

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Kurkela, Jari; Hämäläinen, Jarmo; Leppänen, Paavo H.T.; Shu, Hua; Astikainen, Piia

Title: Passive exposure to speech sounds modifies change detection brain responses in adults

Year: 2019

Version: Accepted version (Final draft)

Copyright: © 2018 Elsevier Inc.

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

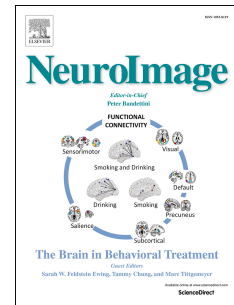
Please cite the original version:

Kurkela, J., Hämäläinen, J., Leppänen, P. H., Shu, H., & Astikainen, P. (2019). Passive exposure to speech sounds modifies change detection brain responses in adults. *NeuroImage*, 188, 208-216. <https://doi.org/10.1016/j.neuroimage.2018.12.010>

Accepted Manuscript

Passive exposure to speech sounds modifies change detection brain responses in adults

L.O. Kurkela Jari, A. Hämäläinen Jarmo, H.T. Leppänen Paavo, Shu Hua, Astikainen Piia



PII: S1053-8119(18)32154-2

DOI: <https://doi.org/10.1016/j.neuroimage.2018.12.010>

Reference: YNIMG 15478

To appear in: *NeuroImage*

Received Date: 27 June 2018

Revised Date: 3 December 2018

Accepted Date: 5 December 2018

Please cite this article as: Kurkela Jari, L.O., Hämäläinen Jarmo, A., Leppänen Paavo, H.T., Hua, S., Piia, A., Passive exposure to speech sounds modifies change detection brain responses in adults, *NeuroImage* (2019), doi: <https://doi.org/10.1016/j.neuroimage.2018.12.010>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 **Passive exposure to speech sounds modifies change detection brain**
2 **responses in adults**

3

4 Kurkela Jari L.O.^{1*}, Hämäläinen Jarmo A.¹, Leppänen Paavo H.T.¹, Shu Hua², Astikainen
5 Piia¹

6

7 ¹ Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

8 ² National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal
9 University, Beijing, China

10

11 * Corresponding author: Department of Psychology, Mattilanniemi 6-8, 40014 University
12 of Jyväskylä, Jyväskylä, Finland; phone: + 358 408054538; e-mail: jari.kurkela@jyu.fi

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33 **ABSTRACT**

34 In early life auditory discrimination ability can be enhanced by passive sound exposure. In
35 contrast, in adulthood passive exposure seems to be insufficient to promote discrimination
36 ability, but this has been tested only with a single short exposure session in humans. We
37 tested whether passive exposure to unfamiliar auditory stimuli can result in enhanced
38 cortical discrimination ability and change detection in adult humans, and whether the
39 possible learning effect generalizes to different stimuli. To address these issues, we
40 exposed adult Finnish participants to Chinese lexical tones passively for 2 h per day on 4
41 consecutive days. Behavioral responses and the brain's event-related potentials (ERPs)
42 were measured before and after the exposure for the same stimuli applied in the exposure
43 phase and to sinusoidal sounds roughly mimicking the frequency contour in speech
44 sounds. Passive exposure modulated the ERPs to speech sound changes in both ignore
45 (mismatch negativity latency, P3a amplitude and P3a latency) and attend (P3b amplitude)
46 test conditions, but not the behavioral responses. Furthermore, effect of passive exposure
47 transferred to the processing of the sinusoidal sounds as indexed by the latency of the
48 mismatch negativity. No corresponding effects in the ERPs were found in a control group
49 that participated to the test measurements, but received no exposure to the sounds. The
50 results show that passive exposure to foreign speech sounds in adulthood can enhance
51 cortical discrimination ability and attention orientation toward changes in speech sounds
52 and that the learning effect can transfer to non-speech sounds.

53

54

55 *Keywords:* perceptual learning, speech sounds, passive exposure, event-related potentials

56

57

58

59

60

61

62

63

64

65

66 1. Introduction

67

68

69 In early infancy, cortical discrimination ability is enhanced even by passive sound
70 exposure alone (e.g., Cheour et al., 1998; Cheour et al., 2002; Kuhl, 2004; Trainor, Lee, &
71 Bosnyak, 2011). In contrast, in adulthood passive sound exposure in absence of training
72 seems to be insufficient to affect the neural-level discrimination ability (Näätänen et al.,
73 1993; Sheehan et al., 2005; Elmer et al., 2017) or behavioral discrimination performance
74 (Wright et al., 2010; 2015). Instead, effects of active discrimination training have been
75 shown in several studies by measuring the mismatch negativity (MMN) (Kraus et al.,
76 1995; Tremblay et al., 1997; Tremblay et al., 1998; Tamminen et al., 2015), P3a (Atienza
77 et al., 2004; Uther et al., 2006; Seppänen et al., 2012) and P2 (Atienza et al., 2002; Reinke
78 et al., 2003; Sheehan et al., 2005) components of event-related potentials (ERPs). These
79 components reflect pre-attentive change detection (MMN) and subsequent attention
80 shifting (P3a) based on a memory trace formed by the learned sound feature (Näätänen et
81 al., 2005; Polich, 2007) and sound feature encoding and stimulus classification (P2) (for a
82 review see Crowley & Colrain, 2004).

83 Even though effects of passive exposure have been studied on brain responses
84 related to pre-attentive change detection, possible effects of passive exposure on attentive
85 change detection of sounds have not been investigated, i.e. effects on N2b and P3b
86 components. 1-hour attentive identification training with speech sounds, however, showed
87 learning-related changes in N2b and P3b (Alain et al., 2010). In another study,
88 identification training resulted in only enhanced P3b responses and no changes in N2b
89 (Ben-David et al., 2011). Similarly, attentive discrimination training with speech sounds
90 resulted in enhanced P3b-like but not N2b-like microstates in electroencephalography
91 (Giroud et al., 2017).

92 Even if previous studies have failed to demonstrate effect of passive exposure on
93 auditory change detection in adults (Näätänen et al., 1993; Sheehan et al., 2005; Elmer et
94 al., 2017; Wright et al., 2010; 2015), passive exposure to sounds seems not to be entirely
95 ineffective either. Perceptual learning on an auditory discrimination task (Wright et al.,
96 2010) or on an identification task (Wright et al., 2015) that is combined with sessions of
97 passive exposure is more efficient than the active training alone as indexed by behavioral
98 responses (Wright et al., 2010; 2015). Furthermore, passive exposure to sounds increases

99 amplitude of the P2 component (Sheehan et al., 2005; Tremblay et al., 2007; Tremblay et
100 al., 2010; Ross et al., 2013). Thus, passive exposure seems to have at least facilitating
101 effect on auditory perceptual learning in adulthood.

102 One possible reason for the failure of the previous studies in demonstrating the effect
103 of passive exposure on discrimination ability can be the short, 1 - 2 hour, exposure time
104 that has been used in previous studies (Näätänen et al., 1993; Sheehan et al., 2005; Elmer
105 et al., 2017). Active training studies have provided training over several days, and this has
106 led to better discrimination ability as indexed by the enhancement of the MMN, P3a and
107 P3b responses (Kraus et al., 1995; Tremblay et al., 1997; Giroud et al., 2017).
108 Furthermore, it has been shown that sleep deprivation hinders the learning-related increase
109 in the MMN amplitude and prevents the appearance of the P3a component (Atienza et al.,
110 2004). Thus, the learning-related changes in cortical responses seem to be sleep-
111 dependent, probably requiring memory consolidation during nocturnal sleep (Alain et al.
112 2015). Based on this assumption, it could be possible that the effects of mere passive
113 exposure emerge if the exposure is expanded on several days, allowing memory
114 consolidation. This has not yet been tested explicitly, however.

115 The evidence on generalization of the auditory learning to stimulus features not
116 encountered during training is scarce. There are some studies showing that frequency or
117 syllable discrimination training generalizes to closely similar untrained stimuli (for a
118 review see Wright & Zhang, 2009). One study applied MMN to study the generalization,
119 and showed that categorization training of labial stop consonant generalizes also to
120 alveolar stop consonant as indicated by the shortened latency and increased amplitude of
121 the MMN to non-trained stimuli (Tremblay et al., 1997).

