Journal of
Cognitive Neuroscience

Auditory attention causes gain enhancement and frequency sharpening at successive stages of cortical processing – evidence from human EEG

SCHOLARONE™
Manuscripts

1 Title: **Auditory attention causes gain enhancement and frequency sharpening at successive**

2 **stages of cortical processing – evidence from human EEG**

3

4 Running head: Successive gain enhancement and sharpening effects

5

6 Authors: Jessica de Boer, Katrin Krumbholz

7 Affiliation: MRC Institute of Hearing Research at the School of Medicine, University of

8 Nottingham.

9

10 Corresponding author: Jessica de Boer, MRC Institute of Hearing Research at the School of

11 Medicine,

12 University of Nottingham, Science Road, University Park, Nottingham, NG7 2RD.

13 E-mail:Jessica.deBoer@nottingham.ac.uk

14

17

18 ABSTRACT

19 Previous findings have suggested that auditory attention causes not only enhancement in

20 neural processing gain, but also sharpening in neural frequency tuning in human auditory

21 cortex. The current study was aimed to reexamine these findings, and investigate whether

22 attentional gain enhancement and frequency sharpening emerge at the same or different

23 processing levels, and whether they represent independent or cooperative effects. For that,

24 we examined the pattern of attentional modulation effects on early, sensory-driven cortical

25 auditory-evoked potentials (CAEPs) occurring at different latencies. Attention was

26 manipulated using a dichotic listening task and was thus not selectively directed to specific

27 frequency values. Possible attention-related changes in frequency tuning selectivity were

28 measured with an EEG adaptation paradigm. Our results show marked disparities in

29 attention effects between the earlier N1 CAEP deflection and the subsequent P2 deflection,

30 with the N1 showing a strong gain enhancement effect, but no sharpening, and the P2

31 showing clear evidence of sharpening, but no independent gain effect. They suggest that

32 gain enhancement and frequency sharpening represent successive stages of a cooperative

33   attentional modulation mechanism, which appears to increase the representational

34   bandwidth of attended versus unattended sounds.

35

36   INTRODUCTION

37   There is manifold evidence that attention causes top-down modulation of sensory-driven, or

38   "exogenous", cortical responses (e.g., Fritz, Shamma, Elhilali, & Klein, 2003; Hillyard, Hink,

39   Schwent, & Picton, 1973; Spitzer, Desimone, & Moran, 1988; reviewed in: Fritz, Elhilali,

40   David, & Shamma, 2007; Mangun & Hillyard, 1995), but the mechanisms underlying this

41   modulation still remain unclear. Two alternative models have been proposed: the "gain

42   enhancement" model assumes that attention increases neuronal responsiveness to the

43   attended stimulus (Hillyard, Vogel, & Luck, 1998; McAdams & Maunsell, 1999), and the

44   "sharpening" model, that attention increases neuronal tuning selectivity (Atiani, Elhilali,

45   David, Fritz, & Shamma, 2009; Fritz et al., 2003; Spitzer et al., 1988). The current study

46   aimed to test whether or how these models apply to the auditory domain. In particular, we

47   wanted to test (i) whether exogenous auditory responses are really affected by attentional

48   sharpening, and, if so, (ii) how gain enhancement and sharpening relate within the context

49   of the auditory processing hierarchy: do they occur at the same or different processing

50   levels, and do they operate cooperatively or independently of one another?

51   Numerous earlier studies have found non-invasively recorded auditory cortical

52   responses to be larger when the evoking sound is attended, rather than unattended

53   (EEG/MEG: Fujiwara, Nagamine, Imai, Tanaka, & Shibasaki, 1998; Hillyard et al., 1973;

54   Hillyard et al., 1998; Woldorff et al., 1993; Woldorff & Hillyard, 1991; fMRI: Jäncke,

55   Mirzazade, & Shah, 1999), and have generally interpreted this finding within the context of a

56   gain enhancement mechanism. More recently, however, it has been suggested that auditory

57   attentional modulation also involves sharpening (Ahveninen et al., 2011; Kauramaki,

58   Jääskeläinen, & Sams, 2007; Okamoto, Stracke, Wolters, Schmael, & Pantev, 2007). To

59   demonstrate sharpening, the previous studies have used paradigms involving "notched

60   noise" (NN) masking, a technique that has been used extensively in behavioral

61   measurements of auditory frequency selectivity (e.g., Glasberg & Moore, 1990). NN masking

62   requires the subject to attend to a fixed-frequency tone, whilst trying to ignore a

63   concurrently presented broadband noise with a spectral notch centered on the tone

64   frequency. When the notch is narrow enough so that the tone response is partially

obscured, or "masked", by the noise response, the size of the unobscured portion of the tone response (over and above the noise response) should depend on the tuning selectivity of the tone-responsive neurons (Sams & Salmelin, 1994), and should thus be sensitive to any sharpening in tuning selectivity induced by attention. Consistent with this expectation, the previous studies have found greater attentional enhancement of the tone response size when the notch was narrower than when it was wider (Kauramaki et al., 2007; Okamoto et al., 2007), or when the masking noise was omitted altogether (Ahveninen et al., 2011). Arguably, however, this finding could also be explained in terms of gain enhancement. This is, because the tone was presented at a fixed intensity and would thus have been less audible when presented in a narrow-notched noise. As a result, the unattended tone response size would have been smaller, and the attentional task would have been more difficult to perform. Earlier findings (Alho, Woods, Algazi, & Näätänen, 1992; Boudreau, Williford, & Maunsell, 2006; Schwent, Hillyard, & Galambos, 1976a, 1976b) suggest that both factors should have led to greater attentional gain enhancement, thus mimicking the effect of attentional sharpening.

To avoid these confounds, the current study manipulated attention and measured tuning selectivity independently using dichotic listening and adaptation, respectively. Tone or noise sequences were presented concurrently to opposite ears ("Ipsi" and "Contra" in Fig. 1A) and subjects were asked to alternately attend to one or other sequence. Cortical auditory-evoked potentials (CAEPs) were recorded in response to the tone sequences, and the tone frequency was varied randomly from trial to trial to vary the degree of adaptation between successive tones. Adaptation refers to the suppression in neuronal response when the same or similar stimulus is presented repeatedly (hence also referred to as "repetition suppression"; Grill-Spector, Henson, & Martin, 2006). Adaptation is ubiquitous across many sensory domains and has become a popular tool for probing functional properties of neuronal populations, particularly in the visual domain (reviewed in Snow, Coen-Cagli, & Schwartz, 2017; Webster, 2015), but to a lesser degree also in the auditory domain (e.g., Briley, Breakey, & Krumbholz, 2013; Edmonds & Krumbholz, 2014; Hewson-Stoate, Schonwiesner, & Krumbholz, 2006; Magezi & Krumbholz, 2010; Salminen, May, Alku, & Tiitinen, 2009). Under the assumption that adaptation is caused by neuronal fatigue (mediated by synaptic depression or somatic after-hyperpolarization; Briley & Krumbholz, 2013; Lanting, Briley, Sumner, & Krumbholz, 2013), the degree of adaptation between two

97   successive tones should depend on the degree of overlap between the neuron populations

98   responsive to the tones, and this, in turn, should depend on the neurons' frequency tuning

99   selectivity. Figure 1B shows predictions of how the adapted tone response sizes might be

100   affected by attentional gain enhancement and sharpening effects. The predictions are based

101   on a simple neuron population model, with model neurons tuned for frequency and subject

102   to activity-dependent adaptation, or fatigue (see Methods for model details). Due to

103   adaptation, the aggregate population response size to the current tone is predicted to

104   increase with increasing frequency separation of the preceding tone, regardless of attention

105   condition (right panels in Fig. 1B). Under the assumption of a pure gain enhancement

106   mechanism (with multiplicative gain; top row in Fig. 1B), attention is predicted to increase

107   the population response size equally across all frequency separations (if response size is

108   expressed in logarithmic units), leaving the shape of the response size function unchanged.

109   In contrast, a pure sharpening mechanism (middle row in Fig. 1B) is predicted to increase

110   the initial slope of the response size function (at small frequency separations), but also, to

111   cause an overall suppression in response size across all frequency separations. The

112   suppression arises, because, as the neurons' tuning selectivity increases, fewer neurons are

113   activated and thus the aggregate population response size decreases. In order to avoid

114   suppression, the sharpening has to be combined with a gain enhancement such that the

115   aggregate response size remains constant (Fig. 1B, bottom row). As a result, the initial slope

116   of the response size function is again predicted to steepen, but the response size now

117   remains unchanged at zero and large frequency separations (when the responses to the

118   successive tones overlap either completely or not at all; see left and middle panels in Fig.

119   1B).

120   ***insert Fig 1 about here***

121   The previous studies that have used NN masking to investigate auditory attentional

122   modulation mechanisms (Ahveninen et al., 2011; Kauramaki et al., 2007; Okamoto et al.,

123   2007) have focused exclusively on the prominent N1 deflection of the CAEPs (Näätänen &

124   Picton, 1987). Here, we also examined the preceding and following P1 and P2 deflections,

125   which, like the N1, are exogenous, and thus presumably represent earlier and later stages of

126   sensory-driven auditory processing. Our results suggest that gain enhancement and

127   sharpening represent cooperative components of a hierarchically distributed auditory

128   attentional modulation mechanism, affecting different sensory-driven processing levels: the

4

129   earliest observed attention effects (in the N1) appeared to be pure gain enhancement

130   effects, whilst sharpening effects appeared to emerge only at later processing levels (in the

131   P2). Our results suggest that gain enhancement and sharpening might work together to

132   increase the representational bandwidth, or "data rate", of attended over unattended

133   auditory information.

