PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2012) **367**, 977–987
doi:10.1098/rstb.2011.0370

Research

Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations

Makio Kashino[1,2,*] and Hirohito M. Kondo[1]

[1]*NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan*
[2]*Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8503, Japan*

Recent studies have shown that auditory scene analysis involves distributed neural sites below, in, and beyond the auditory cortex (AC). However, it remains unclear what role each site plays and how they interact in the formation and selection of auditory percepts. We addressed this issue through perceptual multistability phenomena, namely, spontaneous perceptual switching in auditory streaming (AS) for a sequence of repeated triplet tones, and perceptual changes for a repeated word, known as verbal transformations (VTs). An event-related fMRI analysis revealed brain activity timelocked to perceptual switching in the cerebellum for AS, in frontal areas for VT, and the AC and thalamus for both. The results suggest that motor-based prediction, produced by neural networks outside the auditory system, plays essential roles in the segmentation of acoustic sequences both in AS and VT. The frequency of perceptual switching was determined by a balance between the activation of two sites, which are proposed to be involved in exploring novel perceptual organization and stabilizing current perceptual organization. The effect of the gene polymorphism of catechol-*O*-methyltransferase (COMT) on individual variations in switching frequency suggests that the balance of exploration and stabilization is modulated by catecholamines such as dopamine and noradrenalin. These mechanisms would support the noteworthy flexibility of auditory scene analysis.

Keywords: auditory scene analysis; multi-stable perception; functional magnetic resonance imaging; catechol-*O*-methyltransferase gene polymorphism; genotype; awareness

1. INTRODUCTION

Auditory perception plays indispensable roles in everyday life, such as enabling us to understand what is occurring where, communicate orally, and enjoy music. All these functions depend critically on the listener's internal process of organizing complex acoustic signals into coherent streams that usually correspond to sound sources. This process is called auditory scene analysis [1]. Conscious percepts can be considered a likely interpretation of the external auditory scene.

Over the past decade, significant progress has been made in understanding where and how auditory streams are formed in the brain (for recent reviews [2–4]). A number of studies have identified neural correlates of auditory streaming (AS) consistently in the auditory cortex (AC; or its avian homologue, the field L), with various techniques such as single- or multi-unit recordings for mammals [5–7] and avians [8–10], electroencephalography (EEG) [11–16], magnetoencephalography (MEG) [17–19] and functional magnetic resonance imaging (fMRI) [20,21] for humans.

However, these data do not necessarily mean that streams are formed *in* the AC. A recent study has demonstrated that neural response patterns in the cochlear nucleus (CN; the first nucleus in the auditory pathway) in anaesthetized guinea pigs are consistent with several psychophysical features of the perceptual organization of alternating tones in human listeners [22]. This finding is particularly important, because it suggests a possibility that neural activities corresponding to streams have already been created in subcortical neural sites including the CN, and in extreme cases, the neural correlates of streams observed in the AC may be a simple reflection of those created in the subcortical sites. Alternatively, the neural correlates of streams found in the CN may be created by top-down modulation from the AC through the descending auditory pathway. However, it is not clear which is the case.

Neural correlates of streams have also been found beyond the AC. In an fMRI study, the intraparietal sulcus (IPS), which is implicated in perceptual organization in vision, crossmodal binding and selective attention, showed greater activation when listeners

* Author for correspondence (kashino.makio@lab.ntt.co.jp).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2011.0370 or via http://rstb.royalsocietypublishing.org.

One contribution of 10 to a Theme Issue 'Multistability in perception: binding sensory modalities'.

perceived an ambiguous tone sequence as two streams (S2) than when they perceived the same sequence as a single stream (S1) [23]. The interpretation of this finding is again ambiguous. One possibility is that auditory streams are already formed in or below the AC, and the IPS simply reflects them, for binding with the output of perceptual organization processes in other modalities. Another possibility is that the IPS activation is related to attention, which brings one of the already formed potential streams into awareness. Yet another possibility is that the IPS is related to attention, which modulates the way streams are formed in or below the AC. The critical point here is how selective attention influences streaming, which is a matter of continuing debate [24]. Some psychophysical studies indicate that attention can modulate stream formation under certain conditions [25,26]. On the other hand, EEG studies generally demonstrate that the initial stages of stream formation are stimulus-driven and automatic, and attention may affect only later stages where streams have been formed [11,13,16,27]. In any case, the role of the IPS in AS is not clear.

A related issue that also requires further research is the neural basis of schema-based processes in auditory scene analysis. So far, most studies have been devoted to primitive processes, which are assumed to operate at early levels of auditory information processing to analyse the auditory scene based on a set of simple acoustic rules that are likely to hold for a broad range of sound sources and events in general [1,28,29]. However, it is likely that listeners also use knowledge or schema of specific targets, such as speech, music and environmental sounds, when analysing an everyday auditory scene [24]. But what is the relationship between primitive and schema-based processes? A possibility is that source-specific knowledge stored in higher level brain areas feeds back to lower level brain areas that perform primitive auditory scene analysis. It is also possible that schema-based scene analysis in higher level brain areas is based on feedforward information from lower level brain areas.

Thus, although it is now probable that auditory scene analysis involves broadly distributed neural sites below, in, and beyond the AC, further research is needed to clarify the specific contribution of each neural site, and the functional connectivity and causal relationships among the relevant neural sites, in the formation and selection of streams. Clarifying these points would provide important clues about how a primitive analysis of acoustic features, knowledge of specific sound sources and selective attention interact to achieve auditory scene analysis.

We have addressed these issues through multistability in the perceptual organization of repeated acoustic patterns. Prolonged listening to a repeated triplet-tone sequence (ABA; A and B tones are at different frequencies) produces a series of perceptual switches between S1 and S2 [25,30–32]. Similarly, a series of perceptual changes can be produced by prolonged listening to a repeated word without a pause, which are called verbal transformations (VTs). For instance, 'tress' may be transformed into a variety of verbal forms, such as 'dress', 'stress', 'drest' or even 'Esther' [33,34]. The dissociation between physical stimulation and percepts in these multistability phenomena provides an effective means of identifying neural sites causally involved in the formation and selection of percepts (i.e. streams or verbal forms). We assumed that if some brain area is causally involved in the formation and selection of percepts, then the response in that area should be timelocked to reported perceptual switching within a listener. We also noticed significant individual variation in the number of perceptual switches both in AS and VTs, as reported earlier for visual multistable phenomena such as binocular rivalry [35]. The individual variation could provide an additional clue to the neural correlates of AS. If the magnitude of the response in a certain brain area correlates with the number of perceptual switches across listeners, it would indicate that the area may play a critical role in the formation and selection of percepts.

In a study of bistability in AS for a repeated triplet-tone sequence [36], we combined the use of different frequency differences (Δfs; approx. 2 and 6 semitones centred at 1000 Hz) with an event-related fMRI design to examine whether the temporal dynamics of brain activity differs depending on the direction of the perceptual switches. The dominant percept depends on Δf between high and low tones [24]. The results demonstrated that the activity of the medial geniculate body (MGB) in the thalamus occurred earlier during switching from non-dominant to dominant percepts, whereas that of the AC occurred earlier during switching from dominant to non-dominant percepts, irrespective of Δf. The asymmetry of temporal precedence indicates that the MGB and AC activations play different roles in perceptual switching and depend on perceptual dominance rather than on S1 and S2 percepts *per se*. The results suggest that feedforward and feedback processes in the thalamocortical loop are essential in the formation of percepts in AS.