122 In the present study, we tested two highly novel aspects of auditory perceptual
123 learning: i) Effect of passive speech sound exposure on change detection and attention
124 orienting in ignore and attend test conditions, and ii) if the effect of passive exposure is
125 observed, whether it generalizes to ignored non-speech stimuli. Adult native Finnish
126 participants were exposed to speech sounds (changes in Chinese lexical tones) for a total
127 of 8 hours over 4 days. ERPs were recorded before and after the exposure to the same
128 speech sounds and also to sinusoidal sounds roughly mimicking the pitch contours of the
129 speech sounds. A control group received no exposure but participated only in the ERP
130 recordings at the same time intervals as the experimental group.

131 We expect that the passive exposure would result in modulations in the ERPs,
132 reflecting changes in both pre-attentive and attentive change detection and attention

133 orienting toward changes (MMN, P3a, N2b, and P3b), as the exposure time is longer than
134 in the previous studies (Näätänen et al., 1993; Sheehan et al., 2005; Elmer et al., 2017) and
135 allows memory trace consolidation during the nights between the exposure periods
136 (Stickgold, 2005; Alain et al., 2015). Changes in these ERP components are assumed to
137 occur due to the formation of long-term memory representations of the sounds, making
138 change detection and attention orienting to them more efficient (as in Näätänen et al.,
139 1997; Winkler et al., 1999). We also hypothesized, based on the findings on sound
140 frequency training (Wright & Zhang, 2009), that the effect of passive exposure transfers to
141 the non-speech sounds.

142

143

144 **2. Material and methods**

145

146

147 **2.1 Participants**

148

149 A total of 39 monolingual Finnish-speaking participants (mean age = 23.0 years,
150 standard deviation [SD] = 3.3 years; 32 females and 7 males) volunteered for the study.
151 They were recruited with announcements in the notice boards and e-mail lists of the
152 University of Jyväskylä. The inclusion criteria for the study were an age of 18–30 years,
153 right-handedness, normal hearing measured using audiometry, and self-reported normal
154 vision (or corrected to normal vision). The exclusion criteria for the study were
155 neurological or psychiatric disorders, including sleep problems, and exposure to or training
156 in tonal languages. However, previous exposure during trips to countries where tonal
157 languages are spoken (maximum of 2 weeks) was accepted. Written informed consent was
158 obtained from each participant before inclusion in the study. The experiment was
159 undertaken in accordance with the Declaration of Helsinki, and the ethical committee of
160 the University of Jyväskylä approved the research protocol.

161

162 The participants were divided to two groups, one group of participants were
163 passively exposed to speech sounds ($n = 18$, mean age = 21.7 years, $SD \pm 1.7$) and the
164 other served as a control group ($n = 21$, mean age = 24.1 years, $SD \pm 3.6$). Data was
165 collected in ignore and attend test conditions (described below). From both ignore and
attend conditions data of 3 participants were omitted from statistical analysis due to

166 extensive artifacts in the EEG. After the omission, data of 18 and 21 control group
167 participants and 18 and 15 passive exposure group participants remained for the ignore and
168 attend test conditions, respectively. In the ignore test condition data, 66.6% of the
169 participants in the exposure group and the same portion of the participants in the control
170 group had some musical training or had played an instrument or sang as a hobby. In the
171 attend test condition data, this was the case for 66.6% of the exposure group and 61.9% of
172 the control group participants. All the participants had studied English and Swedish as a
173 foreign language. In addition, in the ignore test condition data, 61.1% of the exposure
174 group and 88.9% of the control group participants had studied an additional language for
175 over 2 years. In the attend test condition data, this was the case for 73.3% of the exposure
176 and 81.0% of the control group participants.

177

178 **2.2 Stimuli**

179

180 We exposed the participants to lexical tones, since Finnish belongs to a quantitative
181 language group, and tonal changes are not part of the phonological system in this
182 language. Therefore, we expected that training effects could be observed. Because
183 discrimination threshold for the lexical tones applied in the study was not known for the
184 Finnish participants, and we did not want the participants to actively listen to the sounds,
185 two levels (large and small) of change were selected to maximize the possibility to find an
186 exposure effect.

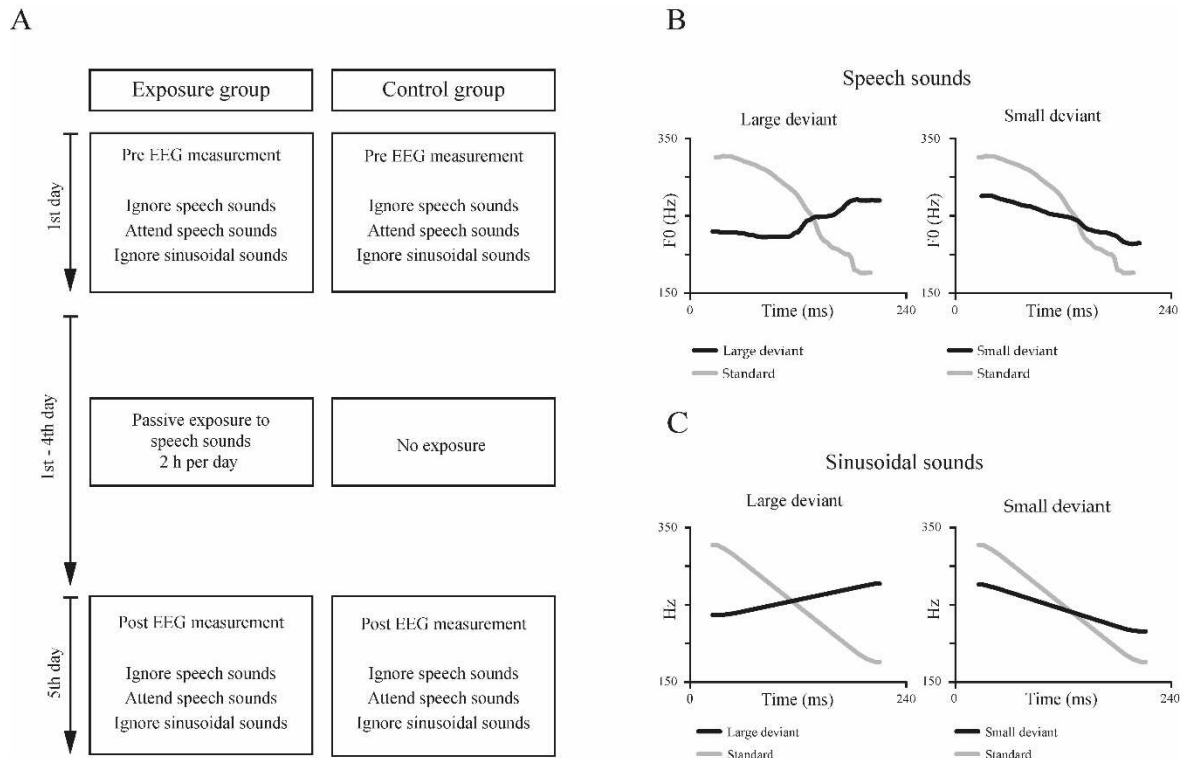
187 The sounds were prepared so that the first phoneme /a/ was spoken by a female
188 native Chinese speaker with rising (i.e., Chinese lexical tone 2) and falling (i.e., Chinese
189 lexical tone 4) pitch contour, and they were recorded at a sampling rate of 44.1 kHz. The
190 sounds were then digitally edited using SoundForge software (SoundForge 9, Sony
191 Corporation, Japan) to modify them to have a duration of 200 ms. To isolate the lexical
192 tones and keep the rest of the acoustic features identical, pitch tier transfer was performed
193 using Praat software (Praat v5.4.06, University of Amsterdam). Pitch tier transfer
194 generated a rising tone and a falling tone, which were identical to each other, except for a
195 pitch contour difference in fundamental frequency (F0). These two tones were taken as the
196 endpoint stimuli to create a continuum of lexical tones with 10 interval steps. A morphing
197 technique was performed in MATLAB (MathWorks, Inc., MA, US), and a STRAIGHT
198 tool (Kawahara et al., 1999) was used to create the three tones applied in the experiment.
199 The repeatedly presented standard sound was the falling tone (Fig. 1A), and deviant

200 sounds were a slightly falling tone (small deviant), and a rising tone (large deviant, Fig.
201 1A) corresponding to the tones 11, 7, and 3, respectively, on the tone continuum. All
202 stimuli were normalized to have the same root mean square intensity. The detailed
203 procedure concerning how the stimuli were generated was reported previously elsewhere
204 (Xi et al., 2010).