134

135   MATERIALS AND METHODS

136   **Participants**

137   23 subjects (7 male; mean age = 23.1, SD = 3.8 years) participated after having given written

138   informed consent. All subjects had hearing thresholds at or below 20 dB HL at all

139   audiometric frequencies (250-8000 Hz), and had no history of audiological or neurological

140   disease. The experimental procedures accorded with the Declaration of Helsinki (Version 6,

141   2008) and were approved by the Ethics Committee of the University of Nottingham School

142   of Psychology, but were not formally pre-registered online in accordance with the 2014

143   amendment to the declaration.

144   **Stimuli and procedure**

145   During the EEG experiment, subjects were comfortably seated in an electrically shielded,

146   sound-attenuating booth (IAC Acoustics, Winchester, United Kingdom). The experiment

147   consisted of four runs with short breaks in between. In three runs, referred to as "active

148   runs", subjects were required to alternately attend to tone or noise sequences, presented to

149   opposite ears, and detect infrequent targets within the attended ear. The to-be-attended

150   ear was indicated by visual instruction and was switched every ~2 min. The ear of

151   presentation of the tone and noise sequences was counterbalanced across subjects. The

152   active runs lasted about 12 min each. In the remaining run, referred to as "passive run", the

153   stimuli were presented passively whilst the subjects watched a silent sub-titled movie of

154   their own choice to remain alert. The duration of the passive run was matched to the total

155   duration for which subjects attended to each ear over the three active runs (i.e., 3×6 min =

156   18 min). The active and passive runs were played consecutively, in counterbalanced order

157   across subjects.

158   The tones ("Ipsi" in Fig. 1A) had a duration of 100 ms, including 20-ms cosine-

159   squared onset and offset ramps, and were presented at a fixed stimulus onset interval (SOI)

160   of 500 ms. A fixed SOI was used, because varying it would have varied the degree of

5

161 adaptation between successive tones (Lanting et al., 2013) and thus confounded the tuning

162 selectivity measurement. The tone frequencies were distributed equally between four

163 different values, which were by 0, 75, 150 and 300 cents higher than 1000 Hz (1000, 1044,

164 1091, 1189 Hz). The tone sequences were pseudo-random de Bruijn sequences consisting of

165 256 items each (lasting ~2 min). They were designed such that not only each frequency

166 individually, but also each possible combination of two, three or four consecutive

167 frequencies occurred an exactly equal number of times (64, 16, 4 and 1, respectively;

168 Brimijoin & O'Neill, 2010).

169   The noise stimuli ("Contra" in Fig. 1A) were generated from equally exciting noise

170 (with equal energy falling in each auditory filter; Glasberg & Moore, 2000), which was box-

171 car filtered between 2000 and 3000 Hz. They had a duration of 200 ms, and were amplitude-

172 modulated with a waxing amplitude envelope consisting of linear onset and offset ramps

173 lasting 150 and 50 ms, respectively. The SOI of the noises was randomized between 666 and

174 966 ms (mean: 816 ms) to decorrelate the onset times of the tones and noises across the

175 two ears.

176   The tone targets were distinguished from the non-target tones by a linearly rising

177 frequency ramp (the non-target tones had a steady frequency; right panel in Fig. 1A). They

178 were presented randomly with a probability of 7.5%, with the constraint that every two

179 successive target tones were separated by at least four non-target tones. The noise targets

180 were time-reversed versions of the non-target noises (non-targets were waxing, and targets

181 were waning noises; Fig. 1A; idea taken from Cusack, Deeks, Aikman, & Carlyon, 2004). They

182 were presented with a probability of 10% and separated by at least two non-target noises.

183 On average, both the tone and noise targets occurred about 20 times within each ~2-min

184 period (targets were presented within both the attended and unattended sequences).

185   All stimuli were generated digitally using Matlab (The Mathworks, Natick, MA, USA)

186 and digital-to-analogue converted with a 24.414-kHz sampling rate and 24-bit amplitude

187 resolution using TDT System 3 (Tucker Davis Technologies, Alachua, FL, USA) consisting of an

188 RP2.1 real-time processor and an HB7 headphone buffer. Both the tone and noise stimuli

189 were presented at a sound pressure level (SPL) of 65 dB using Sennheiser HD-280 Pro

190 circumaural headphones (Sennheiser, Wedemark, Germany).

191 **EEG recordings**

6

192 CAEPs were recorded with 33 Ag/AgCl ring electrodes (EASYCAP, Herrsching, Germany),

193 placed according to the standard 10-20 layout, and a BrainAmp DC EEG amplifier (Brain

194 Products, Gilching, Germany). Skin-to-electrode impedances were maintained below 5 kΩ

195 throughout the recordings. The recording reference was the vertex (Cz) channel and the

196 ground was placed on the central forehead (AFz). The electrode signals were sampled at 500

197 Hz and bandpass-filtered online between 0.1 and 250 Hz using BrainVision Recorder (Brain

198 Products). Only the responses to the non-target tones were analyzed further.

199 **EEG data analysis**

200 The EEG data were first pre-processed using the EEGLAB toolbox (Delorme & Makeig, 2004),

201 which runs under Matlab. First, they were lowpass filtered at 35 Hz using a −48-dB/oct zero-

202 phase IIR filter, and then they were re-referenced to average reference and segmented into

203 500-ms epochs ranging from 100 ms before, to 400 ms after the onsets of the non-target

204 tones. Epochs containing unusually large amplitudes across electrodes (joint probability

205 larger or equal to three standard deviations) were rejected automatically. The remaining

206 epochs were submitted to an independent component analysis (extended infomax

207 algorithm). Components representing eye blinks, lateral eye movements and electro-cardiac

208 activity were removed by manual inspection of the components' temporal traces and scalp

209 topographies.

210 Activity during the baseline period of the tone responses (before the tone onset) was

211 both highly non-stationary and also considerably larger for attended than unattended trials

212 (Fig. 2A), suggesting the presence of longer-lasting endogenous activity from preceding trials

213 (Woldorff, 1993). To minimize the effect of this activity on the analysis of the discernible

214 exogenous deflections (P1, N1 and P2; Fig. 2A), we baseline-corrected each deflection

215 separately, using a different baseline window (referred to as "deflection-specific" baseline

216 correction). All windows were given a minimal duration of only 8 ms. The windows for the

217 N1 and P2 were centered at the peaks of the respective preceding, opposite-polarity

218 deflections (P1 and N1, respectively), thus effectively creating a peak-to-peak difference.

219 This would be expected to minimize any unipolar activity associated with endogenous

220 attentional processing, such as the so-called "processing negativity" (Näätänen, 1990),

221 which would affect opposite-polarity deflections in opposite directions, and thus cancel in

222 the peak-to-peak difference. The window for the P1 was located at the tone onset (around 0

223 ms), close to the P1 deflection start. The baseline correction was performed separately for

224   each subject and analyzed condition. The baseline-corrected deflections will be referred to

225   as $P1_0$, $N1_{P1}$ and $P2_{N1}$ to indicate the differences in baseline window (see Fig. 2B).

226   The $P1_0$, $N1_{P1}$ and $P2_{N1}$ peak amplitudes were measured both from the original

227   sensor data, and also from source waveforms derived from source models fitted to each

228   deflection peak. The sensor data were evaluated at the sensors that showed the largest

229   unattended deflection peaks on average (Fz for the $P1_0$ and $N1_{P1}$, and Cz for the $P2_{N1}$) and

230   referenced to the linked mastoids (average of TP9 and TP10). The source models were fitted

231   to the unattended conditions only (when subjects attended to the noise sequences in the

232   opposite ear or watched a silent movie) to create a spatial filter for exogenous auditory

233   cortical activity. They were implemented in the Brain Electrical Source Analysis software,

234   version 5.3 (BESA, Gräfelfing, Germany), and each consisted of two hemispherically

235   symmetric regional equivalent current dipoles (ECDs; Scherg & Ebersole, 1993), with a four-

236   shell ellipsoidal volume conductor as head model. First, the ECD locations were fitted to a

237   30-ms window centered at the relevant deflection peak in the grand-average response

238   across all subjects and unattended conditions. Then, the ECDs were then re-oriented

239   individually for each subject to maximize the peak source strength along their first dipole

240   direction, and the resulting reoriented first dipole directions were used to extract source

241   waveforms for each individual and condition. The source waveforms showed no significant

242   hemispheric differences, and were thus averaged across hemispheres.

243   The $P1_0$, $N1_{P1}$ and $P2_{N1}$ peak amplitudes were either averaged across all tone

244   frequencies, or evaluated separately for each absolute frequency separation, ΔF, between

245   the current and preceding tones, which could take one of four values (0, 75, 150 or 300

246   cents). On average, the number of trials available for each absolute frequency separation

247   and each subject was 391 (range: 347-414), 479 (409-517), 481 (428-507) and 241 (208-258)

248   when subjects attended to the tone sequences, 397 (374-419), 490 (463-512), 491 (465-508)

249   and 245 (231-260) when they attended to the noise sequences, and 380 (330-406), 469

250   (421-499), 467 (412-506) and 236 (212-248) when they watched a silent movie (passive

251   run).

