
Research Articles: Behavioral/Cognitive

Neural decoding of bistable sounds reveals an effect of intention on perceptual organization

Alexander J. Billig¹, Matthew H. Davis¹ and Robert P. Carlyon¹

¹*MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK, CB2 7EF*

DOI: 10.1523/JNEUROSCI.3022-17.2018

Received: 19 October 2017

Revised: 21 January 2018

Accepted: 6 February 2018

Published: 13 February 2018

Author contributions: A.J.B., M.H.D., and R.C. designed research; A.J.B. performed research; A.J.B. analyzed data; A.J.B., M.H.D., and R.C. wrote the paper.

Conflict of Interest: The authors declare no competing financial interests.

This research was supported by a Medical Research Council Doctoral Training Award Studentship for Alexander J. Billig and by Medical Research Council grant number MC-A060-5PQ70 for Robert P. Carlyon. Alexander J. Billig is currently affiliated with UCL Ear Institute, University College London, United Kingdom, and thanks Timothy D. Griffiths and Ingrid S. Johnsrude for continuing financial support while he wrote this paper.

Corresponding Author: Alexander J. Billig, UCL Ear Institute, 332 Gray's Inn Road, London, WC1X 8EE, United Kingdom. Email: ajbillig@gmail.com

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.3022-17.2018

Alerts: Sign up at www.jneurosci.org/cgi/alerts to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2018 Billig et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1 **Neural decoding of bistable sounds reveals an effect of intention on perceptual**
2 **organization**

3

4 Alexander J. Billig¹, Matthew H. Davis¹, and Robert P. Carlyon¹

5 ¹MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK, CB2

6 7EF

7

8

9 Corresponding Author: Alexander J. Billig, UCL Ear Institute, 332 Gray's Inn Road, London,

10 WC1X 8EE, United Kingdom. Email: ajbillig@gmail.com

11

12 Pages: 38 17 Words in Abstract: 250

13 Figures: 4 18 Words in Introduction: 614

14 Tables: 1 19 Words in Discussion: 1391

15 Multimedia items: 0 20

16 3D models: 0 21

22

23 Conflict of Interest: The authors declare no competing financial interests.

24

25 Acknowledgements: This research was supported by a Medical Research Council Doctoral

26 Training Award Studentship for Alexander J. Billig and by Medical Research Council grant

27 number MC-A060-5PQ70 for Robert P. Carlyon. Alexander J. Billig is currently affiliated

28 with UCL Ear Institute, University College London, United Kingdom, and thanks Timothy

29 D. Griffiths and Ingrid S. Johnsrude for continuing financial support while he wrote this

30 paper.

31

Abstract

32

33 Auditory signals arrive at the ear as a mixture that the brain must decompose into
34 distinct sources, based to a large extent on acoustic properties of the sounds. An
35 important question concerns whether listeners have voluntary control over how many
36 sources they perceive. This has been studied using pure tones H and L presented in
37 the repeating pattern HLH-HLH-, which can form a bistable percept, heard either as
38 an integrated whole (HLH-) or as segregated into high (H-H-) and low (-L--)
39 sequences. Although instructing listeners to try to integrate or segregate sounds
40 affects reports of what they hear, this could reflect a response bias rather than a
41 perceptual effect. We had human listeners (15 males, 12 females) continuously report
42 their perception of such sequences and recorded neural activity using magneto-
43 encephalography. During neutral listening, a classifier trained on patterns of neural
44 activity distinguished between periods of integrated and segregated perception. In
45 other conditions, participants tried to influence their perception by allocating attention
46 either to the whole sequence, or to a subset of the sounds. They reported hearing the
47 desired percept for a greater proportion of time than when listening neutrally.
48 Critically, neural activity supported these reports; stimulus-locked brain responses in
49 auditory cortex were more likely to resemble the signature of segregation when
50 participants tried to hear segregation than when attempting to perceive integration.
51 These results indicate that listeners can influence how many sound sources they
52 perceive, as reflected in neural responses that track both the input and its perceptual
53 organization.

54

55

Significance Statement

56

57 Can we consciously influence our perception of the external world? We address this
58 question using sound sequences that can be heard either as coming from a single
59 source, or as two distinct auditory streams. Listeners reported spontaneous changes in
60 their perception between these two interpretations while we recorded neural activity
61 to identify signatures of such integration and segregation. They also indicated that
62 they could, to some extent, choose between these alternatives. This claim was
63 supported by corresponding changes in responses in auditory cortex. By linking
64 neural and behavioral correlates of perception we demonstrate that the number of
65 objects we perceive can depend not only on the physical attributes of our
66 environment, but also on how we intend to experience it.

67

68

Introduction

69

70 For us to make sense of our environment, the brain must determine which elements of
71 energy arriving at the sensory organs arise from the same source and should therefore
72 be perceptually grouped. In audition, the less rapidly that sequential sounds change in
73 one or more physical quantities, such as frequency, intensity, or spatial location, the
74 more likely they are to be integrated and represented as a single perceptual object or
75 stream (Moore & Gockel, 2012; van Noorden, 1975). The processes that underlie
76 integration and segregation are affected not only by these stimulus features but also
77 by internal states of the listener, such as the degree to which they are attending to the
78 sounds (Carlyon et al., 2001; Sussman et al., 2002; Snyder et al., 2006; Billig and
79 Carlyon, 2015), and by whether the stimuli correspond to a familiar speaker
80 (Johnsrude et al., 2013) or word (Billig et al., 2013). The extent to which observers
81 can voluntarily influence how they perceptually organize the outside world is unclear
82 and bears on questions of whether and how higher-level cognition can influence
83 perception (Fodor, 1983; Pylyshyn, 1999; Firestone and Scholl, 2015; Gross, 2017;
84 Lupyan, 2017).

85

86 A common stimulus for investigating auditory perceptual organization is a repeating
87 pattern of pure tones of high (H) and low (L) frequencies, such as that shown in
88 Figure 1A. For lower frequency separations and presentation rates the sounds tend to
89 be heard as integrated in a single stream that forms a distinctive galloping rhythm. At
90 greater frequency separations and presentation rates the H and L tones typically form
91 two segregated streams (van Noorden, 1975). For a range of stimulus parameters,
92 perception can alternate between the two percepts every few seconds, usually after a

93 longer initial integrated phase (Carlyon et al., 2001; Denham et al., 2013; Pressnitzer
94 & Hupé, 2006; Figure 1B).

95

96

[FIGURE 1 ABOUT HERE]

97

98 For such ambiguous sequences, listeners report being able to exert a degree of control
99 over hearing integration or segregation (van Noorden, 1975; Pressnitzer and Hupé,
100 2006; Micheyl and Oxenham, 2010; Farkas et al., 2016). However, subjective
101 responses may be affected by post-perceptual processes and biases, such as shifts in
102 decision criteria (Green and Swets, 1966) and attempts to meet the perceived aims of
103 the experiment (Orne, 1962). To the extent that they vary with a listener's percept,
104 indirect behavioral or neural measures can bypass such issues. For example, several
105 electro- and magneto-encephalography (EEG/MEG) studies have detected more
106 positive auditory cortical responses approximately 60-100 ms following the onset of
107 the middle tone in such triplets during reports of segregation compared to integration
108 (Gutschalk et al., 2005; Hill et al., 2012; Szalárdy et al., 2013a). We argue that these
109 objective neural measures can also shed light on the neural stages of processing that
110 underlie any genuine effect of intention on perception.

