From Sounds to Words: a neurocomputational model of adaptation, inhibition and memory processes in auditory change detection

Max Garagnani*, Friedemann Pulvermüller

Medical Research Council, Cognition & Brain Sciences Unit
15 Chaucer Rd, Cambridge
CB2 7EF – United Kingdom

Running title: **From Sounds to Words: a model of auditory change detection**

No. of Pages (excl. title page):	38
No. of Figures:	5
No. of Tables:	0
No. of Equations:	6
No. of words (body text):	8451
No. of words (Abstract only):	202
No. of words (Introduction only):	1118

*Corresponding author: Max.Garagnani@mrc-cbu.cam.ac.uk
Tel.: +44 (0)1223 355294, Fax: +44 (0)1223 359062

Abstract

Most animals detect sudden changes in trains of repeated stimuli but only some can learn a wide range of sensory patterns and recognize them later, a skill crucial for the evolutionary success of higher mammals. Here we use a neural model mimicking the cortical anatomy of sensory and motor areas and their connections to explain brain activity indexing auditory change and memory access. Our simulations indicate that while neuronal adaptation and local inhibition of cortical activity can explain aspects of change detection as observed when a repeated unfamiliar sound changes in frequency, the brain dynamics elicited by auditory stimulation with well-known patterns (such as meaningful words) cannot be accounted for on the basis of adaptation and inhibition alone. Specifically, we show that the stronger brain responses observed to familiar stimuli in passive oddball tasks are best explained in terms of activation of memory circuits that emerged in the cortex during the learning of these stimuli. Such memory circuits, and the activation enhancement they entail, are absent for unfamiliar stimuli. The model illustrates how basic neurobiological mechanisms, including neuronal adaptation, lateral inhibition, and Hebbian learning, underlie neuronal assembly formation and dynamics, and differentially contribute to the brain's major change-detection response, the mismatch negativity.

Keywords:

Neurophysiology, mismatch negativity (MMN), long-term memory traces, neural-network simulation, language, electro/magneto-encephalography (EEG/MEG)

1 Introduction

The ability of the brain to automatically detect unexpected, rare events amongst common, frequently repeated ones can be crucial for survival. A well-known index of the human auditory system's ability to detect novelty and change is the Mismatch Negativity (MMN) (Näätänen *et al.*, 1978), an early (latency 100-250ms) event-related response elicited in oddball experiments by infrequent acoustic events (so-called "deviant" stimuli, DEV) presented occasionally among frequently repeated "standard" stimuli (STD). Importantly, the MMN is elicited even in the absence of focused attention, for example when subjects are distracted by a streaming task, demonstrating the automatic nature of the brain mechanisms underlying it (Näätänen, 1990; Schröger *et al.*, 1992).

Several models have been proposed to explain aspects of the brain's ability to automatically detect change (see (Garrido *et al.*, 2009; May & Tiitinen, 2010) for recent reviews). While these different explanations highlight the relevance of different processes, they converge on the importance of short-term mechanisms acting upon, or being driven by, the most recent sensory input. Mechanisms previously hypothesized to underlie the MMN response include cortical inhibition (Näätänen's (1990) "release of tonic inhibition" model), neuronal adaptation (Jääskeläinen *et al.*'s (2004) differential adaptation model), short-term synaptic plasticity (model-adjustment hypothesis (Winkler *et al.*, 1996)) or a combination of these mechanisms. For example, the predictive coding model takes into account both synaptic plasticity and neuronal adaptation (Friston, 2005), whereas models described by May and colleagues (1999; 2010) focus on adaptation and lateral inhibition.

*** Fig 1 about here ***

Closely related to the ability of the brain to automatically detect unexpected, "deviant" sensory events is its capacity to become acquainted with, and recognise, aspects of the environment that occur regularly and with high frequency. However, the ability to learn and later recognise and distinguish large numbers of input patterns, including scenes, specific faces, sounds, and words, requires mechanisms for learning and storage of long-term memory (LTM) traces. Whereas the former (change-detection) capacity is shared by a range of animals, the acquisition of large "vocabularies" is limited to a set of higher vertebrates, and is believed to have played a crucial role in the evolutionary advantage of mammals (Pulvermüller, 1999; Fadiga *et al.*, 2002; Wilson *et al.*, 2004; Fagot & Cook, 2006; Voss, 2009; Pulvermüller & Fadiga, 2010). LTM traces, after having formed, act as long-term representations for patterns of sensory input, i.e., they can be re-activated, thereby signalling the presence of the corresponding elements in the environment. These circuits thus provide a mechanism for the recognition of learned meaningful stimuli. Memory representations can also emerge at more abstract levels, by means of generalisation over structurally similar sequences of sensory patterns (e.g., tone sequences in music, or syntactic structures in language). Recent results indeed support the hypothesis that a modulation of the MMN response reflects the presence and activation of long-term memory traces at different levels (Schröger *et al.*, 1992; Frangos *et al.*, 2005). In the domain of language, familiar speech sounds presented as deviant stimuli in an oddball sequence produce larger MMN responses than unfamiliar speech sounds (Dehaene-Lambertz, 1997; Näätänen *et al.*, 1997), and, similarly, MMNs to familiar and meaningful words are larger than to matched unfamiliar and meaningless pseudowords (Korpilahti *et al.*, 2001; Pulvermüller *et al.*, 2001; Garagnani *et al.*, 2009) (see Fig. 1, panel on the right). A similar difference applies to non-linguistic

familiar vs. unfamiliar speech sounds (Frangos *et al.*, 2005; Jacobsen *et al.*, 2005; Hauk *et al.*, 2006). Even at highly abstract levels, the MMN shows differences between tone sequences respecting sequential regularities and those that do not (Saarinen *et al.*, 1992; Bendixen & Schröger, 2008; Bendixen *et al.*, 2009), and equally between grammatical sentences and ungrammatical or meaningless strings (Shtyrov *et al.*, 2003; Pulvermüller & Shtyrov, 2006; Pulvermüller & Assadollahi, 2007; Shtyrov & Pulvermüller, 2007).

The available evidence, therefore, suggests that the brain response to an unexpected sound results from (at least) two different processes: (1) the automatic detection of auditory change based on the most recent sensory input (within a few seconds), and (2) the activation of previously established memory traces specific to familiar auditory elements, which can emerge in the cortex by means of long-term learning mechanisms (Näätänen *et al.*, 2001; Pulvermüller & Shtyrov, 2006). In the work described here we aimed at investigating the different contributions that these long- and short-term memory mechanisms make to the brain's change detection response. By "long-term mechanisms" we mean long-lasting synaptic changes and plasticity, including long-term potentiation and depression, and the emergence of memory circuits in the cortex; by "short-term mechanisms", we mean neuronal adaptation, lateral inhibition, and, critically, the reverberation of neuronal activity within memory circuits.[1]

It should be clarified here that the present approach – which is rooted in neurobiological theory – views the formation of long-term memory traces in the cortex as the result of correlated activation in sensory and motor brain systems. In particular, following Hebb's postulate (Hebb, 1949), we conjecture that simultaneous

[1] By "reverberation within memory circuits" here we mean repeated forward and backward propagation of firing activity within a set of strongly and reciprocally connected cells – see also (Abeles, 1991).

neuronal activity in sensory and motor systems leads to the strengthening of synaptic links connecting the coactive cells, and to the formation of memory circuits distributed over sensory, motor and mediating "higher" areas (e.g., in prefrontal cortex and "amodal" temporal cortex (Fuster, 2001)). These distributed action-perception circuits are the neurobiological basis of long-term memory traces, and their temporary activation has been proposed to be the mechanism underlying short-term, working or "active" memory (Zipser *et al.*, 1993). In this neurobiological perspective, memory and perception are not entirely separate functions, realised by dedicated components, but processes that emerge in – and are implemented by – networks of neurons governed by the *same* mechanisms. As such, neuronal adaptation, inhibition, activation spreading and synaptic plasticity may all subserve, to different extents, both memory and perception. Hence our choice not to describe the relevant effects as either perceptual or memory-based, and focus on identifying the distinct neural mechanisms at the origins of (and differentially contributing to) the observed cognitive processes and neurophysiological effects.