252   **Statistical analyses**

253   Statistical analyses were conducted using *R* (R Core Team, 2013). Both the behavioral

254   (hit/false-alarm rates and reaction times for target detection) and CAEP data (deflection

255   peak amplitudes) were evaluated with linear mixed-effects models (nlme package; Pinheiro,

256   Bates, DebRoy, Sarkar, & Team, 2017). The CAEP peak amplitudes were first converted to

257   logarithmic units.

258   Homogeneity of variance was tested using Levene's test (car package; Fox &

259   Weisberg, 2011), and normality using quantile-quantile plots of the model residuals. Where

260   variance homogeneity was violated (i.e., the residuals were significantly different across

261   factor levels), each observation was weighted by the inverse of the variance for the

262   respective factor level. This reduces the influence of noisier data points on the model fit.

263   Normality was achieved by log-transformation (applied to the false-alarm rates and reaction

264   times). Any overly influential data points were identified using Cook's distance and

265   excluded.

266   In the models of the CAEP peak amplitudes, the linear frequency separation

267   covariate (ΔF) was shifted downwards by 150 cents (ΔF $\rightarrow$ ΔF $-$ 150 cents) to reduce

268   collinearity with the quadratic covariate (ΔF^2). Next to the fixed effects, all models also

269   contained by-subjects random intercepts and fixed-factor slopes. The fixed effects were

270   fitted using maximum likelihood (ML) estimation, and the random effects using restricted

271   ML (REML) estimation. Random effects were tested using log-likelihood ratio tests. Random

272   effects that failed to produce a significant improvement in model fit were omitted. Fixed

273   effects were evaluated using conditional F-tests following the strategy described in Pinheiro

274   and Bates (2000). Despite some missing data points, the number of data points were

275   sufficiently similar across the various combinations of factor levels to allow type-III

276   (marginal) tests to be evaluated for all included fixed effects. Significant fixed effects were

277   post-hoc tested using Tukey's honestly significant difference (multcomp package; Hothorn,

278   Bretz, & Westfall, 2008).

279   **Neuron population model of attentional modulation effects**

280   Predicted effects of gain enhancement and sharpening were derived with a neuron

281   population model, where each neuron was tuned to a different characteristic frequency, f_c,

282   and subject to activity-dependent adaptation or fatigue, A. The shape of the frequency

283   tuning was defined by a rounded-exponential (roex)

284   function, $W(f) = \left(1 - p\frac{|f_c - f|}{f_c}\right)e^{-p\frac{|f_c - f|}{f_c}}$ (Eq. 1), where f is the frequency of the tone

285   stimulus, $|f_c - f|$ is its absolute separation from the neuron's characteristic frequency (f_c),

286   and p is a parameter that determines the tuning sharpness. p was set such that tuning curve

287 widths corresponded to the equivalent rectangular bandwidths (ERBs) of the auditory

288 frequency filters as determined by behavioral NN masking: $p = \frac{4f_c}{ERB(f_c)}$, where

289 $ERB(f_c) = 24.7(4.37f_c + 1)$ (Glasberg & Moore, 1990). The characteristic frequencies (f_c)

290 varied from 50 to 16000 Hz and were distributed evenly on a cochlear frequency (ERB-rate;

291 Moore & Glasberg, 1983) scale.

292 Adaptation was modelled by multiplying the unadapted response to the current tone

293 frequency, f_0 [given by $W(f_0)$; Eq. 1] with a factor 1-A, where A was proportional to the

294 response to the preceding tone frequency, f_{-1} [given by $W(f_{-1})$]. The degree of adaptation (A)

295 was assumed to decay exponentially over time (t): $A(t) = A(t = 0)e^{-t/\tau}$. The decay time

296 constant, τ, was set to 721.34 ms (compare Briley & Krumbholz, 2013; Roth et al., 1976),

297 which meant that, between successive tone onsets, adaptation decayed by 50% (because

298 $e^{-500/721.34} = 0.5$). The aggregate response size was derived by summing the adapted

299 single-neuron responses across neurons.

300 Attentional gain enhancement was modelled by multiplying the single-neuron tuning

301 functions W (Eq. 1) with a gain factor, $G > 1$. In the simulation shown in Fig. 1B (upper row),

302 G was set to 2 – doubling the attended compared to unattended response size. Attentional

303 sharpening was modelled by dividing the tuning sharpness parameter, p, by a sharpening

304 factor, $S < 1$. In the simulations shown in Fig 1B (middle and bottom rows), S was set to 0.5 –

305 halving the ERBs of the attended compared to unattended tuning functions. If no gain is

306 applied ($G = 1$), halving the ERBs halves the aggregate response sizes (middle row). In order

307 to preserve the aggregate response size (bottom row), G was concurrently raised to 2.

308

309 RESULTS

310 **Behavioral results**

311 During the EEG recordings, subjects either ignored the experimental sounds and watched a

312 silent subtitled movie, or alternately monitored the tone or noise sequences in the different

313 ears for occasional target sounds (frequency-modulated tones and waning noises,

314 respectively; Fig. 1A). In order to match the difficulty in detecting the tone and noise targets,

315 each subject first attended a short pilot session, where the target salience (determined by

316 the frequency or amplitude modulation depth, respectively) was adjusted to yield a ~75%

317 hit rate. Across subjects, the adjusted frequency modulation depth of the tone targets

318 ranged between 100-200 cents, and the amplitude modulation depth of the noise targets

319 ranged between 50-100%.

320   During the experiment proper, the tone targets yielded an actual hit rate close to the

321 adjusted rate (mean $\pm$ standard error: 76.0 $\pm$ 3.1%), and a false alarm rate of 10.0 $\pm$ 2.8 %. In

322 contrast, the actual hit rate for the noise targets was significantly higher [85.0 $\pm$ 2.8%;

323 $F(1,105) = 11.3$, p = 0.0011; here and onwards, statistical tests are based on linear mixed-

324 effects models (LMMs), with F and p values based on conditional F-tests; see Methods], and

325 the false-alarm rate significantly lower [4.4 $\pm$ 1.3%; $F(1,105) = 5.8$ p = 0.0180]. At the same

326 time, however, the noise targets also yielded a longer reaction time [613.6 $\pm$ 26.3 ms vs

327 566.0 $\pm$ 18.0 ms for the tone targets; $F(1,104) = 5.7$ p = 0.0185], suggesting that subjects

328 traded response speed with response accuracy. In the case of the tone sequences, the scope

329 for such speed-accuracy trade-off was limited by the shorter SOI (500 ms vs 816 $\pm$ 150 ms

330 for the noise sequences; see Methods), which limited the reaction time. The presence of

331 speed-accuracy trade-off is supported by the inverse efficiency score (IES), which combines

332 response speed and accuracy measures into a single, overall measure of task performance

333 [IES = RT/(1-PE), where RT is the reaction time and PE is the proportion of errors, i.e., false

334 alarms and missed targets; Townsend & Ashby, 1978], and which was not significantly

335 different between the tone and noise sequences [tones vs noises: 814.4 $\pm$ 124.1 ms vs 829.5

336 $\pm$ 72.2 ms; $F(1,103) = 0.12$ p = 0.7159]. The IES was also not significantly different across the

337 three successive ~12-min measurement runs ["active runs"; main effect of run: $F(1,103) =$

338 0.63, p = 0.4274; interaction between run and sequence type: $F(1,103) = 0.28$, p = 0.5949].

339 **Average CAEPs**

340 The average CAEPs to the non-target tones (averaged across all frequency separations

341 between successive tones; Fig. 2A) exhibited three successive transient deflections, P1, N1

342 and P2, which were clearly discernible, and peaked at similar latencies (around 60, 105 and

343 150 ms), both when the tones were attended (top panel in Fig. 2A), and when they were

344 unattended (i.e., when subjects attended to the noise sequences in the other ear or

345 watched a silent movie; bottom panel in Fig. 2A). Due to the relatively short SOI used (500

346 ms), the CAEPs failed to return to a steady baseline before the subsequent tone onset. As a

347 result, the transient deflections were riding on a background of slowly-varying non-

348 stationary EEG activity from previous trials, which appeared to be particularly evident in the

349 attended condition (Fig. 2A, top). The non-stationarity of this background activity meant

350 that it could not be eliminated by conventional baseline correction, and the use of a fixed

351 SOI (required to control the degree of adaptation between successive tones) meant that it

352 could also not be eliminated by deconvolution-based methods (Lütkenhöner, 2010;

353 Woldorff, 1993). To address this problem, we here opted to baseline-correct each deflection

354 separately, using a baseline window that was both minimal in duration and located close to

355 the respective deflection start (deflection-specific baselining; see Methods). The N1 and P2

356 were baseline-corrected to the respective preceding, opposite-polarity peak – effectively

357 creating a peak-to-peak difference. This would have minimized both the slowly-varying

358 previous-trial baseline, as well as any unipolar endogenous attention-related activity elicited

359 within the current trial (such as the processing negativity; Näätänen, 1990). The baseline-

360 corrected deflections are shown in Fig. 2B (separately for each attention condition) and will

361 be referred to as $P1_0$, $N1_{P1}$ and $P2_{N1}$. Figure 3A shows that they exhibited scalp voltage

362 distributions typical of sources in supratemporal auditory cortex (characterized by a voltage

363 inversion over the temporal bone; Scherg, Vajsar, & Picton, 1989; Vaughan & Ritter, 1970).