111

112 Here we combine subjective and objective measures to demonstrate an effect of
113 intention on perception, reflected in evoked responses in auditory cortex. To do so we
114 measure neural activity with EEG/MEG as participants listen neutrally to HLH-
115 sequences (Figure 1A) and report spontaneous changes in their perception (Figure
116 1B). We derive a univariate marker of perceptual organization in the auditory evoked
117 field at the group level, but also make use of multiple temporal features in the neural

INTENTION AND PERCEPTUAL ORGANIZATION

118 response to train multivariate percept classifiers for each participant. We then study
119 the relative occurrence of these neural signatures when participants actively try to
120 promote integration (by attending to the whole pattern) or segregation (by attending
121 exclusively to tones of one frequency). This allows us to establish whether their
122 reports of successfully influencing their percept are supported by and reflected in
123 stimulus-locked activity in auditory cortex, or are instead more likely to have a post-
124 perceptual locus.
125

126

Materials and Methods

127

Participants

129

130 Data were collected in two separate experimental settings. Twenty-five participants
131 took part in a sound booth (Setting 1), for the purposes of (a) ensuring that stimulus
132 parameters gave rise to integrated and segregated percepts in approximately equal
133 measure, and (b) screening participants before EEG/MEG recording to ensure that
134 they could experience both percepts. Twenty-two of these participants also took part
135 in the EEG/MEG lab (Setting 2), between 1 and 34 days later. Two further
136 participants took part in Setting 2 only, after screening with an online test. All 27
137 participants across both settings were aged 18-40 (mean age = 28.56 years, 12
138 females), right-handed and reported no neurological or developmental disorders. They
139 were recruited from the MRC Cognition and Brain Sciences Unit participant panel or
140 by word of mouth, and were paid for their time. One participant, whose results were
141 not qualitatively different from the remainder of the group, had a threshold of 30 dB
142 HL at 1500 Hz in the left ear. All other participants had normal hearing (<25 dB HL
143 pure tone thresholds over the range of the stimuli, 1000-2000 Hz). All experimental
144 procedures were approved by the Cambridge Psychology Research Ethics Committee.

145

Stimuli

147

148 Sequences of 250 HLH triplets were presented diotically, where H (high) and L (low)
149 were 100-ms pure tones (Figure 1A). The frequency of the H tone was fixed at 1000
150 Hz (Setting 1) or 1017 Hz (Setting 2), except for the final H tone in the final triplet

INTENTION AND PERCEPTUAL ORGANIZATION

151 when it was 250 Hz. The choice of 1017 Hz in Setting 2 was to avoid possible
152 contamination by harmonics of the 50-Hz line noise. The 250-Hz tone was low
153 enough in frequency to be detectable on a low-pass filtered auxiliary channel of the
154 MEG recording set-up in Setting 2, and was included to enable neural recordings to
155 be time-locked to the stimulus. The L tone in a given sequence was lower than that of
156 the H tone by an amount (Δf) of either four or six semitones (both settings). Silent
157 intervals of 50 ms separated tones within a triplet, and silent intervals of 200 ms
158 separated one triplet from the next, giving sequences of 150 s duration. These stimuli
159 were chosen to match Experiment 2 of Gutschalk et al. (2005).

160

161 Filler stimuli lasting a total of 40 s were created to separate experimental sequences
162 from each other. These consisted of 5,005 ms silence, followed by a 100-ms 250-Hz
163 tone (a time-locking signal, with the same purpose as that described in the previous
164 paragraph). This was followed by 1,900 ms of silence, then by 33 pure tones, each of
165 100 ms duration with 50 ms of silence between tones. The frequencies of these tones
166 were selected at random from a log-rectangular distribution from 200-2000 Hz. Their
167 purpose was to interfere with memory of the previous sequence in an effort to
168 minimize context effects, such as those described by Snyder et al. (2009). The filler
169 stimulus continued with 22,945 ms of silence, another 100-ms 250-Hz tone (to warn
170 the participant that the next experimental sequence was about to begin) and a final
171 5,000 ms of silence. All tones in the experimental sequences and filler stimuli
172 included 10-ms linear onset and offset ramps, and were generated digitally at a
173 sample rate of 44100 Hz with 16-bit resolution.

174

175 **Experimental procedures**

INTENTION AND PERCEPTUAL ORGANIZATION

176

177 In Setting 1, participants were seated in a double-walled sound-insulated room and

178 sounds were presented over Sennheiser HD650 headphones at a level of 55 dB SPL.

179 In Setting 2, participants sat under the dewar of a VectorView system (Elekta

180 Neuromag) while MEG and EEG activity was recorded (see “EEG and MEG

181 acquisition and pre-processing” section for details of preparation and recording). In

182 this setting, sounds were presented through tube headphones with silicone inserts at

183 50 dB above the participant’s 1000-Hz pure tone hearing threshold. Using their right

184 hand, participants pressed one computer key (Setting 1) or button box button (Setting

185 2) when hearing an integrated, galloping triplet pattern, and another when hearing the

186 tones segregate into two isochronous sequences (Figure 1B). The screen indicated

187 their most recent response, which corresponded to their current percept. They were

188 told to make a selection as soon as possible after the sequence began, and to make

189 further responses whenever their percept changed. There were four conditions with

190 different instructions. In “Neutral” sequences participants were instructed to let their

191 perception take a natural course. In “Attempt Integration” sequences they tried to

192 promote the integrated percept by attending to the whole pattern. In “Attempt

193 Segregation” sequences they tried to promote the segregated percept by attending

194 either to the H tones (“Attend High”) or the L tones (“Attend Low”).

195

196 The experiment consisted of two (Setting 1) or four (Setting 2) blocks. Each block

197 contained five sequences: two Neutral, one Attempt Integration, one Attend High and

198 one Attend Low. In Setting 1 the order of instruction conditions was the same in both

199 blocks for a given participant; in Setting 2 this order was reversed for the final two

200 blocks. The two Neutral sequences in a block never occurred consecutively, and Δf

INTENTION AND PERCEPTUAL ORGANIZATION

201 alternated between four and six semitones from sequence to sequence. An on-screen
202 message specified the instruction prior to and throughout each trial. Response
203 key/button mapping, and order of instruction and Δf conditions were balanced across
204 participants. Participants relaxed between sequences and took breaks of at least a
205 minute between blocks (while remaining under the dewar in Setting 2). During
206 experimental trials in Setting 2 they were instructed to keep their eyes open and to
207 maintain fixation on a cross in the centre of the screen, or elsewhere if more
208 comfortable, to minimize alpha power and artefacts from eye movements. In Setting
209 2, participants' head positions were checked at the start of each block, and their
210 position adjusted (to minimize loss of MEG signal) if they had dropped by a
211 centimetre or more. Testing lasted approximately 30 min in Setting 1 and 60 min in
212 Setting 2.

213

214 Before the experiment, the concept of streaming was explained using HLH- patterns
215 with Δf of 0, 5 and 12 semitones. Participants practiced reporting their percept while
216 listening neutrally. They were then told that they may be able to influence their
217 percept by attending either to the whole pattern or to one or other sets of tones; these
218 conditions were also practiced. Participants were told that it was far more important to
219 be honest and accurate in their responses than to be successful in their attempts to
220 influence their percept. In Setting 2, practice and experimental blocks occurred after
221 electrode preparation and head position digitization (described in the "Experimental
222 design and statistical analysis - EEG and MEG acquisition and pre-processing"
223 section). The two participants who had not taken part in Setting 1 completed an online
224 training session to familiarize themselves with the stimuli and percept reporting
225 process, and to practice trying to influence their percept. Instructions were repeated in

INTENTION AND PERCEPTUAL ORGANIZATION

226 person immediately prior to the experiment. Those participants who had taken part in
227 Setting 1 more than a week previously also completed the online training as a
228 refresher.

229

230 **EEG and MEG acquisition and pre-processing**

231

232 Magnetic fields were recorded using a VectorView system (Elekta Neuromag) with
233 one magnetometer and two orthogonal planar gradiometers at each of 102 locations.

