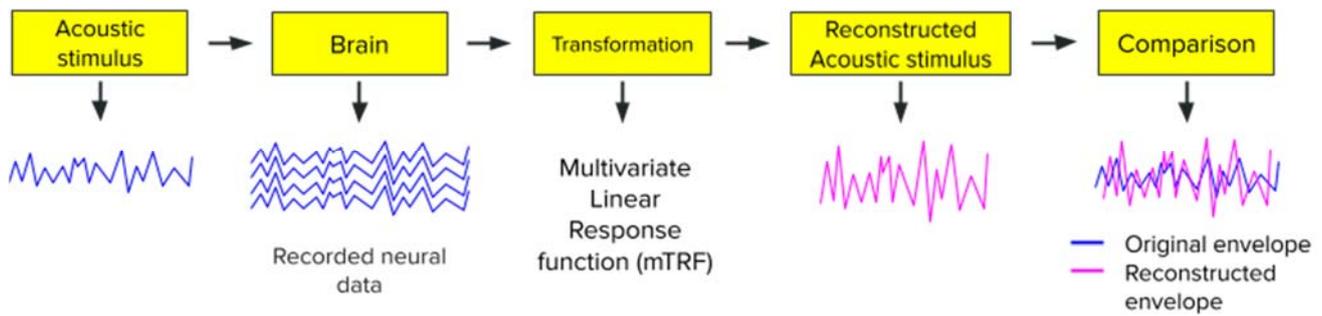


Figure 1. Illustration of the stimulus reconstruction approach. First, participants listen to an acoustic stimulus (i.e., a musical excerpt) while their neural activity is recorded using an EEG. Then, a transformation (i.e., a multivariate temporal response function) is applied to the recorded neural activity in order to reconstruct the original acoustic stimulus. Afterward, a comparison (i.e., a correlation) is computed between the original acoustic envelope and the reconstructed one to estimate the reconstruction quality (i.e., the reconstruction accuracy).

2.3. General procedure

The stimulus reconstruction approach allows one to reconstruct an estimate of the envelope of the auditory stimulus s using electrophysiological neural data d via a linear reconstruction model g [30]. The reconstruction model $g(\tau, n)$ is a temporal response function (see [32] for details) that maps neural data $d(t, n)$ to stimulus $s(t)$ as follows:

$$\hat{s}(t) = \sum_n \sum_{\tau} d(t + \tau, n)g(\tau, n)$$

where \hat{s} denotes the reconstructed stimulus, $d(t, n)$ represents the response of electrode n at time $t=1 \dots T$, and τ are some time lags that represent a window in which the brain's response to the stimulus is supposedly optimal. We defined τ to go from 0.200 ms pre-stimulus to 0.350 ms post-stimulus based on previous works [30,38]. At a sampling rate of 64 Hz, this corresponds to 36 sample shifts (incl. the zero shift).

The model g is estimated by minimizing the mean-squared error between the original stimulus $s(t)$ and the reconstructed one $\hat{s}(t)$:

$$\min e = \sum_t [s(t) - \hat{s}(t)]^2$$

A robust minimizer of the mean-squared error is obtained using the following matrix operations:

$$g = (D^T D + \lambda I)^{-1} D^T s$$

where D is the lagged time series of the response matrix d , and λ is a ridge parameter term introduced to avoid overfitting. Importantly we used a multivariate approach, fully exploiting the 21 EEG channels to reconstruct the stimulus. The model parameters as well as the ridge parameter are generally estimated using the leave-one-out cross validation (LOO) procedure. Once the model parameters have

been tuned, it can be tested on new data (generally, the left out fold), and the end of this process is the reconstructed stimulus \hat{s} . Finally, a reconstruction score is computed, to estimate the reconstruction accuracy of the model, as the Pearson correlation between the reconstructed stimulus \hat{s} and the original stimulus s . The higher the reconstruction score is, the better the reconstruction accuracy.