364 ***insert Fig 2 about here***

365   CAEPs measured at individual sensors may reflect a mixture of contributions from

366 both exogenous and endogenous sources, but only the exogenous contributions represent

367 the *modulatory* attention effects that we aim to investigate. Thus, to maximize these

368 contributions, we analyzed the CAEPs not only in the original sensor space (using the

369 sensors that showed the largest unattended peak amplitude for the respective analyzed

370 deflection; see Methods and Fig. 2B), but also in a source space representing exogenous

371 sources. A different source model was used for each subject and analyzed deflection, based

372 on equivalent dipoles fitted to the respective deflection peak in the individual *unattended*

373 responses (where endogenous contributions should have been minimal; see Methods).

374 Figure 3B shows that the best-fitting sources for all three unattended deflections ($P1_0$, $N1_{P1}$

375 and $P2_{N1}$) localized to the approximate auditory cortex region, and that their average

376 orientations were roughly perpendicular to the supratemporal plane. The goodness of fit

377 ranged between 89.4 and 98.5% for the P1 [mean $\pm$ standard deviation: 96.4 $\pm$ 2.4],

378 between 93.7 and 98.5% for the N1 (97.1 $\pm$ 1.4), and between 88.3 and 98.3% for the P2

379 (95.6 $\pm$ 2.2). The sources were used as spatial filters to extract source waveforms for each

380 individual and condition (see Fig. 3C for the grand-average source waveforms for each

381   attention condition), and the source waveforms were averaged across hemispheres,

382   because no significant hemisphere-specific condition effects were found.

383   ***insert Fig 3 about here***

384   **Attention effects on average CAEPs**

385   Comparison of the average CAEP waveforms between attention conditions (see Figs 2B & 3C

386   for the sensor and source waveforms, respectively) suggests that the $N1_{P1}$, and, to a lesser

387   degree, also the $P2_{N1}$, were enhanced when the tones were attended than when they were

388   unattended, whereas the $P1_0$ seemed to be largely unaffected by attention. The waveforms

389   also suggest that there was little difference between the two unattended conditions (i.e.,

390   when subjects attended to the noise sequences or watched a silent movie, labelled

391   "ignored" and "passive" in Figs 2B & 3C) – for any of the three deflections.

392   These results were confirmed by submitting the average deflection peak amplitudes

393   (in logarithmic units; Fig. 4A) to linear mixed-effects statistical models (LMMs), with

394   attention condition and deflection (if appropriate) as fixed factors. The models were

395   calculated either for successive deflection pairs [$P1_0/N1_{P1}$ and $N1_{P1}/P2_{N1,}$] or for each

396   deflection separately (henceforth referred to as "combined" or "separate LMMs"). Effects

397   that were significant in the current, but not the preceding deflection were interpreted as

398   "emerging" at the level of the current deflection. Both for the sensor, and for the source

399   data, the combined LMM of the $P1_0/N1_{P1}$ peak amplitudes revealed a significant overall

400   (main) effect of attention condition [sensor: $F_{(2,107)} = 3.8$, $p = 0.0254$; source: $F_{(2,105)} =$

401   6.3, $p = 0.0026$], but also showed a significant deflection by attention condition interaction

402   [sensor: $F_{(2,107)} = 15.5$, $p < 0.0001$; source: $F_{(2,105)} = 11.1$, $p < 0.0001$]. The interaction

403   arose, because the attention condition effect was significant only for the $N1_{P1}$ [shown by the

404   respective separate LMMs; sensor: $F_{(2,42)} = 33.0$, $p < 0.001$; source: $F_{(2,42)} = 28.7$, $p <$

405   0.0001], but non-significant for the $P1_0$ [sensor: $F_{(2,43)} = 0.8$, $p = 0.4717$; source: $F_{(2,41)} =$

406   0.4, $p = 0.6672$]. This suggests that the attention condition effect first emerged at the level

407   of the N1. In the combined LMM of the $N1_{P1}/P2_{N1}$ peak amplitudes, the main effect of

408   attention condition was again significant for both the sensor and the source data [sensor:

409     $F_{(2,107)} = 13.3$, $p < 0.001$; source: $F_{(2,105)} = 11.5$, $p < 0.0001$]. In this case, the deflection by

410   attention condition interaction was significant for the source data [$F_{(2,105)} = 5.7$, $p =$

411   0.0045], but non-significant for the sensor data [$F_{(2,107)} = 0.7$, $p = 0.5079$]. Consistent with

412   this, the separate LMM for the $P2_{N1}$ showed a significant attention condition effect for the

413   sensor data [$F(2,42) = 4.1$,   p= 0.0233], but not for the source data [$F(2,41) = 2.12$, p =

414   0.1274]. This suggests that the attention condition effect on the average $P2_{N1}$ peak

415   amplitudes was more labile than for the $N1_{P1}$. For the $N1_{P1}$, the attention condition effect

416   was due to larger peak amplitudes in the attended compared to both unattended (ignored

417   and passive) conditions. This was true for both the sensor (both $p \leq 0.0001$) and source data

418   (both $p < 0.0001$). For the $P2_{N1}$ sensor amplitudes, the difference between the attended and

419   ignored conditions was significant ($p = 0.00971$), but the difference between the attended

420   and the passive conditions was non-significant ($p = 0.22429$; see stars in Fig. 4A). The

421   ignored and passive conditions showed little or no differences between one another – for

422   any deflection and in either the sensor or source data (all $p > 0.4$).

423   **Attention effects on frequency-specific adaptation**

424   To test whether the observed attention effects on the average deflection peak amplitudes

425   were generated by gain enhancement or sharpening, we evaluated the peak amplitudes

426   separately for the different frequency separations, ΔF, from the preceding tone, which were

427   expected to cause different degrees of adaptation (Fig. 1B). In the statistical models (LMMs)

428   frequency separation was included both as a linear (ΔF) and quadratic (ΔF^2) fixed covariate,

429   because, based on the neuron population model predictions (Fig. 1B), the linear covariate

430   alone was not expected to be able to capture the effect of sharpening. For gain

431   enhancement, the model predicted a constant increase in the response size across all

432   frequency separations from the preceding tone. Statistically, this should create a main effect

433   of attention condition, with no interaction with either frequency separation covariate (ΔF or

434   ΔF^2). In contrast, the sharpening mechanism was predicted to cause the response size

435   function to become steeper at small frequency separations, thus making the function more

436   non-linear. Statistically, this should give rise to a significant interaction between ΔF^2 and

437   attention condition. The average peak amplitudes had shown no significant differences

438   between the ignored and passive conditions for any deflection (see Fig. 4A) and the same

439   was also true for the peak amplitudes as a function of frequency separation (Fig. 4B&C).

440   Therefore, the ignored and passive conditions were now merged to a form a single

441   "unattended" condition.

442   ***insert Fig 4 about here***

14

443   The $N1_{P1}$ and $P2_{N1}$ peak amplitudes increased with increasing frequency separation (Fig.

444   4B&C), as predicted by the neuron population model (compare Fig. 1B). The corresponding

445   (separate) LMMs revealed that this increase was significant for both the sensor [main effect

446   of ΔF; $N1_{P1}$: $F(1,241) = 13.8$, $p = 0.003$; $P2_{N1}$: $F(1,243) = 15.2$, $p = 0.0001$] and source data

447   [$N1_{P1}$: $F(1,241) = 5.3$, $p = 0.0221$; $P2_{N1}$: $F(1,243) = 17.9$, $p < 0.0001$]. These results indicate

448   that the N1 and P2 were affected by frequency-specific adaptation. In contrast, the peak

449   amplitudes for the $P1_0$ showed little or no change with frequency separation, for either the

450   sensor [main effects of ΔF and ΔF^2; both $F(1,235) \leq 0.5$, $p \geq 0.4788$] or source data [both

451    $F(1,236) \leq 0.8$, $p \geq 0.3751$], suggesting that the P1 was either not adapted, or that

452   adaptation in the P1 was non-specific to frequency.

453   Figure 4B&C (middle panels) suggests that attention increased the $N1_{P1}$ peak amplitudes

454   about equally across all frequency separations. This finding was statistically confirmed by

455   the non-significance of the interactions between attention condition and both ΔF and ΔF^2 in

456   the separate LMM for the $N1_{P1}$, which applied to both the sensor [both $F(1,241) \leq 1.3$, $p \geq$

457   0.2558] and source data [both $F(1,241) \leq 2.5$, $p \geq 0.1164$], and is consistent with the neuron

458   population model predictions for gain enhancement (compare upper right panel in Fig. 1B).

459   In contrast, the attention effect on the $P2_{N1}$ peak amplitudes depended strongly on

460   frequency separation, with little or no increase at the zero and largest frequency

461   separations (0 and 300 cents; 100 cents correspond to 1 semitone), but large increases at

462   the intervening frequency separations (75 and 150 cents; rightmost panels in Fig. 4B&C).