234 Electric potentials were recorded concurrently using seventy Ag-AgCl sensors
235 arranged in the extended 10-10% configuration, fitted to the scalp using an electrode
236 cap (Easycap) and referenced to an electrode on the nose, with a ground electrode on

237 the right cheek. Head position was continuously monitored using five head position
238 indicator (HPI) coils. Electro-cardiographic (ECG) and horizontal and vertical
239 electro-oculographic (EOG) activity was recorded with three pairs of electrodes. The

240 positions of the EEG sensors, HPI coils and approximately 100 additional head points
241 were digitized with a 3D digitizer (Fastrak Polhemus), relative to three anatomical
242 fiducial points (the nasion and both pre-auricular points). Data were acquired with a

243 sampling rate of 1000 Hz and a high-pass filter of 0.01 Hz. For the magnetometer and
244 gradiometer recordings, the temporal extension of Signal Space Separation in
245 MaxFilter was used to identify bad channels, suppress noise sources, and compensate

246 for head movement. For all sensor types, additional noisy channels were identified
247 and excluded for each participant based on observations during recording and offline
248 visual inspection, as were recording segments containing SQUID jumps, channel

249 pops, and muscle activity. Line noise at 50 Hz and its harmonics was removed using
250 adaptive multitaper regression implemented in the EEGLAB (Delorme and Makeig,

INTENTION AND PERCEPTUAL ORGANIZATION

251 2004; RRID: SCR_007292) plugin CleanLine, after which all activity was
252 downsampled to 250 Hz. Independent components analysis (ICA) was performed in
253 EEGLAB using the Infomax routine (with sub-Gaussian components included) on a
254 version of the data that had been high-pass filtered at 0.5 Hz (6 dB cut-off, 1 Hz
255 transition band, FIR windowed sinc filter) to impose the stationarity assumed by ICA.
256 EEG channels were considerably noisier than magnetometers and gradiometers, and
257 did not improve the quality of the decomposition. They were therefore discarded, and
258 subsequent analyses were restricted to magnetometers and gradiometers only.
259 Components corresponding to eye blinks/movements and cardiac artefacts were
260 identified and projected out of another copy of the data that had been low-pass filtered
261 at 30 Hz (6 dB cut-off, 6.667 Hz transition band, FIR windowed sinc filter) and high-
262 pass filtered at 0.278 Hz (6 dB cut-off, 0.556 Hz transition band, FIR windowed sinc
263 filter). This high-pass filter was selected for reasons explained in the next paragraph.

264

265 The resulting data were divided into 600-ms epochs, each beginning at the start of an
266 HLH- triplet. Epochs beginning less than 1500 ms after a button press (or the start of
267 the sequence) or ending less than 1500 ms before a button press (or the end of the
268 sequence) were excluded from analyses. This minimized neural and muscular activity
269 related to movement, and removed periods around transitions when the reported
270 percept was least likely to be reliable. Baseline correction was not performed due to
271 the repeating nature of the stimulus precluding a sufficient silent period between
272 triplets, which meant that neural responses from one epoch were likely to carry over
273 to the next. Due to the exclusion of epochs close to reported perceptual switches, any
274 such influence should arise solely from triplets with the same (reported) perceptual
275 state, and epoch time can therefore be thought of as circular (see Hill et al. (2012) for

276 a similar approach). The high-pass filter of 0.278 Hz corresponds to a 3600-ms time
277 period, the shortest possible interval between retained epochs corresponding to
278 different perceptual reports. The relatively conservative approach of epoch rejection,
279 necessary for tapping periods that were as perceptually stable as possible, led to a
280 median retention rate of 58% (2900 epochs) per participant, comparable to that in Hill
281 et al. (2012).

282

283 **Dipole fitting**

284

285 Pairs of equivalent current dipoles were fitted to the magnetometer and gradiometer
286 data for each participant separately, using the VB-ECD approach in SPM12 (v6685;
287 RRID: SCR_007037). Reconstructions made use of single shell forward models based
288 on participant-specific T1-weighted structural MRI scans. Sensor positions were
289 projected onto each participant's MRI by minimizing the sum of squared differences
290 between the digitized fiducials and MRI scan fiducials, and between the digitized
291 head shape and the individual scalp mesh. The VB-ECD routine uses a variational
292 Bayes approach to iteratively optimize location and orientation parameters of fitted
293 dipoles. The midpoints of each hemisphere's Heschl's gyrus were used as soft
294 location priors, with no priors for dipole orientation. Fitting was performed separately
295 for magnetometer and gradiometer data, using the mean activity in the 24-ms window
296 centered on the first prominent turning point in the sensor space waveform (peaking
297 40-110 ms after triplet onset), over all epochs. The dipole pair that accounted for the
298 most variance in sensor data out of 20 iterations of the fitting process was selected for
299 each participant and each sensor type. Using these dipoles as spatial filters, further
300 analyses were conducted on the hemisphere-specific source waveforms, and on the

INTENTION AND PERCEPTUAL ORGANIZATION

301 mean waveform across hemispheres. As the polarity of reconstructed waveforms
302 depends on the orientation of the sources with respect to individual anatomy, each
303 participant's source waveforms were inspected and inverted as necessary such that the
304 first prominent turning point (peaking 40-110 ms after triplet onset) was a local
305 maximum. All results were comparable across magnetometers and gradiometers, and
306 are reported for gradiometers only. Fitted dipole pairs accounted for a mean of 91.9%
307 (standard deviation 4.2%) of the variance in the sensor recordings over the fitting
308 window and were located in or close to Heschl's gyrus for all hemispheres (mean
309 MNI coordinates [+/-49 -21 3], standard deviation 6 mm).

310 To verify that our findings were not dependent on the use of location priors in
311 Heschl's gyrus, we performed a separate set of analyses, selecting for each participant
312 the neural component from the ICA that had the maximum back-projected power in
313 the evoked response. Dipoles fitted to these components also had a mean location in
314 Heschl's gyrus, and the reconstructed source waveforms showed qualitatively similar
315 results to those described below. Although for some participants these reconstructed
316 sources were located in regions remote from auditory cortex, their locations were not
317 consistent across participants and not considered further.

318

319 Experimental design and statistical analysis

320

321 Sample size justification

322

323 No published research has used the same approach to test for intention effects with
324 the same stimuli, however effect sizes for two relevant findings can be estimated from
325 previous studies: (a) an intention effect on behavioral streaming measures of $\eta^2=.70$

326 (Pressnitzer & Hupé, 2006) and (b) a percept effect on MEG evoked responses of
327 $\eta^2=.43$ (Gutschalk et al., 2005). Of the two effects, the latter would require the largest
328 sample size to detect, namely 18 participants for 90% power. We tested 24
329 participants in the MEG setting; this allowed for drop-outs and accounted for possible
330 over-estimation of effect sizes due to unreported null findings.

331

332 *Behavioral analyses*

333

334 As there were no significant differences in the mean percentage of segregation
335 reported across the sound booth and EEG/MEG lab settings ($t(15)=1.38$, $p=.189$,
336 $d=0.23$, 95% CI [-0.20 0.04], tested on the 16 participants with no conditions in either
337 setting in which the percentage of segregation was 0 or 100), behavioral data were
338 combined across the two settings. Of the 27 participants, two (both tested only in
339 Setting 1) were excluded from behavioral data analyses. Both had at least one $\Delta f \times$
340 instruction condition with no sequences that met the following criteria: (i) the first
341 reported phase was integrated (ii) at least two completed subsequent phases were
342 reported. These criteria were necessary to allow separate analysis and comparison of
343 the duration of initial-integrated, subsequent-integrated and segregated phases.
344 Percentages of segregation for all remaining participants were logit-transformed and
345 phase durations were log-transformed before being submitted to repeated-measures
346 ANOVAs for analysis as a function of Δf and instruction. These transformations
347 typically produced data with normally distributed residuals. When this was not the
348 case, non-parametric tests were also conducted; these gave rise to the same qualitative
349 pattern of results and are not reported separately. Mean percentages/durations were
350 calculated on the transformed scale, then converted back to percentages/seconds for

INTENTION AND PERCEPTUAL ORGANIZATION

351 reporting. Null hypothesis significance testing was applied, with an alpha value of
352 .05. Degrees of freedom were adjusted for asphericity as appropriate using the
353 Huynh-Feldt correction (uncorrected degrees of freedom are reported for clarity).