205 The sinusoidal sounds were created using SoundForge software (SoundForge 9,
206 Sony Corporation, Japan) and they had the same duration (200 ms) and start and end F0 as
207 the corresponding speech sounds. For the standard sound, the starting frequency of F0 was
208 312 Hz and it gradually decreased to 180 Hz. The large deviant had the starting F0 at 233
209 Hz and it increased gradually to 268 Hz (Fig. 1 B). Lastly, the small deviant had a starting
210 F0 at 268 Hz and it gradually decreased to 215 Hz (Fig. 1B).

211 During the pre- and post-exposure electroencephalogram (EEG) recordings and
212 during the exposure, the sounds were presented in the oddball condition, where a
213 frequently occurring standard stimulus (probability of 0.80) was interspersed with two
214 deviant sounds (large or small, probability of 0.10 for each), using E-prime 1.2
215 (Psychology Software Tools Inc., Sharpsburg, USA) software, resulting in 1000 stimulus
216 presentations with a sound pressure level (SPL) of 70 dB. The inter-stimulus interval (ISI)
217 varied randomly between 440 and 520 ms (offset to onset). The stimuli were delivered in a
218 pseudorandom fashion, with the restriction that consecutive deviant sounds were separated
219 by at least two standard sounds.

220



221

222

223 **Fig. 1. Illustration of the experiment design and fundamental frequency (F0) and pitch applied in the**
 224 **stimuli.** (A) In the pre EEG measurement, stimuli were presented in ignore and attend test conditions for
 225 both groups. After the pre EEG measurement, participants in the exposure group were exposed to speech
 226 sounds for 4 consecutive days, 2 h per day, while control group did not get any exposure. After the exposure
 227 post EEG measurement took place in the same manner as the pre EEG measurement for both groups. (B-C)
 228 In pre- and post-measurements, in separate stimulus blocks the stimuli were speech (phoneme /a/) or
 229 sinusoidal sounds presented in the oddball condition. The same speech sounds were also applied in the
 230 exposure phase. The black lines represent the deviant sounds (large and small deviants), and the gray lines
 231 represent the standard sounds.

232

233

234 2.3 Procedure

235

236 For the exposure group, experimental sessions were conducted on 5 consecutive days
 237 (Fig. 1A). On the first day, a pre-exposure EEG measurement was carried out to determine
 238 the responses to stimuli before the passive exposure. During the experiment, participants
 239 sat in a comfortable chair in a well-lit room with a video connection to the experimenter.
 240 Auditory stimuli were presented via a loudspeaker placed at approximately 50 cm above
 241 the participant's head. Two experimental conditions were applied. First, in separate
 242 stimulation blocks, we played the speech sounds (Fig. 1B) and sinusoidal sounds (Fig. 1C)

243 to the participants and asked them to ignore the sounds and concentrate on a silent movie.
244 Second, participants were instructed to detect changes in the speech sounds (Fig. 1B) and
245 press a button as quickly as possible whenever they detect a deviant sound. They were not
246 informed about the type of changes (e.g., rising or falling pitch).

247 From 2 - 6 hours after the pre-exposure EEG measurement, the first exposure session
248 started. Participants watched silent movies for 2 h, while in the background the same
249 speech sounds as used in the pre-exposure EEG measurement were presented in same
250 oddball paradigm from the loudspeaker placed at approximately 50 cm above the
251 participant's head with sound pressure level (SPL) of 50 dB (Fig. 1A). Participants were
252 instructed to ignore the sounds and pay attention to the movie. Again, no information
253 about the sounds was given to the participants. After every 30 min, a break was taken and
254 participants were asked questions regarding the plot of the movie to keep their focus on it.
255 The second, third, and fourth exposure sessions on the following consecutive days were
256 the same as the first one. A total of 2 h of exposure was given on each day, summing to a
257 total of 8 h for each participant. During the exposure, EEG time-locked to sounds was
258 recorded, but these data are not reported here.

259 On day 5, a post-exposure EEG measurement was performed. The procedure was
260 identical to that in the pre measurement (Fig. 1A).

261 Control participants went through the pre and post EEG measurements the same way
262 as the exposure group on days 1 and 5, but they received no sound exposure between these
263 measurements (Fig. 1A).

264

265 **2.4 EEG measurements**

266

267 Raw EEG was recorded with the Electrical Geodesics Inc. (EGI, Eugene, OR, USA)
268 system with 128-channel sensor nets (Hydrogel GSN 128, 1.0) using Ag-AgCl electrodes.
269 The sampling rate during the pre- and post-tests was 500 Hz, and the data were filtered
270 online from 0.1 Hz to 200 Hz. Impedances for the electrodes were kept below 50 k Ω .

271

272 **2.5 Analysis of behavioral data**

273

274 Responses to the target deviant stimuli were considered hits if they occurred after the
275 offset of the deviant sound and before the onset of the second standard stimulus onset (i.e.

276 the hit response could occur during the post-deviant standard stimulus). The reaction times
277 for the hits were calculated from the onset of the deviant stimulus.

278

279 **2.6 EEG analysis**

280

281 Offline data analysis was performed on Brain Vision Analyzer 2.1 (Brain Products
282 GmbH). An infinite impulse response (IIR) filter for a 1-Hz low cut-off (24 dB/octave)
283 was applied for the continuous EEG data offline. Then, independent component analysis
284 (ICA) was performed to detect and remove eye movement artifacts from the data. Noisy
285 channels were interpolated using a spherical spline model (Perrin et al., 1988); the average
286 of the interpolated channels was seven. Epochs from 100 ms before to 700 ms after the
287 stimulus onset were parsed into segments. The baseline correction was calculated based on
288 a 100-ms prestimulus interval. Epochs including amplitude values outside the range from -
289 200 to 200 μV , activity less than 0.5 μV , and gradients larger than 75 $\mu\text{V}/\text{ms}$ were rejected
290 within 100-ms consecutive intervals for the epoch. Epochs were averaged separately for
291 each deviant type and the standard stimuli that preceded the deviant stimuli. The data were
292 re-referenced offline to an average reference. A 20-Hz high cut-off filter was applied to the
293 averaged segments. Averages that had more than 50/100 accepted epochs were included in
294 the analysis.

295

296 **2.7 Statistical analysis of behavioral data**

297

298 Reaction times and discrimination accuracy were analyzed with a repeated-measures
299 analysis of variance (ANOVA), with the deviant type (large vs. small) and session (pre vs.
300 post) as within-subject factors and the group type as a between-subject factor (exposure vs.
301 control).

302

303 **2.8 Statistical analysis of EEG data**

304

305 For the ignore condition, mean amplitude values were calculated from the time
306 windows of 190–240 ms after stimulus onset for the MMN, of 250–300 ms for the P3a to
307 the speech sounds, and of 300 - 350 ms for the P3a to the sinusoidal sounds. The time
308 windows for the attend condition were 230–280 ms for the N2b and 360–410 ms for the
309 P3b.

310 The mean values for the MMN and P3a in both speech and sinusoidal sounds in the
311 ignore condition were extracted as a mean from three separate electrode clusters, as
312 follows: left frontal (20, 23, 24, 28), mid-frontal (4, 11, 16, 19), and right frontal (3, 117,
313 118, 124), corresponding roughly to the areas of F3, Fz, and F4 in 10-20 system,
314 respectively. For the N2b, the electrode clusters were left central (29, 30, 36, 37), mid-
315 central (5, 6, 12, 11), and right central (87, 104, 105, 111), corresponding to C3, Fz, and
316 C4, respectively. Finally, for the P3b, the left parietal (47, 52, 59, 60), mid-parietal (61, 62,
317 72, 78), and right parietal (85, 91, 92, 98) electrode clusters, corresponding to P3, Pz, and
318 P4, respectively, were selected (Supplementary Fig. 1). The time windows and electrode
319 clusters were selected based on the previous literature (for reviews, see Näätänen et al.,
320 2005; Patel et al., 2005; Polich, 2007) and visual inspection of the topographies and grand
321 averaged waveforms (Supplementary Figs 2 - 5).

322 The latencies were analyzed from the deviant responses only since standard stimuli
323 did not elicit clear responses for all participants. Latencies for the MMN, in both speech
324 and sinusoidal sounds, and N2b were determined as a time point where a minimum
325 amplitude (most negative) value for the deviant response was found from the time window
326 of 150 - 260 ms and 200 - 310 ms, respectively. The latencies for the P3a, in both speech
327 and sinusoidal sounds, and P3b were determined as a time point where a maximum
328 amplitude (most positive) value for the deviant response was found from the time window
329 of 250-350 ms and 340-460 ms, respectively. The electrode clusters for the latency
330 analyses were the same as those applied for the amplitude analysis for each component.