463   This pattern is consistent with the neuron population model predictions for sharpening

464   combined with a commensurate gain enhancement to preserve the aggregate response size

465   (compare bottom right panel in Fig. 1B). Statistically, it was confirmed by the significance of

466   the interaction between attention condition and ΔF^2 in the separate LMM for the $P2_{N1}$,

467   which, again, applied to both the sensor [$F(1,243) = 11.0$, $p = 0.001$] and source data

468   [$F(1,243) = 5.0$, $p = 0.0264$]. The interaction between attention condition and ΔF was non-

469   significant [sensor: $F(1,243) = 0.4$, $p = 0.5292$; source: $F(1,243) = 0.2$, $p = 0.6822$].

470   The difference in the pattern of frequency separation-dependent attention effects

471   between the $N1_{P1}$ and $P2_{N1}$ was statistically confirmed by the three-way interaction

472   between deflection, attention condition and ΔF^2 in the corresponding combined LMM

473   ($N1_{P1}/P2_{N1}$). This interaction, which was significant in the sensor data [$F(1,506) = 8.2$, $p =$

15

474   0.0045] and approached significance in the source data [F(1,506) = 3.2, p = 0.0764],

475   suggesting that sharpening emerges only at the level of the P2. In contrast to the $N1_{P1}$ and

476    $P2_{N1}$, the $P1_0$ peak amplitudes showed no significant attention effects, at any frequency

477   separation, as confirmed by the lack of significant interactions between attention condition

478   and both ΔF or ΔF^2 in the separate LLMs for the $P1_0$ [sensor: both $F(1,235) \leq 0.1$, $p \geq 0.7699$;

479   source: both $F(1,236) \leq 1.4$, $p \geq 0.2427$].

480

481   DISCUSSION

482   The current results suggest that the earliest effects of auditory attentional modulation

483   are mediated by a pure gain enhancement mechanism, and that sharpening emerges only at

484   later processing stages. In the current results, the earliest measured deflection, the P1 –

485   presumed to be generated in primary auditory cortex (Liégeois-Chauvel, Musolino, Badier,

486   Marquis, & Chauvel, 1994; Mäkelä, Hämäläinen, Hari, & McEvoy, 1994; Yvert, Crouzeix,

487   Bertrand, Seither-Preisler, & Pantev, 2001) – was little or not affected by attention. The

488   subsequent N1 showed a strong attention-related enhancement in average peak amplitude,

489   but no differential effects on frequency-specific adaptation, suggesting that the N1 was

490   affected by a pure gain enhancement mechanism. In contrast, the latest measured

491   deflection, the P2, showed a lesser enhancement in average peak amplitude, but a marked

492   increase in the degree of adaptation specificity. Predictions from a neuron population model

493   showed that the pattern of the effects in the P2 was consistent with a sharpening in neural

494   tuning selectivity, combined with a commensurate gain enhancement so that the overall

495   response size remained unchanged.

496   These results are consistent with previous studies that have also found large attentional

497   enhancements in N1 peak amplitude (Hillyard et al., 1973; Neelon, Williams, & Garell,

498   2006a, 2006b), but contradict the conclusion of the previous NN masking studies

499   (Ahveninen et al., 2011; Kauramaki et al., 2007; Okamoto et al., 2007) that attentional

500   enhancement of the N1 is caused by neuronal sharpening. In the NN studies, attention was

501   directed to a specific frequency value and the audibility of the attended stimulus was

502   allowed to vary across conditions. As explained above (Introduction) this would likely have

503   led to variation in the amount of attentional gain enhancement, in a way that would have

504   mimicked the expected effect of sharpening (Alho et al., 1992; Boudreau et al., 2006;

505   Schwent et al., 1976a, 1976b). In the current study, attention was directed to one or other

16

506 ear, and stimulus audibility was fixed across conditions. Our results thus suggest that

507 attention can sharpen selectivity for a feature (here, frequency) even when attention is not

508 selectively focused on a specific feature value. A similar conclusion was reached by Murray

509 and Wojciulik (2004), who used an adaptation paradigm to demonstrate attentional

510 sharpening for visual orientation. In both our and Murray and Wojciulik's studies, the

511 feature in which sharpening was observed (frequency and visual orientation, respectively)

512 was task-relevant (in Murray and Wojciulik's study, subjects had to detect a change in image

513 orientation; in our study, they had to detect a small frequency modulation). It is thus

514 possible that task relevance is a prerequisite for sharpening to occur.

515 The absence of significant attention effects in the earliest, P1, deflection in the current

516 study is consistent with several previous studies (Hillyard et al., 1973; Neelon, Williams, &

517 Garell, 2006a, 2006b) that have also found no significant P1 attention effects. Other studies,

518 however, that have used shorter SOIs, did find significant attention effects in the P1, and

519 even earlier, deflections (Woldorff et al., 1993; Woldorff & Hillyard, 1991), suggesting that

520 the first emergence of attention effects is graded with attentional load.

521 The current finding of a small but significant (in the sensor data) attentional

522 enhancement in the average P2 peak amplitude contrasts with some previous CAEP studies

523 that have found either no significant change (Hillyard et al., 1973) or even a reduction

524 (Hansen & Hillyard, 1980) in the P2 amplitude as a result of attention. The reduction has

525 been attributed to a separate unipolar deflection, termed the "processing negativity" or

526 "Nd", thought to reflect endogenous attention-related processes (Näätänen, 1990). Due to

527 its negative polarity, the Nd would be expected to add to any modulatory enhancement of

528 the N1, but diminish any enhancement of the P2. In the current study, this effect would

529 have been minimized by the deflection-specific baselining procedure used (see Methods).

530 Significant attentional enhancement of the P2 has also been found in intracranial recordings

531 from the auditory temporal region (Neelon et al., 2006a, 2006b), where any influence of the

532 Nd may also have been minimal. The Nd can be demonstrated by calculating the difference

533 wave between attended and unattended responses. In the current study, this was precluded

534 by the experimental design: Difference waves can only be meaningfully calculated when the

535 previous-trial baseline activity in the attended and unattended responses is either the same

536 on average (e.g., Hansen & Hillyard, 1983; Hillyard & Münte, 1984), or can be effectively

537 corrected for (e.g., Woldorff & Hillyard, 1991; Woldorff, 1993). In the current study,

538   attended and unattended trials were temporally separated into different blocks, and so,

539   attended trials were always preceded by attended trials, and unattended trials were always

540   preceded by unattended trials. As a result, the attended responses exhibited a substantially

541   larger previous-trial baseline, on average, than the unattended responses. Correcting for the

542   baseline was also not possible, as this requires a sufficiently variable SOI (Lütkenhöner,

543   2010; Woldorff, 1993). In the current study, the SOI had to be fixed to control the degree of

544   adaptation between successive trials.

545   The N1 and P2 have often been viewed as part of the same component process (the so-

546   called "N1-P2 complex"). However, the marked differences in the pattern of their observed

547   attention effects suggests that, rather than representing a unitary complex, the N1 and P2

548   represent different hierarchical levels of exogenous auditory processing that play distinct

549   functional roles in conscious sound perception. This is supported by previous findings

550   showing that the N1 and P2 differ not only in source structure (Godey, Schwartz, de Graaf,

551   Chauvel, & Liégeois-Chauvel, 2001; Hari, Kaila, Katila, Tuomisto, & Varpula, 1982; Hari et al.,

552   1987; Lütkenhöner & Steinsträter, 1998), but also in functional properties, such as

553   dependence on prior stimulation, general arousal, aging and auditory training (Crowley &

554   Colrain, 2004; Herrmann, Henry, Johnsrude, & Obleser, 2016; Ross, Jamali, & Tremblay,

555   2013; Ross & Tremblay, 2009; Roth et al., 1976; Tremblay, Ross, Inoue, McClannahan, &

556   Collet, 2014).

557   The effect of attention on adaptation, or "repetition suppression", has been investigated

558   by several previous studies – particularly in the visual domain and using fMRI (see Henson &

559   Mouchlianitis, 2007, for review). The results from these studies, however, have been mixed,

560   with some studies finding similar repetition suppression in both attended and unattended

561   conditions (Bentley, Vuilleumier, Thiel, Driver, & Dolan, 2003; Vuilleumier, Schwartz,

562   Duhoux, Dolan, & Driver, 2005), but others finding repetition suppression to be either

563   reduced (Murray & Wojciulik, 2004) or absent in unattended conditions (Eger, Henson,

564   Driver, & Dolan, 2004; Henson & Mouchlianitis, 2007; Yi, Kelley, Marois, & Chun, 2006; Yi,

565   Woodman, Widders, Marois, & Chun, 2004). The previous studies compared responses to

566   repeated versus different stimuli, but, unlike the current study, did not vary the degree of

567   stimulus difference. The current results suggest that the amount of unattended repetition

568   suppression should depend on the relation between the degree of stimulus difference and

569   neuronal tuning selectivity: if we had compared repeated versus different tones with only a

570 single frequency separation, we would have observed similar attended and unattended

571 repetition suppression, if the frequency separation had been greater than 150 cents, but

572 reduced or absent unattended repetition suppression if the frequency separation had been

573 equal to or smaller than 150 cents (see Fig. 4B&C).

