

PHILOSOPHICAL TRANSACTIONS B

Auditory multistability and neurotransmitter concentrations in the human brain

Journal:	<i>Philosophical Transactions B</i>
Manuscript ID	RSTB-2016-0110.R1
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Kondo, Hirohito; NTT Communication Science Laboratories, Human Information Science Laboratory Farkas, Dávid; Hungarian Academy of Sciences Denham, Susan; University of Plymouth Asai, Tomohisa; NTT Communication Science Laboratories, Human Information Science Laboratory Winkler, István; Hungarian Academy of Sciences
Issue Code: Click here to find the code for your issue.:	AUDITORY
Subject:	Neuroscience < BIOLOGY, Behaviour < BIOLOGY, Cognition < BIOLOGY
Keywords:	perceptual organization, individual differences, auditory streaming, verbal transformations, magnetic resonance spectroscopy, neurotransmitters

SCHOLARONE™
Manuscripts

Auditory multistability and neurotransmitter concentrations in the human brain

Hirohito M. Kondo^{1*}, Dávid Farkas^{2,3}, Susan L. Denham⁴, Tomohisa Asai¹,
and István Winkler²

¹ Human Information Science Laboratory, NTT Communication Science Laboratories, NTT Corporation, Atsugi, Kanagawa 243-0198, Japan

² Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar Tudósok körútja 2, 1117 Budapest, Hungary

³ Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Egrý József utca 1, 1111 Budapest, Hungary

⁴ Cognition Institute and School of Psychology, University of Plymouth, Plymouth, Devon PL4 8AA, United Kingdom

Keywords: perceptual organization, individual differences, auditory streaming, verbal transformations, magnetic resonance spectroscopy, neurotransmitters

Summary

Multistability in perception is a powerful tool for investigating sensory-perceptual transformations because it produces dissociations between sensory inputs and subjective experience. Spontaneous switching between different perceptual objects occurs during prolonged listening to a sound sequence of tone triplets or repeated words (termed auditory streaming and verbal transformations, respectively). We used these examples of auditory multistability to examine to what extent neurochemical and cognitive factors influence the observed idiosyncratic patterns of switching between perceptual objects. The concentrations of glutamate-glutamine (Glx) and γ -aminobutyric acid (GABA) in brain regions were measured by magnetic resonance spectroscopy, while personality traits and executive functions were assessed using questionnaires and response inhibition tasks. Idiosyncratic patterns of perceptual switching in the two multistable stimulus configurations were identified using a multidimensional scaling (MDS) analysis. Intriguingly, although switching patterns within each individual differed between auditory streaming and verbal transformations, similar MDS dimensions were extracted separately from the two datasets. Individual switching patterns were significantly correlated with Glx and GABA concentrations in auditory cortex and inferior frontal cortex but not with the personality traits and executive functions. Our results suggest that auditory perceptual organization depends on the balance between neural excitation and inhibition in different brain regions.

Introduction

An essential function of perceptual systems is to parse complex scenes into meaningful components. The sequential integration and segregation of frequency components plays a critical role in auditory scene analysis because sound sources produce information over time [1]. Multistable perceptual phenomena, such as auditory streaming have been used to identify the factors influencing auditory scene analysis. A sequence of repeating tone triplets (ABA, where A and B differ in frequency) is presented in the auditory streaming paradigm [2]. The perceptual organization depends on the frequency difference between the A and B tones and the presentation rate: a small difference and slow presentation favour integration (i.e., all tones experienced as a single stream), whereas a large separation and fast presentation favour segregation (i.e., two separate streams). For intermediate differences and presentation rates, perception tends to switch between the two types of percept [3]. Recent studies using auditory streaming have pointed out that listeners can experience three or more types of perceptual organization, such as integrated, segregated, and combined percepts [4-6]. In verbal transformations, a series of perceptual switches is produced by prolonged listening to a repeated word without a pause [7, 8]. For instance, the stimulus 'tress' may be transformed into a variety of verbal forms, such as 'dress', 'stress', and 'drest' [9]. In daily life, we sometimes have difficulty taking signals from acoustic inputs because of various types of

*Author for correspondence (kondo.hirohito@lab.ntt.co.jp).

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

noise. From the perspective of adaptive behaviour, it is important for the brain to create some possible percepts from ambiguous inputs and fluctuate perceptual interpretations of them. Thus, these paradigms are simple but important for investigating our abilities to identify sound events, communicate with others, and enjoy music.

Multistable stimuli induce successive spontaneous switches between different perceptual objects, in a seemingly random fashion [10]. A recent study has demonstrated that perceptual switching patterns are idiosyncratic: the switching patterns of each participant are more similar to their own switching patterns in different sessions than to those of other participants and this tendency is preserved even when sessions are separated by more than a year [5]. This raises the following question: what are the idiosyncratic factors of perceptual switching?

One possibility is that genetic or anatomical differences are responsible for inter-individual variations in multistable perception. A twin heritability study has revealed that 50% of the variance in the binocular rivalry rate is explained by additive genetic factors [11]. The dopaminergic system has been linked with individual differences of perceptual switching in auditory streaming and verbal transformations [12, 13]. Brain structures, such as regional cortical volumes [14] and interregional connections [15], are associated with spontaneous switching in visual rivalry. However, few studies have examined whether neurochemical factors in the brain contribute to the stability of idiosyncratic switching patterns in an auditory multistable task. Computational models with mutual inhibition and sensory adaptation have been proposed to account for the nonlinear dynamics of perceptual multistability [16-20]. In these models, perceptual switching is driven by adaptation of the winning neural population and lateral inhibition of the competing population [21]. Thus, we can expect that the excitation-inhibition balance plays an important role in perceptual multistability. However, it is essentially difficult to elucidate the excitation-inhibition balance of brain activations because both processes are activity dependent but in fundamentally different ways [22]. To overcome this difficulty, we used magnetic resonance spectroscopy (MRS) to measure the concentration of glutamate-glutamine (Glx) and γ -aminobutyric acid (GABA) within brain regions. In the cerebral cortex, roughly 80% of the neurons are excitatory glutamatergic and 20% are inhibitory GABAergic neurons [23]. The ratio of synaptic neurotransmitters plays a key role in coordinating the pyramidal cell activity and in driving the hemodynamic response. Specifically, MRS studies have demonstrated that the GABA concentration in different cortical areas predicts individual differences in visual awareness [24], visual attention [25], and orientation discrimination [26]. However, it is unclear to what degree neurotransmitter systems are involved in the formation and selection of auditory objects.