354

355 *Univariate neural analyses*

356

357 Epochs in the Neutral condition were averaged for each combination of Δf and
358 reported percept, for each participant. To maximize power, epochs occurring prior to
359 the first percept report were labelled as integrated, and the first integrated phase of a
360 sequence was not considered separately from the remaining integrated phases. The
361 exclusion of any epochs before the first segregated report (in line with some
362 researchers' suggestions to treat these separately (Denham et al., 2013)) led to
363 qualitatively similar results. One participant, who had only five valid epochs in one Δf
364 x percept cell in the Neutral condition, was excluded from subsequent analyses of
365 neural data. All other participants had at least 55 valid epochs per Δf x percept cell in
366 the Neutral condition (the mean across participants of number of epochs in smallest
367 cell was 145). To assess neural activity as a function of percept without stimulus
368 confounds, the timecourses of the two Δf conditions were averaged within each
369 percept before statistical analysis. Percept differences were similarly partialled out of
370 analyses of neural activity as a function of Δf .

371

372 Statistical differences between percepts and frequency separations in the Neutral
373 condition were assessed using a cluster-based permutation method (Maris and
374 Oostenveld, 2007). Within-participants *t*-tests were conducted at each timepoint, and
375 the largest contiguous cluster of values all exceeding a critical *t*-value (corresponding

INTENTION AND PERCEPTUAL ORGANIZATION

376 to an alpha value of .05) was selected for further analysis. Cluster significance was
377 assessed by comparison to a null distribution generated by randomly permuting the
378 labels of condition averages 1000 times within each participant, and using an alpha
379 value of .05.

380

381 Epochs in the Attempt Integration, Attempt Segregation (Attend High), and Attempt
382 Segregation (Attend Low) conditions were averaged within each Δf , without regard to
383 reported percept. Differences over the temporal cluster of interest from the Neutral
384 condition were derived for each of the following contrasts: (a) $\frac{1}{2} * \text{Attempt}$
385 $\text{Segregation (Attend High)} + \frac{1}{2} * \text{Attempt Segregation (Attend Low)} - \text{Attempt}$
386 Integration , (b) $\text{Attempt Segregation (Attend High)} - \text{Attempt Integration}$, (c) Attempt
387 $\text{Segregation (Attend Low)} - \text{Attempt Integration}$, (d) $\text{Attempt Segregation (Attend}$
388 $\text{High)} - \text{Attempt Segregation (Attend Low)}$. In all cases, the two Δf conditions were
389 given equal weight. Paired t -tests were conducted on these differences.

390

391 To test whether effects of intention on univariate neural responses in the non-Neutral
392 conditions were as large as would be expected based on perceptual reports, and under
393 the assumption that the neural signature of percept in the Neutral condition also
394 applied in the non-Neutral conditions, the following calculations were made. The
395 percentages of each percept reported in each non-Neutral condition for each Δf and
396 participant were applied to the relevant mean neural response from the Neutral
397 condition. Simulated and observed values were compared using a paired t -test, with
398 an alpha value of .05.

399

400 *Multivariate neural analyses*

401

402 Epochs were labelled and participants excluded as outlined in the “Experimental
403 design and statistical analysis - Univariate neural analyses” section. Support Vector
404 Machines (SVMs) with linear kernels were trained to classify integrated versus
405 segregated epochs in the Neutral condition for each Δf and participant, using an
406 adapted version of the DDTBOX package (Bode et al., 2017; RRID: SCR_015978) in
407 MATLAB (RRID: SCR_001622). To ensure that the classifiers were unbiased,
408 random sub-sampling within each SVM was used to match the number of epochs
409 across classes. Five-fold cross-validation was applied, and the subsampling and cross-
410 validation process was repeated 100 times. Features were the standardized values of
411 the neural response at the 150 sampled time points of each 600-ms epoch (arising
412 from the 250 Hz sampling frequency), and the cost parameter (C) was set as 1.
413 Classifier performance in the Neutral condition was assessed for each participant by
414 comparing classified versus actual labels and averaging the percent correct over the
415 $5 \times 100 = 500$ iterations, and over Δf conditions. Group classification accuracy was
416 tested against the 50% chance level using a t -test with an alpha value of .05. Feature
417 weights were obtained from the SVM training functions, and corrected using the
418 method of Haufe et al. (2014), which removes strongly weighted but theoretically
419 irrelevant noise features. These were normalized across participants then averaged
420 over Δf for plotting. The 500 trained SVMs for each Δf and participant were also used
421 to classify all epochs in the non-Neutral conditions, regardless of percept report. The
422 percentage classified as segregated was compared across non-Neutral conditions
423 using within-participants t -tests with an alpha value of .05, for the same contrasts as
424 outlined in the “Univariate neural analyses” section.

425

INTENTION AND PERCEPTUAL ORGANIZATION

426 To test whether task-related differences in the percentage of epochs classified as
427 segregated was as high as would be expected based on subjective reports, it was
428 necessary to take into account the accuracy of the trained classifiers in the Neutral
429 condition. The percentage of reports of segregation for each participant, frequency
430 separation, and task was multiplied by $(\text{Neutral classification accuracy} - 50)/50$ (i.e.
431 the Neutral classification accuracy above chance, as a proportion from -1 to 1). The
432 expected task-related difference in the percentage of epochs classified as segregated
433 was derived for each participant and frequency separation, by taking the average of
434 these adjusted percentages of segregated reports over the two Attempt Segregation
435 conditions and subtracting the adjusted percentage of segregated reports in the
436 Attempt Integration condition. These expected difference values were then averaged
437 over frequency separations, and compared to the observed differences using a paired
438 t -test, with an alpha value of .05.
439

440

Results

441

Behavioral results

443

444 As shown in Figure 2A, segregation was reported for a greater proportion of time for
445 the larger than for the smaller Δf for all tasks. This arose from a combination of
446 shorter initial integrated phases (Figure 2B), shorter subsequent integrated phases
447 (Figure 2C), and longer segregated phases (Figure 2D). All of these effects were
448 statistically significant (Figure 2A: $F(1,24) = 34.89, p < .001, \eta^2_p = .59, 95\% \text{ CI } [.48$
449 $.72]$. Figure 2B: $F(1,24) = 46.58, p < .001, \eta^2_p = .66, 95\% \text{ CI } [.58 .77]$. Figure 2C:
450 $F(1,24) = 15.28, p < .001, \eta^2_p = .39, 95\% \text{ CI } [.18 .58]$. Figure 2D: $F(1,24) = 11.28, p < .001,$
451 $\eta^2_p = .32, 95\% \text{ CI } [.09 .58]$).

452

453

[FIGURE 2 ABOUT HERE]

454

455 Importantly, the percentage of time each percept was reported was also affected by
456 the task instructions ($F(1,24) = 51.55, p < .001, \eta^2_p = .68, 95\% \text{ CI } [.58 .80]$; Figure 2A).
457 This effect was reflected in extended phases of the intended percept (although not to a
458 significant extent for non-initial integrated phases) and shortened phases of the
459 unintended percept, in comparison to the Neutral condition (Figure 2B, 2C, 2D; see
460 Table 1 for statistics). Focusing on tones of a single frequency to promote segregation
461 had a larger effect on the percentage of time hearing segregation than trying to hold
462 the three tones in a triplet together (the black lines are closer to the blue lines than to
463 the green lines in Figure 2A; $t(24) = 3.03, p = .006, d = 0.67, 95\% \text{ CI } [0.17 1.17]$).