574 Previous studies from the visual (Summerfield, Trittschuh, Monti, Mesulam, & Egner,

575 2008; Summerfield, Wyart, Johnen, & de Gardelle, 2011) and auditory (Todorovic, van Ede,

576 Maris, & de Lange, 2011; Wacongne et al., 2011) domains have demonstrated that

577 repetition suppression is not only determined by the local stimulus context (locally

578 preceding stimuli), but is also modulated by prior expectation, such that the amount of

579 repetition suppression is reduced when stimulus repetition is unexpected. This is contrary to

580 the idea of bottom-up neuronal fatigue, and has been taken to suggest that repetition

581 suppression may instead reflect the action of a hierarchical predictive coding mechanism,

582 which combines bottom-up stimulus representations with prior, top-down stimulus

583 expectations (e.g., Friston, 2005; Knill & Pouget, 2004). Within this predictive coding

584 framework, it has been hypothesized that attention may modulate the top-down stimulus

585 expectations – increasing expectation for attended over unattended stimuli (Friston, 2009;

586 Rao, 2005). Several recent studies have interpreted their findings within the context of this

587 hypothesis (Chennu et al., 2013; Hsu, Hämäläinen, & Waszak, 2014; Kok, Rahnev, Jehee,

588 Lau, & de Lange, 2012). The current study, however, suggests an alternative, or at least

589 complementary, explanation. This is, because all stimuli and all stimulus transitions

590 (including higher-order transitions between non-consecutive stimuli) were perfectly

591 balanced (see Methods), and thus presumably equally expected – and attention was also

592 distributed equally across all stimuli. This excludes an explanation in terms of top-down

593 expectation, and instead suggests that attention modulates bottom-up representational

594 properties.

595 The P2 amplitude showed little or no attention-related change when the frequency

596 separation from the preceding tone was either zero or larger. According to the neuron

597 population model predictions, this suggest that the P2 was affected by a combination of

598 sharpening and gain enhancement, and that the amount of gain enhancement matched the

599 degree of sharpening, such that the overall response size remained unchanged. This

600 suggests that gain enhancement and sharpening are distinct but cooperative components of

601 a hierarchically distributed attentional modulation mechanism, which adaptively adjusts the

602 representational bandwidth of auditory cortical processing in accordance with attentional

603 demand. Sharpening increases representational resolution, but, without a commensurate

604 enhancement in gain, this would lead to decrease in representational accuracy (because

605 fewer channels would be activated, or each channel would be activated less strongly). By

606 combining and matching gain enhancement and sharpening effects, the auditory system can

607 increase representational resolution whilst, at the same time, maintaining representational

608 accuracy. And by cascading the gain enhancement and sharpening effects across different

609 processing levels – presumably with different limitations on representational resources

610 (Ahissar & Hochstein, 2004), the system retains the ability to quickly switch attention to

611 new, or currently unattended, sounds.

612

613 ACKNOWLEDGEMENTS

REFERENCES

617 Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning.
618   *Trends in Cognitive Sciences, 8*(10), 457-464.
619 Ahveninen, J., Hämäläinen, M., Jääskeläinen, I. P., Ahlfors, S. P., Huang, S., Lin, F. H., et al. (2011).
620   Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from
621   noise. *Proceedings of the National Academy of Sciences of the United States of America,*
622   *108*(10), 4182-4187.
623 Alho, K., Woods, D. L., Algazi, A., & Näätänen, R. (1992). Intermodal selective attention. II. Effects of
624   attentional load on processing of auditory and visual stimuli in central space.
625   *Electroencephalography and Clinical Neurophysiology, 82*(5), 356-368.
626 Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Task difficulty and performance
627   induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive
628   fields. *Neuron, 61*(3), 467-480.
629 Bentley, P., Vuilleumier, P., Thiel, C. M., Driver, J., & Dolan, R. J. (2003). Effects of attention and
630   emotion on repetition priming and their modulation by cholinergic enhancement. *Journal of*
631   *Neurophysiology, 90*(2), 1171-1181.
632 Boudreau, C. E., Williford, T. H., & Maunsell, J. H. (2006). Effects of task difficulty and target
633   likelihood in area V4 of macaque monkeys. *Journal of Neurophysiology, 96*(5), 2377-2387.
634 Briley, P. M., Breakey, C., & Krumbholz, K. (2013). Evidence for pitch chroma mapping in human
635   auditory cortex. *Cerebral Cortex, 23*(11), 2601-2610.
636 Briley, P. M., & Krumbholz, K. (2013). The specificity of stimulus-specific adaptation in human
637   auditory cortex increases with repeated exposure to the adapting stimulus. *Journal of*
638   *Neurophysiology, 110*(12), 2679-2688.

639 Brimijoin, W. O., & O'Neill, W. E. (2010). Patterned tone sequences reveal non-linear interactions in
640 auditory spectrotemporal receptive fields in the inferior colliculus. *Hearing Research, 267*(1-
641 2), 96-110.
642 Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., et al. (2013).
643 Expectation and attention in hierarchical auditory prediction. *The Journal of Neuroscience,*
644 *33*(27), 11194-11205.
645 Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent
646 component process: age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732-744.
647 Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and
648 time course of selective attention on auditory scene analysis. *Journal of Experimental*
649 *Psychology. Human Perception and Performance, 30*(4), 643-656.
650 Edmonds, B. A., & Krumbholz, K. (2014). Are interaural time and level differences represented by
651 independent or integrated codes in the human auditory cortex? *Journal of the Association*
652 *for Research in Otolaryngology, 15*(1), 103-114.
653 Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2004). BOLD repetition decreases in object-
654 responsive ventral visual areas depend on spatial attention. *Journal of Neurophysiology,*
655 *92*(2), 1241-1247.
656 Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second ed.). Los Angeles, CA:
657 Sage.
658 Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of*
659 *London. Series B, Biological Sciences, 360*(1456), 815-836.
660 Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences,*
661 *13*(7), 293-301.
662 Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention--focusing the
663 searchlight on sound. *Current Opinion in Neurobiology, 17*(4), 437-455.
664 Fritz, J. B., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of
665 spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience, 6*(11),
666 1216-1223.
667 Fujiwara, N., Nagamine, T., Imai, M., Tanaka, T., & Shibasaki, H. (1998). Role of the primary auditory
668 cortex in auditory selective attention studied by whole-head neuromagnetometer. *Brain*
669 *Research. Cognitive Brain Research, 7*(2), 99-109.
670 Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise
671 data. *Hearing Research, 47*(1-2), 103-138.
672 Glasberg, B. R., & Moore, B. C. J. (2000). Frequency selectivity as a function of level and frequency
673 measured with uniformly exciting notched noise. *The Journal of the Acoustical Society of*
674 *America, 108*(5 Pt 1), 2318-2328.
675 Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., & Liégeois-Chauvel, C. (2001). Neuromagnetic
676 source localization of auditory evoked fields and intracerebral evoked potentials: a
677 comparison of data in the same patients. *Clinical Neurophysiology, 112*(10), 1850-1859.
678 Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of
679 stimulus-specific effects. *Trends in Cognitive Sciences, 10*(1), 14-23.
680 Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory
681 attention. *Electroencephalography and Clinical Neurophysiology, 49*(3-4), 277-290.
682 Hansen, J. C., & Hillyard, S. A. (1983). Selective attention to multidimensional auditory stimuli.
683 *Journal of Experimental Psychology: Human Perception and Performance, 9*(1), 1-19.
684 Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982). Interstimulus interval dependence of
685 the auditory vertex response and its magnetic counterpart: implications for their neural
686 generation. *Electroencephalography and Clinical Neurophysiology, 54*(5), 561-569.
687 Hari, R., Pelizzone, M., Mäkelä, J. P., Hällström, J., Leinonen, L., & Lounasmaa, O. V. (1987).
688 Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts.
689 *Audiology, 26*(1), 31-43.