2.4. Training and testing the model

In this study, we wanted to see if we can reconstruct a trial with a model trained on every other monodic trial. To do so, we used a custom Python script where an initial LOO was performed to separate training and test phase; and because we introduced a ridge parameter (λ), we used a second LOO nested in the training phase to estimate the optimal value of the ridge parameter.

At the end of the test phase, a reconstruction \hat{s}_i is obtained for each musical excerpt s_i , based on a model trained on all the other musical excerpts. This procedure allowed us to make sure that the musical excerpt we wanted to reconstruct was only used to validate the model.

Because we had 32 monodic trials, we trained separate models for each trial type. Therefore, this resulted in 32 models for monodic trials for each participant.

2.5. Reconstruction accuracy

The reconstruction accuracy corresponds to the Pearson correlation between the reconstructed envelopes and the original musical envelopes. This led to 32 r values per participant.

2.6. Statistical analysis

In order to see whether the behavioral indicators were good predictors of the reconstruction accuracy of monodic stimuli, we fitted linear mixed models (LMMs) on single trial data using R statistical software and the lme4 package [33–36]. To take into account the measures dependency induced by our experimental design, LMMs included Subject and Stimulus as random effects. The model assumptions of non-collinearity of the predictors as well as the linearity, homoscedasticity and normality of the residuals were systematically verified using visual inspections of the residuals QQ-plot and histograms. In addition, we reported 95% confidence intervals (CIs), and p -values were computed using a Wald t -distribution approximation.

3. Results

Participant responses concerning the familiarity of the musical excerpts were rather consistent with our initial stimulus categorizations of familiar versus unfamiliar music. Participants answered unfamiliar at 67% when the stimuli were unfamiliar, and they answered familiar at 80% when the stimuli were familiar. Because there was no systematic categorization of musical excerpts across participants, we used individual participants' responses to categorize musical excerpts as familiar or not.

The overall reconstruction accuracy, namely, the correlation between the EEG-reconstructed envelopes and the original musical envelopes, was 0.063 (± 0.08). This estimate was significantly different from the null distribution ($t(40) = 14.32$, $p < 0.001$; Cohen's $d = 2.24$).

In order to explore the predictive power of the behavioral indicators (familiarity, mind wandering and repetition) on the reconstruction accuracy of monodic trials, we started by computing a stepwise backward regression (SBR). This model selection procedure involves starting with a model containing all the possible predictors (the “full model”) and then iteratively removing the least useful predictor one-at-a-time until no further predictor can be deleted without a statistically significant loss of fit. Therefore, our full model was a model with Stimulus and Subject as random effects and Familiarity, Mind wandering and Repetition, as well as all the possible interactions between them, as fixed effects (Formula: Reconstruction ~ Familiarity * Mind wandering * Repetition * (1|Subject) + (1|Stimulus)).

The SBR indicated that the best model to predict the reconstruction accuracy of monodic trials was a model with Stimulus and Subject as random effects and Familiarity as fixed effect (Formula: Reconstruction ~ Familiarity + (1|Subject) + (1|Stimulus); see Table 2 for full details).

Table 2. Summary table for the stepwise backward regression model.

Backward reduced random-effect table							
	Eliminated	npar	logLik	AIC	LRT	Df	Pr (>Chisq)
<none>		7	1462.6	-2911.2			
Subject	0	6	1428.6	-2845.3	67.946	1	< 2.2e-16 ***
Stimulus	0	6	1457.7	-2902.1	11.098	1	0.0008642 ***
Backward reduced fixed-effect table							
<i>degree of freedom method: Satterthwaite</i>							
	Eliminated	Sum Sq	Mean Sq	NumDF	DenDF	F-value	Pr (>F)
Familiarity*Mind wandering*Repetition	1	0.0037	0.0037	1	1294.83	0.634	0.425
Mind wandering*Repetition	2	0.00062	0.00062	1	1283.22	0.106	0.744
Familiarity*Repetition	3	0.0021	0.0021	1	1297.88	0.36	0.548
Repetition	4	0.00683	0.00683	1	1228.62	1.173	0.278
Familiarity*Mind wandering	5	0.01955	0.01955	1	1280.28	3.357	0.067 .
Mind wandering	6	0.00061	0.00061	1	39	0.105	0.747
Familiarity	0	0.03773	0.03773	1	454.34	6.4675	0.011 *