To investigate the constituents of the brain response to auditory change we implemented biologically grounded neural-network models that reproduced important structural and functional properties of relevant cortical areas. In a series of simulation studies carried out on such models we systematically manipulated short- and long-term mechanisms, and analysed the effects of the (combined and independent) presence (or absence) of these mechanisms on the resultant MMN response. We ran two sets of simulations: in Experiment 1, neurophysiological principles applied but no long-term memory mechanisms were implemented. The resultant networks (Fig. 2B) were thus "tabula rasa" models, i.e. contained no LTM traces. In Experiment 2, mechanisms of synaptic plasticity and long-term memory formation were included.

The resultant networks (now including memory traces – Fig. 2C) were applied to model effects of familiarity on the MMN response. In both Experiments, the influence of basic neuronal mechanisms of short-term memory (adaptation and inhibition) was examined by systematically varying their availability.

*** Fig 2 about here ***

2 Material and Methods

2.1 Experiment 1 – MMN to frequency change in auditory areas

To isolate the contributions of short-term mechanisms to the MMN response, as it is elicited, for example, by frequency deviants in auditory oddball stimulation, we used tabula rasa networks with the 3-area structure depicted in Fig. 2B, modelling the three main auditory areas, primary auditory cortex, auditory belt and parabelt (A1, AB, PB, Fig. 2A). Each model area is comprised of two layers of 25-by-25 units each, one consisting of excitatory cells, the other of inhibitory ones (not depicted in Fig. 2). Each network unit, or "node", represents a cluster of real neurons, and is realised as a graded-response leaky-integrator cell having sigmoid activation function with threshold φ (see Appendix A for details). Higher- (lower)-than-average cell activation levels produce an increase (decrease) in the threshold φ, effectively modelling neural (or "spike-rate") adaptation; the impact of adaptation on the activation of a cell ("adaptation strength") is determined, in the model, by the parameter α (see Appendix A, Eq. (A.3)), which was modulated across the simulations. Within- (recurrent) and between-area synaptic connections are not "all-to-all" but random, patchy and topographic, as typically found in the mammalian cortex (Gilbert & Wiesel, 1983; Braitenberg & Schüz, 1998). Local reciprocal

connections between excitatory and inhibitory layers of each area realise lateral inhibition (Braitenberg & Schüz, 1998), as follows: each inhibitory cell I receives excitatory input from all excitatory cells situated within an overlying 5x5 neighbourhood and projects back to the single excitatory cell E located directly above it. The parameter w_I, indicating the weight of the projection I→E (identical across all cells and areas) was manipulated in the simulations to modulate local inhibition strength. An area-specific self-regulatory mechanism, termed "global" inhibition, was also implemented, in order to prevent the overall network activation from falling into non-physiological states (total saturation or inactivity). For further details of the model, see (Garagnani *et al.*, 2008).

2.1.1 Materials

The presentation of a "sound" to a network was simulated by simultaneously activating a predetermined pattern of 17 cells in area A1 (2.72% of all A1 units). There were six pairs of randomly generated standard (STD) and deviant (DEV) stimulus patterns. The probability that any two patterns shared one or more cells was approximately 0.38. All networks used for this experiment (having the 3-area structure illustrated in Fig. 2B and described above) had their synaptic links and weights initialised at random. Long-term learning mechanisms were not implemented to examine the influence of inhibition and adaptation mechanisms separately, assuming unfamiliar, new sounds are processed by an untrained, "naïve" auditory cortex.

2.1.2 Design

To investigate predictions of adaptation, local inhibition, and combined adaptation-inhibition theories of the MMN, we used four networks of different types. In the first

network, adaptation and local inhibition mechanisms were both removed, simulating a situation in which cells do not adapt to the input stimulus and the local connections implementing lateral inhibition are absent (α=0, w_I=0). In the second network, neuronal adaptation was effective (α=10) and local inhibition absent (w_I=0), whereas the opposite (α=0, w_I=1.0) was true in the third one; finally, both mechanisms were active (α=10, w_I=1.0) in the fourth network.

Oddball experiments were simulated as follows. Each "trial" started with a baseline of six simulation time steps, after which a stimulus was presented for four time steps. The baseline was also the inter stimulus interval, ISI. The oddball sequence consisted of 80% STD trials intermixed with 20% DEV (or critical) trials. The order was pseudo-random, with successive DEV trials separated by 2-to-6 intervening STD trials. Network's output was recorded from the penultimate STD stimulus' offset to the onset of the stimulus following a critical trial. For each of six stimulus pairs, ten simulated evoked responses were collected in each of the 4 randomly initialised networks, producing a total of 60 trials per stimulus type per network.

2.1.3 Evaluation

For each of the four networks, we computed and plotted the average (across trials and stimuli) of the total network response (sum of the output of all excitatory cells in areas A1, AB and PB) to DEV and pre-deviant STD stimuli, and the difference (DEV – STD), or MMN, over a period of 14 simulation time-steps.[2] Statistical analyses were carried out by means of paired two-sample *t*-tests on the difference values (DEV -

[2] As ISI was 6 simulation time-steps, the network's output during the last 4 steps of the ISI following the STD stimulus was used to plot both the 4 steps preceding the stimulus onset in the evoked DEV response's baseline, and the last 4 steps of the evoked STD response.

STD) obtained from the 60 trials collected. Error bars on the plots give the standard error of the mean (SE).

2.2 Experiment 2 – Long-term memory MMN in the language cortex

Here we set out to investigate the neurobiological mechanisms underlying the formation of memory traces, and the effect of such memory traces' activation on neurophysiological responses. As a paradigm case we chose the learning of the words of a language, contrasting it with the lack of formation of such processing devices for meaningless unfamiliar pseudowords. More precisely, we aimed at providing a mechanistic explanation of experimental data showing larger brain responses to familiar words than to unknown pseudowords (Korpilahti *et al.*, 2001; Pulvermüller *et al.*, 2001). Words and familiar speech sounds are auditory objects, but are produced by articulatory movements. Because in typical language development speech production leads to auditory perception of the self-produced sounds, Hebbian learning entails a coupling of specific motor and auditory circuits in distributed sensorimotor circuits (Fry, 1966; Pulvermüller & Preissl, 1991; Pulvermüller, 1999)**.** The presence in the cortex of strong links associating speech sounds with corresponding articulations has been confirmed by a significant body of experimental evidence (Pulvermüller, 1999; Fadiga *et al.*, 2002; Wilson *et al.*, 2004; Pulvermüller & Fadiga, 2010). Therefore, the model of the auditory cortex used in Experiment 1 was extended here by adding three areas (Fig. 2A, 2C) that mimic the function of inferior-frontal motor, premotor and prefrontal cortex (areas M1, PM and PF).

2.2. 1 Materials

The extended network structure used in Experiment 2 is shown in Fig. 2C, and included the superior-temporal area model used in Experiment 1. Three inferior-

frontal areas were added, with connectivity reflecting major features of the known neuroanatomical links within and between inferior frontal areas (see Methods, Experiment 1, and (Garagnani *et al.*, 2008)). Neuroanatomical links between inferior-frontal and superior-temporal areas were added to mimic long-distance connections by way of the extreme capsule and the arcuate fascicle, which have been documented to be present in macaques and even richer developed in humans (e.g., Pandya & Yeterian, 1985; Catani *et al.*, 2005; Petrides & Pandya, 2009). To enable the network to develop memory traces for words, learning was allowed by simulating long-term potentiation and depression (LTP/LTD) mechanisms (Malenka & Nicoll, 1999) (see Appendix A).

We built a set of 12 to-be-memorised sensorimotor patterns, thought to represent 12 words (W). Each sensorimotor pattern p_{SM} comprised one auditory/acoustic pattern p_{A1} and one motor/articulatory neural pattern p_{M1}, identifying, respectively, 17 specific cells in area A1 (the auditory "word stimulus") and 17 (different) cells in M1 (the corresponding motor pattern), which were co-activated during the learning of the word-pattern p_{SM}. Co-activated cells in A1 and M1 can be thought to represent the neural correlates of acoustic and articulatory phonetic features of a word, respectively. Twelve not-previously-learnt sensorimotor patterns, or "pseudowords" (PW), identical in size to the word patterns, were generated by randomly selecting and re-combining sub-parts of word patterns. To control for the degree of overlap[3] between auditory patterns of Ws and PWs, the stimuli were built in such a way that they could be arranged into four sets of six pairs – (W, W), (W, PW), (PW, W), and (PW, PW) – such that the overlap between any two paired patterns was constant (≈8.3%, see Appendix C). After learning, the auditory patterns (neural units in A1) were used to

[3] The portion of shared active cells.

stimulate the network.