331 Statistical analyses were performed on IBM SPSS Statistics v. 24 (IBM corporation,
332 NY, USA). The mean amplitude values were analyzed separately for each component with
333 a repeated-measures ANOVA with stimulus type (standard vs. deviant), deviant type (large
334 vs. small), electrode cluster (left vs. mid vs. right) and session (pre vs. post) as within-
335 subject factors and a between-subject factor group (exposure vs. control). The mean
336 deviant response latency values were analyzed separately for each component with
337 repeated-measures ANOVA, with deviant type (large vs. small), electrode cluster (left vs.
338 mid vs. right) and session (pre vs. post) as within-subject factors and a between-subject
339 factor group (exposure vs. control). Huynh-Feldt-corrected degrees of freedom were used
340 whenever the sphericity assumption was violated. The corrected p-values are reported, but
341 the degrees of freedom are reported as uncorrected. Parietal eta square (η_p^2) was used as an
342 index of the effect size estimate. Here we give a complete report of only interaction effects

343 that contain session x group effect since our focus is on the effect of the passive exposure.
344 Other effects are reported in the supplementary materials.

345 Repeated measures of ANOVAs and paired t-tests (two-tailed with Bonferroni
346 correction) were used to further investigate the interaction effects. For the four-way
347 interactions, repeated measures of ANOVAs with session x group interaction effects were
348 investigated first, and then continued with session related interactions separately for each
349 group. For the three-way interactions session related interactions were investigated
350 separately for each group.

351 P-values and confidence intervals (CIs) of 95% are reported after performing a
352 bootstrapping with 1,000 permutations. Cohen's d with pooled standard deviation was
353 used as an index of the effect size estimate.

354 Whenever a statistically meaningful interaction effects of session x group was found
355 in the ANOVA, Pearson correlation coefficients (two-tailed) were calculated between
356 behavioral responses (reaction time, accuracy) and the corresponding ERP amplitude and
357 latency values of the deviant responses from the post measurement. For the correlations, P-
358 values, 99% CIs, and correlation coefficients are reported based on 1,000 permutations in
359 bootstrapping. The threshold for statistical significance was $p < 0.05$.

360

361

362 **3. Results**

363

364

365 **3.1 Attend condition for speech sounds**

366

367 **3.1.1 Behavioral results**

368

369 There were no interactions including session x group for the reaction times or for the
370 accuracy of the behavioral responses. Detailed results for the behavioral responses are
371 reported in the supplementary materials S1.1.

372

373

374

375

376 3.1.2 N2b component

377

378 There were no interactions including session x group for the N2b amplitude or
379 latency (Table 1). The responses to deviant and standard sounds in N2b time window are
380 reported in supplementary Fig. 2.

381

382 3.1.3 P3b component

383

384 For the P3b amplitude, an interaction effect of deviant type x stimulus type x session
385 x group was found (Table 1). The following ANOVAs separately for the deviant responses
386 (deviant type x session x group) (Supplementary Table 1) or separately for small and large
387 deviant responses (stimulus type x session x group) revealed no session x group
388 interactions (Supplementary Table 2). However, the subsequent ANOVA (stimulus type x
389 deviant type x session) performed separately for each group, revealed a significant
390 interaction effect of stimulus type x session in the passive exposure group, $F_{1,14} = 4.97$, $p =$
391 0.043 , $\eta_p^2 = 0.26$ (Supplementary Table 3). There was no session-related interaction effect
392 in the control group. For the passive exposure group, subsequent t-tests were conducted
393 where amplitude values between the pre and post measurements were compared separately
394 for the standard and deviant sounds. These revealed that the deviant responses became
395 significantly more positive from the pre measurement (2.98 ± 1.55) to post measurement
396 (3.61 ± 1.20), $t(14) = 2.78$, $p = 0.032$, 95% CI $[-1.09, -0.21]$, $d = 0.45$ (Fig. 2 and Fig. 3),
397 but there was no change in the standard responses. There was no exposure effect for the
398 latencies of the P3b response (Table 1).

399

400

401

402

403

404

405

406

407

408

409 **Table 1.** Summary of the significant effects in the repeated measures of ANOVA for the attend condition
 410 (speech sounds). * marks exposure-related effect. Degrees of freedom (df), F-values (F), P-values (P), and
 411 parietal eta squared (η_p^2) for effect sizes are reported.

Component/variable	Effect	df	F	P	η_p^2
N2b/ amplitude	Stimulus type	1,34	37.9	0.0001	0.53
	Deviant type	1,34	5.2	0.030	0.13
	Session	1,34	57.9	0.0001	0.63
	Electrode x Group	2,33	3.5	0.043	0.17
	Deviant type x Stimulus type	1,34	38.8	0.003	0.23
	Stimulus type x Electrode	2,33	6.4	0.004	0.28
	Session x Electrode	2,33	4.8	0.014	0.22
	Stimulus type x electrode x session	2,33	4.9	0.013	0.23
	N2b (deviant)/ latency	Deviant type	1,34	77.36	0.0001
Session		1,34	9.46	0.004	0.22
Electrode		2,33	9.91	0.0001	0.38
Deviant type x Session		1,34	4.17	0.049	0.11
P3b/ amplitude	Deviant type	1,34	32.18	0.0001	0.47
	Stimulus type	1,34	97.03	0.0001	0.74
	Session	1,34	4.68	0.038	0.12
	Electrode	2,33	22.08	0.0001	0.57
	Electrode x Group	2,33	4.26	0.023	0.21
	Deviant type x Stimulus type	1,34	34.93	0.0001	0.51
	Stimulus type x Session	1,34	5.37	0.027	0.14
	Stimulus type x Electrode	2,33	33.64	0.0001	0.67
	Session x Electrode	2,33	5.05	0.012	0.23
	Stimulus type x Electrode x Group	2,33	9.36	0.001	0.36
	Deviant type x Stim type x Electrode	2,33	6.42	0.004	0.28
	Deviant type x Stim type x Session x Group*	1,34	6.50	0.015	0.16
	P3b (deviant)/ latency	Deviant type	1,34	8.99	0.005
Session		1,34	59.89	0.0001	0.64
Electrode		1,34	4.08	0.027	0.20

412

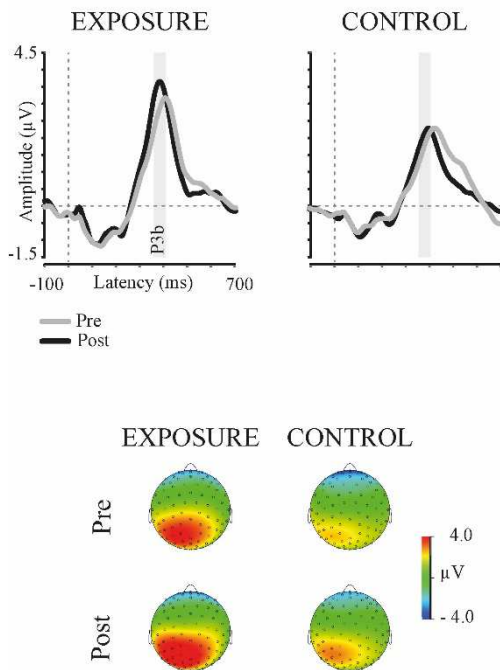
413 **3.1.4 Correlations between ERPs in attend condition and behavioral responses**

414

415 In the exposure group, there was a marginally significant correlation between the
 416 post measurement P3b amplitude and reaction times for the small deviant, $r = -0.497$, $p =$
 417 0.059 , 99% CI $[-0.80, -0.11]$; the larger the response amplitude was, the faster the reaction
 418 time became. Other correlations were non-significant.

419

DEVIANT RESPONSES



420

421

422 **Fig. 2. Grand averaged P3b responses in the exposure group, n = 15 and the control group, n = 21.** The

423 gray lines represent responses to deviant sounds from the pre measurement, and the black lines signify the

424 responses to deviant sounds from the post measurement. Grand averaged waveforms are presented as mean

425 values from the collapsed electrode clusters (left, middle, and right parietal; see Supplementary Fig. 1 and 3).

426 In lower panel, grand averaged scalp topographies of the deviant responses as a mean amplitude value of the

427 analyzed time window at 360–410 ms from the 128 electrodes for the P3b for the exposure group and control

428 group are shown.