Another approach is to search for high-level processes related to switching patterns for each individual [27-29]. Although the frontal areas are probably responsible for perceptual switching in different multistable stimuli [30-32], there is little evidence for a link between personality traits and switching patterns, except for the results of one recent study [33]. We assessed personality traits and executive functions to investigate whether cognitive abilities are associated with individual differences in perceptual switching. Perceptual organization may also be influenced by the big five traits (measured by the Big Five Inventory [34]), which are considered to be the basic broad domains of personality. Behavioural impulsivity and flexibility have been measured by using the UPPS Impulsive Behaviour Inventory [35] and ego-resiliency scale [36, 37], respectively. Inhibitory control is thought to be one of the important executive functions [38]. The present study focused on inhibition of a prepotent response to test the possible role of cognitive-level inhibition processes in perceptual organization. We extracted several variables for auditory streaming and verbal transformations and computed correlations between perceptual variables, neurotransmitter measures, personality traits, and response inhibition abilities.

Methods

Participants

We recruited thirty-four Japanese participants (24 males and 10 females; 21-60 years, $M_{\text{age}} = 36.3$, $SD_{\text{age}} = 11.2$) for this study. The score of the Edinburgh Handedness Inventory [39] was 94.8 ± 5.2 , indicating that they were strongly right-handed. None had any history of neurological or psychiatric disorders. All gave written informed consent, which was approved by the Ethics and Safety Committees of NTT Communication Science Laboratories and ATR-Promotions. On the basis of the performance of catch trials (see Data Analysis section for details), we excluded 11 participants from subsequent analyses. One additional participant was discarded because she did not complete the verbal transformations task. Thus, the reported results were derived from 22 participants (16 males and 6 females; 21-58 years, $M_{\text{age}} = 32.6$, $SD_{\text{age}} = 10.7$).

MRS Data Acquisition

MRS data were acquired with a Siemens MAGNETOM Trio 3-T MRI scanner with a 12-channel receive-only head coil. Head motion was minimized with comfortable padding around the participant's head. For assessment of cortical thickness and volume, anatomical images were obtained with a T1-weighted pulse sequence (isotropic voxel size of 1 mm^3). To minimize confounding factors, we acquired MR spectra at a fixed time during the day, which was one and a half hours from 1:00 to 2:30 p.m.

MR spectra were acquired from four $3 \times 3 \times 3 \text{ cm}^3$ voxels of interest: the auditory cortex (AC), inferior frontal cortex (IFC), prefrontal cortex (PFC), and anterior cingulate cortex (ACC) (Supplementary Figure 1a). Voxels were positioned by using internal landmarks in order to achieve a consistent position between participants. The AC voxel included Heschl's gyrus (Brodmann area: BA 41) and the anterior part of the temporal plane (BA 42). The IFC voxel

1
2
3 included the pars opercularis (BA 44) and pars triangularis (BA 45) of the inferior frontal gyrus. The PFC voxel was
4 located at the anterior part of the middle frontal gyrus (BA 46). The ACC voxel (including portions of BAs 32 and 9) was
5 located superior to the genu of the corpus callosum and centred on the interhemispheric fissure. All voxels except the
6 ACC one were angled parallel to the brain surface of the left hemisphere. For each participant, voxels were separately
7 placed to exclude cerebral spinal fluid from the ventricles or the cortical surface.

8 Four consecutive runs were acquired from the different voxels for each participant. Before each run, we
9 carefully carried out manual shimming (approximately 5 min) of the magnetic field in the voxel to avoid line broadening.
10 MR spectra were obtained by using a GABA spectral editing sequence. For each spectrum, 64 spectral averages of 1024
11 data points were acquired with a repetition time of 1500 ms and an echo time of 68 ms, resulting in scan duration of 3
12 min 18 sec. We used the short duration to reduce effects of head motion on MR spectra. An editing pulse with bandwidth
13 of 44 Hz was applied at 1.9 ppm (on) and 7.5 ppm (off) in interleaved scans. Differences in the edited spectra yielded the
14 Glx and GABA peaks (Supplementary Figure 1b). The unsuppressed water signal was also acquired from the same voxel.

15 GANNET and in-house software was used to quantify total Glx and GABA in the difference spectra [40]. All
16 spectra were phase aligned with reference to water, frequency aligned to creatine, and modelled with a simple Gaussian
17 function. The final results were expressed as the ratio of Glx and GABA signal areas (peaks at 3.76 and 3.00 ppm,
18 respectively) relative to the unsuppressed water signal area (W). The Glx/W and GABA/W concentrations were
19 quantified in institutional unit (i.u.) [41]. The Glx/W and GABA/W concentrations (mean \pm standard error) were $1.23 \pm$
20 0.03 i.u. and 1.71 ± 0.06 i.u. for AC; 1.19 ± 0.05 i.u. and 1.55 ± 0.08 i.u. for IFC; 1.13 ± 0.05 i.u. and 1.43 ± 0.08 i.u. for
21 PFC; and 1.06 ± 0.04 i.u. and 1.30 ± 0.05 i.u. for ACC. The fit errors related to the Glx/W and GABA/W values were
22 $13.2 \pm 0.8\%$ and $8.5 \pm 0.4\%$ for AC; $15.7 \pm 0.7\%$ and $11.7 \pm 0.5\%$ for IFC; $15.2 \pm 0.7\%$ and $11.3 \pm 0.5\%$ for PFC; and
23 $16.6 \pm 0.6\%$ and $9.9 \pm 0.5\%$ for ACC.

24 Task Procedures

25 Right after the MRS data acquisition, participants performed the multistability tasks and response inhibition
26 tasks and filled out the personality questionnaires in a quiet room. The behavioural tasks were controlled by the Cogent
27 2000 Toolbox running under MATLAB (Mathworks Inc.) on a PC, whereas the personality questionnaires, consisting of
28 the Big Five Inventory, the UPPS Impulsive Behaviour Inventory, and ego-resiliency scale, were pen-and-paper tests.
29 The order of these tests was randomized across participants.

30 For auditory streaming, the stimuli and task procedures were identical to those used in our previous studies [33,
31 42]. The streaming stimulus was a 4-min-long sequence of a repeating ABA- pattern, where the frequency of the A tone
32 was 400 Hz and the frequency difference between the A and B tones was four semitones. The tones had a duration of 75
33 ms, and the stimulus onset asynchrony was 150 ms. Participants were instructed to continuously report their perception
34 by holding down the assigned arrow key on a computer keyboard for as long as they perceived the tones in the same way.
35 Participants were given four alternatives for categorizing their perception: Integrated (ABA-ABA-), Segregated (A-A-
36 and -B---B--), Combined (AB-- and --A- or -BA- and A---), and None (none of the above possibilities). The five test
37 blocks were preceded by training blocks, where participants practiced reporting their perception confidently and
38 precisely.