2.2.2 Design

In the first part of this experiment, the network learned sensorimotor (A1-M1) patterns. Each word pattern was presented to the network two thousand times. Stimulus duration was two time-steps; stimuli were followed by an ISI of variable length[4] during which input activity was driven by white noise. As described in detail by Garagnani et al. (2008), the presence of Hebbian learning mechanisms in this architecture induces the emergence of model correlates of lexical representations – cell assemblies (CAs) for words – as strongly connected distributed circuits that associate the paired "sensory" (A1) and "motor" (M1) activation patterns through neural elements located in areas linking auditory and motor cortex. The present network successfully learnt sensorimotor patterns for word pattern.

In the second part of the study, the resulting trained network was used to simulate the neurophysiological responses of the language cortex to W and PW stimuli. In each "trial", the model's auditory area A1 was stimulated either by a familiar (W) or unfamiliar (PW) pattern. An oddball paradigm was implemented in which 93% of STD trials were intermixed with 7% DEV trials. DEV trials were always preceded by 6-to-10 STD stimuli. Other features of the simulation of the oddball experiments were the same as in Experiment 1.

We employed 4 networks of 3 different types (see Experiment 1) which, respectively, included adaptation only, local inhibition only, and both adaptation and inhibition mechanisms together; networks lacking both adaptation and inhibition were

[4] During training, we enforced stimulus presentation to occur always at the same initial level of network activation. This required the length of the ISI to be varied (dynamically) in proportion to the amount of activation that the previous stimulus had produced in the network.

omitted, because this type of model was shown in Experiment 1 to be unable to model basic MMN responses. To control for overall degree of stimulus suppression (or activation) two networks having combined adaptation-inhibition were used: while the first, the "full scale" version, had adaptation and inhibition parameters set to values applied in the adaptation- and inhibition-only networks (α=10, w_I=1.0), the second "reduced" version controlled for the overall level of dysfacilitation by halving both values (α=5, w_I=0.5). Four experiments were run on each of these 4 networks, in which the lexicality[5] of standard and deviant stimuli was varied systematically (congruent lexicality of standard and deviant stimuli: W-W, PW-PW; and incongruent lexicality: W-PW, PW-W). All other parameters and features were identical to Experiment 1. To assess the variability of the network response due to the random elements of the simulations (i.e., neuronal background noise, jittering of the number of STD before a DEV) each of these 16 experiments was repeated four times, producing a total of 240 trials per stimulus type per lexical context per network.

2.2.3 Evaluation

The evaluation was identical to that of Experiment 1, except that average STD, DEV and MMN responses (and standard errors) were computed across 240 trials for each of the 4 possible lexical combinations (see above) and for each of the 4 networks.

3 Results

3.1 Experiment 1 – MMN to frequency change in auditory areas

As shown in Figure 3.(A), in absence of neuronal adaptation and local inhibition mechanisms the network does not bring about a MMN – in fact, a larger response to

[5] The lexical status of a linguistic item (words are lexical items, pseudowords are not).

the STD than to the DEV stimulus emerged (time-step 8: $t(59)=6.95$, $p<0.001$), a behaviour which is contrary to extant experimental evidence. In all three remaining cases – inhibition-only, adaptation-only and combined inhibition-adaptation – the network exhibits a clear MMN, with the responses to the DEV significantly larger than those to the STD stimuli (at time-step 8, all t_{239} values > 6, all p values < 0.001). Note that the largest MMN amplitude was produced by the last type of network.

*** Fig 3 about here ***

Figure 4 plots the average area-specific activations contributing to the simulated N1-to-the-STD and MMN responses[6] as extracted from the output of inhibition-only networks. The simulated activation peaks ("source strengths") in areas A1 and AB (panels (A)-(C)) were submitted to a weighted-averaging procedure which computed the "centre of gravity" (or, more appropriately, centre of mass) of the simulated sources, yielding an output similar to that of the Equivalent Current Dipole (ECD) estimation (see Appendix B for details). As visible from panel (D), the changes in relative activation of the same auditory areas lead to distinct estimates of the underlying sources' locations, thus explaining, in part, the well-known differences between ECD loci of the MMN and N1 to the STD responses (Tiitinen *et al.*, 1993; Korzyukov *et al.*, 1999).

*** Fig 4 about here ***

*** Fig 5 about here ***

[6] Here, the N1 is calculated as the largest peak of the simulated response to the STD, whereas the MMN was calculated as the largest peak of the difference wave, DEV−STD.

3.2 Experiment 2 – Long-term memory MMN in the language cortex

Figure 5 shows the STD, DEV and resultant MMN responses of 6-area networks realising adaptation, inhibition, and full and reduced combined adaptation-inhibition to "speech" stimuli (familiar auditory "word", W, and unfamiliar "pseudoword", PW, patterns) presented to the primary "auditory" area (A1).

The adaptation-only network results (first column on the left of Figure 5) show that STD responses tend to be larger for Ws (red curves) than for PWs (blue curves) at time steps 7 and 8. A similar trend is evident for DEV responses. As a result, the MMN responses depend on the lexical status of both STD and DEV stimuli. As shown in the two bottom plots, when the context is PW STDs the DEV responses to Ws are larger than those to PWs. Similarly, in the context of STDs that are Ws, the MMNs elicited by Ws are larger than those to PWs. MMNs to Ws and PWs presented in incongruent contexts (i.e., W in PW STD context and PW in W STD context) are also different, with W MMNs being larger than PW MMNs (Fig. 5, bottom panel of column 1, time-step 7, t_{239}=5.46, p<0.001). This is consistent with neurophysiological evidence (see Introduction and Discussion below). However, in congruent contexts there is no difference between simulated MMN responses to Ws in W STD context and PWs in PW STD context, ($F < 1$, n.s.), which is difficult to reconcile with pre-existing empirical data.

The inhibition-only network (second column of Figure 5) shows marginally smaller STD responses than the adaptation model, but some relatively large MMNs. Once again, W STD responses tend to be larger than PW STD ones and a similar difference is present for DEV stimuli, although the onset of differences to DEVs now appears already at time step 6 (t(239)=1.95, p<0.03). Numerically, the lexicality difference (W minus PW response) seems to be larger in the DEV than in the STD responses,

resulting in a general lexicality effect in MMN responses: MMNs to Ws are now larger than MMNs to PWs regardless of the familiarity of context or standard stimuli (Fig. 5, column 2, bottom panel, time-steps 6: t_{239}=1.95, p<0.03, and 7). In addition, word MMNs were larger in W context than in PW context at time step 5 (t_{239}=2.42, p<0.02).

The third and fourth columns of Fig. 5 plot the results of the networks with combined adaptation and inhibition. As in the inhibition-only model, a relatively small (reduced values, top panel of column 3), or even marginal (maximal values, top panel of column 4) lexicality effect emerged in the STD responses (time step 7). The enlargement of the DEV response to W relative to that to PW stimuli was more substantial at time-step 6 for the reduced values model, and at steps 6 and 7 for the maximal values model. This resulted in a generalised lexicality effect on the MMN: W MMNs were larger than PW MMNs, regardless of the familiarity of the context or STD stimuli (Fig. 5, bottom panels, column 3, time-step 7: t_{239}=1.79, p<0.05; column 4, steps 6: t_{239}=3.31, p<0.01 and 7).

In sum, all models reproduced the larger DEV responses to Ws than PWs, which led to a word enhancement of the MMN (i.e., W MMN > PW MMN) given the STD context was kept constant. Also a lexicality effect on the MMN in incongruent contexts was found throughout. However, the adaptation-only model failed to produce a larger MMNs to familiar, compared with unfamiliar, stimuli when these were presented in congruent contexts.

4 Discussion

4.1 Experiment 1 – MMN to frequency change in auditory areas

We used a neuronal circuit model fashioned according to the neuroanatomy of the auditory system to identify and explain the contribution of short-term mechanisms to the change-detection response in the human brain. Mechanisms of neuronal adaptation and lateral inhibition, acting independently (Fig. 3B and 3C) or together (Fig. 3D) in randomly and sparsely connected networks are sufficient to produce a change detection response (i.e., larger responses to deviants compared with standard stimuli) as seen, for example, in the MMN elicited by frequency change. The simulation in which inhibition worked in synergy with adaptation produced the largest MMNs.