The best model clearly differed from a “null model” containing only the random effects (Likelihood Ratio Test: Chisq = 6.5, $p = 0.01$). Importantly, the best model indicated that familiarity affected the reconstruction accuracy of monodic stimuli ($\beta = -0.01$, 95% CI [$3.02e-03$, 0.02], $t(1307) = 2.54$, $p = 0.011$; *Std. beta* = 0.16, 95% CI [0.04 , 0.29]). More precisely, reconstruction of familiar stimuli was higher than reconstruction of less familiar ones (see Figure 2A). This model significantly predicted around 11% of the variance of the reconstruction score (*conditional* $R^2 = 0.11$). However, the part of explained variance related to the fixed effect (Familiarity) alone was around 0.7% (*marginal* $R^2 = 6.52^{-3}$) which means that the vast majority of the variance explained by this model is due to the random effects alone, namely, the Stimulus and the Subject.

Moreover, the SBR indicated a trend towards an interaction between the Familiarity and the Mind wandering (see Figure 2B). Therefore, we computed a LMM for each level of the Mind wandering Predictor to better understand the behavior of the effect of stimulus familiarity.

We found a significant effect of familiarity on the reconstruction of monodic stimuli for the High Mind wandering level ($\beta = -0.02$, 95% CI $[-0.04, -4.95e-03]$, $t(475) = -2.54$, $p = 0.011$; Std. beta = -0.26 , 95% CI $[-0.47, -0.06]$) but not for the Low Mind wandering level ($\beta = -0.01$, 95% CI $[-0.02, 2.57e-03]$, $t(827) = -1.57$, $p = 0.118$; Std. beta = -0.13 , 95% CI $[-0.29, 0.03]$). Indeed, the reconstruction accuracy of monodic stimuli for the participants who experienced more mind wandering was higher for familiar stimuli than for less familiar ones (see Figure 2B), while this was not the case for the participants who experienced less mind wandering (see Figure 2B)