464 However, there was no effect of attending to the high versus the low tones during
465 segregated listening ($t(24)=1.29$, $p=.208$, $d=0.27$, 95% CI [-0.15 0.69]).

466

467 [TABLE 1 ABOUT HERE]

468

469 **Univariate neural results**

470

471 Neural responses time-locked to the onset of each HLH- triplet were extracted for
472 each percept in the Neutral condition independent of Δf (Figure 3A), averaging over
473 the dipoles in bilateral auditory cortices (Figure 3B). A univariate analysis revealed a
474 time window 216-288 ms post triplet onset (66-138 ms post L tone onset) during
475 which epochs reported as segregated evoked a significantly more positive response
476 than those reported as integrated, independent of Δf (cluster-based permutation test,
477 $p=.001$; window-specific test, $t(22)=4.18$, $p<.001$, $d=0.32$, 95% CI [0.12 0.52]; Figure
478 3C). When based on single dipoles, the size of the percept effect in this time window
479 did not differ between the left and right hemispheres ($t(22)=1.55$, $p=.135$, $d=0.35$,
480 95% CI [-0.13 0.83]).

481

482 [FIGURE 3 ABOUT HERE]

483

484 The effect of intention on the neural response in this window was determined by
485 subtracting the mean over all epochs during attempts at integration from the mean
486 over all epochs during attempts at segregation, regardless of reported percept. The
487 group difference was significantly greater than zero ($t(22)=3.14$, $p=.005$, $d=0.17$, 95%
488 CI [0.05 0.29]; Figure 3D, middle), paralleling the percept comparison in the Neutral

489 condition (Figure 3D, left) and supporting participants' reports that they heard more
490 segregation when they tried to do so than when they tried to hear integration. This
491 effect is unlikely to be driven by attention-related modulations of neural responses to
492 particular tones independent of perceptual organization; the conditions in which
493 attention was focused on the H or the L tones did not differ significantly from each
494 other ($t(22)=0.23$, $p=.819$, $d=0.01$, 95% CI [-0.11 0.14]; Figure 3D, right).
495 Importantly, there was also no evidence for a residual response bias; the magnitude of
496 the neural difference in the non-Neutral conditions was similar to that expected if all
497 reports in those conditions were accurate ($t(22)=0.38$, $p=.710$, $d=0.08$, 95% CI [-0.35
498 0.51]).

499

500 **Multivariate neural results**

501

502 The difference waveform in the Neutral condition (Figure 3C) indicated that multiple
503 time windows might be informative in distinguishing between integrated and
504 segregated percepts, beyond the 216-288 ms window determined from the univariate
505 analysis. To make use of information across the entire epoch, we sought multivariate
506 temporal patterns that distinguished between integrated and segregated percepts at a
507 single-trial level, and which were allowed to vary across participants. Linear Support
508 Vector Machines (SVMs) trained for each Δf and participant (Figure 4A) achieved
509 classification accuracy significantly above chance ($t(22)=6.11$, $p<.001$, $d=1.77$, 95%
510 CI [1.17 2.86]; Figure 4C); this was driven by responses in multiple time windows,
511 including that identified in the univariate analysis (Figure 4B). When based on single
512 dipoles, classifier performance did not differ between the left and right hemispheres
513 ($t(22)=1.02$, $p=.321$, $d=0.22$, 95% CI [-0.21 0.68]).

514

515

[FIGURE 4 ABOUT HERE]

516

517 The SVMs trained on Neutral epochs were then used to classify epochs in the other
518 conditions. Paralleling the univariate results, a greater percentage of epochs were
519 classified as segregated when participants attempted segregation than when they tried
520 to integrate the sounds ($t(22)=3.87$, $p<.001$, $d=1.12$, 95% CI [0.63 1.77]; Figure 4D,
521 left). Again, this was not driven by epochs in which tones of one particular frequency
522 were attended; the percentage of epochs classified as segregated was similar whether
523 participants attended to high or low tones ($t(22)=0.45$, $p=.657$, $d=0.13$, 95% CI [-0.59
524 0.63]; Figure 4D, right).

525

526 The task-related difference in the percentage of epochs classified as segregated (mean
527 2.5%) was more than an order of magnitude smaller than the difference in reported
528 proportions (mean 36.6%). This discrepancy was due to non-perfect classifier
529 performance; although accuracy was above chance (50%), the mean was only 53.2%
530 and the maximum across participants 59.1%. After taking into account the accuracy
531 of each classifier, the task-related difference in the percentage of epochs classified as
532 segregated was no different from that expected if all percept reports in the non-
533 Neutral conditions were accurate ($t(22)=0.55$, $p=.586$, $d=0.08$, 95% CI [-0.24 0.42]).
534 In line with the univariate analysis, there was therefore no evidence for a residual
535 response bias.

536

537 The effect of intention determined by the multivariate analysis was larger and more
538 reliable than that from the univariate analysis. The more flexible approach was able to

INTENTION AND PERCEPTUAL ORGANIZATION

539 exploit the data of participants whose neural activity did not align with the group
540 percept signature in the 216-288 ms time window. For example, one participant's
541 percept in the Neutral condition could be decoded above chance based on the activity
542 at a range of timepoints, including an effect in the opposite direction from that of the
543 group around 216 ms post triplet onset (Figure 4B, dashed orange trace; Figure 4C,
544 orange circle).
545

546

Discussion

547

548 Our findings demonstrate that listeners can exert intentional control over how many
549 objects they perceive in an ambiguous auditory scene. Differences in auditory cortical
550 responses during attempts to hear repeating patterns of pure tones as an integrated
551 whole versus segregated streams were consistent with signatures of these percepts
552 obtained during a neutral listening condition. These differences supported listeners'
553 subjective reports that they could, to some extent, "hear what they want to hear".

554

Indexing low-level perception

556

557 We argue that the activity measured during neutral listening relates to the percept
558 rather than to decisions made during the process of reporting it. The inherent
559 uncertainty of localization based on MEG precludes ascribing a primary versus non-
560 primary auditory cortical locus; our source reconstruction appears consistent with
561 either of these. However, it seems unlikely that post-perceptual decision-related
562 activity would originate from auditory regions and be so consistently timed from the
563 onset of each stimulus. Furthermore, we excluded epochs surrounding button presses
564 to minimize the contribution of activity relating to motor planning or execution. We
565 therefore take the neutral neural signature to reflect perceptual experience. The use of
566 bistable stimuli to probe perception also avoided acoustic confounds. Although we
567 presented stimuli with two different frequency separations, leading to different
568 reported proportions of segregation (c.f. Gutschalk et al., 2005), the key comparisons
569 of neural activity were between alternative percepts of identical sounds.

570

INTENTION AND PERCEPTUAL ORGANIZATION

571 Our interpretation of activity in the non-Neutral conditions assumes that the neural
572 response carried more information about perception than about the instructions
573 themselves, which differed in terms of how listeners were to attend to the sounds.
574 Selective attention is known to affect the evoked response to tones even when
575 perceptual organization is stable (Hillyard et al., 1973; Näätänen et al., 1978). Such
576 modulations would presumably be maximally different across the two sub-conditions
577 in which participants attended exclusively to either the H or L tones, rather than
578 between one of these sub-conditions and the case when listeners attended to all of the
579 tones. However, we found no difference for attention to the H versus the L tones over
580 the time window of interest in the univariate analysis, nor in the percentage of epochs
581 classified as segregated in the multivariate analysis. We therefore argue that attention
582 alone (without concomitant changes in perceptual organization) cannot account for
583 the observed neural effects.