Our results converge with previous neurocomputational results by May *et al.* (1999) on the conclusion that a combined inhibition-adaptation model best explains experimental data on the MMN to frequency deviants. May and colleagues used a tonotopically organized array of artificial neural units as a model of the auditory cortex and investigated the effects of deviant stimulus frequency on MMN amplitude and peak latency. As the non-monotonic shape of the curve plotting MMN peak latency against the frequency difference between the deviant and the standard stimulus was reproduced by the combined adaptation-inhibition model but not by the adaptation-only one, the authors concluded that the frequency MMN is best explained by synergistic action of adaptation and lateral inhibition. Our results are consistent with these earlier findings, and, in addition, contribute novel evidence, namely, that the inhibition-only model produces a frequency MMN very similar to that produced by the adaptation-only model (Fig. 3). Thus, the results of our simulations are in line with (and integrate) two previous accounts which explained the emergence of an

MMN to frequency change either on the basis of release of tonic inhibition (Näätänen, 1990) or neuronal adaptation (Jääskeläinen *et al.*, 2004).

To account for the stronger anterior superior-temporal sources observed for the MMN response compared with N1 responses, previous adaptation models (Jääskeläinen *et al.*, 2004; May & Tiitinen, 2010) assumed different frequency tuning of neuronal responses for anterior and posterior areas of the superior-temporal cortex (non-specific in posterior parts, sharply tuned in anterior parts). However, these assumptions are difficult to maintain in view of current knowledge about adaptation in the auditory system. In contrast with these assumptions, neurophysiological evidence indicates clear and sharply tuned tonotopy in A1 (which lies medially), but that (*i*) anterior and posterior sections of the lateral belt in the superior-temporal gyrus include *both* sharply and broadly tuned cell populations (Petkov *et al.*, 2006), and (*ii*) neurons in the auditory belt generally respond optimally to *more complex* patterns, including band-passed noise bursts and species-specific monkey calls (Rauschecker & Tian, 2000). Therefore, a general anterior-to-posterior gradient in frequency tuning appears in lack of support; furthermore, if such a gradient were present, existing evidence would suggest a broader tuning when moving *away* from the primary auditory cortex and into the belt, and not the other way around, as these previous models assumed. Thus, the anterior shift of the MMN generators compared with N1 sources cannot be explained by such local functional differences.

The results of Experiment 1 explain the shift of the MMN generators relative to those of the N1 on the basis of well-documented neurobiological mechanisms (lateral inhibition and propagation of neuronal firing activity – e.g., see (Matthews, 2001)) which apply uniformly across cortical areas. Fig. 3 illustrates how the differential activations of the auditory core (A1) and belt (AB) areas can lead to different

equivalent current dipole loci underlying MMN and N1, even in absence of adaptation, and, crucially, without assuming any frequency tuning gradient. In fact, this result can be explained purely on the basis of stimulus-specific inhibition occurring in the auditory areas during oddball stimulation. More precisely, as the STD stimulus is repeatedly presented to A1, two processes take place at the same time: (*i*) the response of the activated cells in A1 and AB is inhibited; (*ii*) because of this attenuation, the input from A1 into AB is also reduced. The cumulative effect of (*i*) and (*ii*) leads to a twofold decrease of the AB response, resulting in a particularly weak N1's anterior source. Thus, when a new (DEV) stimulus activates "fresh" cells in both A1 and AB, the relative increase in response is larger in AB than in A1 (panels (A) and (B) of Fig. 3), explaining the shift of the MMN generators. This general argument applies equally if adaptation is used instead of (or in addition to) inhibition.

Although the model used in Experiment 1, including only lateral inhibition, adaptation and activation spreading mechanisms, goes a long way in replicating a number of experimental results on auditory change detection, it cannot explain *all* aspects of the real MMN response (see also Sec. 4.4). In particular, the tabula rasa networks used here do not include mechanisms for learning and long-term memory (LTM). Because of this, they cannot reproduce MMN differences between, e.g., learned tone sequences and unfamiliar ones (Näätänen *et al.*, 1993), MMN dynamics mimicking perceptual discrimination, which emerge and vanish with the ability to perceive the relevant distinctions (Kujala *et al.*, 2001), the enhanced brain responses to learned familiar speech sounds and words as compared with unfamiliar sounds and pseudowords (Näätänen *et al.*, 1997; Pulvermüller *et al.*, 2001), the MMN seen to violations of highly abstract familiar patterns such as melodic stereotypes or syntactic

rules (Saarinen *et al.*, 1992; Pulvermüller & Shtyrov, 2003; Shtyrov *et al.*, 2003; Bendixen & Schröger, 2008; Bendixen *et al.*, 2009), and memory-related subcomponents of the MMN response to pitch, duration and intensity change (Jacobsen & Schröger, 2001; Jacobsen *et al.*, 2003; Jacobsen & Schroger, 2003). In order to replicate and explain the above MMN effects, attributable only to LTM (Näätänen *et al.*, 2005; Näätänen *et al.*, 2010), a model must be equipped with memory representations, or long-term synaptic plasticity (i.e., learning) mechanisms that enable their formation.

4.2 Experiment 2 – Long-term memory MMN in the language cortex

Experiment 2 investigated the mechanisms underlying long-term memory effects manifest in the MMN using a neurobiological model of the left language cortex in which memory circuits for words had emerged by means of Hebbian learning. The results obtained with memory networks endowed with both adaptation and inhibition were fully consistent with experimental evidence (according to which MMNs to familiar sounds and words are larger than those to matched unfamiliar items) regardless of the context in which stimuli are presented (e.g., Näätänen *et al.*, 1997; e.g., Pulvermüller *et al.*, 2001). This general pattern also arose in the memory network with inhibition but lacking adaptation mechanisms. In contrast, the memory plus adaptation model failed to reproduce the enhanced MMN to familiar items in congruent contexts: deviant word stimuli in word context did not yield larger MMNs than deviant pseudowords in pseudoword context, which is in contrast with experimental evidence (Korpilahti *et al.*, 2001; Shtyrov *et al.*, 2005). These results suggest that the contribution of local inhibition may be more important for eliciting a normal pattern of cognitive MMN effects than the functionality of adaptation mechanisms. Nonetheless, the activation of, and reverberation of excitation within,

neuronal circuits functioning as memory traces is necessary in both combined adaptation-inhibition and inhibition models to explain the word enhancement of the MMN.[7] In all simulations, standard responses were somewhat larger to words than to pseudowords, thereby modelling an effect reported experimentally (e.g., Jacobsen *et al.*, 2004). However, in most networks with inhibition, this familiarity difference in standard responses was small compared with the respective difference in DEV responses. We speculate that the lack of lexical enhancement observed in absence of inhibition is the result of the failure of lateral inhibition mechanisms to "protect" the excitation within a memory trace from the surrounding background noise. Indeed, when a memory circuit receives excitation from sensory input, lateral inhibition leads to automatic suppression of its (noisy) neighbouring cells, allowing excitation within the circuit to propagate and reverberate undisturbed. This type of local "neuronal fencing" does not occur, e.g., in adaptation-only networks.

There are additional results emerging from Experiment 2 which have not been reported (or investigated) in previous experimental research. All networks including adaptation produced larger word-related MMNs in pseudoword (incongruent) context than in word (congruent) context (Fig. 5, columns 1, 3 and 4 from the left, time-step 7: all t values > 3.2, all p values < 0.001). In addition, all networks except the maximal-values combined adaptation-inhibition model showed stronger late (time-steps 7-8) responses to pseudowords in congruent than incongruent context. Finally, only the inhibition-only network yielded stronger earlier (time-step 5-6) word MMNs in congruent context than against a background of pseudoword standards. These new predictions can be addressed empirically by future experimental research on the

[7] Note that all effects obtained in these simulations were activation contrasts induced by stimuli that were "perceptually" equivalent and, therefore, could not have yielded any difference in the networks employed in Experiment 1.

mechanisms underlying MMN to familiar and unfamiliar stimuli, and the results of such investigations may be used to weigh the different model types against each other.

Note that regardless of the familiarity of the stimuli, MMN responses were present in all simulations of Experiment 2 already at time-step 5 (all t values > 8, all p values < 0.001). In other words, the generation of a change detection response (as observed in the tabula rasa networks of Experiment 1) is replicated and extended here to models that contained memory traces.