429

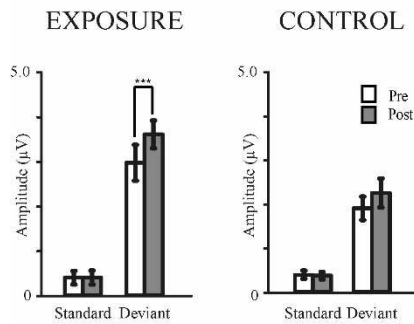
430

431

432

433

P3b AMPLITUDE



434

435

436 **Fig. 3. Passive exposure enhanced the P3b amplitude.** The mean amplitude values of P3b averaged across
 437 the three electrode clusters (left, middle, and right parietal) in 360–410 ms time window. The white bars
 438 indicate the amplitude values at the pre measurement, and the gray bars represent those at the post
 439 measurement. *** indicates a statistically significant difference ($p < 0.05$) and error bars indicate the
 440 standard error of the mean. Post exposure responses to the deviant sounds were enhanced compared to those
 441 in pre measurement. No such effect was found in the control group.

442

443

444 **3.2 Ignore condition for speech sounds**

445

446 **3.2.1 MMN component**

447

448 There was no exposure effect for the amplitude of the MMN response. However, for
 449 the MMN latency, an interaction effect of session x group was found (Table 2). Separate t-
 450 tests for the groups comparing the latencies between pre and post measurements showed
 451 that the deviant response latencies became shorter from the pre to post measurements in
 452 the exposure group, while no such changes in latencies were found in the control group
 453 (Fig. 4 and 5).

454

455 **3.2.2 P3a component**

456

457 For the amplitude of the P3a component, an interaction effect for stimulus type x
 458 session x group was found (Table 2). Subsequent ANOVA (stimulus type x session)
 459 performed separately for each group revealed that there was an interaction effect of
 460 stimulus type x session in the exposure group, $F_{1,17} = 5.66$, $p = 0.029$, $\eta_p^2 = 0.25$, while no
 461 session-related main or interaction effects were observed in the control group

462 (Supplementary Table 4). Separate t-tests for standard and deviant responses comparing
 463 the amplitude change from pre to post measurement were conducted for the exposure
 464 group. Responses to deviant sounds increased in amplitude toward a positive polarity and
 465 the same was observed for the responses to the standard sounds (Fig. 4 and 5).

466 The passive exposure affected the latencies of the deviant responses in the P3a time
 467 window, as indicated by the deviant type x session x group interaction effect (Table 2).
 468 Subsequent ANOVA (deviant type x session) performed separately for each group
 469 revealed an interaction effect of deviant type x session in the exposure group, $F_{1,17} = 9.01$,
 470 $p = 0.008$, $\eta_p^2 = 0.35$. No session-related main or interaction effects were found in the
 471 control group. Subsequent t-tests comparing latencies separately for the deviant types in
 472 the exposure group showed that after the exposure, the latency of P3a to the large change
 473 was shorter ($273.6 \text{ ms} \pm 14.9$) than it was before the exposure ($294.0 \text{ ms} \pm 24.5$), $t(17) =$
 474 2.92 , $p = 0.02$, 95% CI [8.0, 34.0], $d = 1.00$ (Fig. 5). No effect was found for the small
 475 change (Fig. 5).

476

477 **Table 2.** Summary of the significant effects in the repeated measures of ANOVA for the ignore condition
 478 where the speech sounds were presented. * marks exposure-related effect. Degrees of freedom (df), F-values
 479 (F), P-values (P), and parietal eta squared (η_p^2) for effect sizes are reported.

480

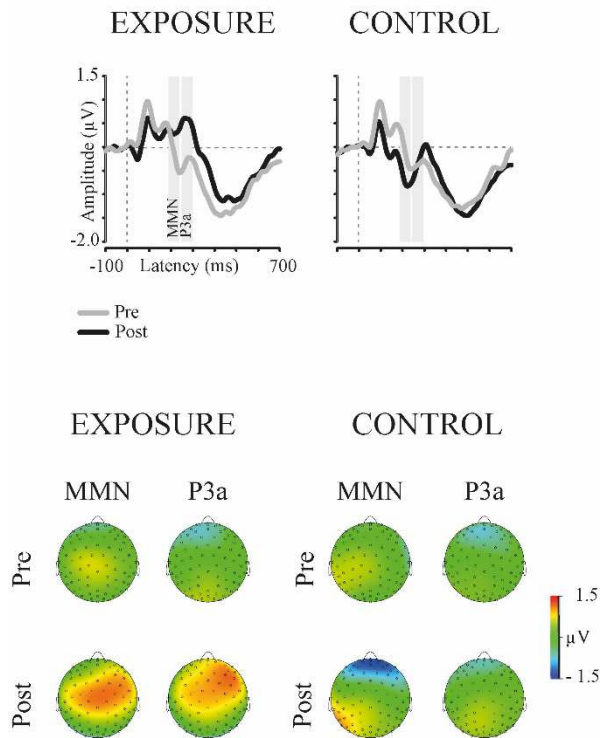
Component/variable	Effect	df	F	P	η_p^2
MMN/ amplitude	Stimulus type	1,34	6.73	0.014	0.17
MMN (deviant)/ latency	Electrode	2,33	5.14	0.011	0.24
	Deviant type	1,34	6.92	0.013	0.17
	Deviant type x Session	1,34	4.17	0.049	0.11
	Session x Group*	1,34	6.71	0.014	0.17
P3a/ amplitude	Deviant type	1,34	4.27	0.047	0.11
	Stimulus type	1,34	4.34	0.045	0.11
	Session	1,34	6.91	0.013	0.17
	Electrode	2,33	4.59	0.017	0.22
	Stimulus type x Group	1,34	46.57	< 0.0001	0.58
	Session x Group*	1,34	21.12	< 0.0001	0.38
	Deviant type x Group	1,34	13.76	0.001	0.29
	Dev type x Stim type x Group	1,34	12.83	0.001	0.27
	Stimulus type x Session x Group*	1,34	9.90	0.003	0.23
	Stimulus type x Electrode x Group	2,33	4.95	0.013	0.23
P3a (deviant)/ latency	Session x Group*	1,34	4.78	0.036	0.12
	Deviant type x Session x Group*	1,34	6.64	0.014	0.16

481

482

483

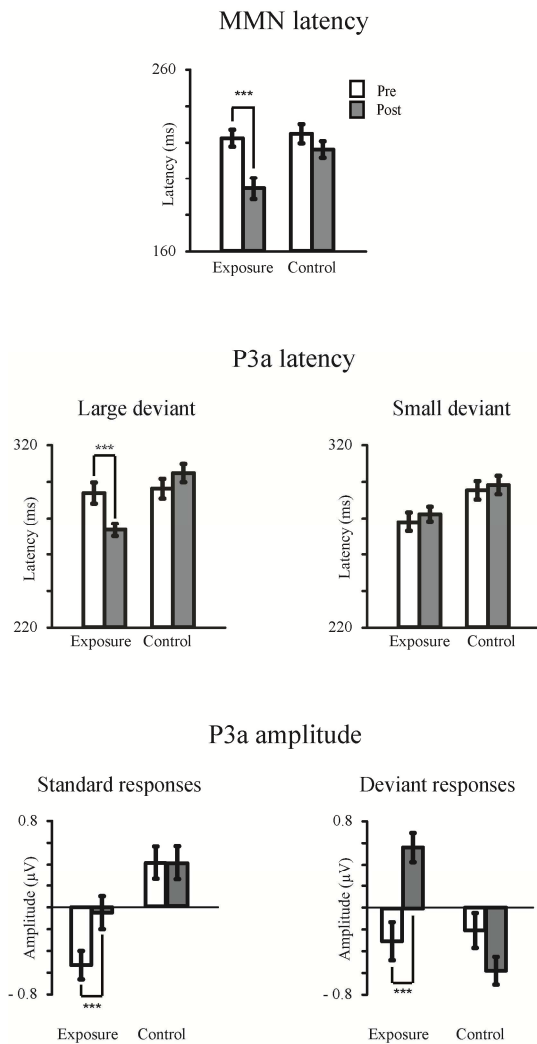
DEVIANT RESPONSES



484

485 **Fig. 4. Grand averaged deviant responses in the time windows of the mismatch negativity (MMN) and**
 486 **P3a responses in the exposure group, n = 18, and in the control group, n = 18.** Grand average waveforms
 487 to deviant responses (small and large deviant averaged) are presented as mean values from the collapsed
 488 electrode clusters (left frontal, middle frontal, right frontal; see Material and methods and Supplementary
 489 Fig. 1 and 4). The gray lines represent responses in the pre measurement, while black lines represent
 490 responses in the post measurement for the exposure group and the control group. The light gray bars
 491 represent the time windows analyzed for the MMN and P3a responses (190–240 ms and 250–300 ms post
 492 stimulus onset, respectively). In lower panel, the grand averaged scalp topographies of the responses to
 493 deviant sounds as a mean amplitude value of the analyzed time window from the 128 electrodes for the
 494 MMN and P3a for the exposure group and the control group are shown.