39 For the verbal transformations, the stimulus consisted of repetitions of the word "banana", spoken by a female
40 native Japanese speaker [32]. Participants heard the word for 340 ms without gaps while refraining from silently
41 repeating it. They were asked to continuously indicate the word they heard by holding down the assigned arrow keys on a
42 computer keyboard. The alternatives to be marked were Banana, Nappa ("vegetables" in English), some Nonsense word,
43 or some Other word (i.e., actual Japanese words other than Banana and Nappa), and None (undecided). Five 4-min blocks
44 were conducted for verbal transformations for comparability with the auditory streaming task.

45 Three inhibition tasks were employed to assess participants' ability to inhibit automatic or prepotent responses.
46 Although the term of "inhibition" is commonly used to describe a wide variety of functions, the concept of inhibition in
47 this study is restricted to the controlled suppression of dominant responses. In the antisaccade task, a visual cue was
48 followed by a target, and participants had to suppress stimulus-driven attention induced by the cue. The stop-signal task
49 was similar to a go/nogo task, but required them to inhibit an already initiated motor response. Thus, in addition to
50 response accuracies, latencies to stop a response could be calculated. In the Stroop task, participants had to resolve name-
51 colour conflicts and provide a response verbally. The detailed procedures of personality questionnaires and inhibition
52 tasks are described in Supplementary Task Procedures and Supplementary Figure 2. The scores for all items of the
53 personality scales were summed to produce the subscale scores. The reliabilities reached a satisfactory level (range of
54 0.73 to 0.88). A summary of descriptive statistics for the personality scales and inhibition tasks is shown in
55 Supplementary Tables 1 and 2.

56 Data Analysis

57 For the auditory streaming paradigm, catch trials were used to exclude participants who likely did not fully
58 understand the instructions (see Supplementary Data Analysis for details). Each of the possible perceptual alternatives
59 were promoted in the catch trials. On the basis of catch-trial performance, we excluded participants from further analysis
60 if their correct identification of the disambiguated Integrated, Segregated, or Combined percept fell below 30% or their
composite correct score was less than 60% [42]. The data for 11 participants were removed, making the sample size 22.

1
2
3
4 For each multistability task, the time-series data of phase durations were collected from all test blocks. Phases
5 shorter than 300 ms were excluded from data analyses [43], resulting in 99.7% of the analysed data. Results of a previous
6 study suggested that the first block should be discarded from the analysis because the switching pattern in this block
7 substantially differs from that in the rest of the blocks [42]. The following analyses were based on blocks two to five
8 (Supplementary Figure 3a). In addition, the first phase was removed from each block, because the first phase duration is
9 known to be longer than that of the subsequent ones [3, 44]. Individual switching patterns were characterised by using
10 transition matrices, which represent the conditional probabilities of perceptual switches [6]. We modelled time-series
11 switching data using a Markov chain of four types of percepts in auditory streaming and five types of percepts in verbal
12 transformations. The parameters of a Markov chain were used to calculate transition probabilities from one to the other
13 percepts. The transition matrices include several measures of switching patterns, such as switch numbers, phase durations,
14 and percept proportions.

15 The number of perceptual switches and the durations of each percept were extracted from transition matrices.
16 For auditory streaming, the number of switches across blocks was 35.8 ± 24.6 (mean \pm standard error). The phase
17 durations were 14.7 ± 13.3 s for Integrated, 7.5 ± 5.8 s for Segregated, 6.6 ± 4.9 s for Combined, and 1.5 ± 1.4 s for None.
18 The proportions of the percepts were $49.0 \pm 18.7\%$ for Integrated, $27.3 \pm 11.8\%$ for Segregated, $23.4 \pm 23.9\%$ for
19 Combined, and $0.3 \pm 0.5\%$ for None. For verbal transformations, the number of switches was 32.5 ± 18.8 . The phase
20 durations were 19.3 ± 49.8 s for Banana, 9.2 ± 8.0 s for Nappa, 10.4 ± 7.5 s for Nonsense, 10.2 ± 6.2 s for Others, and 4.5
21 ± 1.8 s for None. The proportions of the percepts were $26.8 \pm 20.1\%$ for Banana, $31.7 \pm 15.5\%$ for Nappa, $19.0 \pm 13.7\%$
22 for Nonsense, $20.7 \pm 16.6\%$ for Other percepts, and $1.8 \pm 0.0\%$ for None.

23 The time to discover all percepts was calculated by simulating the switching patterns using the transition
24 matrices [33]. Short discovery times suggest that all alternatives are relatively easy to perceive for the given participant,
25 whereas long discovery times suggest that some perceptual alternatives are less viable. We ran the simulation 1,000 times
26 for each participant until all percepts were discovered. The median time of the simulations was used as the "time to
27 discover all percepts" variable of the participant. The times to discover all percepts were 24.4 ± 33.4 s for auditory
28 streaming and 197.7 ± 180.4 s for verbal transformations.

29 The transition matrices were suitable for comparing individuals based on multidimensional scaling (MDS) of
30 their switching patterns (see also Supplementary Data Analysis). A scree test was performed to decide on the number of
31 dimensions in the MDS (Supplementary Figure 3b). A smaller stress value (less than 0.05) indicates that the dataset is
32 well represented by the corresponding number of MDS dimensions [45]. To interpret the MDS dimensions, we assessed
33 the extent to which the coordinates of each participant were correlated with the following variables: for auditory
34 streaming, the number of switches, the time to discover all percepts, the durations and proportions of the Integrated,
35 Segregated, and Combined percepts; for verbal transformations, the durations and proportions of the Banana, Nappa,
36 Nonsense, and Others.

37 We calculated significance levels by Spearman's rank order correlations using two methods. The first approach
38 was to estimate the probability of obtaining the correlation by a random projection of the factor values. The probability
39 was determined by permuting the factor vectors 10,000 times and establishing the proportion of random correlations that
40 were higher than the one obtained empirically (p_{perm}). The second approach was to control for the family-wise error rate
41 in this randomization context by registering the highest absolute correlation between the given MDS dimension and each
42 perceptual variable in each permutation run. The distribution of these maximal coefficients was then used to compute the
43 p -value of the observed correlations (p_{fwe}). Variables were sorted into five families: perceptual measures (eight variables
44 in auditory streaming and ten variables in verbal transformations); Glx measures in the four brain regions (AC, IFC, PFC,
45 and ACC); GABA measures derived from the same brain regions as the Glx measures; personality traits (ten variables of
46 the personality scales); measures from the inhibition tasks (accuracy of antisaccade trials, stop-signal reaction time, and
47 reaction time difference in the Stroop task).

48 Results

49 Idiosyncratic switching patterns were first checked by using the intra-individual difference calculated from the
50 Kullback-Leibler (KL) distances between participant's own transition matrices across blocks (Supplementary Figure 3c).
51 Then inter-individual difference was assessed by computing the KL distances between the participant's and every other
52 participants' transition matrices. The relationship between a participant's intra- and inter-individual consistency was
53 tested by Wilcoxon's signed rank test. The hypothesis was that the intra-individual distances would be smaller than the
54 inter-individual ones. Thus, we used a one-tailed test for each participant (α -level = 0.05).