Looking at the results of Experiment 1 alone, one may be tempted to conclude that adaptation, inhibition and activation spreading mechanisms are sufficient to explain *all* aspects of the (real) brain response to frequency change. However, in view of the results of Experiment 2 and of previous simulations, we believe that this conclusion would not to be entirely accurate. In particular, we have previously shown that the presentation of not-represented stimuli (the model analogues of uncommon sounds, pseudowords, or other not previously learned items) to a model with memory circuits leads to the partial activation of such circuits (Garagnani *et al.*, 2008; Garagnani, 2009). As mature brains are not tabula rasa entities but are equipped with a range of long-term memory traces, we conjecture that the same partial activations may occur in the brain when stimulated with pseudowords or other unfamiliar items (such as tone pips). The driving forces behind these partial memory trace activations would be (1) the physical similarity between familiar and unfamiliar stimuli – e.g., the fact that unfamiliar noise bursts and narrow-band noise patterns are part of the acoustic structure of speech stimuli (Rauschecker & Scott, 2009) – and (2) the strong internal connection of memory circuits, which allow efficient propagation of neural activity. Thus, an explanation of the MMN response based purely on adaptation and inhibition

and which ignores the role of long-term memory mechanisms seems incomplete, not just in the case of pseudoword stimuli, but also for simple tone sequences. This position is both motivated by our simulations and consistent with experimental evidence from studies investigating elementary acoustic stimuli (e.g., see (Jacobsen & Schröger, 2001; Jacobsen & Schroger, 2003; Näätänen *et al.*, 2010)) and meaningless pseudowords (Garagnani *et al.*, 2009; Shtyrov *et al.*, 2010).

4.3 MMN = N1?

The present results also enable us to address the long-debated issue of whether the MMN is in fact an enhancement of the N1 component of the event-related brain response. The N1 to the standard stimulus obtained in the oddball task and the MMN (which adds to it in the deviant response) are "the same" in the sense that both components are the consequence of the same neurofunctional principles at work across the neuronal substrate, and no additional mechanisms (such as prewired neurons with *a-priori* change detection abilities or differential frequency tuning) are necessary to explain them. Crucially, however, MMN and standard-elicited N1 responses involve the activation of different, although overlapping, neuronal populations, and these activations are caused by (common) underlying processes that contribute differentially to the two responses. In particular, the combined results of Experiments 1 and 2 indicate that the simulated MMN responses to learned, familiar items (words) and to not previously learnt ones (pseudowords) are strongly driven by the activation of both "fresh", unsuppressed cells *and* long-term memory circuits; in contrast, the N1 responses to the repeated standard stimulus did not show strong memory trace related effects, which, if present, were relatively small in comparison to the size of the MMN response (see Simulation 2, Figure 5). The N1 to the standard stimulus was invariably attenuated because of adaptation and local inhibition

mechanisms. However, activation of unattenuated cells in areas A1 and AB accounts for only part of the MMN: the other significant contribution comes from activity quickly propagating (and reverberating) in unsuppressed memory circuits, which are not restricted to the auditory system but can extend to distant (for example, frontal) areas. In summary, this suggests that the spreading of excitation through strong links in "fresh", distributed memory circuits plays an essential role in explaining the observed MMN difference between familiar and novel stimuli. In this sense, MMN elicited by deviant stimuli and N1 to repeated standard stimuli are fundamentally distinct.

A further argument in favour of fundamental differences between MMN and N1 relates to the cortical sources underlying these responses. Experiment 1 showed that adaptation and inhibition mechanisms could underlie the observed source shift in the temporal lobe. However, the additional activation in cortical areas distant from primary auditory cortex is best explained by distributed memory circuits activated by specific types of stimuli. For language stimuli, these circuits link together neurons in inferior-frontal and superior-temporal cortex and, because the link between these areas is more strongly developed in the left than in the right hemisphere (Catani *et al.*, 2005), this explains why brain responses to language stimuli, and especially their inferior-frontal sources, tend to be left lateralised (Shtyrov *et al.*, 2000; Näätänen *et al.*, 2001; Pulvermüller & Shtyrov, 2006). For nonlinguistic stimuli, there may be no clear MMN laterality or even laterality to the right. In earlier simulation studies, we have highlighted the role of frontotemporal memory circuits in explaining the spatiotemporal activation dynamics of frontal and temporal sources in language processing (Garagnani *et al.*, 2008). Models solely relying on adaptation or local inhibition have difficulty explaining such specific effects in distant cortical areas.

It should be added that, depending on the paradigm adopted, the N1 response can be elicited either by the repeated presentation of the same stimulus (as in the classic oddball paradigm) or by a range of different stimuli. The latter design is preferred in psycholinguistic studies trying to avoid the response-reducing effect of stimulus repetition (Pulvermüller & Shtyrov, 2006). The N1 is thus often elicited by changing stimuli, for example, by large sets of different words (Pulvermüller *et al.*, 1995; Sereno *et al.*, 1998; Pulvermüller *et al.*, 2009). In this case, the activation of long-term memory circuits can also contribute to the N1 response. Indeed, in these repetition-free designs it was observed that effects of word frequency and semantic word properties were present in late parts of the N1 (or N160) response. Furthermore, even the latency of the responses that reflected lexical status and semantic properties of words was the same in the non-repeat N100 and the MMN experiments (Pulvermüller *et al.*, 2009). This suggests that the family of N1 responses includes variants that are similar to the MMN: N1 to unexpected stimuli and MMN responses can both contain significant contributions from the activation of memory traces, whereas N1 to repeated standards is only marginally affected by it.

4.4 Towards the neurobiology of change detection

We used neural network models mimicking structure and function of nerve cell circuits in specific cortical areas to simulate and explain different facets of the brain dynamics underlying auditory change detection. In Experiment 1 (adopting tabula rasa networks lacking memory representations) local inhibition and neuronal adaptation were found to explain equally well the short-term effects seen in the MMN response, especially the fact that rare deviant stimuli produce larger brain responses than frequently repeated standard stimuli from which they differ in frequency. A further finding was that these two processes synergistically interacted to produce the clearest

(simulated) change-detection response. This finding makes good sense, as in the intact brain both of these mechanisms are at work together. As mentioned earlier in the discussion, however, a randomly connected network endowed with inhibition and adaptation but lacking long-term memory mechanisms cannot account for the experimental result that familiar speech sounds and words produce enhanced MMNs compared with similar unfamiliar sounds and meaningless pseudowords. Instead, as illustrated by Experiment 2, these data can be explained in terms of activation of stimulus-specific memory circuits which, due to their strong internal connections, effectively amplify sensory-elicited activity. Memory circuits with such characteristics spontaneously emerge in randomly and sparsely connected networks (as those used here) in presence of Hebbian synaptic learning rules, solely as a consequence of repeated activation of specific sets of cells (Garagnani *et al.*, 2008). Such circuits may emerge in word learning: when humans use new words, neurons in inferior frontal motor, premotor and prefrontal cortex (which control the articulations) and neurons in superior temporal neurons in primary auditory, auditory belt and parabelt areas – stimulated by the self generated sounds – are active together, leading to Hebbian learning and cell assembly formation across areas. Such cell assembly formation is a plausible correlate of word learning in early infancy, but can, in principle, occur throughout life. Hebbian memory circuits can emerge for static sounds or narrowly defined series of perceptual inputs, for example, tone sequences, phonemes, syllables, words or the sounds of non-verbal actions (e.g., whistling, finger clicks), but also for highly abstract sequences, including syntactic rules (Pulvermüller & Knoblauch, 2009). The present simulation results on memory trace neuronal dynamics, highlighted here in the context of brain indexes of word and pseudoword processing, may therefore explain data on brain responses to learned familiar and

unfamiliar patterns at various levels of combinatorial complexity, including the MMN to syntactic and abstract rule violations (Saarinen *et al.*, 1992; Shtyrov *et al.*, 2003; Bendixen & Schröger, 2008; Bendixen *et al.*, 2009).

We do not intend to claim that all auditory object representations that the brain constructs during the parsing of auditory scenes necessarily consist of learned memory circuits; the formation of perceptual representations that predict future sound events – as observed in auditory streaming or gestalt perception – can be explained, for example, by general (probably inborn) principles of perception (Carlyon, 2004; Winkler *et al.*, 2009). However, these general principles may be underpinned by neuronal circuits similar to the learned memory traces that our models developed.