495



496

497

498 **Fig. 5. Mismatch negativity latency and P3a latency and amplitude are enhanced after the passive**
 499 **exposure.** The white and gray bars represent the mean values from the pre and post measurements,
 500 respectively. Error bars indicate the standard error of the mean. *** indicates a statistically significant (p
 501 $< .05$) difference. The latency of the response to the deviant sounds in the MMN time window shortened
 502 from the pre measurement (222.45 ms \pm 19.61) to the post measurement (194.57 ms \pm 24.81), $t(17) = 4.79$,
 503 $p = 0.00$, 95% CI [17.44, 39.78], $d = 1.25$), in the exposure group, while there were no changes in the
 504 control group. In addition, in the P3a time window, deviant response latencies became significantly shorter
 505 for the large deviant after the exposure, but not for the small deviant, and there were no changes in the
 506 control group. Standard response amplitudes were more positive after the exposure ($-0.05 \mu\text{V} \pm 0.64$)
 507 compared to the pre measurement ($-0.53 \mu\text{V} \pm 0.55$), $t(17) = 3.18$, $p = 0.026$, 95% CI $[-0.79, -0.19]$, $d = 0.80$.
 508 Also deviant response amplitudes became more positive after the exposure ($0.57 \mu\text{V} \pm 0.57$) compared to the
 509 pre measurement ($-0.30 \mu\text{V} \pm 0.74$), $t(17) = 4.70$, $p = 0.002$, 95% CI $[-1.23, -0.53]$, $d = 1.32$.

510

511

512

513 3.3 Ignore condition for sinusoidal sounds

514 The transfer effect to non-exposed sound features was tested by presenting
 515 sinusoidal sounds roughly mimicking the pitch contours of the speech sounds in a passive
 516 oddball condition (Fig. 1C). The transfer effect was investigated for the components and
 517 variables showing group x session interaction effects in the ignore condition where speech
 518 sounds were presented, i.e. for the MMN latency, P3a amplitude and P3a latency.

519 3.3.1 MMN component

520 For the MMN latency, there was a significant interaction effect of electrode cluster x
 521 session x group (Table 3). Subsequent ANOVA (electrode cluster x session) performed
 522 separately for each group revealed a significant main effect session for both groups $F_{1,17} =$
 523 15.03 , $p = 0.001$, $\eta_p^2 = 0.47$; $F_{1,17} = 5.90$, $p = 0.027$, $\eta_p^2 = 0.26$ (Supplementary table 5).
 524 Post hoc paired samples t-tests comparing latencies between pre and post measurements
 525 separately for groups showed that in the exposure group the latencies got significantly
 526 shorter from pre measurement ($234.9 \text{ ms} \pm 10.58$) to post measurement ($215.27 \text{ ms} \pm$
 527 21.44), $t(17) = 3.88$, $p = 0.014$, 95% CI [10.43 , 29.74], $d = 1.16$). There were no changes
 528 in the latencies in the control group.

529 **Table 3.** Summary of the significant effects in the repeated measures of ANOVA for the sinusoidal sounds in
 530 the ignore condition. * marks exposure-related effect. Degrees of freedom (df), F-values (F), P-values (P),
 531 and Partial eta squared (η_p^2) for effect size are reported.

Component/variable	Effect	df	F	P	η_p^2
MMN (deviant)/ Latency	Electrode	2,33	6.89	0.003	0.29
	Session x Group*	1,34	19.57	< 0.001	0.37
	Session x Electrode x Group*	2,33	4.52	0.018	0.22
P3a/ Amplitude	Stimulus	1,34	37.9	< 0.001	0.53
	Electrode	2,33	8.4	0.001	0.34
	Stimulus x Group	1,34	12.4	0.001	0.27
	Stimulus x Session	1,34	4.2	0.047	0.11
	Stimulus x Session x Electrode x Group*	2,33	4.0	0.027	0.20
	P3a (deviant)/ latency	Deviant type	1,34	7.36	0.01
	Deviant type x Group	1,34	5.65	0.023	0.14

532

533 3.3.2 P3a component

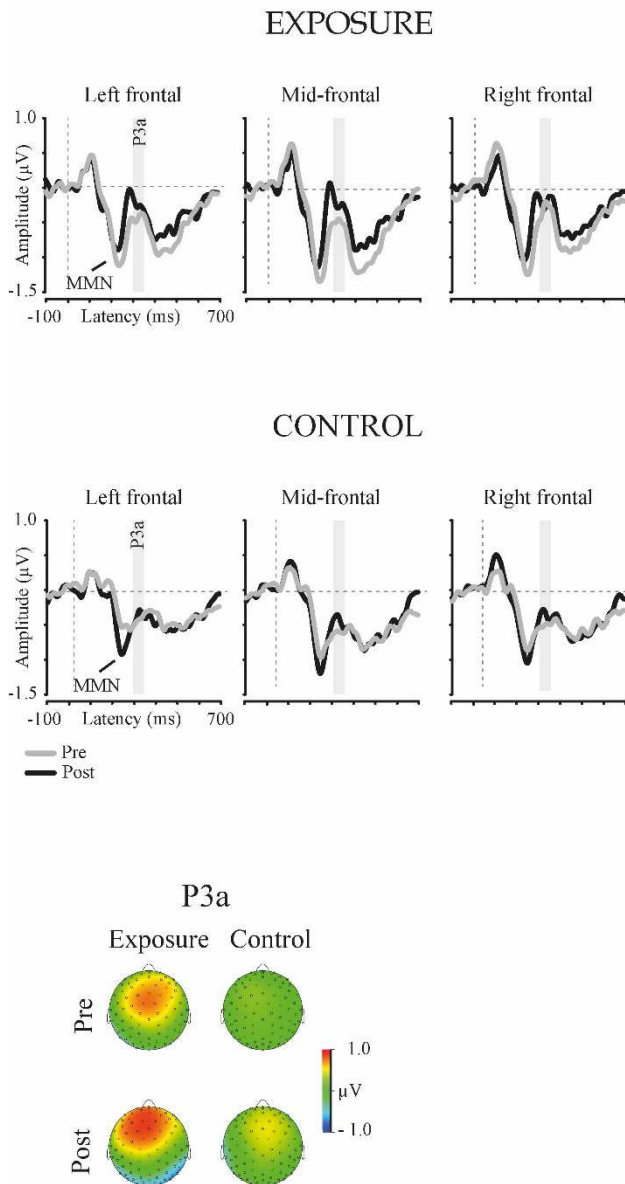
534 For the P3a amplitude, there was an interaction effect of stimulus type x electrode
 535 cluster x session x group (Table 3). The following ANOVAs (electrode cluster x session x
 536 group) for the deviant responses (Supplementary table 6) or (electrode cluster x session x

537 group) performed separately for each electrode cluster revealed no session x group
538 interaction effects (Supplementary table 7) Subsequent ANOVA (stimulus type x electrode
539 cluster x session) performed separately for each group revealed significant interaction
540 effect of stimulus type x electrode cluster x session in the control group, while there was
541 no session-related effects in the exposure group (Supplementary Table 8). T-tests
542 comparing deviant and standard responses separately from pre measurement to post
543 measurement in each electrode cluster was performed in the control group. They did not
544 reveal any significant results.

545

546

DEVIANT RESPONSES



547

548 **Fig. 6. Grand averaged MMN and P3a responses to sinusoidal sounds.** Grand averaged waveforms of
 549 responses reflecting MMN and P3a in the exposure group ($n = 18$) and the control group ($n = 18$). The gray
 550 lines represent responses to deviant sounds (large and small deviant types averaged) from the pre
 551 measurement, while black lines represent responses to deviant sounds from the post measurement for the
 552 exposure group and the control group. The light gray bars represent the time windows applied in the analysis
 553 of for the P3a amplitude (mean amplitude value between 300–350 ms). The mean scalp topographies of the
 554 differential response (standard subtracted from the deviant, deviant types averaged) from the 128 electrodes
 555 for the P3a for the exposure group and the control group from the analyzed time window. Although there
 556 was a significant stimulus type \times electrode cluster \times session \times group effect for the P3a amplitude, post-hoc
 557 tests did not reveal any generalization of the exposure effect. Please note that the MMN amplitude was not
 558 investigated since there was no exposure effect for it.