In a study of VTs for the repeated word 'banana' [37], we conducted an event-related fMRI analysis, which revealed that the left inferior frontal cortex (IFC), anterior cingulate cortex (ACC) and left prefrontal cortex (PFC) were activated when there were perceptual changes from one verbal form to another, but not when there were tone pips superimposed on the repeated word sequence. The number of perceptual changes showed positive and negative correlations with the signal intensity in the left IFC and the left ACC, respectively. The results suggest that the active generation of verbal forms may be linked with articulatory gestures for speech production, and that the frequency of perceptual switches is determined by a balance between the activations of the two brain regions. Structural equation modelling (SEM) demonstrated that individual differences in the number of perceptual changes depend on negative feedback from the ACC to the IFC via the posterior insular cortex (PIC). These findings suggest that distributed frontal areas are involved in the formation and selection of the percepts underlying VTs. The areas identified in this study largely overlap those found in an fMRI study on VTs produced by the mental rehearsal of a repeated word [38], and in an event-related intracerebral EEG study of perceptual switching in VTs for two implanted

epileptic patients [39]. The findings of these three VT studies are consistent with the dual-stream model of speech processing, which assumes that a ventral auditory stream maps sounds onto meaning whereas a dorsal stream maps sounds onto articulatory-based representations [40], and a general model for auditory–motor transformations in which the dorsal stream is characterized as the 'do-pathway' [41].

Our two studies, using a quite similar paradigm except for the stimuli, demonstrated the involvement of different brain networks in perceptual switching, but both highlighted the importance of the interaction of distributed sites. In the present paper, we re-analyse the data from those two studies to examine further the commonality and differences between AS (as an example of primitive scene analysis) and VTs (as an example of schema-based scene analysis). The analysis revealed the critical involvement of the cerebellum (Cb) in AS and the caudate nucleus (Cd) in VTs, neither of which were considered in our previous studies. We discuss this new finding in terms of the idea of motor involvement in the formation of percepts, not only in VTs as pointed out earlier, but also in AS.

To gain further insights into the neural basis of perceptual switching, we also conducted a new experiment to examine how neurotransmitters affect the individual variation in the number of perceptual switches in the two tasks, by means of the gene polymorphism of catechol-*O*-methyltransferase (COMT), which plays an important role in the degradation of catecholamines such as dopamine and noradrenalin [42]. A functional single-nucleotide polymorphism of the gene for COMT results in a methionine to valine mutation at position 158 (Val158Met). The Val variant catabolizes catecholamine at up to four times the rate of its methionine counterpart, resulting in significantly lower synaptic catecholamine levels following neurotransmitter release. Although the genetic effects of COMT on perception are unclear, a recent EEG study showed that the amplitude of the N100 component was smaller for Met/Met individuals than for Val/Met and Val/Val individuals during an auditory task, suggesting that the COMT genotype is associated with poor sensory gating of auditory stimuli [43]. It is possible that auditory multi-stable perception is also modulated by the COMT genotype. The results will be discussed in terms of the balance between exploration and stabilization in the formation and selection of percepts.

The two ideas, the motor involvement and the exploration–stabilization opponency, point to new directions for research on the formation and selection of auditory percepts.

2. BEHAVIOURAL, FUNCTIONAL IMAGING AND GENOTYPING EXPERIMENTS

(a) *Method for behaviour–neuroimaging experiments*

The methods used for the behaviour–neuroimaging experiments are described in the electronic supplementary material. In the AS task, the stimuli were 225 repetitions of a triplet tone that comprised high and low tones with intervals of silence. Two Δfs were used in the experiment in Kondo & Kashino [36], but only the results for the Δf of 2 semitones are used for analysis here, to match the number of participants with the VT task. In the VT task, the stimuli were 265 repetitions of the word 'banana'. In both tasks, the participants were instructed to listen to the sound sequence and indicate by a button press whenever they detected perceptual changes. We assigned 24 participants to five 90 s runs of AS or VTs (12 participants for each task).

Table 1. Activated brain areas derived from conjunction and subtraction analyses. Coordinates (x, y, z) indicate the voxel of maximal significance in each brain region (false discovery rate, FDR <0.01). L, left; R, right. See figure 1 legend for the abbreviations.

region		x	y	z	t-value
(*a*) *conjunction analysis*					
PIC	L	−38	0	8	6.56
	R	36	−6	8	6.09
AC	L	−52	−24	12	4.22
	R	60	−16	10	4.50
MGB	L	−14	−26	0	5.34
	R	12	−22	0	4.69
(*b*) *subtraction analysis* (*AS minus VT*)					
PIC	L	−46	−20	18	5.30
MGB	R	18	−26	0	4.12
Cb	L	−16	−56	−20	3.90
(*c*) *subtraction analysis* (*VT minus AS*)					
PFC	L	−40	38	20	5.44
	R	40	42	32	3.81
ACC	L	−6	42	28	4.58

(b) *Common and different brain activations*

We first performed a conjunction analysis to identify any common activation for the two tasks. The AC, MGB and PIC were activated bilaterally during perceptual switches (table 1*a*). AC activations were localized along the Heschl gyrus and extended to the posterior part of the superior temporal cortex. These results indicate that auditory-related areas play an essential role in the formation of auditory percepts regardless of stimulus type. It is unlikely that these activations reflect changes in physical inputs because we used a tone detection task in the triplet or word sequences as a baseline in both tasks.

We identified task-dependent activations timelocked with perceptual switches in AS and VTs (figure 1). AC activations in VTs were widely spread on the planum temporale, compared with those in AS, and the insular activation distribution was larger for the former. We found activations of the frontal and subcortical areas, as well as the auditory-related areas, in VTs. In particular, speech-specific perceptual changes may be modulated by activations of the prefrontal cortex (PFC), IFC, ACC and Cd.

Next we conducted a subtraction analysis to directly compare activations between the two tasks (figure 2). The left PIC, right MGB and anterior lobe (lobules IV and V) of the left Cb were activated in the contrast of AS minus VTs (table 2*b*), whereas the bilateral PFC and left ACC were activated in the contrast of VTs minus AS (table 2*c*).