Recent animal studies have shown MMN-like responses (derived from stimulus specific adaptation) not just in primary auditory cortex (Ulanovsky *et al.*, 2003) but also in the medial geniculate body (Anderson et al., 2009; Antunes *et al.*, 2010) and inferior colliculus (Malmierca *et al.*, 2009) of the rat. In addition, deviance-related responses have been detected also in the human auditory brainstem (Slabu *et al.*, 2010; Grimm *et al.*, 2010) very early in time (30-40 ms after stimulus onset). While the modelling of subcortical structures was not the main focus of this work[8] and represents an interesting future direction, we speculate that the generation of change detection responses in such structures may be driven by mechanisms of adaptation and inhibition similar to those implemented here. Furthermore, the contribution of learned memory traces to the MMN should be, according to our model and theory, a cortical function.

[8] Though note that our networks do contain area-specific inhibitory feedback links (see (Garagnani *et al.*, 2008) for details) implementing global regulatory functions that are thought to be mediated, in the brain, by cortico-striato-thalamic loops (Miller & Wickens, 1991; Wickens, 1993).

The neurobiological instantiations of memory traces as distributed circuits spanning different cortical areas allowed us to explain cognitive long-term memory related effects that are well known to be present in the MMN response. Local (or lateral) inhibition mechanisms, introducing competition between learned memory circuits, were also found to be necessary for replicating the full pattern of MMN responses to familiar and unfamiliar stimuli. Neuronal adaptation did not serve this function, whereas both adaptation and inhibition interacting together with the memory circuits also provided a full explanation of the crucial data considered here.

The present simulations set the stage for an integration of cognitive models of the MMN and N1 with neurobiological models that view action-perception circuits as the basis of memory in the cortex. Evidence from animal research indicates that long-term memory units in the cortex are best thought of as distributed neuronal assemblies bridging frontal (action-related) and posterior (perception-related) areas (Fuster, 1995; , 1997; , 2003). Direct neurophysiological recordings in monkeys show activity of frontal and temporal neurons in memory tasks, thus suggesting a role of the distributed long-term circuits also in short term or working memory (Goldman-Rakic, 1995). The temporary activation of a circuit would accordingly be the neurobiological basis of working memory. This is precisely the mechanism by which temporary working memory processes contact long-term memory in our present simulations (see also (Zipser *et al.*, 1993)). By using a biologically realistic network model to account for MMN dynamics and the related cognitive processes of change detection, automatic attention reorientation, and memory trace access, we hope to contribute to the integration of cognitive and neurobiological theory.

4.5 Conclusions

The present neurobiologically grounded model of memory and perception provides the first unified account for neurophysiological data on both basic MMN effects as they are seen in frequency deviance detection and cognitive phenomena, such as the enhanced MMNs to meaningful words as compared with unfamiliar pseudowords. The model allows integrating recent evidence on familiarity-related enhancement of the MMN with previous experimental and computational data on auditory change detection. The results demonstrate that, while the activation and reverberation of long-term memory traces for familiar items realised as distributed neuronal circuits (Pulvermüller, 1999) explains the enhanced MMN responses to familiar stimuli compared with unfamiliar ones, short-term neuronal phenomena (adaptation and inhibition) can explain other features of the brain response observed generally to acoustic change. Hence, only the combined presence of basic cortical mechanisms – local inhibition, spike-rate adaptation, and long-term synaptic plasticity yielding memory circuits and reverberatory short-term memory activity therein – is sufficient to provide the brain with two critical skills that played a key role in its evolutionary success, namely, the ability to detect change automatically and the capacity to learn, store and recognise patterns that are familiar and meaningful. These mechanisms are not restricted to the learning, extraction and storage of simple stimulus regularities, but can cover vocabularies of learned signs and even highly abstract sequential patterns including syntactic rules.

References

Abeles, M. (1991) *Corticonics - Neural circuits of the cerebral cortex*. Cambridge University Press, Cambridge.

Artola, A. & Singer, W. (1993) Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends in Neurosciences*, **16**, 480-487.

Bendixen, A. & Schröger, E. (2008) Memory trace formation for abstract auditory features and its consequences in different attentional contexts. *Biol Psychol*, **78**, 231-241.

Bendixen, A., Schröger, E. & Winkler, I. (2009) I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *J Neurosci*, **29**, 8447-8451.

Braitenberg, V. & Schüz, A. (1998) *Cortex: statistics and geometry of neuronal connectivity*. Springer, Berlin.

Carlyon, R.P. (2004) How the brain separates sounds. *Trends Cogn Sci*, **8**, 465-471.

Catani, M., Jones, D.K. & Ffytche, D.H. (2005) Perisylvian language networks of the human brain. *Annals of Neurology*, **57**, 8-16.

Dehaene-Lambertz, G. (1997) Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, **8**, 919-924.

Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, **15**, 399-402.

Fagot, J.l. & Cook, R.G. (2006) Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences*, **103**, 17564-17567.

Frangos, J., Ritter, W. & Friedman, D. (2005) Brain potentials to sexually suggestive whistles show meaning modulates the mismatch negativity. *Neuroreport*, **16**, 1313-1317.

Friston, K. (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*, **360**, 815-836.

Fry, D.B. (1966) The development of the phonological system in the normal and deaf child. In Smith, F., Miller, G.A. (eds.) *The genesis of language*. MIT Press, Cambridge, MA, pp. 187-206.

Fuster, J.M. (1995) *Memory in the cerebral cortex*. MIT Press, Cambridge, MA.

Fuster, J.M. (1997) Network memory. *Trends in Neurosciences*, **20**, 451-459.

Fuster, J.M. (2001) The prefrontal cortex--an update: time is of the essence. *Neuron*, **30**, 319-333.

Fuster, J.M. (2003) *Cortex and mind: Unifying cognition*. Oxford University Press, Oxford.

Garagnani, M. (2009) Understanding language and attention: brain-based model and neurophysiological experiments *MRC Cognition and Brain Sciences Unit*. University of Cambridge, Cambridge, pp. 127.

Garagnani, M., Shtyrov, Y. & Pulvermüller, F. (2009) Effects of Attention on what is known and what is not: MEG Evidence for Functionally Discrete Memory Circuits. *Front Hum Neurosci*, **3**, 10.

Garagnani, M., Wennekers, T. & Pulvermüller, F. (2008) A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *Eur J Neurosci*, **27**, 492-513.

Garrido, M.I., Kilner, J.M., Stephan, K.E. & Friston, K.J. (2009) The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, **120**, 453-463.

Gilbert, C.D. & Wiesel, T.N. (1983) Clustered Intrinsic Connections in Cat Visual-Cortex. *Journal of Neuroscience*, **3**, 1116-1133.

Goldman-Rakic, P.S. (1995) Cellular basis of working memory. *Neuron*, **14**, 477-485.

Hauk, O., Shtyrov, Y. & Pulvermüller, F. (2006) The sound of actions as reflected by mismatch negativity: rapid activation of cortical sensory-motor networks by sounds associated with finger and tongue movements. *Eur J Neurosci*, **23**, 811-821.

Hebb, D.O. (1949) *The organization of behavior*. John Wiley, New York.

Hubel, D. (1995) *Eye, brain, and vision*. Scientific American Library, New York.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levanen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H. & Belliveau, J.W. (2004) Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci U S A*.

Jacobsen, T., Horenkamp, T. & Schroger, E. (2003) Preattentive memory-based comparison of sound intensity. *Audiol Neurootol*, **8**, 338-346.

Jacobsen, T., Horvath, J., Schröger, E., Lattner, S., Widmann, A. & Winkler, I. (2004) Pre-attentive auditory processing of lexicality. *Brain Lang*, **88**, 54-67.

Jacobsen, T. & Schroger, E. (2003) Measuring duration mismatch negativity. *Clin Neurophysiol*, **114**, 1133-1143.

Jacobsen, T. & Schröger, E. (2001) Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, **38**, 723-727.

Jacobsen, T., Schröger, E., Winkler, I. & Horvath, J. (2005) Familiarity Affects the Processing of Task-irrelevant Auditory Deviance. *J Cogn Neurosci*, **17**, 1704-1713.

Korpilahti, P., Krause, C.M., Holopainen, I. & Lang, A.H. (2001) Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, **76**, 332-339.

Korzyukov, O., Alho, K., Kujala, A., Gumenyuk, V., Ilmoniemi, R.J., Virtanen, J., Kropotov, J. & Näätänen, R. (1999) Electromagnetic responses of the human auditory cortex generated by sensory-memory based processing of tone-frequency changes. *Neurosci Lett*, **276**, 169-172.