55 Using the transition matrices from auditory streaming, 17 of the 22 participants (77.3%) were distinguishable
56 from the rest of the participants (Supplementary Table 3). Using the number of switches variable, 15 of the 22
57 participants (68.2%) were separable from the rest. For verbal transformations, the transition matrices allowed 15 of the 22
58 participant (68.2%) to be distinguished from the rest of the participants, whereas by the number of switches, 16 of the 22
59 participants (72.7%) were separable from the rest. In both auditory streaming and verbal transformations, 14 participants
60 (63.6%) had characteristic transition matrices, although only ten participants (45.5%) had characteristic switch numbers.
The transition matrices were chosen for the MDS because it led to similar results reflecting characteristic switching
patterns for each individual.

The MDS was used to examine relationships between individual cases in the auditory streaming dataset (Table
1). A two-dimensional solution gave an acceptable fit to the data (stress = 0.028) (Supplementary Figure 3b). The first

dimension was positively related to the proportion of the Combined percept ($r_s = 0.986$), whereas it was negatively related to the proportion of the Integrated and Segregated percepts ($r_s = -0.744$ and -0.666). The first dimension also showed negative correlations with the duration of Integrated and Segregated percepts ($r_s = -0.614$ and -0.604). Because Integrated and Segregated percepts are the most frequent ones in a classic streaming paradigm [2], our interpretation is that the first dimension is strongly affected by the presence of the additional option of describing one's perception in terms of the Combined percept. Thus, the first dimension was named the 'exploration-exploitation' axis. This idea is consistent with the evidence that participants with high scores on the first dimension discovered all percepts quickly ($r_s = -0.495$). The second dimension was positively related to the proportion of the Integrated percept ($r_s = 0.540$) and negatively to that of the Segregated percept ($r_s = -0.572$). Thus, the second dimension was termed the 'integration-segregation' axis. Taken together, these results suggest that the listener's task to categorize their perception in terms of multiple alternatives produces distinct individual differences in auditory streaming and that these differences can be mapped on two dimensions.

Table 1. MDS dimensions derived from auditory streaming variables.

Measure	First dimension			Second dimension		
	r_s	p_{perm}	p_{fwe}	r_s	p_{perm}	p_{fwe}
Duration of Integrated	-0.614	0.003	0.020	0.312	0.160	0.568
Duration of Segregated	-0.604	0.002	0.024	-0.085	0.703	0.999
Duration of Combined	0.208	0.352	0.872	0.022	0.923	1.000
Proportion of Integrated	-0.774	< 0.001	< 0.001	0.540	0.011	0.063
Proportion of Segregated	-0.666	< 0.001	0.007	-0.572	0.007	0.036
Proportion of Combined	0.986	< 0.001	< 0.001	-0.172	0.439	0.940
Number of switches	0.422	0.057	0.244	-0.211	0.344	0.863
Time to discover all percepts	-0.495	0.023	0.111	0.605	0.002	0.022

Note. Values indicated in bold are significant ($p < 0.05$, $N = 22$). r_s , Spearman's rank order correlation coefficient; p_{perm} , permuted p -value; p_{fwe} , family-wise error controlled p -value.

We identified the neurotransmitter measures associated with switching patterns of auditory streaming (Supplementary Table 4). The Glx in the AC was negatively correlated with the 'exploration-exploitation' dimension ($r_s = -0.550$, $p_{\text{perm}} = 0.007$, $p_{\text{fwe}} = 0.039$), but the other variables were not. This suggests that higher Glx concentration in this region is related to the 'exploitation' property of auditory streaming. We further examined the relationship between neurotransmitter measures and perceptual variables. The Glx in the AC was correlated positively with the proportion of the Segregated percept ($r_s = 0.761$, $p_{\text{perm}} = 0.001$, $p_{\text{fwe}} = 0.001$) and negatively with that of the Combined percept ($r_s = -0.520$, $p_{\text{perm}} = 0.014$, $p_{\text{fwe}} = 0.050$) (Figure 1a). This confirms that participants with a higher Glx concentration in the AC experience more Segregated and fewer Combined percepts. It has been found that neural responses in the AC can account for important features of auditory streaming [46-49]. Here we argue that the formation and selection of the Combined percept also requires other brain areas. One possible candidate is the IFC, because the GABA measured there was related positively to the proportion of the Combined percept ($r_s = 0.446$, $p_{\text{perm}} = 0.041$, $p_{\text{fwe}} = 0.138$) and negatively to the duration of the Segregated percept ($r_s = -0.425$, $p_{\text{perm}} = 0.050$, $p_{\text{fwe}} = 0.185$) (Figure 1b). Thus, it is possible that the 'exploration-exploitation' property of auditory streaming is supported by a balance between Glx and GABA concentrations in different brain regions.

MDS was also used to examine what separates individuals' switching patterns in verbal transformations (Table 2). A three-dimensional solution was chosen (stress = 0.014) (Supplementary Figure 3b). The first dimension was related positively to the proportion of Banana ($r_s = 0.816$) and negatively to that of Others ($r_s = -0.906$). The second and third dimensions were negatively correlated with the proportion of Nappa ($r_s = -0.966$) and Nonsense ($r_s = -0.868$). These results indicate that the proportion of each percept, rather than its duration, is associated with individual differences in verbal transformations.

Neurotransmitter measures did not show any significant correlation with the MDS dimensions of verbal transformations (Supplementary Table 5). We then investigated the relationship between neurotransmitter measures and perceptual variables. The Glx concentration in the AC was related positively to the duration of Banana ($r_s = 0.542$, $p_{\text{perm}} = 0.011$, $p_{\text{fwe}} = 0.042$) and Nappa ($r_s = 0.581$, $p_{\text{perm}} = 0.005$, $p_{\text{fwe}} = 0.021$) (Figure 2a) and negatively to the number of switches ($r_s = -0.539$, $p_{\text{perm}} = 0.010$, $p_{\text{fwe}} = 0.010$). The Glx concentration in the AC showed a negative correlation with the time to discover all percepts ($r_s = -0.539$, $p_{\text{perm}} = 0.010$, $p_{\text{fwe}} = 0.157$). Given that Banana and Nappa are the most common percepts in this variant of verbal transformations, the pattern of correlations described is partly consistent with

that obtained for auditory streaming. In addition, the GABA in the IFC was related to the duration of Others ($r_s = -0.433$, $p_{\text{perm}} = 0.042$, $p_{\text{fwe}} = 0.167$) (Figure 2b). Thus, there is the possibility that Glx and GABA concentrations in the AC and IFC contribute to durations of verbal forms.