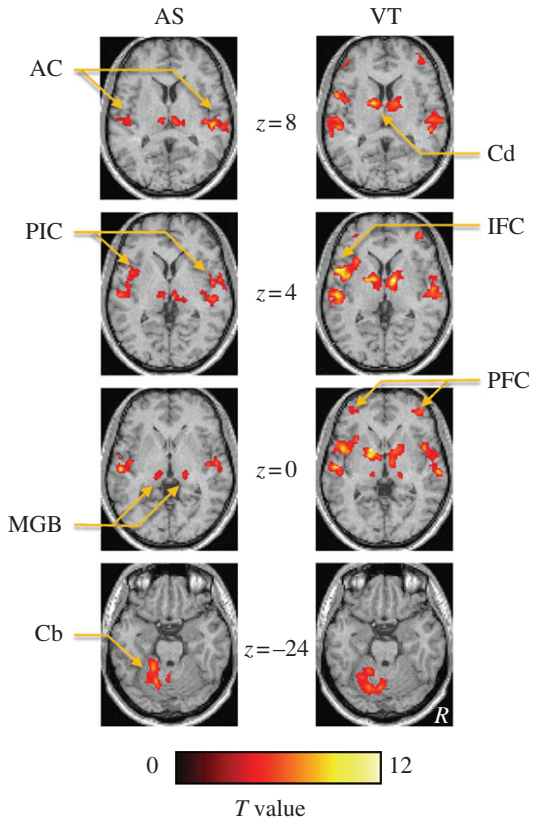

Figure 1. Auditory streaming (AS) and verbal transformation (VT) activations ($n = 12$ for each). Coloured areas on the horizontal plane of the Montreal Neurological Institute template indicate significant activations timelocked with perceptual switches ($p < 0.001$, uncorrected). AC, auditory cortex; Cb, cerebellum; Cd, caudate nucleus; IFC, inferior frontal cortex; MGB, medial geniculate body; PFC, prefrontal cortex; PIC, posterior insular cortex.

(c) *Brain–behaviour relationship in number of perceptual switches*

We found that brain areas related to individual variations in the number of perceptual switches differed between the two tasks (figure 3). In AS, the magnitude of the left Cb activation increased with an increase in the number of perceptual switches, whereas that of the right MGB showed a corresponding decrease. In VTs, the magnitude of the left IFC activation increased with an increase in the number of perceptual switches, whereas that of the left ACC activation decreased. These results indicate that different networks modulate the number of perceptual switches in the AS and VT tasks, and in each network the balance between the two sites affects the number of perceptual switches.

It should be noted that the number of perceptual switches is related in a non-trivial way to the sensitivity of participants to parametric manipulations. Moreover, some of the correlations shown in figure 3 are not significant. Further studies are necessary to establish the brain–behaviour relationship in the number of perceptual switches.

(d) *Functional networks underlying perceptual switching*

We created inter-region networks to further examine the effective connectivity between brain areas in

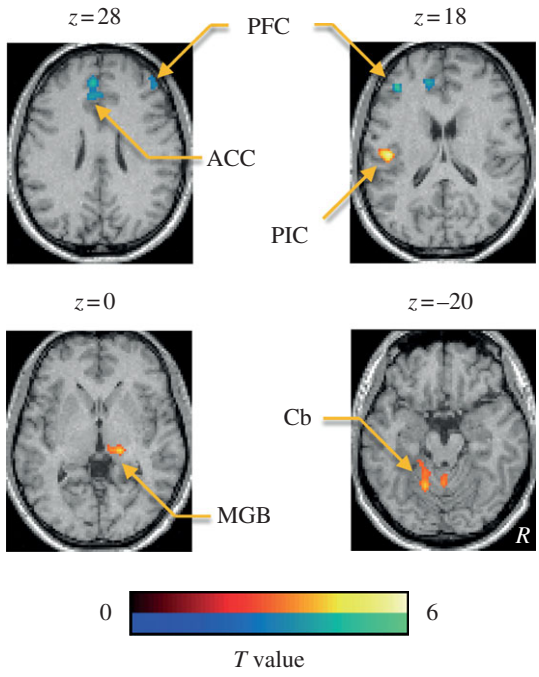

Figure 2. Results of subtraction analyses. The orange areas indicate greater activation for AS than for VTs, whereas blue indicates greater activation for VTs than for AS. A lax threshold was chosen for display purposes ($p < 0.005$, uncorrected). ACC, anterior cingulate cortex. See figure 1 legend for the abbreviations.

Table 2. Fit indices for the best-fitting models. Non-significant χ^2 statistics indicate a good fit to time-series data for brain areas. The data for each group are derived from four individuals. A lower SRMR value and higher CFI value both represent a good fit. SRMR, standardized root mean-squared residual; CFI, Bentler comparative fit index.

group	χ^2	d.f.	p	SRMR	CFI
auditory streaming					
individuals with high frequency	68.30	3	0.001	0.090	0.87
individuals with low frequency	79.22	3	0.001	0.104	0.77
verbal transformations					
individuals with high frequency	69.41	5	0.001	0.071	0.87
individuals with low frequency	171.40	5	0.001	0.094	0.84

modulating the number of perceptual switches. We concentrated on task-dependent activations in AS and VTs and used SEM analysis to compare a best-fitting model of inter-region networks for groups with high- and low-frequency switching. SEM analysis provides the synchronization strength of signal changes between brain areas as path coefficients and estimates fit indices of structural models (for details, see [36,37]).

We estimated a best-fitting model that would account for signal changes of local maxima in regions of interest [44,45]. Most values of fit indices in the selected model did not reach the standard criteria for statistical significance. However, the values were

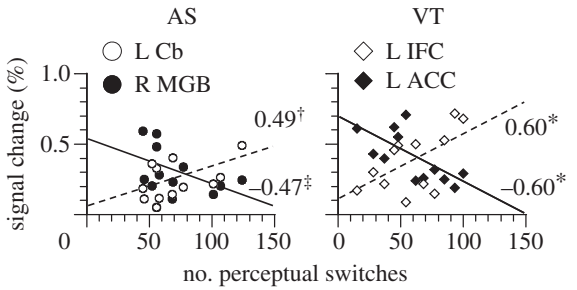

Figure 3. Correlations between the number of perceptual switches and signal changes. Circles indicate individual data ($n = 12$). Signal changes represent average magnitudes of haemodynamic response during perceptual switches. $^{*}p < 0.05$, $^{\dagger}p < 0.10$ $^{\ddagger}p < 0.20$. See figures 1 and 2 legends for the abbreviations.

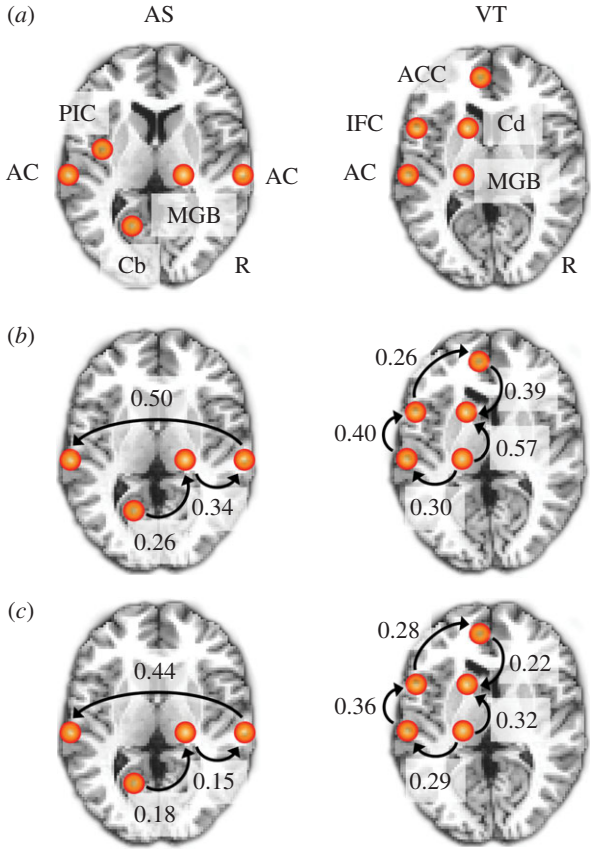

Figure 4. (*a*) Default setting and best fitting models in AS and VTs. Inter-region networks for groups consisting of individuals with (*b*) high- and (*c*) low-frequency perceptual switching. Signal changes in brain areas are standardized into zero mean and unit variance for each 90 s run, and the standardized data are set as observed variables (sample size, $n = 900$). The models are computed using the maximum-likelihood method. All of the path coefficients are significant ($p < 0.01$).