Kujala, T., Kallio, J., Tervaniemi, M. & Näätänen, R. (2001) The mismatch negativity as an index of temporal processing in audition. *Clin Neurophysiol*, **112**, 1712-1719.

Malenka, R.C. & Nicoll, R.A. (1999) Neuroscience - Long-term potentiation - A decade of progress? *Science*, **285**, 1870-1874.

Matthews, G.G. (2001) *Neurobiology: molecules, cells and systems*. Blackwell Science.

May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., Taylor, J.G. & Näätänen, R. (1999) Frequency change detection in human auditory cortex. *J Comput Neurosci*, **6**, 99-120.

May, P.J. & Tiitinen, H. (2010) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, **47**, 66-122.

Miller, R. & Wickens, J.R. (1991) Corticostriatal cell assemblies in selective attention and in representation of predictable and controllable events: a general statement of corticostriatal interplay and the role of striatal dopamine. *Concepts in Neuroscience*, **2**, 65-95.

Näätänen, R. (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, **13**, 201-288.

Näätänen, R., Gaillard, A.W. & Mäntysalo, S. (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, **42**, 313-329.

Näätänen, R., Jacobsen, T. & Winkler, I. (2005) Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology*, **42**, 25-32.

Näätänen, R., Kujala, T. & Winkler, I. (2010) Mechanisms underlying conscious perception in audition: a unique window to central auditory processing openend by the mismatch neagticity (MMN) and related responses. *Psychophysiology*, **. In press.**

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J. & Alho, K. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, **385**, 432-434.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M. & Paavilainen, P. (1993) Development of a memory trace for a complex sound in the human brain. *NeuroReport*, **4**, 503-506.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P. & Winkler, I. (2001) 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, **24**, 283-288.

Pandya, D.N. & Yeterian, E.H. (1985) Architecture and connections of cortical association areas. In Peters, A., Jones, E.G. (eds.) *Cerebral cortex. Vol. 4. Association and auditory cortices*. Plenum Press, London, pp. 3-61.

Petkov, C.I., Kayser, C., Augath, M. & Logothetis, N.K. (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol*, **4**, e215.

Petrides, M. & Pandya, D.N. (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol*, **7**, e1000170.

Pulvermüller, F. (1999) Words in the brain's language. *Behavioral and Brain Sciences*, **22**, 253-336.

Pulvermüller, F. & Assadollahi, R. (2007) Grammar or serial order?: Discrete combinatorial brain mechanisms reflected by the syntactic Mismatch Negativity. *Journal of Cognitive Neuroscience*, **in press**.

Pulvermüller, F. & Fadiga, L. (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, **11**, 1-11.

Pulvermüller, F. & Knoblauch, A. (2009) Discrete combinatorial circuits emerging in neural networks: a mechanism for rules of grammar in the human brain? *Neural Netw*, **22**, 161-172.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R.J. & Näätänen, R. (2001) Memory traces for words as revealed by the mismatch negativity. *Neuroimage*, **14**, 607-616.

Pulvermüller, F., Lutzenberger, W. & Birbaumer, N. (1995) Electrocortical distinction of vocabulary types. *Electroencephalography and Clinical Neurophysiology*, **94**, 357-370.

Pulvermüller, F. & Preissl, H. (1991) A cell assembly model of language. *Network: Computation in Neural Systems*, **2**, 455-468.

Pulvermüller, F. & Shtyrov, Y. (2003) Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimage*, **20**, 159-172.

Pulvermüller, F. & Shtyrov, Y. (2006) Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, **79**, 49-71.

Pulvermüller, F., Shtyrov, Y. & Hauk, O. (2009) Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, **110**, 81-94.

Rauschecker, J.P. & Scott, S.K. (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*, **12**, 718-724.

Rauschecker, J.P. & Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A*, **97**, 11800-11806.

Romanski, L.M., Bates, J.F. & Goldman-Rakic, P.S. (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol*, **403**, 141-157.

Saarinen, J., Paavilainen, P., Schöger, E., Tervaniemi, M. & Näätänen, R. (1992) Representation of Abstract Attributes of Auditory-Stimuli in the Human Brain. *Neuroreport*, **3**, 1149-1151.

Schröger, E., Näätänen, R. & Paavilainen, P. (1992) Event-related potentials reveal how non-attended complex sound patterns are represented by the human brain. *Neuroscience Letters*, **146**, 183-186.

Sereno, S.C., Rayner, K. & Posner, M.I. (1998) Establishing a time line for word recognition: evidence from eye movements and event-related potentials. *NeuroReport*, **13**, 2195-2200.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R.J. & Näätänen, R. (2000) Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, **12**, 657-663.

Shtyrov, Y., Kujala, T. & Pulvermüller, F. (2010) Interactions between language and attention systems: early automatic lexical processing? *Journal of Cognitive Neuroscience*, **22**, 1465-1478.

Shtyrov, Y., Pihko, E. & Pulvermüller, F. (2005) Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? *Neuroimage*, **27**, 37-47.

Shtyrov, Y. & Pulvermüller, F. (2007) Early activation dynamics in the left temporal and inferior-frontal cortex reflect semantic context integration. *Journal of Cognitive Neuroscience*, **19**, 1633-1642.

Shtyrov, Y., Pulvermüller, F., Näätänen, R. & Ilmoniemi, R.J. (2003) Grammar processing outside the focus of attention: an MEG study. *Journal of Cognitive Neuroscience*, **15**, 1195-1206.

Tiitinen, H., Alho, K., Huotilainen, M., Ilmoniemi, R.J., Simola, J. & Näätänen, R. (1993) Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity. *Psychophysiology*, **30**, 537-540.

Uppenkamp, S., Johnsrude, I.S., Norris, D., Marslen-Wilson, W. & Patterson, R.D. (2006) Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage*, **31**, 1284-1296.

Voss, J.L. (2009) Long-term associative memory capacity in man. *Psychon Bull Rev*, **16**, 1076-1081.

Wickens, J.R. (1993) *A theory of the striatum*. Pergamon Press, Oxford.

Wilson, S.M., Saygin, A.P., Sereno, M.I. & Iacoboni, M. (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci*, **7**, 701-702.

Winkler, I., Denham, S.L. & Nelken, I. (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci*, **13**, 532-540.

Winkler, I., Karmos, G. & Näätänen, R. (1996) Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Res*, **742**, 239-252.

Zipser, D., Kehoe, B., Littlewort, G. & Fuster, J. (1993) A spiking network model of short-term active memory. *Journal of Neuroscience*, **13**, 3406-3420.

Appendix A

Each cell or "node" of the network represents a cortical column of approximately 0.25mm^2 size (Hubel, 1995), containing ~$2.5 \cdot 10^4$ neurons (Braitenberg & Schüz, 1998)[9]. The state of each cell x is uniquely defined by its membrane potential $V(x,t)$, representing the average of the sum of all excitatory and inhibitory postsynaptic potentials acting upon neural pool (cluster) x at time t. The membrane potential $V(x,t)$ at time t of a model cell x with membrane time-constant τ is governed by the equation:

$$\tau \cdot \frac{dV(x,t)}{dt} = -V(x,t) + V_{In}(x,t) \qquad \text{(A.1)}$$

where $V_{In}(x,t)$ is the total input to x (sum of all excitatory and inhibitory synaptic inputs to cell x at time t; inhibitory synapses are given a negative sign).

The output of an excitatory cell x at time t is defined as follows:

$$\mathrm{O}(x,t) = \begin{cases} 0 & \text{if } V(x,t) \le \varphi \\ (V(x,t) - \varphi) & \text{if } 0 < (V(x,t) - \varphi) \le 1 \\ 1 & \text{otherwise} \end{cases} \qquad \text{(A.2)}$$

$\mathrm{O}(x,t)$ represents the average (graded) firing rate (number of action potentials per time unit) of cluster x at time t; it is a piecewise-linear sigmoid function of the cell's membrane potential $V(x,t)$, clipped into the range [0, 1] and with slope 1 between the lower and upper thresholds φ and $\varphi + 1$. The output $\mathrm{O}(x,t)$ of an inhibitory cell is 0 if $V(x,t) < 0$, and $V(x,t)$ otherwise.

[9] These figures are meant to provide only an estimate of the grain of the model; as noted by Hubel (1995), the size of a macrocolumn (or "module") varies substantially between cortical layers (ranging from 0.1mm^2 in layer 4C to 4mm^2 in layer 3) and cortical areas (*ibid.*, p.130).