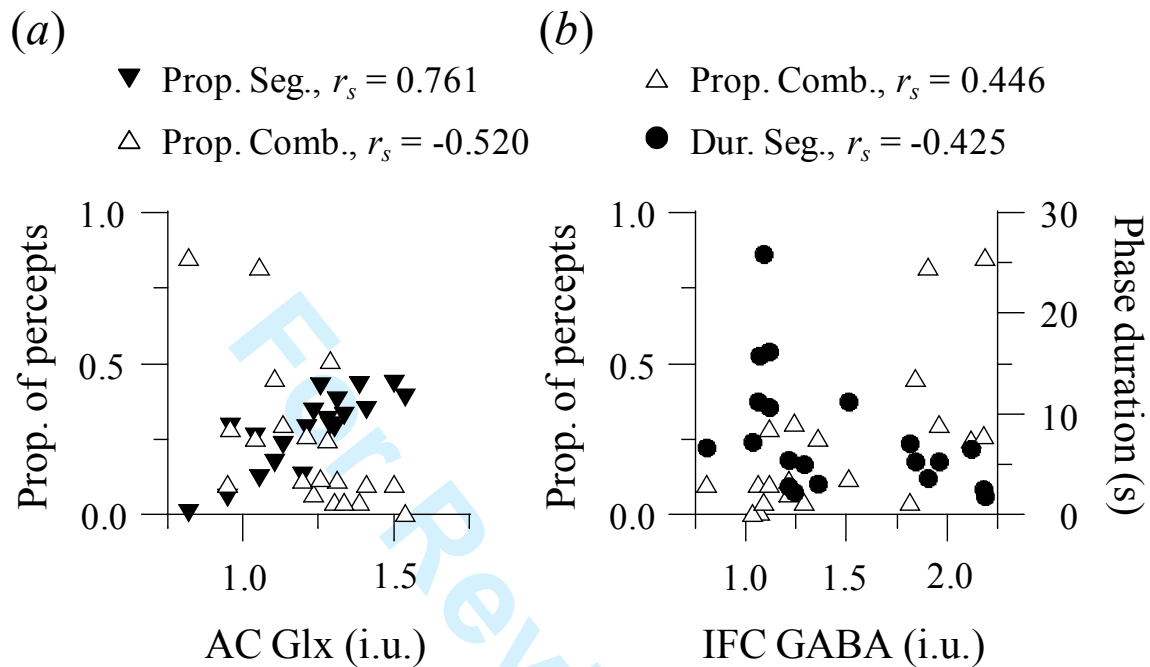


Figure 1. Scatter plots of neurotransmitter concentrations and auditory streaming variables. Symbols indicate individual data. Grubbs test did not reveal any outliers for the variables.

The first MDS dimension of auditory streaming was not related to the first ($r = 0.132$, $p = 0.559$), second ($r = 0.167$, $p = 0.457$), and third MDS dimensions of verbal transformations ($r = -0.147$, $p = 0.514$). The same was true for the second MDS dimension of auditory streaming for the first ($r = 0.017$, $p = 0.940$), second ($r = 0.089$, $p = 0.692$), and third MDS dimensions of verbal transformations ($r = 0.380$, $p = 0.081$). Thus, an MDS position in one multistability task does not appear to be related to the MDS position in the other task. For the personality scales and response inhibition tasks, we did not find any measure that significantly correlated with any of the MDS dimensions of the switching patterns (Supplementary Tables 6 to 9).

Table 2. MDS dimensions derived from verbal transformations variables.

Measure	First dimension			Second dimension			Third dimension		
	r_s	p_{perm}	p_{fwe}	r_s	p_{perm}	p_{fwe}	r_s	p_{perm}	p_{fwe}
Duration of Banana	0.318	0.145	0.677	0.287	0.193	0.771	0.095	0.671	1.000
Duration of Nappa	-0.036	0.872	1.000	-0.271	0.222	0.820	0.012	0.958	1.000
Duration of Nonsense	-0.064	0.778	1.000	0.091	0.685	1.000	0.007	0.972	1.000
Duration of Others	-0.419	0.058	0.344	0.041	0.854	1.000	0.069	0.761	1.000
Proportion of Banana	0.816	< 0.001	< 0.001	0.250	0.261	0.876	0.355	0.104	0.553
Proportion of Nappa	0.142	0.518	0.995	-0.966	< 0.001	< 0.001	-0.080	0.732	1.000
Proportion of Nonsense	-0.191	0.397	0.965	0.403	0.068	0.384	-0.868	< 0.001	< 0.001
Proportion of Others	-0.906	< 0.001	< 0.001	0.029	0.898	1.000	0.414	0.056	0.353
Number of switches	-0.027	0.909	1.000	-0.187	0.405	0.972	-0.177	0.433	0.977
Time to discover all percepts	0.465	0.031	0.218	-0.169	0.445	0.982	0.110	0.622	0.999

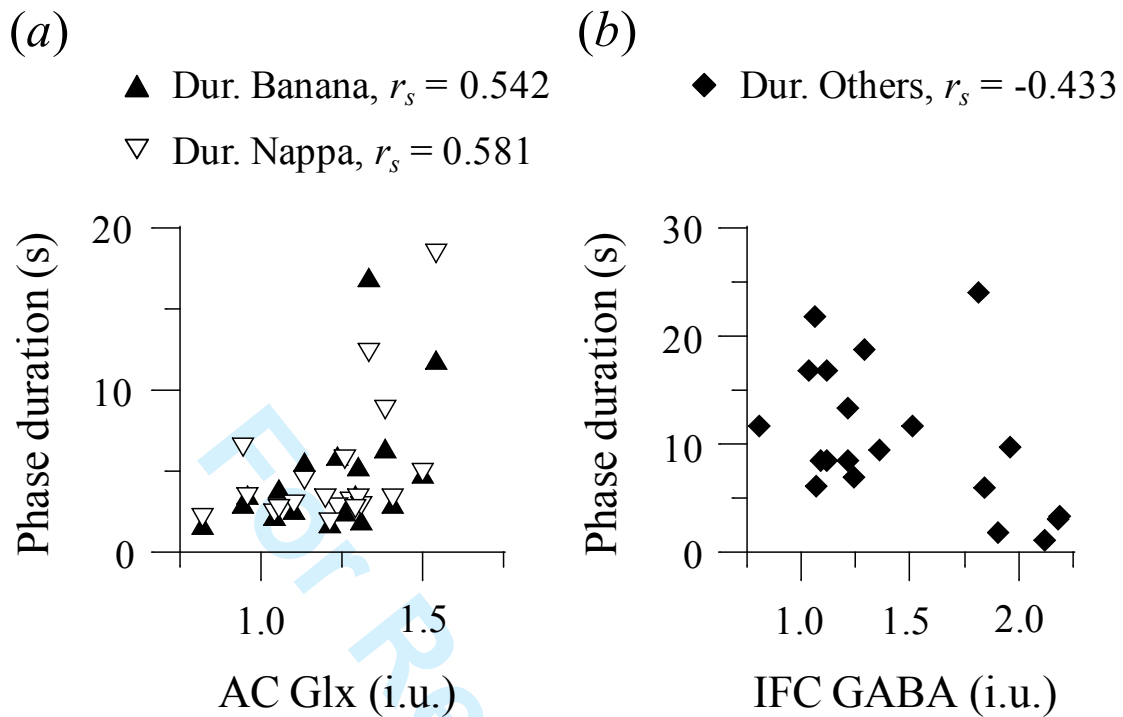


Figure 2. Scatter plots of neurotransmitter concentrations and verbal transformations variables. Symbols indicate individual data. Grubbs test did not reveal any outliers for the variables.