close to the standard criteria and better than those for other possible models (table 2). First, we postulated a network model consisting of the AC, MGB, PIC and Cb in AS (figure 4*a*). We removed PIC activity from the network for simplicity and obtained a best-fitting model for each group. The magnitude of the path coefficient from the right MGB to the right AC was greater for the high-frequency group than for the

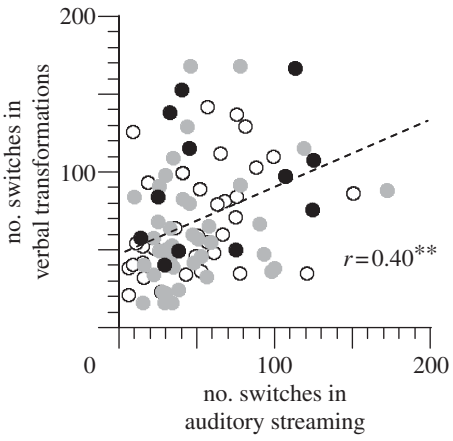

Figure 5. Correlation of the number of perceptual switches between AS and VTs. Circles and diamonds indicate individual data (12 Met/Met, 45 Val/Met and 35 Val/Val individuals). The stimuli and task procedure used in this experiment are identical to those used in the fMRI experiment (see text). Black circles, Met/Met; grey circles, Val/Met; open circles, Val/Val. $^{**}p < 0.01$.

low-frequency group: 0.34 (95% confidence interval 0.28–0.42) and 0.15 (0.07–0.24). Second, we assumed a network model consisting of the ACC, AC, IFC, Cd and MGB in VTs (figure 4*b*). The magnitude of the path coefficient from the left MGB to the left Cd was greater for the high-frequency group than for the low-frequency group: 0.57 (0.50–0.63) and 0.32 (0.25 to 0.37). The magnitude of the path coefficient from the left ACC to the left Cd was greater for the high-frequency group than for the low-frequency group: 0.39 (0.32–0.45) and 0.22 (0.15–0.28). The results confirm that pivotal areas affecting the perceptual switching number differ for AS and VTs. For both tasks, however, it should be noted that interactions of widely distributed sites in cortical and subcortical areas contribute to individual differences in the formation and selection of auditory percepts.

(e) *Genotype–behaviour relationship*

Thus far, the analyses of individual differences in the number of perceptual switches have been performed for different groups of listeners for the AS and VT tasks. To see whether the perceptual switching numbers for the two tasks are correlated for the same group of listeners, we conducted an additional experiment [42]. We newly recruited 92 participants (see the electronic supplementary material for technical details). They performed the two tasks (AS and VTs) using the same stimuli and procedure as used for the experiments described above. We found that the numbers of perceptual switches were 50.8 ± 3.9 (mean ± s.e.m.) for AS and 68.4 ± 4.5 for VTs, $t_{91} = 3.76$, $p < 0.001$. There was a significant positive correlation between the numbers of perceptual switches in the two tasks, $r = 0.40$, $p < 0.01$ (figure 5). This indicates that perceptual switching in AS and VTs shares a common neural mechanism to a certain extent.

To obtain further clues to the neural bases of the individual variation in auditory perceptual switching, we examined how neurotransmitter functioning

influences the number of perceptual switches. We focused on the functional polymorphism of the COMT Val^{158}Met gene, which plays an important role in the degradation of catecholamines such as dopamine and noradrenalin. We identified the COMT polymorphism of all the participants, and categorized them into three groups on the basis of their alleles: 18 Met/Met, 42 Val/Met and 32 Val/Val individuals. An ANOVA revealed that the number of perceptual switches in the AS task was greater for the Met/Met group (mean $\pm$ s.e.m.: 77.7 ± 14.0) than for the Val/Met group (46.7 ± 4.2) or the Val/Val group (44.2 ± 5.1); $F_{2,89} = 5.72$, $\eta^2 = 0.013$, partial $\eta^2 = 0.0617$, $p < 0.01$. In the VT task, the number of perceptual switches was greater for the Met/Met group (90.2 ± 14.8) than for the Val/Met group (60.5 ± 5.7) or the Val/Val group (66.2 ± 5.9); $F_{2,89} = 3.23$, $\eta^2 = 0.007$, partial $\eta^2 = 0.0358$, $p < 0.05$. These results indicate that the numbers of perceptual switches both in AS and VTs are modulated by catecholamines such as dopamine and noradrenalin. The values of η^2 and partial η^2 could qualify as 'small to medium' size effects, according to Cohen's arbitrary scale [46].

3. DISCUSSION

(a) *Distributed sites and distinct networks*

Here, we seek to reveal the neural sites and their interactions causally involved in the formation and selection of auditory percepts, taking advantage of the covariation between reported perceptual switching in multistable phenomena and neural activations measured by fMRI within a listener. We also examined the commonality and difference between primitive and schema-based scene analysis processes by comparing AS and VTs. In both AS and VTs, neural activity was found to be timelocked to perceptual switching in the AC, MGB and PIC. These sites are considered to constitute the core of auditory scene analysis. In addition, the Cb played a significant role in AS, whereas the frontal loop consisting of the IFC, ACC, DLPFC and Cd was involved in VTs. Apparently, the formation and selection of auditory percepts involve functional networks widely distributed in cortical and subcortical areas, and the contribution of each network depends on stimulus types and tasks.

The interaction between the AC and MGB in AS has already been discussed in detail in Kondo & Kashino [36]. The new analysis revealed that the thalamocortical loop is also involved in VTs. A SEM analysis demonstrated that the MGB is functionally connected to the Cb in AS and to the Cd in VT, and the strength of those connections depended on the frequency of the perceptual switches. These findings suggest a possibility that the MGB acts as a hub in the formation of auditory percepts, communicating directly with those subcortical sites outside the auditory pathway in addition to the AC.

In the following sections, we discuss the functional significance of certain neural sites, networks and neurotransmitters. The discussion is based on the concept of predictive coding. Predictive coding, or essentially similar ideas, has a long history in perception research [47,48] and in signal processing technologies. It is now widely used in conjunction with a Bayesian inference framework in various domains in cognitive neuroscience, because it provides convincing explanations for important aspects of cognitive processes, such as robustness, efficiency and plasticity [49,50]. The basic concept of the predictive coding approach to perception is that perception is a process that generates testable hypotheses about the causes of its sensory input, based both on prior knowledge (or in Bayesian terms, the prior probability of the hypotheses) and the current sensory input. The predictions produced by the hypotheses are then compared with the sensory input. The most probable hypothesis (or the hypothesis with the highest posterior probability) given the input serves as percepts. The prediction error, or the difference between the prediction and the sensory input, is important because it indicates the occurrence of a new event, or the inappropriateness of the current hypothesis, leading to a switch to a new hypothesis with the smallest prediction error.