559

560 4. Discussion

561

562

563 Here we show in adult humans that passive exposure to foreign speech sounds for 4
564 consecutive days, 2 h per day, enhanced the neural discrimination ability and attention
565 orientation toward changes in the speech sounds as indexed by ERPs recorded in ignore
566 and attend test conditions. The effect of passive exposure to auditory change detection
567 mechanism has earlier been found only in infants (Cheour et al., 1998; Cheour et al., 2002;
568 Kuhl, 2004; Trainor, Lee, & Bosnyak, 2011). In the attend test condition, effect of passive
569 exposure was demonstrated as enhanced P3b amplitude. In the ignore test condition,
570 effects of passive exposure were demonstrated as shortened latency of the MMN and
571 enhanced amplitude and shortened latency of the P3a.

572 The learning effect generalized to some extent to novel sounds: the latency of the
573 MMN shortened to the sinusoidal sounds not encountered during the exposure phase. This
574 effect was demonstrated only in the exposure group, not in the control group.

575 Effects of auditory perceptual learning have rarely been tested for attentive change
576 detection. Here we showed that passive exposure enhanced the amplitude of the P3b and
577 there was trend towards significant correlation between the enhanced P3b amplitude and
578 shortened behavioral reaction times to small deviant. Previously, it has been shown that
579 when the perceptual task becomes easier, the P3b amplitude increases (Isreal et al., 1985).
580 Our results are also in line with one previous study which showed that active training to
581 discriminate speech sounds enhances the microstates related to the P3b component
582 accompanied by improvements in behavioral reaction times (Giroud et al., 2017). In the
583 light of the context-updating model (Polich, 2007), passive exposure seems to ease
584 comparison process between the representation of the standard sound in memory and the
585 deviant sound input, which is also reflected as shortened reaction times.

586 N2b was the other component that was investigated in the attend test condition.
587 Here, the amplitude of the N2b was not enhanced, nor was its latency shortened due to
588 passive exposure. In prior studies applying attentive training, in line with our findings,
589 N2b was not enhanced during identification task (Ben-David et al. 2011) or during
590 discrimination task (Giroud et al. 2017). However, a study that had longer practice period
591 than in study by Ben-David et al. (2011) reported that enhancement in ability to identify
592 speech sounds were followed by increased N2b (Alain et al. 2010). It remains thus unclear

593 whether perceptual learning requires attentive training to modulate the N2b, and if so,
594 whether the training should be identification training, instead of discrimination training.

595 No exposure-related effects were found for behavioral responses measured in the
596 attentive test condition. Reaction times decreased from pre to post measurement in both
597 groups, and the decrease was larger for the small deviant than for the large deviant. The
598 detection of the large deviant was more accurate (97.0%) than the detection of small
599 deviant (88.4%) but there was no change from pre to post test in accuracy. Sometimes
600 neural changes related to auditory learning precede its behavioral indices (Tremblay et al.,
601 1998). It is possible that the passive exposure applied here should have been more
602 extended in order to induce changes at behavioral level.

603 In the ignore test condition, the peak latency of the MMN response to the speech
604 sounds was significantly shorter after the exposure compared with that before the
605 exposure, while no latency changes were observed in the control group. Changes in MMN
606 latency have been interpreted as enhanced discrimination ability due to enhanced memory
607 traces of deviant sounds in studies where attentive training has been found to modulate the
608 latency of MMN (Kraus et al., 1995; Tremblay et al., 1997; Tremblay et al., 1998). Our
609 results show that similar modulation of the MMN can be induced by mere passive
610 exposure. However, the MMN amplitude did not show any changes due to the passive
611 exposure, although it typically increases in studies utilizing active training (Näätänen et al.,
612 1993; Kraus et al., 1995; Tremblay et al., 1997). Also in our previous study in rats, 36-h
613 passive exposure to spectro-temporal changes in speech sound /a/ enhanced the mismatch
614 response amplitude (Kurkela et al., 2016).

615 Even though the amplitude of the MMN response was not changed due to passive
616 exposure, the following P3a component's amplitude was enlarged. Also the latency of P3a
617 was shortened, for the large deviant. Previously, active training of discriminating pitch
618 (Seppänen et al., 2012), tone sequences (Atienza et al., 2004) or to learning to use Morse
619 code (Uther et al., 2006) have led to increased P3a amplitude. Our results demonstrate that
620 mere passive exposure suffices changes in the involuntary attention-shifting mechanism
621 which the P3a is typically linked to.

622 Here passive exposure expanded over four consecutive days, totaling in eight hours
623 of exposure. Studies applying attentive training have usually also spread the training
624 sessions over several days (6-9 days) (for example, Kraus et al., 1995; Tremblay et al.,
625 1997). Previous studies of passive exposure have applied only short exposure session
626 during a single day, and in these works, no effects of passive exposure in ERPs reflecting

627 sound discrimination have been demonstrated (Näätänen et al., 1993; Sheehan et al., 2005;
628 Elmer et al., 2017). It can be assumed that we were able to demonstrate the effect of
629 passive exposure in the ERPs related to change detection and attention shifting because the
630 exposure was relatively long-lasting. In addition, the exposure that was extended for four
631 consecutive days could have facilitated memory consolidation for the exposed sounds
632 during the nocturnal sleep, allowing the emergence of the exposure effect. Previous
633 studies applying attentive training have shown that sleep deprivation prevents the
634 emergence or enhancement of the MMN and the P3a (Alain et al., 2015; Ross et al., 2015;
635 Atienza et al., 2004). It thus seems that nocturnal sleep is a crucial factor for the
636 emergence of learning-related enhancement in the change detection, probably due to
637 memory consolidation for the learned sounds.

638 We also investigated the generalization of the perceptual learning from speech
639 sounds to sinusoidal sounds. This was studied for the MMN latency and P3a latency and
640 amplitude since these showed the effect of passive exposure. For the sounds that were
641 mimicking the pitch contours of the speech sounds, the peak latency of the deviant
642 responses in the MMN time window was significantly shorter in the exposure group after
643 the exposure than before the exposure. No such effect was found in the control group. This
644 pattern of results can be interpreted to reflect transfer of learning at a neural level due to
645 passive exposure to speech sounds. Our result related to generalization is in line with the
646 results of previous behavioral studies showing that learning to discriminate sound
647 frequencies or syllables generalizes to closely similar sounds (for a review, see Wright et
648 al. 2008). Furthermore, our findings are in line with a study, where attentive training
649 induced changes in the MMN (Tremblay et al., 1997). The learning effect also transferred
650 to novel speech stimuli, i.e. from one place of articulation (labial) to another (alveolar).
651 Our results extend this finding by showing that similar transfer effect can be induced by
652 mere passive exposure and from speech to non-speech sounds.

653 The generalization effect was not observed for the P3a amplitude or latency. It thus
654 seems that in the case of passive exposure latency changes are more sensitive than
655 amplitude changes to reflect generalization effect.

656 In summary, passive exposure to foreign speech sounds for 2 h for 4 consecutive
657 days induced plastic cortical changes related to change detection and attention shift
658 mechanisms. As indexed by ERPs, this was demonstrated in the attend test condition by
659 increased P3b amplitude and in the ignore test condition by the shortened latency of the
660 MMN and P3a as well as increased amplitude of the P3a component. In addition, the

661 latency of the MMN shortened to the sinusoidal sounds not encountered during the
662 exposure, reflecting generalization of the learning effect. For the first time, these results
663 demonstrated that mere passive exposure to sounds can induce plastic changes related to
664 change detection in the adult human brain, which was previously thought to happen only in
665 infancy during the sensitive period. Changes in brain responses occurred from 8 h of
666 exposure. This encourages testing the effectiveness of passive exposure in real-life
667 language learning situations.

668

669 **Disclosure statement**

670 The authors report no conflicts of interest.

671 **Acknowledgements**

672

673 The study was supported by the Academy of Finland (project number 273134 for
674 Piia Astikainen) and by the Natural Science Foundation of China (grant number
675 81461130018 for Hua Shu).

676 **5. References**

677

678 Alain C, Campeanu S, Tremblay K (2010) Changes in sensory evoked responses coincide
679 with rapid improvement in speech identification performance. *J Cogn Neurosci*
680 22:392-403. doi: 10.1162/jocn.2009.21279.