Discussion

The present results demonstrated that Glx and GABA concentrations in the AC and IFC were related to idiosyncratic switching patterns of auditory streaming and verbal transformations. In contrast, we found no evidence that neurotransmitter concentrations in PFC and ACC contributed to individual differences in the switching patterns. This lack of correlation is consistent with recent neuroimaging evidence in visual bistable perception [50]. Thus, the inter-individual variation of auditory multistability can be linked to the balance of glutamatergic and GABAergic signalling between different brain regions. We did not find any correlation between perceptual switching patterns, personality traits, and response inhibition abilities. Thus, cognitive factors probably have a limited effect on auditory multistability.

We acquired the following two dimensions from the transition matrices of auditory streaming: 'exploration-exploitation' and 'integration-segregation'. The MDS results in this study are consistent with those obtained in a recent study [33]. The dimensions probably reflect a general principle of switching patterns in auditory streaming, beyond language and culture, because similar results are obtained in laboratories in different countries. More importantly, Glx concentrations in the AC were correlated with the proportion of the Integrated and Segregated percepts (i.e., more exploitation), whereas GABA concentrations in the IFC were correlated with the proportion of the Combined percept and the short time to discover all percepts (i.e., more exploration). This suggests that switching patterns are based on the balance of Glx and GABA concentrations between sensory and suprasensory areas. Although most neuroimaging studies have focused on the role of AC in perceptual organization [47, 48, 51], several researchers have argued that the intraparietal sulcus mediates the figure-ground segregation in auditory scenes [52, 53]. Thus, there is the possibility that an interaction between different cortical areas is responsible for individual differences in auditory multistability. The previous studies mentioned above allowed only two choices for participants to report their percepts of sound sequences, whereas the present study allowed them four alternatives. The latter possibly involves schema-based processes to a larger degree for classifying one's experience. Given that the IFC is associated with perceptual classification, the IFC involvement in perceptual organization should be larger when the classification part of the task becomes more complicated (i.e., when neither of the common categories fit one's perception).

We found different contributions of the AC and IFC to verbal transformations, as well as to auditory streaming. Glx concentrations in the AC were correlated with the durations of Banana and Nappa percepts, whereas GABA concentrations in the IFC were correlated with the proportion of the Other percept. A simple interpretation is that the IFC involvement in verbal transformations depends on speech-specific mechanisms [32, 54]. However, several researchers have argued that the IFC is associated with the generation of perceptual objects even in vision. For the apparent motion quartet, IFC activity occurs earlier during spontaneous perceptual switches but not during stimulus-driven changes [55]. Thus, the temporal precedence of the activation indicates that the IFC participates in initiating the formation of perceptual objects.

We did not find any significant correlation between auditory multistability and personality scales. In contrast, a recent study demonstrated that ego-resiliency is linked to switching patterns of auditory streaming [33]. In this study, a post-hoc power analysis showed that more than 70 participants were needed to achieve a statistical power of 80% for significant correlations between ego-resiliency scores and switching patterns of auditory streaming (at the α -level of 0.05). This suggests that the lack of correlation in the current study was possibly due to low statistical power. Also we did not find any correlation between auditory multistability and response inhibition. Some theoretical models have postulated that perceptual switching is determined by the dynamics of mutual inhibition between neural populations representing each percept [17, 18]. Our results suggest that neurochemical inhibition is not directly related to inhibition of a prepotent response at the cognitive level.

Previous studies have demonstrated that the dopamine tone of individuals is related to idiosyncratic switching patterns of auditory streaming and verbal transformations [12, 13]. However, it is unclear what brain areas are influenced by the different neurotransmitter systems. The present study suggests that AC Glx and IFC GABA concentrations are associated with auditory multistability. This is consistent with findings in the literature of auditory scene analysis, which has shown that perceptual organization involves an interaction between distributed neural circuits below, in, and beyond AC [51, 56, 57]. Further, our findings have clinical implications in that dysfunctions of the GABAergic and glutamatergic systems impact auditory perceptual organization.

Additional Information

Acknowledgments

We thank Keith Heberlein, Mark A. Brown, and Heiko Meyer for providing a GABA spectral editing sequence (a work-in-progress version), and Yasuhiro Shimada and Takanori Kochiyama for helping with imaging data collection.

Ethics

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics and Safety Committees of NTT Communication Science Laboratories and ATR-Promotions.

Data Accessibility

Supplementary Material includes a supplementary text, three figures, and nine tables. The datasets supporting this article are available at <https://figshare.com/s/0531ebc55884a85dd38d>.

Authors' Contributions

H.M.K., D.F., S.L.D., and I.W. designed the study; H.M.K. and T.A. collected data; H.M.K., D.F., S.L.D., and T.A. analysed data; H.M.K., D.F., and I.W. interpreted data; H.M.K., D.F., S.L.D., and I.W. wrote the manuscript. All authors helped draft the manuscript and approved the final version of the article.

Competing Interests

H.M.K. is the Guest Editor of the special issue.

Funding

This research was supported by the Hungarian Academy of Sciences (Lendület Project LP-36/2012 to D.F. and I.W.).