In the field of auditory scene analysis, the predictive coding approach has been applied mainly to the extraction of temporal regularities in acoustic sequences [30,50–52], but it could provide a unified framework that accounts for both the primitive and schema-based processes of auditory scene analysis, if we regard the primitive process as predictive coding based on the statistical structures of general acoustic events in the real world, and the schema-based process as predictive coding using the generative models of specific acoustic events such as voices or musical instruments. However, the neural bases of predictive coding in the formation of auditory percepts remain unclear. The experimental results reported here provide some clues.

(b) *Possible roles of motor-based predictions for segmentation of acoustic sequence*

The critical contribution of the Cb to AS is a new finding of the present analysis. The previous analysis [36] did show significant activation timelocked to perceptual switching in the Cb, but we did not discuss its functional role, owing to ambiguity in interpretation. The activation, mainly found in the left hemisphere of the Cb, could be a motor response artefact, because the listeners were asked to press a button with their left thumb when they perceived perceptual switching. The present re-analysis has made this possibility less likely, because the Cb activation was also consistently shown in the subtraction of tone detection from perceptual switching and of AS from VTs, in all of which the listeners' tasks were similar (i.e. a button press with the left thumb at the point of perceptual switching or tone detection). Therefore, we conclude that the Cb is causally involved in perceptual switching in AS.

Before discussing the functional role of the Cb in AS, we would like to point out an important difference between visual and auditory multistability phenomena. Multistable stimuli in vision, such as ambiguous figures and binocular rivalry, do not usually have a temporal structure, as is obvious for static images. Moving stimuli, such as ambiguous moving plaids, simply keep

moving continuously with no meaningful temporal 'chunks'. Prolonged presentation of these static or moving stimuli would provide no new information, or additional ambiguity, to the interpretation or identification of targets. On the other hand, multistable stimuli in hearing have a distinctive (and often unambiguous) temporal structure when presented without repetition. And when they are presented repeatedly without a pause, a new temporal structure is introduced. For example, a sequence of repeated triplets of ABA tones, which has been widely used in AS studies [25], is often referred to as 'bistable', but it can be segmented in multiple ways such as ABA-ABA-. . ., A-A-A-A-. . ., B---B---, A-AB/A-AB, which have actually been reported by some listeners. Here, the essence of ambiguity is not only in spectral grouping (i.e. S1 or S2), but also in the segmentation of the temporal structure (i.e. distinctive rhythm patterns). Similarly, in VTs, a clear, unambiguous utterance (e.g. 'tress') can produce percepts of other verbal forms (e.g. 'stress' and 'Esther') when repeated without pause [33,34]. Apparently, the repetition introduces a new temporal structure with ambiguity in the temporal segmentation of word boundaries.

The segmentation of temporal sequences is not essential only in the perception of artificial acoustic stimuli such as repeated triplets and words. The sounds we encounter in daily life, including speech, music and environmental sounds often have repetitive structures to some extent, which serve as units of perception. The organization of temporal sequences into recognizable chunks, such as words or syllables in speech and rhythm or beats in music, is a fundamental function of auditory information processing. The temporal structure can be hierarchical, and inherently ambiguous. In a sense, repeated sequences used in auditory multistability experiments can be considered extremely simplified versions of music, speech or environmental sounds. The temporal organization of sounds is also important when we selectively attend to a particular target and detect a new sound in a multisource acoustic environment, where the extraction of temporal regularity and deviation from it provide important cues [30,51–53].

Now we propose that the Cb plays an essential role in the segmentation of acoustic sequences, especially in the detection of periodicity or rhythm in the range of hundreds of milliseconds. Anatomically, the Cb is connected reciprocally with the MGB and AC [54,55]. Functionally, the Cb has been implicated in temporal processing [56], controlling motor timing especially on shorter timescales (millisecond) [57,58], and synchronization to rhythm [59,60]. More fundamentally, the Cb performs feedforward and error-correction computations [61,62] and sensory–motor integration [63,64]. The accurate timing of body movements is based on a feedforward prediction of the timing of an upcoming movement, and the use of sensory feedback information to modify and correct subsequent movements. Moreover, when subjects perform purely auditory perceptual tasks, neuroimaging studies consistently show cerebellar (left lateral crus I area) activity [65]. Therefore, it is plausible that the Cb generates an accurate prediction about subsecond periodicity, and compares it with sensory information from the auditory system, to extract the temporal regularities of acoustic sequences.

Cb activation was found in VTs (figure 1), but did not remain in the subtraction analysis (figure 2). This may be due to the difference between the stimuli and tasks of AS and VTs. The stimulus for AS is simpler and isochronous, whereas the stimulus for VTs has a more complex spectrotemporal pattern and requires speech-specific processing in the frontal network. The involvement of the frontal network in VTs has already been discussed in Kondo & Kashino [37], as well as in other studies [38,39]. However, the involvement of the Cd has not been reported. Below we focus on the functional significance of the Cd.

The Cd is a part of the basal ganglia, which has parallel feedback loops with several cortical areas. It is crucially involved in motor control, cognition and emotional processing. The Cd is implicated in the automatic execution of overlearned movement patterns and the planning of non-routine movement patterns [66]. Although, the Cd and Cb both process rhythmic information [67,68], their roles are different. The Cb acts as a precision clock to mediate analysis of absolute, duration-based timing information, whereas the Cd, as a part of a striato-thalamo-cortical network, processes relative, beat-based timing information [69]. The Cb mainly processes timing information on short timescales (less than 1 s), whereas the Cd processes timing information on longer timescales (1 s and above) [70]. The Cd is involved in the processing of complex temporal structures and the perception of predictable cues (regular beats, metre, temporal chunks, etc.) [68,71]. Moreover, the Cd is implicated in expectation. The Cd codes for breaches in expectation, and points towards a distributed network involved in detecting, signalling and adjusting behaviour and expectations towards violated predictions [72,73]. The Cd, together with the anterior and posterior cingulate and thalamus, is involved with target detection and novelty processing of auditory stimuli [74].

In summary, the Cd plays significant roles in the production and learning of motor sequences, the generation of predictions based on motor models and the detection of prediction errors. These functions are somewhat similar to those of the Cb, but the Cd covers longer, more complex temporal structures than the Cb. Here, we propose that the Cd provides the basic segmentation, or chunking, of speech. Finding the boundaries of phonetic segments and words in continuous speech is usually effortless and automatic in native or well-learned languages, but not in unfamiliar languages. For example, non-native speakers of English who are familiar with American English can find British English hard to understand owing to its different rhythmic structure. Speech is less periodic and temporally more variable than music. Finding appropriate segmentations requires the processing of relative timing on longer timescales rather than accurate periodicity detection or duration judgement on shorter timescales. These requirements provide a good fit with the functions of the Cd. We propose that the Cd contributes to the basic segmentation of speech based on coarse spectrotemporal information from the MGB. The segmentation then provides a

constraint for phonetic analysis in the dorsal pathway in which detailed spectrotemporal information from the AC is matched with predictions generated by the frontal areas including the IFC. This hypothesis is consistent with the present finding that the connections between the MGB and Cd and between the Cd and ACC are stronger in listeners who reported frequent switches than in listeners who reported less frequent switches in VTs. Clearly, further experimental support is necessary.