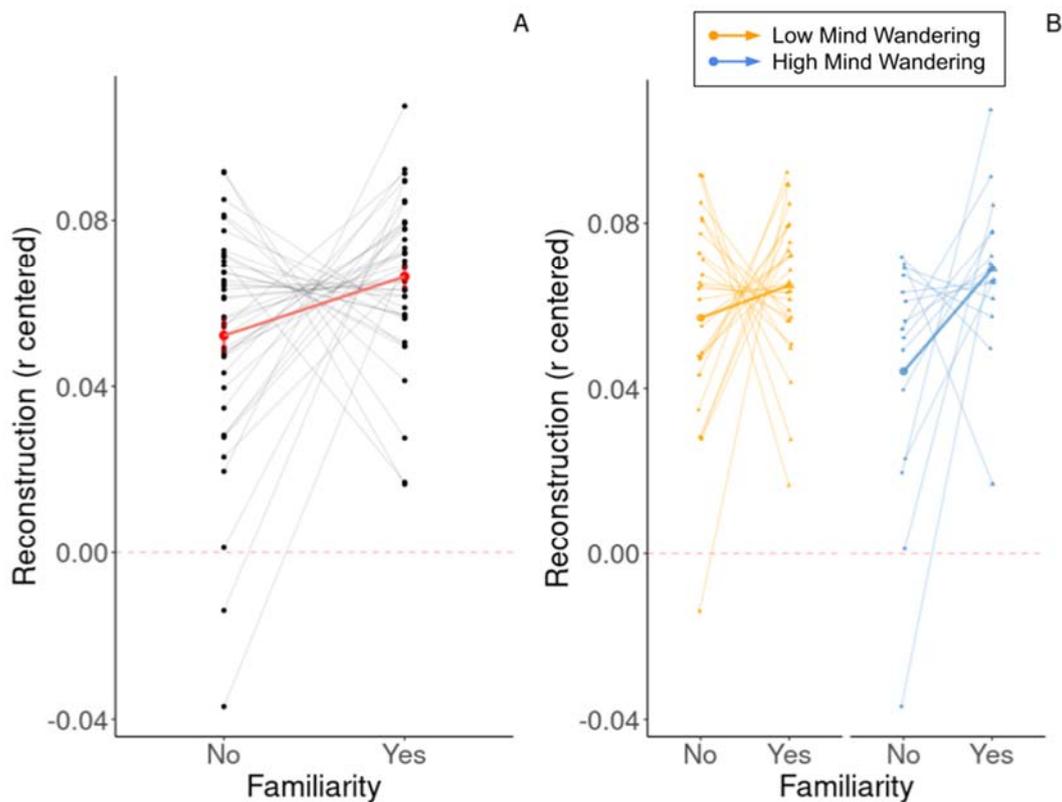


Figure 2. Illustration of the effect of familiarity on reconstruction accuracy and the effects of Familiarity on reconstruction accuracy within the Mind Wandering indicators' levels. A) Average reconstruction accuracy when excerpts were familiar and unfamiliar. Average reconstruction was higher for familiar than unfamiliar excerpts ($\beta = 0.01$, 95% CI $[3.02e-03, 0.02]$, $p = 0.011$). B) Average reconstruction when excerpts were familiar and unfamiliar and when mind wandering was low or high. Average reconstruction was higher for familiar excerpts for participants with high mind wandering ($\beta = 0.02$, 95% CI $[0.04, 8.64e-03]$, $p = 0.003$), but it was similar for familiar and unfamiliar excerpts for participants with low mind wandering ($\beta = -6.96e-03$, 95% CI $[-0.02, 5.19e-03]$, $p = 0.262$). Error bars indicate plus or minus one standard error of the mean. Large points indicate the group averages, and small points indicate individual data.

4. Discussion and conclusion

In this work, we hypothesized that reconstruction accuracy would be higher for familiar excerpts compared to less familiar ones. Our analysis confirmed this hypothesis and revealed a positive main effect of familiarity on reconstruction accuracy (Figure 1A). This result corroborates the findings of previous works showing an enhancement of cortical tracking for familiar song utterances [17] and familiar natural music [18,19]. One possible reason why the cortical tracking of familiar stimuli is enhanced compared to unfamiliar ones is because they are more predictable, as they elicit more precise expectations due to the presence of preexisting memory traces [24,37]. Some recent results suggest that expectations may directly impact cortical tracking of music and that, in particular, long-term experience better captures the neural correlates of melody compared to short-term statistical regularities [38,39]. It has also been shown that stimulus predictability could have an influence on auditory scene analysis (e.g., [40]), although most of the studies investigating the influence of predictability on auditory scene analysis used paradigms in which concurrent sounds were present or where sound sequences contained deviant sounds.

Interestingly, one may argue that unfamiliar stimuli could have elicited greater neural responses, leading to better cortical tracking, because they are more prone to induced expectations violation than familiar ones, which could be in line with the results of Kumagai and colleagues [20]. However, extending the proposition of Weineck and colleagues [19], it is likely that our unfamiliar stimuli did not elicit expectation violation because they were classical Western music excerpts that were composed following the common rules of the genre. Thus, even if participants were not familiar with the particular excerpt, they were probably familiar with Western music in some way. Consequently, it is probable that familiar music excerpts, similarly to more intelligible or familiar speech (e.g., [41,42]), induced better cortical tracking because memory priors were higher compared to unfamiliar music and because unfamiliar music excerpts did not induce violation of these priors.

Besides an explanation in terms of better predictions for familiar stimuli, it is also possible that familiarity modulated cortical tracking because familiar music excerpts elicited greater engagement/attention of the participants.

That being said, while the effect of familiarity was significant, its predictive power was rather weak (0.7% of the reconstruction accuracy), suggesting that, in our experiment, the familiarity with the musical excerpt probably does not massively impact its cortical tracking. However, because most previous papers in the field do not report the predictive power, it remains difficult to assess the global impact of familiarity on cortical tracking.