681 Alain C, Da Zhu K, He Y, Ross B (2015) Sleep-dependent neuroplastic changes during
682 auditory perceptual learning. *Neurobiol Learn Mem*, 118:133–142. doi:
683 10.1016/j.nlm.2014.12.001.

684 Atienza M, Cantero JL, Dominguez-Marin E. (2002) The time course of neural changes
685 underlying auditory perceptual learning. *Learning & Memory* 9:138-150. doi:
686 10.1101/lm.46502.

687 Ben-David BM, Campeanu S, Tremblay KL, Alain C (2011) Auditory evoked potentials
688 dissociate rapid perceptual learning from task repetition without learning.
689 *Psychophysiology* 48:797-807. doi: 10.1111/j.1469-8986.2010.01139.x.

690 Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Näätänen R (1998)
691 Development of language-specific phoneme representations in the infant brain. *Nat*
692 *Neurosci* 1:351–3. doi: 10.1038/1561.

693 Cheour M, Martynova O, Näätänen R, Erkkola R, Sillanpää M, Kero P, Raz A, Kaipio
694 ML, Hiltunen J, Aaltonen O, Savela J, Hämäläinen H (2002) Speech sounds learned
695 by sleeping newborns. *Nature* 415:599–600. doi: 10.1038/415599b.

696 Crowley KE, Colrain IM (2004) A review of the evidence for P2 being an independent
697 component process: Age, sleep and modality. *Clin Neurophysiol*, 115:732–744. doi:
698 10.1016/j.clinph.2003.11.021.

699 Elmer S, Hausheer M, Albrecht J, Kühnis J (2017) Human brainstem exhibits higher
700 sensitivity and specificity than auditory-related cortex to short-term phonetic
701 discrimination learning. *Sci Rep* 7:7455. doi: 10.1038/s41598-017-07426-y.

702 Giroud N, Lemke U, Reich P, Matthes KL, Meyer M (2017) Longitudinal auditory
703 learning facilitates auditory cognition as revealed by microstate analysis. *Biol*
704 *Psychol* 123: 25–36. doi: 10.1016/j.biopsycho.2016.11.007.

705 Kawahara H, Masuda-Katsuse I, de Cheveigne A (1999) Restructuring speech
706 representations using a pitch-adaptive time–frequency smoothing and an

- 707 instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure
708 in sounds. *Speech Commun* 27:187–207. doi: 10.1016/S0167-6393(98)00085-5.
- 709 Kraus N, McGee T, Carrell TD, King C, Tremblay K, Nicol T (1995) Central auditory
710 system plasticity associated with speech discrimination training. *Cognitive Neurosci*
711 7:25–32. doi: 10.1162/jocn.1995.7.1.25.
- 712 Kuhl PK (2004) Early language acquisition: cracking the speech code. *Nature Rev*
713 *Neurosci* 5:831–843. doi: 10.1038/nrn1533.
- 714 Kurkela JLO Lipponen A, Hämäläinen JA, Näätänen R, Astikainen P (2016) Passive
715 exposure to speech sounds induces long-term memory representations in the auditory
716 cortex of adult rats. *Scientific reports*, 6:38904. doi: 10.1038/srep38904.
- 717 Näätänen R, Schröger E, Karakas S, Tervaniemi M, Paavilainen P (1993) Development of
718 a memory trace for a complex sound in the human brain. *NeuroReport* 4(5):503–506.
- 719 Näätänen R, et al. (1997) Language-specific phoneme representations revealed by electric
720 and magnetic brain responses. *Nature*, 385:432. doi: 10.1038/385432a0.
- 721 Patel SH, Azzam PN (2005) Characterization of N200 and P300: Selected studies of the
722 Event-Related Potential. *Int J Med Sci* 2:147–154. doi:10.7150/ijms.2.147.
- 723 Polich J (2007) Updating P300: An integrative theory of P3a and P3b. *Clin Neurophysiol*
724 118:2128–2148. doi: 10.1016/j.clinph.2007.04.019.
- 725 Reinke KS, He Y, Wang C, Alain C (2003) Perceptual learning modulates sensory evoked
726 response during vowel segregation. *Cogn Brain Res* 17:781-791. doi:
727 10.1016/S0926-6410(03)00202-7.
- 728 Ross B, Jamali S, Tremblay KL (2013) Plasticity in neuromagnetic cortical responses
729 suggests enhanced auditory object representation. *BMC Neurosci* 14:151. doi:
730 10.1186/1471-2202-14-151.
- 731 Sams M, Alho K, Näätänen R. (1983) Sequential effects on the erp in discriminating two
732 stimuli. *Biol Psychol* 17:41–58. doi: 10.1016/0301-0511(83)90065-0.
- 733 Seppänen M, Pesonen AK, Tervaniemi M (2012) Music training enhances the rapid
734 plasticity of P3a/P3b event-related brain potentials for unattended and attended target
735 sounds. *Atten Percept Psychophys* 74:600–612. doi: 10.3758/s13414-011-0257-9.
- 736 Sheehan KA, McArthur GM, Bishop DVM (2005) Is discrimination training necessary to
737 cause changes in the P2 auditory event-related brain potential to speech sounds?
738 *Cogn Brain Res* 25:547–553. doi: 10.1016/j.cogbrainres.2005.08.007.
- 739 Stickgold R. (2005) Sleep-dependent memory consolidation. *Nature* 437:1272-1278. doi:
740 10.1038/nature04286.

- 741 Tamminen H, Peltola MS, Kujala T, Näätänen R. (2015) Phonetic training and non-native
742 speech perception—New memory traces evolve in just three days as indexed by the
743 mismatch negativity (MMN) and behavioural measures. *Int J Psychophysiol*, 97: 23-
744 29. doi: 10.1016/j.ijpsycho.2015.04.020.
- 745 Tremblay K (2007) Training-related changes in the brain: Evidence from human auditory-
746 evoked potentials. *Semin Hear* 28(2):120–132. doi: 10.1055/s-2007-973438.
- 747 Tremblay K, Kraus N, Carrell TD, McGee T (1997) Central auditory system plasticity:
748 Generalization to novel stimuli following listening training. *J Acoust Soc Am*
749 102(6):3762–3773. doi: 10.1121/1.420139.
- 750 Tremblay K, Kraus N, McGee T (1998) The time course of auditory perceptual learning:
751 neurophysiological changes during speech-sound training. *Neuroreport* 9:3557–
752 3560. doi: 10.1097/00001756-199811160-00003.
- 753 Tremblay KL, Inoue K, McClannahan K, Ross B (2010) Repeated stimulus exposure alters
754 the way sound is encoded in the human brain. *PloS One* 5:e10283. doi:
755 10.1371/journal.pone.0010283.
- 756 Uther M, Kujala A, Huotilainen M, Shtyrov Y, Näätänen R (2006). Training in Morse
757 code enhances involuntary attentional switching to acoustic frequency: Evidence
758 from ERPs. *Brain Res* 1073:417–424. doi: 10.1016/j.brainres.2005.12.047.
- 759 Winkler et al. (1999) Brain responses reveal the learning of foreign language phonemes.
760 *Psychophysiology*, 36:638-642. doi: /10.1111/1469-8986.3650638.
- 761 Wright BA, Baese-Berk MM, Marrone N, Bradlow AR (2015) Enhancing speech learning
762 by combining task practice with periods of stimulus exposure without practice. *J.*
763 *Acoust. Soc. Am.*, 138:928-937. doi: 10.1121/1.4927411.
- 764 Wright BA, Sabin AT, Zhang Y, Marrone N, Fitzgerald MB (2010). Enhancing perceptual
765 learning by combining practice with periods of additional sensory stimulation. *J*
766 *Neurosci*, 30:12868–12877. doi: 10.1523/JNEUROSCI.0487-10.2010.
- 767 Wright BA, Zhang Y (2009) A review of the generalization of auditory learning. *Philos*
768 *Trans R Soc Lond B Biol Sci*, 364:301–311. doi: 10.1098/rstb.2008.0262.
- 769 Xi J, Zhang L, Shu H, Zhang Y, Li P (2010) Categorical perception of lexical tones in
770 Chinese revealed by mismatch negativity. *Neuroscience* 170:223–231. doi:
771 10.1016/j.neuroscience.2010.06.077.
- 772 Zhang L, Xi J, Wu H, Shu H, Li P (2012) Electrophysiological evidence of categorical
773 perception of Chinese lexical tones in attentive condition. *Neuroreport* 23:35–9. doi:
774 10.1097/WNR.0b013e32834e4842.