(c) *Individual variation as a balance between exploration and stabilization*

A feature of the present study is the focus on individual variation in multistable phenomena. Although different neural sites were identified as affecting individual variation in the frequency of perceptual switches for AS and VTs, there are two aspects common to both phenomena. First, the frequency of perceptual switches is determined by a balance of two neural sites. The Cb promotes switching and the MGB suppresses it in AS, whereas the IFC promotes switching and the ACC suppresses it in VTs (figure 3). Apparently, perceptual switching is controlled by an opponency of two factors: the exploration for a new perceptual organization (i.e. the interpretation of auditory scene) and the stabilization of the current perceptual organization. Second, the neural sites promoting switching are presumably related to the generation of motor-based predictions. In AS, the Cb is assumed to generate rhythm patterns that may match the periodicities of the input acoustic sequence. In VTs, the IFC is assumed to generate hypotheses concerning phonetic segments that may match the spectrotemporal patterns of the input acoustic sequence. This is reasonable if we consider the nature of predictive coding. The generation of predictions is critical in terms of maintaining the robustness of perception in ambiguous situations. At the same time, there remains a risk that percepts (the selected prediction) are dissociated from actual acoustic events. The contribution of predictions generated by internal models will be larger in determining percepts when the acoustic input is more ambiguous. This makes it necessary for the perceptual system to switch to a new hypothesis (the interpretation of the auditory scene) more frequently. Therefore, it would be natural for the frequency of switching to be correlated with the activity in the neural sites that generate predictions.

Next, we discuss the genotype–behaviour relationship in the frequency of perceptual switches. The finding that the frequencies of perceptual switches are significantly correlated between AS and VTs within listeners suggests the involvement of neurotransmitters that modulate the functions of networks involved in both these phenomena. The involvement of neurotransmitters in visual multistability phenomena has been suggested previously [35].

Here, we examined the effect of the polymorphisms of the COMT Val158Met gene, which plays an important role in the degradation of catecholamines, such as dopamine and noradrenalin. The dopaminergic pathways mainly project to the frontal cortex, including the Cd and ACC, whereas the noradrenergic pathways project to various areas of the brain. Thus, it is possible that dopamine affects VTs more than AS, but it is difficult to differentiate the effects of dopamine and noradrenalin on AS and VTs based only on the present results.

The polymorphisms of the COMT Val158Met gene have been shown to affect cognitive functions, especially processing capacity and efficiency in the prefrontal cortex: performance tends to be better for Met (methionine) allele carriers than for Val (valine) allele carriers in the Wisconsin Card Sorting Test [75] and the n-back task [76,77]. The COMT polymorphisms also affect sensory processing. An EEG experiment showed that high prefrontal efficiency as suggested by the COMT Met/Met genotype is associated with poor sensory gating of auditory stimuli [43].

Of particular relevance to the present study are recent findings showing that the COMT polymorphisms are involved in the control of the 'exploration–exploitation trade-off', that is, the dilemma involved in keeping a balance between two antagonistic constraints of the stable maintenance of the current choice and the flexible switching to novel and potentially important choices for maximizing reward [78]. COMT is associated with a particular type of 'directed exploration', in which exploratory decisions are made in proportion to Bayesian uncertainty about whether other choices might produce outcomes that are better than the present state [79]. It was also shown that COMT genotypes modulate event-related potential in an auditory oddball task, which is related to novelty processing [80].

Other lines of research have also demonstrated the involvement of catecholamines in the exploration–exploitation trade-off. Pupil dilation, which is a physiological marker of the level of noradrenalin released from the locus coeruleus (LC), has been shown to correlate with the control state, suggesting the involvement of the LC-noradrenalin system in the regulation of the exploration–exploitation trade-off [81]. A study indicates that pupil dilation correlates with perceptual switching in ambiguous stimuli, suggesting that noradrenalin could play a critical role in perceptual multistability [82] (note that the validity of this study is controversial [83]).

As discussed above, persisting with the current perceptual interpretation is not necessarily an appropriate strategy when the sensory input is ambiguous. The flexible exploration of alternative interpretations is necessary, but too much exploration would make perception unstable. We propose that catecholamines, such as dopamine and noradrenalin, may control the balance between exploration and stabilization in the formation and selection of auditory percepts, as has been shown in behavioural decision-making research. The present results suggest that the balance between exploration and stabilization in perception is to some extent determined genetically. Such diversity may contribute to the adaptation of perception to the environment as a whole species.

4. CONCLUSIONS

The re-analysis of our previous neuroimaging studies of auditory multistability phenomena, namely, AS [36] and VTs [37], revealed that the formation and selection of auditory percepts involve widely distributed networks

in the brain. In addition to the thalamocortical loop contributing to both AS and VTs, we identified specific contributions of the Cb in AS and the frontal network consisting of the IFC, ACC, PFC and Cd in VTs. Based on this finding, we propose that motor-based predictions produced by subcortical sites may play crucial roles in the segmentation of temporal sequences not only in speech perception, as suggested in previous studies, but also in the AS of simple acoustic patterns. We also examined individual variations in the frequency of perceptual switches, and found that the frequency of perceptual switching was determined by a balance between the activation of two sites, which are proposed to be involved in exploring novel perceptual organization and stabilizing current perceptual organization. The Cb promotes switching and the MGB suppresses it in AS, whereas the IFC promotes switching and the ACC suppresses it in VTs. We conducted an additional experiment to examine the effect of the gene polymorphism of COMT on individual variations in switching frequency. The results suggest that the balance between exploration and stabilization is modulated by catecholamines such as dopamine and noradrenalin.

The present study raises many questions, rather than solving them. Obviously, further research is needed to evaluate the two hypotheses proposed here, namely, the motor involvement in the segmentation of acoustic sequences, and the opponency between exploration and stabilization in auditory scene analysis. The present study has several limitations. First, the neuroimaging studies reported here used fMRI, which has poor temporal resolution. To analyse the causal relationship between neural sites, fine temporal information is necessary. Complementary studies using EEG or MEG would provide such information. Second, the behavioural analyses of perceptual switching were based on listeners' subjective reports. Development of objective measures of perceptual switching would enable us to examine the covariation of perception and neural responses even in animals. Investigations with a new methodology free from these limitations would promote a further understanding of how auditory percepts emerge in the brain.

We thank Brian Moore and two anonymous reviewers for their thoughtful comments on an earlier version of this manuscript.