584

585 Another important feature of our design was the simultaneous collection of percept
586 reports and neural data, allowing us to draw direct associations between the two.
587 Some previous studies have inferred integration or segregation using measures
588 sensitive to stimulus manipulations that also affect perceptual organization, such as
589 the mismatch negativity (Sussman et al., 1999; Winkler et al., 2006; Carlyon et al.,
590 2010) or performance on a deviant detection task (Carlyon et al., 2010; Micheyl and
591 Oxenham, 2010; Billig et al., 2013; Spielmann et al., 2014). However such measures
592 are influenced by additional factors (Divenyi and Danner, 1977; Spielmann et al.,
593 2013, 2014; Sussman et al., 2013; Szalárdy et al., 2013b), and the degree to which
594 they, in isolation, can provide a reliable indication of perceptual organization over the
595 course of sustained bistable stimulation is unclear.

596

597 **Implications for auditory scene analysis**

598

599 The more positive response for segregation compared to integration from 66-138 ms
600 after the onset of the L tone was consistent with previous findings (Gutschalk et al.,
601 2005; Hill et al., 2012; Szalárdy et al., 2013a). It may in part reflect an increased P1m
602 response to the L tone during segregation, due to a release from adaptation by
603 responses to the previous H tone as neuronal receptive fields narrow and segregation
604 occurs (Fishman et al., 2001; Gutschalk and Dykstra, 2014). However, our results do
605 not depend on this interpretation; given the continuous stimulation paradigm it is not
606 clear how the observed differences relate to responses to individual tones.
607 Furthermore, our participant-specific classification analysis indicated that this time
608 window was not the most diagnostic of percept for all individuals. Variability across
609 listeners may arise from distinct listening strategies, or reflect differences in how
610 multiple components from repeated sounds summate to an aggregate measured signal.
611 Multivariate techniques such as representational similarity analysis have provided
612 insight into the fine spatial patterns representing stimulus information in the brain
613 (Haxby et al., 2001; Kriegeskorte et al., 2008). Here we applied a different form of
614 multivariate analysis – classification in the temporal domain - to reveal individualized
615 percept-specific patterns in neural activity (see also Wilbertz et al., 2017, Reichert et
616 al., 2014, for classification of bistable visual perception).

617

618 We observed effects of percept and intention when analyzing responses generated by
619 neural sources in bilateral auditory cortex. Functional magnetic resonance imaging
620 has also revealed greater responses in precuneus and right intraparietal sulcus during

INTENTION AND PERCEPTUAL ORGANIZATION

621 segregation compared to integration (Cusack, 2005; Hill et al., 2011). Our analysis of
622 precisely stimulus-locked responses would have been insensitive to more temporally
623 diffuse effects that such studies may have tapped. Further evidence for the
624 involvement in streaming of a network beyond auditory cortex comes from activity
625 during perceptual reversals (as opposed to during stable periods of integration or
626 segregation) in inferior colliculus, thalamus, insula, supramarginal gyrus, and
627 cerebellum (Kashino and Kondo, 2012; Kondo and Kashino, 2009; Schadwinkel and
628 Gutschalk, 2010). How these regions support or reflect either spontaneous reversals or
629 voluntary switches remains to be established.

630

631 A distinction has been drawn between primitive and schema-based processes of
632 perceptual organization (Bregman, 1990). Primitive processes automatically partition
633 a scene based on its physical properties, whereas schema-based processes select
634 elements based on attention or prior knowledge. One might expect different neural
635 instantiations of the outcomes of these processes; however, we found the same
636 segregation signature regardless of whether listeners allowed their perception to take a
637 natural course, deliberately attended to the H tones, or deliberately attended to the L
638 tones. We argue that the neural realization of an auditory scene may not only consist
639 of distinct representations of attended and unattended streams of differing
640 fidelity (Mesgarani and Chang, 2012; Puvvada and Simon, 2017) but also mark
641 whether any segregation has occurred at all (c.f. Gandras et al. 2017; Szalárdy,
642 Winkler, et al. 2013).

643

644 We asked participants to try to influence their percept by attending either to a subset
645 of the tones, or to all of them. The former approach may succeed by narrowing

646 receptive fields of auditory cortical neurons such that different populations respond to
647 the tones of each frequency (Fritz et al., 2007; Ahveninen et al., 2011), or by
648 introducing a perceived loudness difference between H and L tones (van Noorden,
649 1975; Dai et al., 1991). In contrast, repeatedly shifting attention across frequencies
650 may promote integration by disrupting these effects. The size of the change in reports
651 from the Neutral to the Attempt Segregation condition was greater than that from the
652 Neutral to the Attempt Integration condition. This was not the case in previous studies
653 (Pressnitzer and Hupé, 2006; Micheyl and Oxenham, 2010), a fact that may reflect
654 differences in stimuli, or in how instructions were interpreted. We also note that in
655 our experiment, volitional control similarly affected reported durations of intended
656 and unintended phases, whereas Pressnitzer and Hupé (2006) found that phases of
657 unwanted percepts were curtailed to a greater degree than target phases were
658 extended. Listeners in that study may have used additional strategies to shorten
659 segregated phases, such as briefly diverting attention away from the tone sequence
660 (Carlyon et al., 2003). Phase duration distributions have informed modelling of
661 auditory scene analysis (Mill et al., 2013; Rankin et al., 2015) and prompted parallels
662 to be drawn between different forms of bistability across sensory modalities
663 (Pressnitzer and Hupé, 2006). We emphasize that the interaction between stimulus
664 characteristics and high-level factors such as attention, which may differ across
665 bistable phenomena, must be considered in general accounts of how the brain handles
666 perceptual ambiguity (van Ee et al., 2005; Kogo et al., 2015).

667

668 **Summary**

669

INTENTION AND PERCEPTUAL ORGANIZATION

670 Auditory bistability offers a powerful means of understanding how cognitive states,
671 such as listening goals, attention, and prior knowledge, influence perception, while
672 controlling for stimulus differences. Linking subjective reports with neural measures
673 on a trial-by-trial basis allows us to tap into low-level processes, as opposed to post-
674 perceptual decisions. This method identifies signatures of perceptual experience in
675 auditory cortex to demonstrate that listeners can not only use attention to enhance the
676 representation of a subset of sounds, but also intentionally alter the number of distinct
677 objects heard to make up the auditory scene.
678