References

- [1] Bregman, A.S. 1990 *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA, MIT Press.
- [2] van Noorden, L.P.A.S. 1975 *Temporal coherence in the perception of tone sequences*. Eindhoven, The Netherlands, PhD thesis, Eindhoven University of Technology.
- [3] Pressnitzer, D. & Hupé, J.M. 2006 Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* **16**, 1351-1357. (doi:10.1016/j.cub.2006.05.054).
- [4] Bendixen, A., Denham, S.L., Gyimesi, K. & Winkler, I. 2010 Regular patterns stabilize auditory streams. *J. Acoust. Soc. Am.* **128**, 3658-3666. (doi:10.1121/1.3500695).
- [5] Denham, S., Böhm, T.M., Bendixen, A., Szalárdy, O., Kocsis, Z., Mill, R. & Winkler, I. 2014 Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. *Front. Neurosci.* **8**, 25. (doi:10.3389/fnins.2014.00025).
- [6] Denham, S., Bendixen, A., Mill, R., Tóth, D., Wennekers, T., Coath, M., Böhm, T., Szalárdy, O. & Winkler, I. 2012 Characterising switching behaviour in perceptual multi-stability. *J. Neurosci. Methods* **210**, 79-92. (doi:10.1016/j.jneumeth.2012.04.004).
- [7] Warren, R.M. 1961 Illusory changes of distinct speech upon repetition – The verbal transformation effect. *Br. J. Psychol.* **52**, 249-258.
- [8] Warren, R.M. & Gregory, R.L. 1958 An auditory analogue of the visual reversible figure. *Am. J. Psychol.* **71**, 612-613.
- [9] Warren, R.M. 1968 Verbal transformation effect and auditory perceptual mechanisms. *Psychol. Bull.* **70**, 261-270.
- [10] Leopold, D.A. & Logothetis, N.K. 1999 Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* **3**, 254-264.
- [11] Miller, S.M., Hansell, N.K., Ngo, T.T., Liu, G.B., Pettigrew, J.D., Martin, N.G. & Wright, M.J. 2010 Genetic contribution to individual variation in binocular rivalry rate. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2664-2668. (doi:10.1073/pnas.0912149107).
- [12] Kondo, H.M., Kitagawa, N., Kitamura, M.S., Koizumi, A., Nomura, M. & Kashino, M. 2012 Separability and commonality of auditory and visual bistable perception. *Cereb. Cortex* **22**, 1915-1922. (doi:10.1093/cercor/bhr266).
- [13] Kashino, M. & Kondo, H.M. 2012 Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. *Philos. Trans. R. Soc. B. Biol. Sci.* **367**, 977-987. (doi:10.1098/rstb.2011.0370).
- [14] Kanai, R., Bahrami, B. & Rees, G. 2010 Human parietal cortex structure predicts individual differences in perceptual rivalry. *Curr. Biol.* **20**, 1626-1630. (doi:10.1016/j.cub.2010.07.027).
- [15] Genç, E., Bergmann, J., Singer, W. & Kohler, A. 2011 Interhemispheric connections shape subjective experience of bistable motion. *Curr. Biol.* **21**, 1494-1499. (doi:10.1016/j.cub.2011.08.003).