REFERENCES

1 Bregman, A. S. 1990 *Auditory scene analysis: the perceptual organisation of sound*. Cambridge, MA: MIT Press.
2 Shamma, S. A. & Micheyl, C. 2010 Behind the scenes of auditory perception. *Curr. Opin. Neurobiol.* **20**, 361–366. (doi:10.1016/j.conb.2010.03.009)
3 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)
4 Snyder, J. S. & Alain, C. 2007 Toward a neurophysiological theory of auditory stream segregation. *Psychol. Bull.* **133**, 780–799. (doi:10.1037/0033-2909.133.5.780)
5 Fishman, Y. I., Reser, D. H., Arezzo, J. C. & Steinschneider, M. 2001 Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear. Res.* **151**, 167–187. (doi:10.1016/S0378-5955(00)00224-0)
6 Fishman, Y. I., Arezzo, J. C. & Steinschneider, M. 2004 Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am.* **116**, 1656–1670. (doi:10.1121/1.1778903)
7 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)
8 Bee, M. A. & Klump, G. M. 2004 Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol.* **92**, 1088–1104. (doi:10.1152/jn.00884.2003)
9 Bee, M. A. & Klump, G. M. 2005 Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav. Evol.* **66**, 197–214. (doi:10.1159/000087854)
10 Itatani, N. & Klump, G. M. 2009 Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J. Neurophysiol.* **101**, 3212–3225. (doi:10.1152/jn.91333.2008)
11 Alain, C., Arnott, S. R. & Picton, T. W. 2001 Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 1072–1089. (doi:10.1037/0096-1523.27.5.1072)
12 Alain, C., Reinke, K., He, Y., Wang, C. & Lobaugh, N. 2005 Hearing two things at once: neurophysiological indices of speech segregation and identification. *J. Cogn. Neurosci.* **17**, 811–818. (doi:10.1162/0898929053747621)
13 Snyder, J. S., Alain, C. & Picton, T. W. 2006 Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* **18**, 1–13. (doi:10.1162/089892906775250021)
14 Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L. & Alain, C. 2009 Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology* **46**, 1208–1215. (doi:10.1111/j.1469-8986.2009.00870.x)
15 Sussman, E., Ritter, W. & Vaughan Jr, H. G. 1999 An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* **36**, 22–34. (doi:10.1017/S0048577299971056)
16 Winkler, I., Takegata, R. & Sussman, E. 2005 Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Brain Res. Cogn. Brain Res.* **25**, 291–299. (doi:10.1016/j.cogbrainres.2005.06.005)
17 Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T. & Kaneko, S. 2001 Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Res.* **897**, 222–227. (doi:10.1016/S0006-8993(01)02224-7)
18 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)
19 Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C. & Melcher, J. R. 2007 Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *J. Neurosci.* **27**, 13 074–13 081. (doi:10.1523/JNEUROSCI.2299-07.2007)
20 Deike, S., Gaschler-Markefski, B., Brechmann, A. & Scheich, H. 2004 Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* **15**, 1511–1514. (doi:10.1097/01.wnr.0000132919.12990.34)
21 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)
22 Pressnitzer, D., Sayles, M., Micheyl, C. & Winter, I. M. 2008 Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* **18**, 1124–1128. (doi:10.1016/j.cub.2008.06.053)
23 Cusack, R. 2005 The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* **17**, 641–651. (doi:10.1162/0898929053467541)
24 Moore, B. C. J. & Gockel, H. E. 2012 Properties of auditory stream formation. *Phil. Trans. R. Soc. B* **367**, 919–931. (doi:10.1098/rstb.2011.0355)
25 van Noorden, L. P. A. S. 1975 *Temporal coherence in the perception of tone sequences*. Eindhoven, The Netherlands: Eindhoven University of Technology.
26 Carlyon, R. P. 2004 How the brain separates sounds. *Trends Cogn. Sci.* **8**, 465–471. (doi:10.1016/j.tics.2004.08.008)
27 Sussman, E. S., Horváth, J., Winkler, I. & Orr, M. 2007 The role of attention in the formation of auditory streams. *Percept. Psychophys.* **69**, 136–152. (doi:10.3758/BF03194460)
28 Darwin, C. J. 1997 Auditory grouping. *Trends Cogn. Sci.* **1**, 327–333. (doi:10.1016/S1364-6613(97)01097-8)
29 Moore, B. C. J. & Gockel, H. 2002 Factors influencing sequential stream segregation. *Acta Acust. United Ac.* **88**, 320–332.
30 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)
31 Kashino, M., Okada, M., Mizutani, S., Davis, P. & Kondo, H. M. 2007 The dynamics of auditory streaming: psychophysics, neuroimaging, and modeling. In *Hearing—from sensory processing to perception* (eds B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Upperkamp & J. Verhey), pp. 275–283. Berlin, Germany: Springer.
32 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)
33 Basirat, A., Schwartz, J.-L. & Sato, M. 2012 Perceptuo-motor interactions in the perceptual organization of speech: evidence from the verbal transformation effect. *Phil. Trans. R. Soc. B* **367**, 965–976. (doi:10.1098/rstb.2011.0374)
34 Warren, R. M. & Gregory, R. L. 1958 An auditory analogue of the visual reversible figure. *Am. J. Psychol.* **71**, 612–613. (doi:10.2307/1420267)
35 Pettigrew, J. D. & Miller, S. M. 1998 A 'sticky' interhemispheric switch in bipolar disorder? *Proc. R. Soc. Lond. B* **265**, 2141–2148. (doi:10.1098/rspb.1998.0551)
36 Kondo, H. M. & Kashino, M. 2009 Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* **29**, 12 695–12 701. (doi:10.1523/JNEUROSCI.1549-09.2009)
37 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)
38 Sato, M., Beciu, M., Lœvenbruck, 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)
39 Basirat, A., Sato, M., Schwartz, J. L., Kahane, P. & Lachaux, J. P. 2008 Parieto-frontal gamma band activity during the perceptual emergence of speech forms. *Neuroimage* **42**, 404–413. (doi:10.1016/j.neuroimage.2008.03.063)
40 Hickok, G. & Poeppel, D. 2007 The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402. (doi:10.1038/nrn2113)
41 Warren, J. E., Wise, R. J. & Warren, J. D. 2005 Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci.* **28**, 636–643. (doi:10.1016/j.tins.2005.09.010)
42 Kondo, H. M., Kitagawa, N., Kitamura, M., Nomura, M. & Kashino, M. In press. Separability and commonality of auditory and visual bistable perception. *Cereb. Cortex.* (doi:10.1093/cercor/BHR266)
43 Majic, T., Rentzsch, J., Gudlowski, Y., Ehrlich, S., Juckel, G., Sander, T., Lang, U. E., Winterer, G. & Gallinat, J. 2011 COMT Val108/158Met genotype modulates human sensory gating. *Neuroimage* **55**, 818–824. (doi:10.1016/j.neuroimage.2010.12.031)
44 Büchel, C. & Friston, K. J. 1997 Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb. Cortex.* **7**, 768–778. (doi:10.1093/cercor/7.8.768)
45 Horwitz, B., Tagamets, M.-A. & McIntosh, A. R. 1999 Neural modeling, functional brain imaging, and cognition. *Trends Cogn. Sci.* **3**, 91–98. (doi:10.1016/S1364-6613(99)01282-6)
46 Cohen, J. 1969 *Statistical power analysis for the behavioural sciences*. New York, NY: Academic Press.
47 Gregory, R. L. 1998 *Eye and brain*, 5th edn. Oxford, UK: Oxford University Press.
48 Helmholtz, H. V. 1860 *Treatise on physiological optics*, vol. 3 (eds J. P. C. Southall). New York, NY: Dover.
49 Friston, K. 2010 The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138. (doi:10.1038/nrn2787)
50 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)
51 Andreou, L.-V., Kashino, M. & Chait, M. 2011 The role of temporal regularity in auditory segregation. *Hear. Res.* **280**, 228–235. (doi:10.1016/j.heares.2011.06.001)
52 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)
53 Winkler, I., Denham, S. L. & Nelken, I. 2009 Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* **13**, 532–540. (doi:10.1016/j.tics.2009.09.003)
54 Halverson, H. E., Lee, I. & Freeman, J. H. 2010 Associative plasticity in the medial auditory thalamus and cerebellar interpositus nucleus during eyeblink conditioning. *J. Neurosci.* **30**, 8787–8796. (doi:10.1523/JNEUROSCI.0208-10.2010)
55 Pastor, M. A., Vidaurre, C., Fernández-Seara, M. A., Villanueva, A. & Friston, K. J. 2008 Frequency-specific coupling in the cortico-cerebellar auditory system. *J. Neurophysiol.* **100**, 1699–1705. (doi:10.1152/jn.01156.20078)
56 Ivry, R. B. & Spencer, R. M. 2004 The neural representation of time. *Curr. Opin. Neurobiol.* **14**, 225–232. (doi:10.1016/j.conb.2004.03.013)
57 Buhusi, C. V. & Meck, W. H. 2005 What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* **6**, 755–765. (doi:10.1038/nrn1764)
58 Lewis, P. A. & Miall, R. C. 2003 Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Curr. Opin. Neurobiol.* **13**, 250–255. (doi:10.1016/S0959-4388(03)00036-9)