679 **References**

680

- 681 Ahveninen J, Hämäläinen MS, Jääskeläinen IP, Ahlfors SP, Huang S, Lin F-H, Raij
 682 T, Sams M, Vasios CE, Belliveau JW (2011) Attention-driven auditory cortex
 683 short-term plasticity helps segregate relevant sounds from noise. *Proc Natl Acad*
 684 *Sci* 108:4182–4187.
- 685 Billig AJ, Carlyon RP (2015) Automaticity and primacy of auditory streaming:
 686 Concurrent subjective and objective Measures. *J Exp Psychol Hum Percept*
 687 *Perform* 42:339–353.
- 688 Billig AJ, Davis MH, Deeks JM, Monstrey J, Carlyon RP (2013) Lexical influences
 689 on auditory streaming. *Curr Biol* 23:1585–1589.
- 690 Bode S, Feuerriegel D, Bennett D, Alday PM (2017) The Decision Decoding
 691 ToolBOX (DDTBOX) – A multivariate pattern analysis toolbox for event-related
 692 potentials.
- 693 Boehnke SE, Phillips DP (2005) The relation between auditory temporal interval
 694 processing and sequential stream segregation examined with stimulus laterality
 695 differences. *Percept Psychophys* 67:1088–1101.
- 696 Bonnel AM, Possamaï CA, Schmitt M (1987) Early modulation of visual input: a
 697 study of attentional strategies. *Q J Exp Psychol* 39:757–776.
- 698 Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*.
 699 Cambridge, MA: MIT Press.
- 700 Carlyon RP, Cusack R, Foxton JM, Robertson IH (2001) Effects of attention and
 701 unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept*
 702 *Perform* 27:115–127.
- 703 Carlyon RP, Plack CJ, Fantini DA, Cusack R (2003) Cross-modal and non-sensory
 704 influences on auditory streaming. *Perception* 32:1393–1402.
- 705 Carlyon RP, Thompson SK, Heinrich A, Pulvermüller F, Davis MH, Shtyrov Y,
 706 Cusack R, Johnsrude IS (2010) Objective measures of auditory scene analysis.
 707 In: *The Neurophysiological Bases of Auditory Perception* (Lopez-Poveda EA,
 708 Palmer AR, Meddis R, eds).
- 709 Cusack R (2005) The intraparietal sulcus and perceptual organization. *J Cogn*
 710 *Neurosci* 17:641–651.
- 711 Dai HP, Scharf B, Buus S (1991) Effective attenuation of signals in noise under
 712 focused attention. *J Acoust Soc Am* 89:2837–2842.
- 713 Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of
 714 single-trial EEG dynamics. *J Neurosci Methods* 134:9–21.
- 715 Denham SL, Gyimesi K, Stefanics G, Winkler I (2013) Perceptual bistability in
 716 auditory streaming: How much do stimulus features matter? *Learn Percept* 5:73–
 717 100.
- 718 Divenyi PL, Danner WF (1977) Discrimination of time intervals marked by brief
 719 acoustic pulses of various intensities and spectra. *Percept Psychophys* 21:125–
 720 142.
- 721 Farkas D, Denham SL, Bendixen A, Winkler I (2016) Assessing the validity of
 722 subjective reports in the auditory streaming paradigm. *J Acoust Soc Am*
 723 139:1762–1772.
- 724 Firestone C, Scholl BJ (2015) Cognition does not affect perception: Evaluating the
 725 evidence for “top-down” effects. *Behav Brain Sci* 4629:1–77.
- 726 Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of

- 727 auditory stream segregation in primary auditory cortex of the awake monkey.
728 *Hear Res* 151:167–187.
- 729 Fodor JA (1983) *The Modularity of Mind*. Cambridge, MA: MIT Press.
- 730 Fritz JB, Elhilali M, David S V, Shamma SA (2007) Does attention play a role in
731 dynamic receptive field adaptation to changing acoustic salience in A1? *Hear*
732 *Res* 229:186–203.
- 733 Gandras K, Grimm S, Bendixen A (2017) Electrophysiological correlates of speaker
734 segregation and foreground-background selection in ambiguous listening
735 situations. *J Neurosci*.
- 736 Green BF, Anderson LK (1956) Color coding in a visual search task. *J Exp Psychol*
737 51:19–24.
- 738 Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics*. New York,
739 NY: Wiley.
- 740 Gross S (2017) Cognitive penetration and attention. *Front Psychol* 8:1–12.
- 741 Gutschalk A, Dykstra AR (2014) Functional imaging of auditory scene analysis. *Hear*
742 *Res* 307:98–110.
- 743 Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005)
744 Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci*
745 25:5382–5388.
- 746 Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, Bießmann F
747 (2014) On the interpretation of weight vectors of linear models in multivariate
748 neuroimaging. *Neuroimage* 87:96–110.
- 749 Haxby J V., Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001)
750 Distributed and overlapping representations of faces and objects in ventral
751 temporal cortex. *Science* (80-) 293:2425–2430.
- 752 Hill KT, Bishop CW, Miller LM (2012) Auditory grouping mechanisms reflect a
753 sound’s relative position in a sequence. *Front Hum Neurosci* 6:1–7.
- 754 Hill KT, Bishop CW, Yadav D, Miller LM (2011) Pattern of BOLD signal in auditory
755 cortex relates acoustic response to perceptual streaming. *BMC Neurosci* 12:85.
- 756 Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective
757 attention in the human brain. *Science* (80-) 182:177–180.
- 758 Johnsrude IS, Mackey A, Hakyemez H, Alexander E, Trang HP, Carlyon RP (2013)
759 Swinging at a cocktail party: Voice familiarity aids speech perception in the
760 presence of a competing voice. *Psychol Sci* 24:1995–2004.
- 761 Kashino M, Kondo HM (2012) Functional brain networks underlying perceptual
762 switching: auditory streaming and verbal transformations. *Philos Trans R Soc B*
763 367:977–987.
- 764 Kogo N, Hermans L, Stuer D, van Ee R, Wagemans J (2015) Temporal dynamics of
765 different cases of bi-stable figure–ground perception. *Vision Res* 106:7–19.
- 766 Kriegeskorte N, Mur M, Bandettini P a. (2008) Representational similarity analysis -
767 connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- 768 Lupyan G (2017) Changing what you see by changing what you know: The role of
769 attention. *Front Psychol* 8:1–15.
- 770 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-
771 data. *J Neurosci Methods* 164:177–190.
- 772 Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker
773 in multi-talker speech perception. *Nature* 485:233–236.
- 774 Micheyl C, Oxenham AJ (2010) Objective and subjective psychophysical measures of
775 auditory stream integration and segregation. *J Assoc Res Otolaryngol* 11:709–
776 724.

- 777 Mill RW, Böhm TM, Bendixen A, Winkler I, Denham SL (2013) Modelling the
778 emergence and dynamics of perceptual organisation in auditory streaming. *PLoS*
779 *Comput Biol* 9:e1002925.
- 780 Moore BCJ, Gockel HE (2012) Properties of auditory stream formation. *Philos Trans*
781 *R Soc B* 367:919–931.
- 782 Näätänen R, Gaillard AW, Mäntysalo S (1978) Early selective-attention effect on
783 evoked potential reinterpreted. *Acta Psychol (Amst)* 42:313–329.
- 784 Orne MT (1962) On the social psychology of the psychological experiment: With
785 particular reference to demand characteristics and their implications. *Am*
786 *Psychol* 17:776–783.
- 787 Pressnitzer D, Hupé J-M (2006) Temporal dynamics of auditory and visual bistability
788 reveal common principles of perceptual organization. *Curr Biol* 16:1351–1357.
- 789 Puvvada KC, Simon XJZ (2017) Cortical Representations of Speech in a Multitalker
790 Auditory Scene. 37:9189–9196.
- 791 Pylyshyn Z (1999) Is vision continuous with cognition? The case for cognitive
792 impenetrability of visual perception. *Behav Brain Sci* 22:341-365-423.
- 793 Rankin J, Sussman E, Rinzel J (2015) Neuromechanistic model of auditory bistability.
794 *PLOS Comput Biol* 11:e1004555.
- 795 Reichert C, Fendrich R, Bernarding J, Tempelmann C, Hinrichs H, Rieger JW (2014)
796 Online tracking of the contents of conscious perception using real-time fMRI.
797 *Front Neurosci* 8:116.
- 798 Schadwinkel S, Gutschalk A (2010) Transient bold activity locked to perceptual
799 reversals of auditory streaming in human auditory cortex and inferior colliculus.
800 *J Neurophysiol* 105:1977-1983.
- 801 Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates
802 of auditory stream segregation. *J Cogn Neurosci* 18:1–13.
- 803 Snyder JS, Holder WT, Weintraub DM, Carter OL, Alain C (2009) Effects of prior
804 stimulus and prior perception on neural correlates of auditory stream segregation.
805 *Psychophysiology* 46:1208–1215.
- 806 Spielmann MI, Schröger E, Kotz SA, Bendixen A (2014) Attention effects on
807 auditory scene analysis: insights from event-related brain potentials. *Psychol Res*
808 78:361–378.
- 809 Spielmann MI, Schröger E, Kotz SA, Pechmann T, Bendixen A (2013) Using a
810 staircase procedure for the objective measurement of auditory stream integration
811 and segregation thresholds. *Front Psychol* 4.
- 812 Sussman ES, Chen S, Sussman-Fort J, Dinces E (2013) The five myths of MMN:
813 Redefining how to use MMN in basic and clinical research. *Brain Topogr*
814 27:553–564.
- 815 Sussman ES, Ritter W, Vaughan HG (1999) An investigation of the auditory
816 streaming effect using event-related brain potentials. *Psychophysiology* 36:22–
817 34.
- 818 Sussman ES, Winkler I, Huotilainen M, Ritter W, Näätänen R (2002) Top-down
819 effects can modify the initially stimulus-driven auditory organization. *Brain Res*
820 *Cogn Brain Res* 13:393–405.
- 821 Szalárdy O, Böhm TM, Bendixen A, Winkler I (2013a) Event-related potential
822 correlates of sound organization: Early sensory and late cognitive effects. *Biol*
823 *Psychol* 93:97–104.
- 824 Szalárdy O, Winkler I, Schröger E, Widmann A, Bendixen A (2013b) Foreground-
825 background discrimination indicated by event-related brain potentials in a new
826 auditory multistability paradigm. *Psychophysiology* 50:1239–1250.