Here, we also hypothesized that the mind wandering level will negatively affect the reconstruction accuracy of monodic excerpts. Inconsistent with our hypothesis, we found that reconstruction accuracy was similar for all the participants, no matter if they experienced a low or a high level of mind wandering during the experiment. Although this result may sound surprising because of the detrimental effect of mind wandering on task performance (e.g., [43]), we used a retrospective measurement of mind wandering requiring the participants to estimate how frequently their mind wandered during each trial. This approach may suffer from forgetting and mental-aggregation biases [44].

The original result of this work is that the effect of familiarity was modulated by the global level of mind wandering. Indeed, participants that experienced a high level of mind wandering during the experiment yielded higher reconstruction accuracy for familiar excerpts than for unfamiliar ones. By contrast, participants that experienced a low level of mind wandering did not show such a familiarity

effect. This could indicate that when attention is mostly dedicated to the task at hand (i.e., listening to the musical excerpt), familiarity may not enhance the cortical tracking, possibly because a large amount of the available attentional resources can be devoted to actively listening to the excerpts, independently of whether it is familiar or unfamiliar. However, when attention is directed to unrelated thoughts rather than to the task at hand, leading the target musical excerpt to become “unattended,” the familiarity may compensate for this reduction of attentional resources devoted to the treatment of the musical excerpt. In such a context, familiar excerpts seem to somehow better “capture” attentional resources, leading to a more robust cortical representation. This idea could find support in some results suggesting that 1) although the underlying mechanism is still unclear, a cortical representation of the unattended sound can occur without direct attention (see, for example, [12] and [45]), and 2) familiar excerpts, because they are more predictable, are less computationally demanding [37], which suggests that they may be processed with less attentional resources.

Interestingly, we found that the reconstruction accuracy was similar for the first and the second presentation of the musical excerpts, indicating a lack of short-term knowledge effect on cortical representation of monodic musical stimuli. This result is in line with the work of Di Liberto and colleagues [9], who showed no effect of repetition on cortical tracking of monodic synthetic music, but not with the work of Madsen and collaborators [18], who found a decrease in cortical tracking for familiar music across repetitions.

Overall, our results indicate that stimulus familiarity favors neural tracking. Moreover, this familiarity effect is present with long-term knowledge but not with short-term knowledge of the stimuli. Importantly, this long-term familiarity effect on neural stimuli representations is modulated by attention and is maximal when participants are not well focused on the task at hand. Further work is required to better understand how this effect of familiarity on cortical representation is driven by long-term knowledge about the musical excerpts.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

J.B. is funded by Oticon Medical (research agreement n°14294-CRI04) and Region Provence-Alpes Côte-d’Azur. D.S. was supported by APA foundation (RD-2016-9), ANR-16-CE28-0012-01 (RALP), ANR-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

References

1. Chandrasekaran B, Kraus N (2010) The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology* 47(2): 236–246. <https://doi.org/10.1111/j.1469-8986.2009.00928.x>
2. Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat Neurosci* 15(4): 511–517. <https://doi.org/10.1038/nn.3063>