59 Zatorre, R. J., Chen, J. L. & Penhune, V. B. 2007 When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* **8**, 547–558. (doi:10.1038/nrn2152)
60 Thaut, M. H. 2003 Neural basis of rhythmic timing networks in the human brain. *Ann. N. Y. Acad. Sci.* **999**, 364–373. (doi:10.1196/annals.1284.044)
61 Mauk, M. D. & Buonomano, D. V. 2004 The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340. (doi:10.1146/annurev.neuro.27.070203.144247)
62 Ohyama, T., Nores, W. L., Murphy, M. & Mauk, M. D. 2003 What the cerebellum computes. *Trends Neurosci.* **26**, 222–227. (doi:10.1016/S0166-2236(03)00054-7)
63 Bloedel, J. 1992 Functional heterogeneity with structural homogeneity: how does the cerebellum operate? *Behav. Brain Sci.* **15**, 666–678.
64 Bower, J. M. 1995 The cerebellum as a sensory acquisition controller. *Hum. Brain Mapp.* **2**, 255–256. (doi:10.1002/hbm.460020407)
65 Petacchi, A., Laird, A. R., Fox, P. T. & Bower, J. M. 2005 Cerebellum and auditory function: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain. Mapp.* **25**, 118–128. (doi:10.1002/hbm.20137)
66 Jankowski, J., Scheef, L., Hüppe, C. & Boecker, H. 2009 Distinct striatal regions for planning and executing novel and automated movement sequences. *Neuroimage* **44**, 1369–1379. (doi:10.1016/j.neuroimage.2008.10.059)
67 Bengtsson, S. L. & Ullén, F. 2006 Dissociation between melodic and rhythmic processing during piano performance from musical scores. *Neuroimage* **30**, 272–284. (doi:10.1016/j.neuroimage.2005.09.019)
68 Thaut, M. H., Demartin, M. & Sanes, J. N. 2008 Brain networks for integrative rhythm formation. *PLoS ONE* **3**, e2312. (doi:10.1371/journal.pone.0002312)
69 Teki, S., Grube, M., Kumar, S. & Griffiths, T. D. 2011 Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* **31**, 3805–3812. (doi:10.1523/JNEUROSCI.5561-10.2011)
70 Nenadic, I., Gaser, C., Volz, H. P., Rammsayer, T., Häger, F. & Sauer, H. 2003 Processing of temporal information and the basal ganglia: new evidence from fMRI. *Exp. Brain. Res.* **148**, 238–246.
71 Kotz, S. A., Schwartze, M. & Schmidt-Kassow, M. 2009 Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* **45**, 982–990. (doi:10.1016/j.cortex.2009.02.010)
72 Schiffer, A. M. & Schubotz, R. I. 2011 Caudate nucleus signals for breaches of expectation in a movement observation paradigm. *Front. Hum. Neurosci.* **5**, 38. (doi:10.3389/fnhum.2011.00038)
73 Langner, R., Kellermann, T., Boers, F., Sturm, W., Willmes, K. & Eickhoff, S. B. 2011 Modality-specific perceptual expectations selectively modulate baseline activity in auditory, somatosensory, and visual cortices. *Cereb. Cortex* **21**, 2850–2862. (doi:10.1093/cercor/bhr083)
74 Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B. & Liddle, P. F. 2001 Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. *Psychophysiology* **38**, 133–142. (doi:10.1111/1469-8986.3810133)
75 Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., Goldman, D. & Weinberger, D. R. 2001 Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc. Natl Acad. Sci. USA* **98**, 6917–6922. (doi:10.1073/pnas.111134598)
76 Goldberg, T. E. & Weinberger, D. R. 2004 Genes and the parsing of cognitive processes. *Trends Cogn. Sci.* **8**, 325–335. (doi:10.1016/j.tics.2004.05.011)
77 Weinberger, D. R., Egan, M. F., Bertolino, A., Callicott, J. H., Mattay, V. S., Lipska, B. K., Berman, K. F. & Goldberg, T. E. 2001 Prefrontal neurons and the genetics of schizophrenia. *Biol. Psychiatry* **50**, 825–844. (doi:10.1016/S0006-3223(01)01252-5)
78 Cohen, J. D., McClure, S. M. & Yu, A. J. 2007 Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Phil. Trans. R. Soc. Lond. B* **362**, 933–942. (doi:10.1098/rstb.2007.2098)
79 Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. 2009 Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat. Neurosci.* **12**, 1062–1068. (doi:10.1038/nn.2342)
80 Marco-Pallarés, J. *et al.* 2010 Neurophysiological markers of novelty processing are modulated by COMT and DRD4 genotypes. *Neuroimage* **53**, 962–969. (doi:10.1016/j.neuroimage.2010.02.012)
81 Jepma, M. & Nieuwenhuis, S. 2011 Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory. *J. Cogn. Neurosci.* **23**, 1587–1596. (doi:10.1162/jocn.2010.21548)
82 Einhäuser, W., Stout, J., Koch, C. & Carter, O. 2008 Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proc. Natl Acad. Sci. USA* **105**, 1704–1709. (doi:10.1073/pnas.0707727105)
83 Hupé, J. M., Lamirel, C. & Lorenceau, J. 2009 Pupil dynamics during bistable motion perception. *J. Vis.* **9**(7), 10. (doi:10.1167/9.7.10)