690 Henson, R. N., & Mouchlianitis, E. (2007). Effect of spatial attention on stimulus-specific
691   haemodynamic repetition effects. *NeuroImage, 35*(3), 1317-1329.
692 Herrmann, B., Henry, M. J., Johnsrude, I. S., & Obleser, J. (2016). Altered temporal dynamics of
693   neural adaptation in the aging human auditory cortex. *Neurobiology of Aging, 45*, 10-22.
694 Hewson-Stoate, N., Schonwiesner, M., & Krumbholz, K. (2006). Vowel processing evokes a large
695   sustained response anterior to primary auditory cortex. *The European Journal of
696   Neuroscience, 24*(9), 2661-2671.
697 Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with
698   event-related brain potentials. *Attention, Perception, & Psychophysics, 36*(2), 185-198.
699 Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention
700   in the human brain. *Science, 182*(4108), 177-180.
701 Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism
702   of selective attention: electrophysiological and neuroimaging evidence. *Philosophical
703   Transactions of the Royal Society of London. Series B, Biological Sciences, 353*(1373), 1257-
704   1270.
705 Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models.
706   *Biometrical Journal, 50*(3), 346-363.
707 Hsu, Y. F., Hämäläinen, J. A., & Waszak, F. (2014). Repetition suppression comprises both attention-
708   independent and attention-dependent processes. *NeuroImage, 98*, 168-175.
709 Jäncke, L., Mirzazade, S., & Shah, N. J. (1999). Attention modulates activity in the primary and the
710   secondary auditory cortex: a functional magnetic resonance imaging study in human
711   subjects. *Neuroscience Letters, 266*(2), 125-128.
712 Kauramaki, J., Jääskeläinen, I. P., & Sams, M. (2007). Selective attention increases both gain and
713   feature selectivity of the human auditory cortex. *PloS One, 2*(9), e909.
714 Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and
715   computation. *Trends in Neurosciences, 27*(12), 712-719.
716 Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of
717   prediction in silencing sensory signals. *Cerebral Cortex, 22*(9), 2197-2206.
718 Lanting, C. P., Briley, P. M., Sumner, C. J., & Krumbholz, K. (2013). Mechanisms of adaptation in
719   human auditory cortex. *Journal of Neurophysiology, 110*(4), 973-983.
720 Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials
721   recorded from the auditory cortex in man: evaluation and topography of the middle latency
722   components. *Electroencephalography and Clinical Neurophysiology, 92*(3), 204-214.
723 Lütkenhöner, B. (2010). Baseline correction of overlapping event-related responses using a linear
724   deconvolution technique. *NeuroImage, 52*(1), 86-96.
725 Lütkenhöner, B., & Steinsträter, O. (1998). High-precision neuromagnetic study of the functional
726   organization of the human auditory cortex. *Audiology & Neuro-Otology, 3*(2-3), 191-213.
727 Magezi, D. A., & Krumbholz, K. (2010). Evidence for opponent-channel coding of interaural time
728   differences in human auditory cortex. *Journal of Neurophysiology, 104*(4), 1997-2007.
729 Mäkelä, J. P., Hämäläinen, M., Hari, R., & McEvoy, L. (1994). Whole-head mapping of middle-latency
730   auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology,
731   92*(5), 414-421.
732 Mangun, G. R., & Hillyard, S. A. (1995). Mechanisms and models of selective attention. In M. D. Rugg
733   & M. G. H. Coles (Eds.), *Electrophysiology of Mind: Event-related Brain Potentials and
734   Cognition* (pp. 40-85). New York, NY: Oxford University Press.
735 McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of
736   single neurons in macaque cortical area V4. *The Journal of Neuroscience, 19*(1), 431-441.
737 Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter
738   bandwidths and excitation patterns. *The Journal of the Acoustical Society of America, 74*(3),
739   750-753.

740 Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral
741   occipital complex. *Nature Neuroscience, 7*(1), 70-74.
742 Näätänen, R. (1990). The Role of Attention in Auditory Information-Processing as Revealed by Event-
743   Related Potentials and Other Brain Measures of Cognitive Function. *Behavioral and Brain*
744   *Sciences, 13*(2), 201-232.
745 Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to
746   sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-
747   425.
748 Neelon, M. F., Williams, J., & Garell, P. C. (2006a). The effects of attentional load on auditory ERPs
749   recorded from human cortex. *Brain Research, 1118*(1), 94-105.
750 Neelon, M. F., Williams, J., & Garell, P. C. (2006b). The effects of auditory attention measured from
751   human electrocorticograms. *Clinical Neurophysiology, 117*(3), 504-521.
752 Okamoto, H., Stracke, H., Wolters, C. H., Schmael, F., & Pantev, C. (2007). Attention improves
753   population-level frequency tuning in human auditory cortex. *The Journal of Neuroscience,*
754   *27*(39), 10383-10390.
755 Pinheiro, J., & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer
756 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2017). nlme: Linear and Nonlinear Mixed
757   Effects Models. from https://CRAN.R-project.org/package=nlme
758 R Core Team. (2013). A Language and Environment for Statistical Computing. from http://www.R-
759   project.org
760 Rao, R. P. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport,*
761   *16*(16), 1843-1848.
762 Ross, B., Jamali, S., & Tremblay, K. L. (2013). Plasticity in neuromagnetic cortical responses suggests
763   enhanced auditory object representation. *BMC Neuroscience, 14*, 151.
764 Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in
765   young and older listeners. *Hearing Research, 248*(1-2), 48-59.
766 Roth, W. T., Krainz, P. L., Ford, J. M., Tinklenberg, J. R., Rothbart, R. M., & Kopell, B. S. (1976).
767   Parameters of temporal recovery of the human auditory evoked potential.
768   *Electroencephalography and Clinical Neurophysiology, 40*(6), 623-632.
769 Salminen, N. H., May, P. J., Alku, P., & Tiitinen, H. (2009). A population rate code of auditory space in
770   the human cortex. *PloS One, 4*(10), e7600.
771 Sams, M., & Salmelin, R. (1994). Evidence of sharp frequency tuning in the human auditory cortex.
772   *Hearing Research, 75*(1-2), 67-74.
773 Scherg, M., & Ebersole, J. S. (1993). Models of brain sources. *Brain Topography, 5*(4), 419-423.
774 Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked
775   potentials. *Journal of Cognitive Neuroscience, 1*(4), 336-355.
776 Schwent, V. L., Hillyard, S. A., & Galambos, R. (1976a). Selective attention and the auditory vertex
777   potential. Effects of signal intensity and masking noise. *Electroencephalography and Clinical*
778   *Neurophysiology, 40*(6), 615-622.
779 Schwent, V. L., Hillyard, S. A., & Galambos, R. (1976b). Selective attention and the auditory vertex
780   potential. I. Effects of stimulus delivery rate. *Electroencephalography and Clinical*
781   *Neurophysiology, 40*(6), 604-614.
782 Snow, M., Coen-Cagli, R., & Schwartz, O. (2017). Adaptation in the visual cortex: a case for probing
783   neuronal populations with natural stimuli *F1000Research, 6*.
784 Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and
785   neuronal performance. *Science, 240*(4850), 338-340.
786 Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural
787   repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11*(9),
788   1004-1006.

789 Summerfield, C., Wyart, V., Johnen, V. M., & de Gardelle, V. (2011). Human Scalp
790   Electroencephalography Reveals that Repetition Suppression Varies with Expectation.
791   *Frontiers in Human Neuroscience, 5*, 67.
792 Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural
793   adaptation to repeated sounds in the auditory cortex: an MEG study. *The Journal of*
794   *Neuroscience, 31*(25), 9118-9123.
795 Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems.
796   In J. Castellan & F. Restle (Eds.), *Cognitive Theory* (Vol. 3, pp. 199-239). Hillsdate, NJ: Erlbaum
797   Associates.
798 Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2
799   response a biomarker of learning? *Frontiers in Systems Neuroscience, 8*, 28.
800 Vaughan, H. G., Jr., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the
801   human scalp. *Electroencephalography and Clinical Neurophysiology, 28*(4), 360-367.
802 Vuilleumier, P., Schwartz, S., Duhoux, S., Dolan, R. J., & Driver, J. (2005). Selective attention
803   modulates neural substrates of repetition priming and "implicit" visual memory:
804   suppressions and enhancements revealed by FMRI. *Journal of Cognitive Neuroscience, 17*(8),
805   1245-1260.
806 Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011).
807   Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings*
808   *of the National Academy of Sciences of the United States of America, 108*(51), 20754-20759.
809 Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science, 1*, 547-567.
810 Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs:
811   analysis and correction. *Psychophysiology, 30*(1), 98-119.
812 Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993).
813   Modulation of early sensory processing in human auditory cortex during auditory selective
814   attention. *Proceedings of the National Academy of Sciences of the United States of America,*
815   *90*(18), 8722-8726.
816 Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective
817   listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology,*
818   *79*(3), 170-191.
819 Yi, D. J., Kelley, T. A., Marois, R., & Chun, M. M. (2006). Attentional modulation of repetition
820   attenuation is anatomically dissociable for scenes and faces. *Brain Research, 1080*(1), 53-62.
821 Yi, D. J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored
822   stimuli: dissociable effects of perceptual and working memory load. *Nature Neuroscience,*
823   *7*(9), 992-996.
824 Yvert, B., Crouzeix, A., Bertrand, O., Seither-Preisler, A., & Pantev, C. (2001). Multiple supratemporal
825   sources of magnetic and electric auditory evoked middle latency components in humans.
826   *Cerebral Cortex, 11*(5), 411-423.
827

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

Figure 1: Experimental design and neuron population model predictions. (A) Dichotic attention task: one ear received rapid sequences of sine tones varying pseudo-randomly between four different frequencies ("Ipsi"), and the other received sequences of waxing noises ("Contra"). Subjects were instructed to attend to either sequence alternately (~2 min cycle) and detect rare target sounds (shown in red; non-targets are shown in black) in the attended sequence. The tone targets were distinguished by an upward-sweeping frequency trajectory, and the noise targets by a time-reversed (waning) amplitude envelope (see sound examples on the right; tone frequencies and stimulus durations not to scale). (B) Neuron population model predictions of attentional gain enhancement and sharpening (see Methods). The top and middle rows show the predicted effects of gain enhancement and sharpening separately, and the bottom row shows the effect of sharpening combined with a gain increase to counteract suppression in the aggregate response size. The left and middle columns show simulated population activity distributions to the adapting and adapted tones (open and filled outlines), plotted as a function of neuron characteristic frequency in quasi-logarithmic (cochlear frequency) units. The right column shows the aggregate sizes of the adapted responses as a function of the adaptor-probe frequency separation. The unattended responses are shown in gray (replotted across rows), and the attended responses in color.