3. Mesgarani N, David SV, Fritz JB, et al. (2009) Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J Neurophysiol* 102(6): 3329–3339. <https://doi.org/10.1152/jn.91128.2008>
4. Nourski KV, Reale RA, Oya H, et al. (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. *J Neurosci* 29(49): 15564–15574. <https://doi.org/10.1523/JNEUROSCI.3065-09.2009>
5. Pasley BN, David SV, Mesgarani N, et al. (2012) Reconstructing speech from human auditory cortex. *PLoS Biol* 10(1). <https://doi.org/10.1371/journal.pbio.1001251>
6. Kubanek J, Brunner P, Gunduz A, et al. (2013) The Tracking of Speech Envelope in the Human Cortex. *PLoS ONE* 8(1). <https://doi.org/10.1371/journal.pone.0053398>
7. Doelling KB, Poeppel D (2015) Cortical entrainment to music and its modulation by expertise. *P Natl Acad Sci USA* 112(45): E6233–E6242. <https://doi.org/10.1073/pnas.1508431112>
8. Baltzell LS, Srinivasan R, Richards VM (2017) The effect of prior knowledge and intelligibility on the cortical entrainment response to speech. *J Neurophysiol* 118(6): 3144–3151. <https://doi.org/10.1152/jn.00023.2017>
9. Di Liberto GM, Pelofi C, Shamma S, et al. (2020) Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening. *Acoust Sci Technol* 41(1): 361–364. <https://doi.org/10.1250/ast.41.361>
10. Harding EE, Sammler D, Henry MJ, et al. (2019) Cortical tracking of rhythm in music and speech. *NeuroImage* 185: 96–101. <https://doi.org/10.1016/j.neuroimage.2018.10.037>
11. Belo J, Clerc M, Schön D (2022) Attentional inhibition ability predicts neural representation during challenging auditory streaming. *BioRxiv Preprint*.
12. O’Sullivan JA, Power AJ, Mesgarani N, et al. (2015) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb Cortex* 25(7): 1697–1706. <https://doi.org/10.1093/cercor/bht355>
13. Zion Golumbic EM, Ding N, Bickel S, et al. (2013) Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron* 77(5): 980–991. <https://doi.org/10.1016/j.neuron.2012.12.037>
14. Connine CM, Mullennix J, Shernoff E, et al. (1990) Word Familiarity and Frequency in Visual and Auditory Word Recognition. *J Exp Psychol Learn* 16(6): 1084–1096. <https://doi.org/10.1037/0278-7393.16.6.1084>
15. Skuk VG, Schweinberger SR (2013) Gender differences in familiar voice identification. *Hearing Res* 296: 131–140. <https://doi.org/10.1016/j.heares.2012.11.004>
16. Blank H, Wieland N, von Kriegstein K (2014) Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neurosci Biobehav Rev* 47: 717–734. <https://doi.org/10.1016/j.neubiorev.2014.10.022>
17. Vanden Bosch der Nederlanden CM, Joanisse MF, Grahn JA, et al. (2022) Familiarity modulates neural tracking of sung and spoken utterances. *NeuroImage* 252: 119049. <https://doi.org/10.1016/j.neuroimage.2022.119049>
18. Madsen J, Margulis EH, Simchy-Gross R, et al. (2019) Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity and training. *Sci Rep* 9(1): 1–8. <https://doi.org/10.1038/s41598-019-40254-w>