- [16] Wilson, H.R. 2003 Computational evidence for a rivalry hierarchy in vision. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14499-14503. (doi:10.1073/pnas.2333622100).
- [17] Moreno-Bote, R., Rinzel, J. & Rubin, N. 2007 Noise-induced alternations in an attractor network model of perceptual bistability. *J. Neurophysiol.* **98**, 1125-1139. (doi:10.1152/jn.00116.2007).
- [18] Noest, A.J., van Ee, R., Nijs, M.M. & van Wezel, R.J.A. 2007 Percept-choice sequences driven by interrupted ambiguous stimuli: a low-level neural model. *J. Vis.* **7**, 1-14. (doi:10.1167/7.8.10).
- [19] Klink, P.C., van Ee, R. & van Wezel, R.J.A. 2008 General validity of Levelt's propositions reveals common computational mechanisms for visual rivalry. *PLOS ONE* **3**, e3473. (doi:10.1371/journal.pone.0003473).
- [20] Huguet, G., Rinzel, J. & Hupé, J.M. 2014 Noise and adaptation in multistable perception: noise drives when to switch, adaptation determines percept choice. *J. Vis.* **14**, 19. (doi:10.1167/14.3.19).
- [21] Hohwy, J., Roepstorff, A. & Friston, K. 2008 Predictive coding explains binocular rivalry: an epistemological review. *Cognition* **108**, 687-701. (doi:10.1016/j.cognition.2008.05.010).
- [22] Logothetis, N.K. 2008 What we can do and what we cannot do with fMRI. *Nature* **453**, 869-878. (doi:10.1038/nature06976).
- [23] Rubenstein, J.L.R. & Merzenich, M.M. 2003 Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255-267. (doi:10.1046/j.1601-183X.2003.00037.x).
- [24] van Loon, A.M., Knapen, T., Scholte, H.S., St. John-Saaltink, E., Donner, T.H. & Lamme, V.A.F. 2013 GABA shapes the dynamics of bistable perception. *Curr. Biol.* **23**, 823-827. (doi:10.1016/j.cub.2013.03.067).
- [25] Kihara, K., Kondo, H.M. & Kawahara, J.I. in press Differential contributions of GABA concentration in frontal and parietal regions to individual differences in attentional blink. *J. Neurosci.*
- [26] Edden, R.A.E., Muthukumaraswamy, S.D., Freeman, T.C.A. & Singh, K.D. 2009 Orientation discrimination performance is predicted by GABA concentration and gamma oscillation frequency in human primary visual cortex. *J. Neurosci.* **29**, 15721-15726. (doi:10.1523/JNEUROSCI.4426-09.2009).
- [27] Snyder, J.S., Gregg, M.K., Weintraub, D.M. & Alain, C. 2012 Attention, awareness, and the perception of auditory scenes. *Front. Psychol.* **3**, 15. (doi:10.3389/fpsyg.2012.00015).
- [28] Moore, B.C.J. & Gockel, H.E. 2012 Properties of auditory stream formation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 919-931. (doi:10.1098/rstb.2011.0355).
- [29] Bendixen, A. 2014 Predictability effects in auditory scene analysis: a review. *Front. Neurosci.* **8**, 60. (doi:10.3389/fnins.2014.00060).
- [30] Windmann, S., Wehrmann, M., Calabrese, P. & Güntürkün, O. 2006 Role of the prefrontal cortex in attentional control over bistable vision. *J. Cogn. Neurosci.* **18**, 456-471. (doi:10.1162/089892906775990570).
- [31] Knapen, T., Brascamp, J., Pearson, J., van Ee, R. & Blake, R. 2011 The role of frontal and parietal brain areas in bistable perception. *J. Neurosci.* **31**, 10293-10301. (doi:10.1523/JNEUROSCI.1727-11.2011).
- [32] Kondo, H.M. & Kashino, M. 2007 Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage* **36**, 123-130. (doi:10.1016/j.neuroimage.2007.02.024).
- [33] Farkas, D., Denham, S.L., Bendixen, A., Tóth, D., Kondo, H.M. & Winkler, I. 2016 Auditory multi-stability: idiosyncratic perceptual switching patterns, executive functions and personality traits. *PLOS ONE* **11**, e0154810. (doi:10.1371/journal.pone.0154810).
- [34] John, O.P. & Srivastava, S. 1999 The big-five trait taxonomy: history, measurement, and theoretical perspectives. In *Handbook of Personality: Theory and Research* (eds L.A. Pervin & O.P. John), pp. 102-138, 2nd ed.
- [35] Whiteside, S.P. & Lynam, D.R. 2001 The five factor model and impulsivity: using a structural model of personality to understand impulsivity. *Pers. Individ. Differ.* **30**, 669-689.
- [36] Block, J.H. & Kremen, A.M. 1996 IQ and ego-resiliency: conceptual and empirical connections and separateness. *J. Pers. Soc. Psychol.* **70**, 349-361.
- [37] Farkas, D. & Orosz, G. 2015 Ego-resiliency reloaded: a three-component model of general resiliency. *PLOS ONE* **10**, e0120883. (doi:10.1371/journal.pone.0120883).
- [38] Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A. & Wager, T.D. 2000 The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognit. Psychol.* **41**, 49-100.
- [39] Oldfield, R.C. 1971 The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97-113.
- [40] Edden, R.A.E., Puts, N.A.J., Harris, A.D., Barker, P.B. & Evans, C.J. 2014 Gannet: a batchprocessing tool for the quantitative analysis of gamma-aminobutyric acid-edited MR spectroscopy spectra. *J. Magn. Reson. Imaging* **40**, 1445-1452. (doi:10.1002/jmri.24478).
- [41] Mullins, P.G., McGonigle, D.J., O'Gorman, R.L., Puts, N.A.J., Vidyasagar, R., Evans, C.J., Cardiff Symposium on MRS of GABA & Edden, R.A.E. 2014 Current practice in the use of MEGA-PRESS spectroscopy for the detection of GABA. *Neuroimage* **86**, 43-52. (doi:10.1016/j.neuroimage.2012.12.004).
- [42] Farkas, D., Denham, S.L., Bendixen, A. & Winkler, I. 2016 Assessing the validity of subjective reports in the auditory streaming paradigm. *J. Acoust. Soc. Am.* **139**, 1762-1772. (doi:10.1121/1.4945720).
- [43] Moreno-Bote, R., Shpiro, A., Rinzel, J. & Rubin, N. 2010 Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J. Vis.* **10**, 1-18. (doi:10.1167/10.11.1).
- [44] Denham, S.L. & Winkler, I. 2006 The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* **100**, 154-170. (doi:10.1016/j.jphysparis.2006.09.012).
- [45] Kruskal, J.B. 1964 Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1-27.
- [46] Micheyl, C., Tian, B., Carlyon, R.P. & Rauschecker, J.P. 2005 Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* **48**, 139-148. (doi:10.1016/j.neuron.2005.08.039).
- [47] Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M. & Oxenham, A.J. 2005 Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* **25**, 5382-5388. (doi:10.1523/JNEUROSCI.0347-05.2005).
- [48] Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B. & Wilson, E.C. 2007 The role of auditory cortex in the formation of auditory streams. *Hear. Res.* **229**, 116-131. (doi:10.1016/j.heares.2007.01.007).
- [49] Wilson, E.C., Melcher, J.R., Micheyl, C., Gutschalk, A. & Oxenham, A.J. 2007 Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J. Neurophysiol.* **97**, 2230-2238. (doi:10.1152/jn.00788.2006).
- [50] Brascamp, J., Blake, R. & Knapen, T. 2015 Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nat. Neurosci.* **18**, 1672-1678. (doi:10.1038/nn.4130).
- [51] Kondo, H.M. & Kashino, M. 2009 Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* **29**, 12695-12701. (doi:10.1523/JNEUROSCI.1549-09.2009).
- [52] Cusack, R. 2005 The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* **17**, 641-651. (doi:10.1162/0898929053467541).
- [53] Teki, S., Chait, M., Kumar, S., von Kriegstein, K. & Griffiths, T.D. 2011 Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci.* **31**, 164-171. (doi:10.1523/JNEUROSCI.3788-10.2011).
- [54] Sato, M., Beciu, M., Loevenbruck, H., Schwartz, J.L., Cathiard, M.A., Segebarth, C. & Abry, C. 2004 Multistable representation of speech forms: a functional MRI study of verbal transformations. *Neuroimage* **23**, 1143-1151. (doi:10.1016/j.neuroimage.2004.07.055).
- [55] Sterzer, P. & Kleinschmidt, A. 2007 A neural basis for inference in perceptual ambiguity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 323-328. (doi:10.1073/pnas.0609006104).
- [56] Kondo, H.M., Pressnitzer, D., Toshima, I. & Kashino, M. 2012 Effects of self-motion on auditory scene analysis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6775-6780. (doi:10.1073/pnas.1112852109).
- [57] Schadwinkel, S. & Gutschalk, A. 2011 Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J. Neurophysiol.* **105**, 1977-1983. (doi:10.1152/jn.00461.2010).

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

For Review Only

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

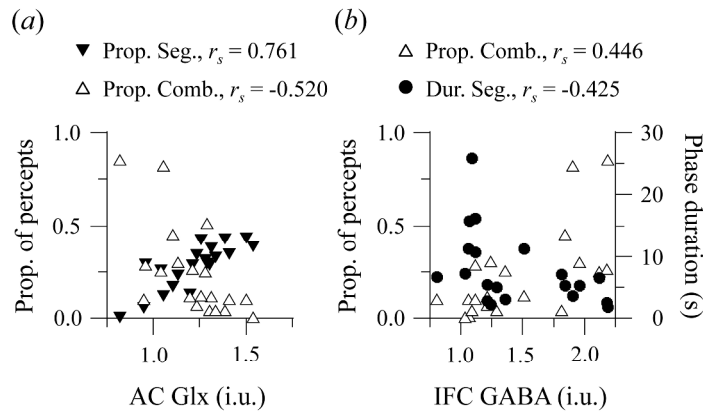


Figure 1

Figure 1

297x420mm (300 x 300 DPI)

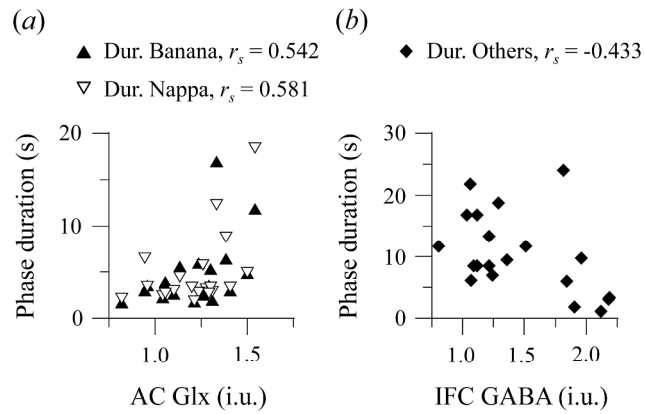


Figure 2

Figure 2

297x420mm (300 x 300 DPI)