43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
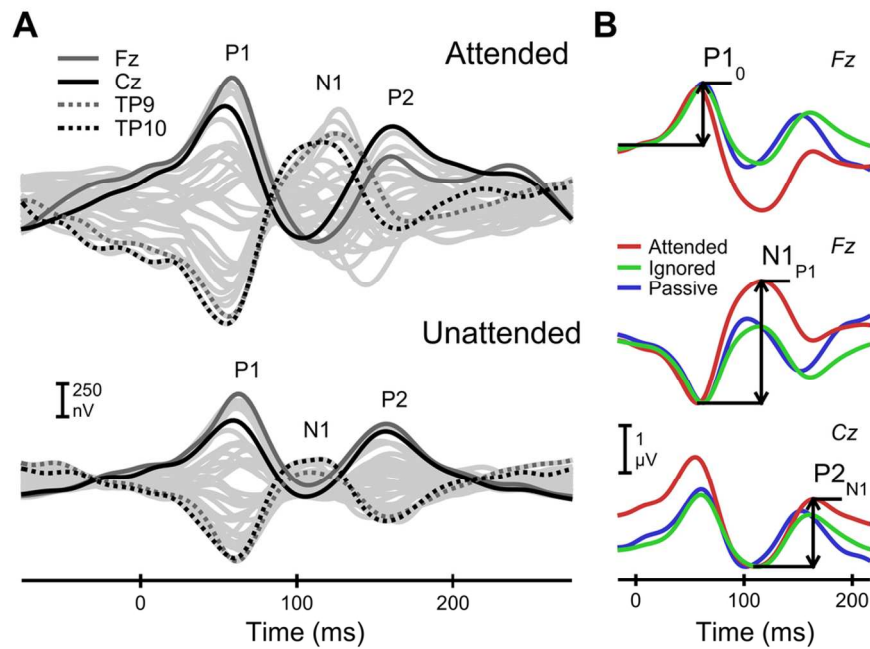58
59
60

91x103mm (300 x 300 DPI)

Figure 2: Cortical auditory-evoked potentials (CAEPs). (A) Grand-average CAEPs across subjects and tone frequencies in the attended (top) and unattended (bottom) conditions, shown as a butterfly plot of all 33 sensors (gray lines). The colored lines highlight the sensors with the largest positive (fronto-central and central; Fz and Cz) and negative (left and right mastoids; TP9 and TP10) deflection amplitudes. No baseline correction was applied to these responses. (B) Baseline-corrected grand-average CAEPs for the P1 (P10; top), N1 (N1P1; middle) and P2 (P2N1; bottom) and each attention condition (attended, ignored and passive; see legend). The waveforms shows the sensors with the largest positive amplitude for each deflection (Fz for P10 and N1P1, Cz for the P2N1), referenced to the average of the mastoid sensors (TP9 and TP10). The vertical arrows show the baseline-corrected peak amplitudes for the attended condition.
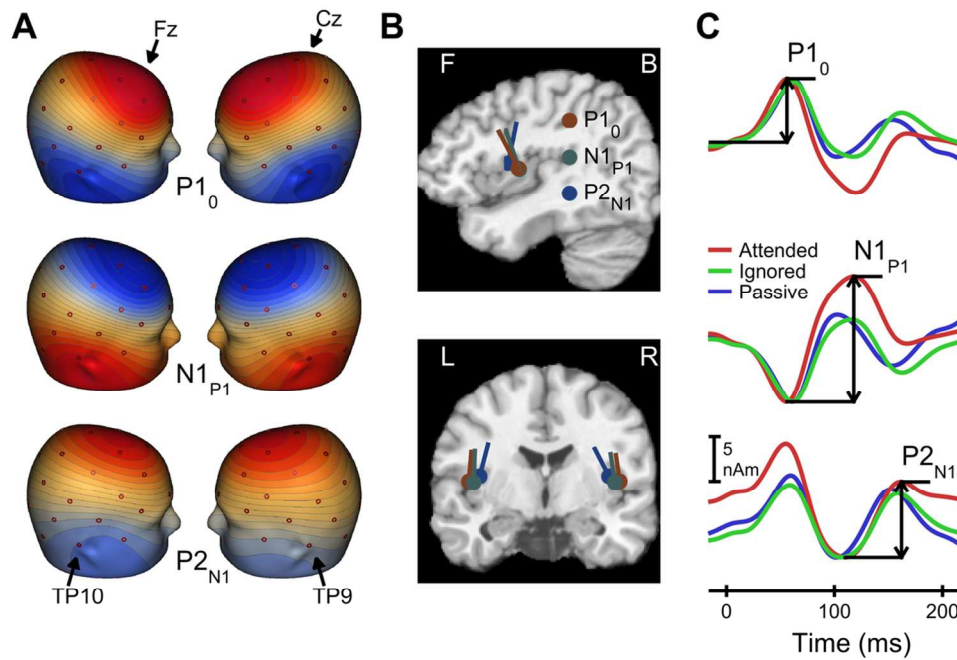
54x37mm (600 x 600 DPI)

Figure 3: CAEP source analysis. (A) Scalp voltage distributions at the peak latencies of the baseline-corrected P10, N1P1 and P2N1 deflections, averaged over all unattended trials (ignored and passive conditions). (B) Source locations of symmetric equivalent current dipole (ECD) pairs, fitted to each deflection peak (see Methods). The dipole locations (dots) and orientations (short lines) are shown on sagittal (top) and coronal (bottom) slices of the single-subject MNI template brain. (C) Grand-average source waveforms for the P10 (top), N1P1 (middle) and P2N1 (bottom) deflections in each attention condition. The vertical arrows show the baseline-corrected peak amplitudes for the attended condition.
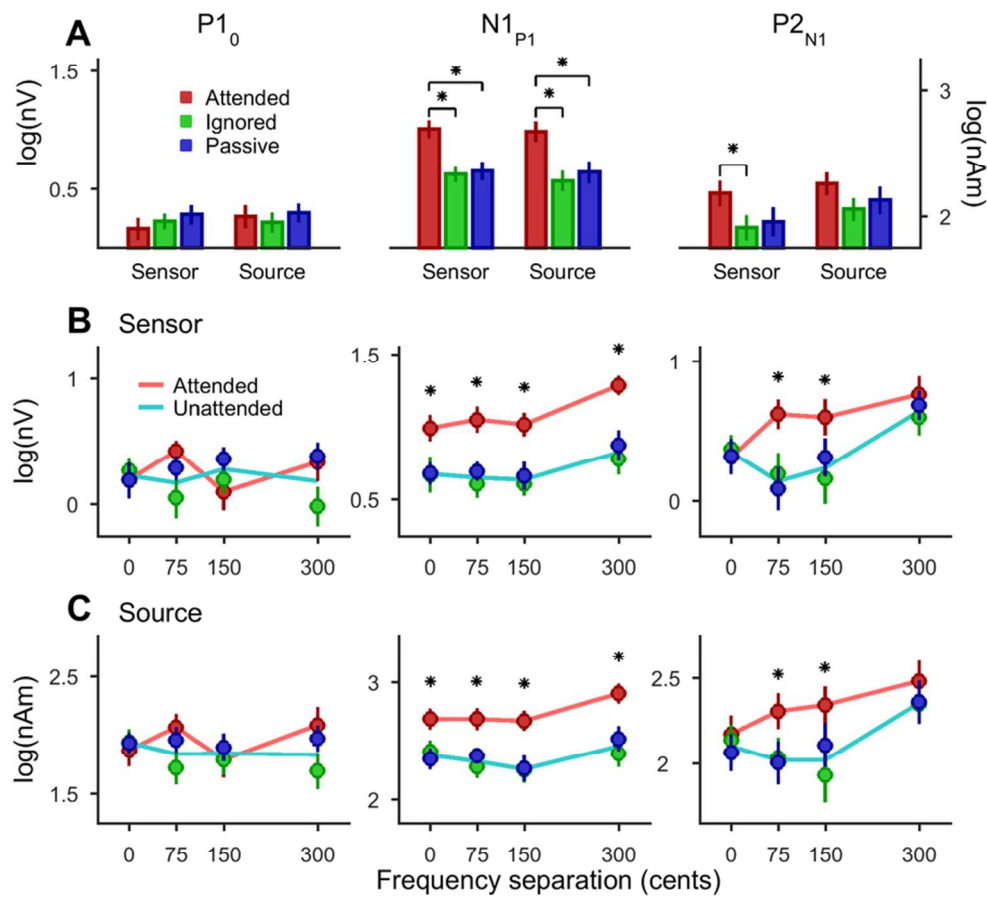
54x37mm (600 x 600 DPI)

Figure 4: Attention effects on CAEP peak amplitudes. (A) Attention effect of the average deflection peak amplitudes. Different attention conditions are indicated by different colors (see legend), and different deflections are shown in different panels. In each panel, the left set of bars shows the results for the sensor data, and the right set of bars shows the results for the source data. (B) Attention effect on the sensor-based deflection peak amplitudes as a function of the frequency separation from the preceding tone (expressed as absolute value in cents, where 100 cents = 1 semitone). The different attention conditions are shown by different colors as in A. (C) Same as in (B), but for the source-based deflection peak amplitudes. In all panels, the mean deflection peak amplitudes are expressed in logarithmic units, and the error bars show the logarithmic standard error of the mean.

78x76mm (300 x 300 DPI)