19. Weineck K, Wen OX, Henry MJ (2022) Neural entrainment is strongest to the spectral flux of slow music and depends on familiarity and beat salience. *Elife* 11: e75515. <https://doi.org/10.7554/eLife.75515>
20. Kumagai Y, Matsui R, Tanaka T (2018) Music familiarity affects EEG entrainment when little attention is paid. *Front Hum Neurosci* 12: 1–11. <https://doi.org/10.3389/fnhum.2018.00444>
21. Seli P, Kane MJ, Smallwood J, et al. (2018) Mind-Wandering as a Natural Kind: A Family-Resemblances View. *Trends Cogn Sci* 22(6): 479–490. <https://doi.org/10.1016/j.tics.2018.03.010>
22. Barron E, Riby LM, Greer J, et al. (2011) Absorbed in thought: The effect of mind wandering on the processing of relevant and irrelevant events. *Psychol Sci* 22(5): 596–601. <https://doi.org/10.1177/0956797611404083>
23. Kam JWY, Xu J, Handy TC (2014) I don't feel your pain (as much): The desensitizing effect of mind wandering on the perception of others' discomfort. *Cogn Affect Behav Ne* 14(1): 286–296. <https://doi.org/10.3758/s13415-013-0197-z>
24. Patou F, Py H, Backus B, et al. (2019) Oticon Medical's ecosystem of lab- and field investigational platforms for behavioral and cognitive research in CI Peripheral neural interfacing Behavior.
25. Gramfort A, Luessi M, Larson E, et al. (2013) MEG and EEG data analysis with MNE-Python. *Front Neurosci* 7(7 DEC): 1–13. <https://doi.org/10.3389/fnins.2013.00267>
26. Crosse MJ, Zuk NJ, Di Liberto GM, et al. (2021) Linear Modeling of Neurophysiological Responses to Naturalistic Stimuli : Methodological Considerations for Applied Research. PsyArXiv. <https://doi.org/10.31234/osf.io/jbz2w>
27. Greenlaw KM, Puschmann S, Coffey EBJ (2020) Decoding of Envelope vs. Fundamental Frequency During Complex Auditory Stream Segregation. *Neurobiol Language* 1(3): 268–287. https://doi.org/10.1162/nol_a_00013
28. Biesmans W, Vanthornhout J, Wouters J, et al. (2015) Comparison of speech envelope extraction methods for EEG-based auditory attention detection in a cocktail party scenario. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015-Novem, 5155–5158. <https://doi.org/10.1109/EMBC.2015.7319552>
29. Available from: <https://github.com/mcdermottLab/pycochleagram>
30. O'Sullivan JA, Power AJ, Mesgarani N, et al. (2015) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb Cortex* 25(7): 1697–1706. <https://doi.org/10.1093/cercor/bht355>
31. Crosse MJ, Di Liberto GM, Bednar A, et al. (2016) The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Front Hum Neurosci* 10: 1–14. <https://doi.org/10.3389/fnhum.2016.00604>
32. Lalor EC, Power AJ, Reilly RB, et al. (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J Neurophysiol* 102(1): 349–359. <https://doi.org/10.1152/jn.90896.2008>
33. R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: <https://www.R-project.org/>.
34. Bates D, Mächler M, Bolker B, et al. (2015) Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 67(1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
35. Kuznetsova A, Brockhoff PB, Christensen RHB (2014) Package 'lmerTest'. Test for random and fixed effects for linear mixed effect models (lmer objects oflme4 package). R package Version 2. Available from: <https://github.com/runchaubo/lmerTestR>.

36. Millard SP, EnvStats-An R (2013) An R package for environmental statistics. New York: Springer-Verlag.
37. Huron D (2006) Sweet anticipation: Music and the psychology of expectation (pp. xii, 462). The MIT Press.
38. Di Liberto GM, Pelofi C, Bianco R, et al. (2020). Cortical encoding of melodic expectations in human temporal cortex. *ELife* 9: 1–26. <https://doi.org/10.7554/eLife.51784>
39. Kern P, Heilbron M, De Lange FP, et al. (2022) Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. *BioRxiv Preprint*.
40. Bendixen A (2014) Predictability effects in auditory scene analysis: A review. *Front Neurosci* 8: 1–16. <https://doi.org/10.3389/fnins.2014.00060>
41. Verschueren E, Vanthornhout J, Francart T (2020) The effect of stimulus choice on an EEG-Based objective measure of speech intelligibility. *Ear Hearing* 41(6): 1586–1597. <https://doi.org/10.1097/AUD.0000000000000875>
42. Soni S, Tata MS (2021) Brain electrical dynamics in speech segmentation depends upon prior experience with the language. *Brain Lang* 219: 104967. <https://doi.org/10.1016/j.bandl.2021.104967>
43. Smallwood J (2013) Distinguishing how from why the mind wanders: A process-occurrence framework for self-generated mental activity. *Psychol Bull* 139(3): 519–535. <https://doi.org/10.1037/a0030010>
44. Kane MJ, Smeekens BA, Meier ME, et al. (2021) Testing the construct validity of competing measurement approaches to probed mind-wandering reports. *Behav Res Methods* 53: 2743. <https://doi.org/10.3758/s13428-021-01685-4>
45. Hausfeld L, Riecke L, Valente G, et al. (2018) Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage* 181: 617–626. <https://doi.org/10.1016/j.neuroimage.2018.07.052>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).