INTENTION AND PERCEPTUAL ORGANIZATION

- 827 van Ee R, Van Dam LCJ, Brouwer GJ (2005) Voluntary control and the dynamics of
828 perceptual bi-stability. *Vision Res* 45:41–55.
- 829 van Noorden LPAS (1975) Temporal Coherence in the Perception of Tone
830 Sequences. Univeristy of Technology, Eindhoven, The Netherlands.
- 831 Wilbertz G, van Kemenade BM, Schmack K, Sterzer P (2017) fMRI-based decoding
832 of reward effects in binocular rivalry. *Neurosci Conscious* 3:1–10.
- 833 Winkler I, Van Zuijen TL, Sussman ES, Horváth J, Näätänen R (2006) Object
834 representation in the human auditory system. *Eur J Neurosci* 24:625–634.
- 835
- 836

837 **Legends**

838

839 **Table 1. Statistics for effects of intention on phase durations (compared to the**
840 **Neutral condition)**

841

842 **Figure 1. Stimulus parameters and percept reporting**

843 (A) Two triplets of a stimulus sequence consisting of high (H) and low (L) tones with
844 a frequency separation (Δf) of 4 or 6 semitones. The H tone frequency was 1000 Hz
845 for soundbooth testing and 1017 Hz for testing with electro-/magneto-
846 encephalography.

847 (B) Illustrative changes in perceptual organization with corresponding button press
848 reports. During integration (blue), H and L tones are perceived as belonging to a
849 single pattern, whereas during segregation (green) they form two separate perceptual
850 streams. Perception typically alternates every few seconds after a longer initial
851 integrated phase.

852

853 **Figure 2. Behavioral analyses**

854 Effects of frequency separation (Δf) and task on (A) the percentage of time reporting
855 segregation, (B) the duration of initial integrated phases, (C) the duration of
856 subsequent integrated phases, and (D) the duration of segregated phases. Phase
857 durations are plotted on a log scale. Black squares: listen neutrally, blue triangles:
858 attempt integration, green crosses: attempt segregation by attending to high tones,
859 green circles: attempt segregation by attending to low tones. Error bars: 95% within-
860 participants confidence intervals.

861

862 Figure 3. Univariate neural analyses

863 (A) Group timecourse (mean and 95% within-participants confidence intervals) of
864 neural activity for integrated (blue) and segregated (green) reports in the Neutral
865 condition, across frequency separations. Activity is projected through a spatial filter
866 based on dipoles in bilateral auditory cortex fitted to the sensor data separately for
867 each participant. The timing of each tone in the triplet is indicated below the plot.

868 (B) Mean and 95% within-participants confidence interval of the fitted dipole
869 locations for the activity in Figure 3A. Sources are shown on a template brain, with
870 coordinates in MNI space. Mean reconstructed sources lie in bilateral posteromedial
871 Heschl's gyrus.

872 (C) *t*-values for the Neutral Segregated minus Integrated group difference wave,
873 across frequency separations. Dashed red lines indicate the critical *t*-values at $p=.05$,
874 and the shaded red area represents the largest supra-threshold cluster.

875 (D) Differences in neural activity averaged over the time window of interest for
876 Neutral Segregated minus Integrated (left), Attempt Segregation minus Attempt
877 Integration (middle), and Attend High minus Attend Low (right), across frequency
878 separations. Filled circles correspond to individual participants, with mean and 95%
879 confidence intervals shown in red. The orange circle represents a single participant
880 also highlighted in Figures 4B, 4C, and 4D, for comparison of results across
881 univariate and multivariate approaches.

882

883 Figure 4. Multivariate neural analyses

884 (A) Schematic illustration of the classification approach. Linear Support Vector
885 Machines (SVMs) are trained for each participant and frequency separation in the
886 Neutral condition (left panel) to find the hyperplane (dashed line) that optimally

INTENTION AND PERCEPTUAL ORGANIZATION

887 separates epochs reported as integrated (blue circles) and segregated (green circles).
888 The SVMs are then applied to epochs (white circles) in the Attempt Integration
889 (middle panel) and Attempt Segregation (right panel) conditions.

890 **(B)** Directed feature weights for classification in the Neutral condition, across
891 frequency separations. A positive weight at a given timepoint reflects that a more
892 positive neural response contributes to a segregated classification. The mean across
893 the group is plotted in red with 95% confidence intervals in pink. One peak in the
894 mean trace lies in the significant 216-288 ms window from the univariate analysis.
895 However, classification can make use of different features (timepoints) for different
896 participants. The dashed orange trace corresponds to one participant whose neural
897 activity is dissimilar to the group mean (both for feature weights plotted here, and raw
898 activity in the 216-288 ms window shown in Figure 3D, left) and whose data did not
899 contribute to the univariate effect of intention (Figure 3D, middle). Perception can
900 nonetheless be decoded for this participant (Figure 4C), contributing to the
901 multivariate effect of intention (Figure 4D, left). The timing of each tone in the triplet
902 is indicated below the plot.

903 **(C)** Classification accuracy for Segregated versus Integrated epochs in the Neutral
904 condition, across frequency separations. Filled circles correspond to individual
905 participants, with mean and 95% confidence intervals shown in red. The orange circle
906 represents a single participant also highlighted in Figures 4B, 4C, and 4D, for
907 comparison of results across univariate and multivariate approaches, as described
908 above. Chance classification accuracy is at 50% (green dashed line).

909 **(D)** Differences in percentage of epochs classified as segregated for Attempt
910 Segregation minus Attempt Integration (left), and Attend High minus Attend Low
911 (right), across frequency separations. Filled circles correspond to individual

INTENTION AND PERCEPTUAL ORGANIZATION

912 participants, with mean and 95% confidence intervals shown in red. The orange circle
913 represents a single participant also highlighted in Figures 4B, 4C, and 4D, for
914 comparison of results across univariate and multivariate approaches.

Table 1. Statistics for effects of intention on phase durations (compared to the Neutral condition)

Task	Phases	<i>t</i> -value	<i>p</i> -value	<i>d</i>	<i>d</i> 95% CI
Attempt Integration	Initial integrated	3.75	.001	0.40	[0.15 0.65]
	Subsequent integrated	1.75	.094	0.17	[-0.03 0.37]
	Segregated	3.90	<.001	0.56	[0.22 0.90]
Attempt Segregation (Attend High)	Initial integrated	5.30	<.001	0.70	[0.36 1.05]
	Subsequent integrated	4.95	<.001	0.85	[0.42 1.28]
	Segregated	2.62	.015	0.35	[0.05 0.64]
Attempt Segregation (Attend Low)	Initial integrated	3.70	.001	0.63	[0.23 1.03]
	Subsequent integrated	4.13	<.001	0.66	[0.27 1.04]
	Segregated	3.67	.001	0.66	[0.24 1.08]