Neuronal adaptation is realised (in excitatory cells only) by allowing the threshold φ in Eq. (A.2) to be cell-specific and vary in time. More precisely:

$$\varphi(x, t) = \alpha \cdot \omega(x,t) \qquad \text{(A.3)}$$

where $\omega(x,t)$ is the time-average of the cell's recent output and α is the "adaptation strength" (see below for parameter values used in the simulations).

For any excitatory cell x, the approximate time-average $\omega(x,t)$ of its output $O(x,t)$ is estimated by integrating Eq. (A.4) below, assuming initial average $\omega(x,0)=0$:

$$\tau_A \cdot \frac{d\omega(x,t)}{dt} = -\omega(x,t) + O(x,t) \qquad \text{(A.4)}$$

The low-pass dynamics of the cells (Eq. (A.1), (A.2) and (A.4)) are integrated using the Euler scheme with step size Δt, where $\Delta t = 0.5$ (in arbitrary units of time). Other parameter values are reported below.

The learning rule used to simulate synaptic plasticity is based on the Artola-Bröcher-Singer model of LTP/LTD (Artola & Singer, 1993). In the implementation, we discretized the continuous range of possible synaptic efficacy changes into two possible levels, $+\Delta w$ and $-\Delta w$ (with $\Delta w \ll 1$ and fixed). We defined as "active" any link from a cell x such that the output $O(x,t)$ of cell x at time t is larger than θ_{pre}, where $\theta_{pre} \in]0,1]$ is an arbitrary threshold representing the minimum level of presynaptic activity required for LTP (or LTD) to occur. Thus, given any two cells x and y linked with weight $w_t(x,y)$, the new weight $w_{t+1}(x,y)$ is calculated as follows:

$$w_{t+1}(x,y) = \begin{cases} w_t(x,y)+\Delta w & \text{if } O(x,t) \geq \theta_{pre} \text{ and } V(y,t) \geq \theta_+ \\ w_t(x,y)-\Delta w & \text{if } O(x,t) \geq \theta_{pre} \text{ and } \theta_- \leq V(y,t) < \theta_+ \\ w_t(x,y)-\Delta w & \text{if } O(x,t) < \theta_{pre} \text{ and } V(y,t) \geq \theta_+ \\ w_t(x,y) & \text{otherwise} \end{cases} \qquad \text{(A.5)}$$

Parameter values used for the simulations are:

(Eq. (A.1)) Excitatory cells: τ = 2.5 (in simulation time-steps);

Inhibitory cells: τ = 5 (in simulation time-steps);

(Eq. (A.3)) No adaptation: $\alpha = 0$;

Average adaptation: $\alpha = 5$;

Maximum adaptation: $\alpha = 10$;

(Eq. (A.4)) Time constant for computing gliding-average of cell activity:

$\tau_A = 15$ (in simulation time-steps);

(Eq. (A.5)) Post-synaptic potential thresholds for LTP/LTD: $\theta_- = 0.15$, $\theta_+ = 0.25$;

Pre-synaptic output activity required for synaptic change: $\theta_{pre} = 0.05$;

Learning rate: $\Delta w = 0.0005$.

Appendix B

In general, the centre of mass R of a system of particles is defined as the average of their positions, r_i, weighted by their masses, m_i:

$$R = \frac{\sum m_i r_i}{\sum m_i} \qquad \text{(B.1)}$$

Assuming additive effects of sources of neuronal activity and a localisation procedure that finds a point source of this centre (e.g., the equivalent source dipole method), the "masses" m_i correspond to the activation values in areas A1 and AB emerging from the simulations, and their positions, to the locations of A1 and AB. The area-specific simulated peak activations (in arbitrary units) for N1 and MMN responses were:

$$\text{A1:} \quad m_{\text{N1,A1}} = 5.38; \; m_{\text{MMN, A1}} = 2.65$$

$$\text{AB:} \quad m_{\text{N1,AB}} = 1.15; \; m_{\text{MMN, AB}} = 1.19$$

Without any loss of generality, we can assume the positions of A1 and AB to be r_1= +L and r_2= –L (with respect to the "centre location" C=(A1–AB)/2, Fig. S2.(D)). Thus, by applying Eq. (B.1) above, the locations of the centres of mass (yellow circles in Figure 3) of the simulated N1 and MMN responses are $C_{N1} \approx 0.648 \cdot L$ and $C_{MMN} \approx 0.380 \cdot L$, respectively; the corresponding total strengths are given by $m_{\text{N1, A1}} + m_{\text{N1, AB}} = 6.53$ and $m_{\text{MMN,A1}} + m_{\text{MMN,AB}} = 3.84$ (in arbitrary units).

Appendix C

The algorithm $\mathcal{A}$ for generating pseudoword patterns described in (Garagnani *et al.*, 2008) takes a set S of n (word, W) patterns (each defined on a grid of size 25-by-25 cells) as input and produces a new set $S' = \mathcal{A}(S)$ as output containing n (pseudoword, PW) patterns of the same size, as follows: (1) each word pattern $x_i \in S$ ($i \in \{1, .., n\}$) is divided into 25 five-by-five squares; (2) every new pseudoword pattern $y \in S'$ is built by combining 25 squares taken at random (uniform probability distribution) from the x_i patterns, and preserving all the original squares' positions. As a result, each new pattern y ends up containing, on average, $25/n$ squares from each pattern $x_i \in S$, with $i \in \{1,.., n\}$. Thus, given a set S of n patterns, the new set $S' = \mathcal{A}(S)$ contains patterns that overlap (on average) by $1/n^{th}$ with the patterns in S.

Here, we applied algorithm $\mathcal{A}$ to a set S_1 of 12 randomly generated patterns and produced a second set $S_2 = \mathcal{A}(S_1)$ of patterns (overlapping by $1/12 \approx 8.33\%$ with those in S_1). These two sets were then used as follows: first, we split each set S_i into two (randomly chosen) halves of six patterns, called S_i^W and S_i^{PW} (for $i \in \{1,2\}$). Second, we trained the network with the set $W = S_1^W \cup S_2^W$; this changed the lexical status of the patterns in W to that of "words". Third, we used $S_1^{PW} \cup S_2^{PW}$ as pseudoword pattern set, and formed four sets of pairs (STD, DEV) – to be used for oddball stimulation – in the following way: we generated six (l=W, r=W) pairs using $l \in S_1^W$ and $r \in S_2^W$, six (l=W, r=PW) pairs using $l \in S_1^W$, $r \in S_2^{PW}$; six (l=PW, r=W) pairs using $l \in S_1^{PW}$, $r \in S_2^W$; and six (l=PW, r=PW) pairs using $l \in S_1^{PW}$, $r \in S_2^{PW}$. Because $S_2 = \mathcal{A}(S_1)$, any two patterns l, r such that $l \in S_1$ and $r \in S_2$ overlap (on average) by

~8.33%. Thus, any pattern pair in any of the above four sets also exhibits (on average) the same, fixed amount of overlap (8.33%).[10]

[10] The actual overlap between any two binary patterns can only be one of a finite set of values, containing only the ratios between number of overlapping cells and pattern size (e.g., if size = 17 cells, 1 cell overlap ≈ 5.9%, 2 cells overlap ≈ 11.8%, 3 cells overlap ≈ 17.6%, etc.).

Figure Legends

Figure 1. **Mismatch Negativity (MMN) responses**. **Left**: responses to rare, occasional (deviant) stimuli placed among frequently repeated ones (standard), and resulting MMN (shaded area) [*adapted from* Shtyrov et al. (2005), their Fig. 4]. **Right**: MMNs produced by familiar (word) and unfamiliar (pseudoword) deviant items; note the enhanced MMN response to words compared to pseudowords [*adapted from* Pulvermüller et al. (2001), their Fig. 4].

Figure 2. **Model architectures and their brain-structural basis. (A)** Six different areas of the left perisylvian language cortex are indicated in different colours. M1= primary motor; PM=pre-motor; PF=pre-frontal; A1=auditory core, AB=belt; PB=parabelt. Black arrows indicate long-distance cortico-cortical connections between auditory (Wernicke's) and motor (Broca's) association areas. Experimental evidence (Romanski *et al.*, 1999; Petkov *et al.*, 2006) indicates that the auditory system consists of three main areas, A1, AB and PB; although these systems have been studied mainly in macaques, homologous areas of the human auditory cortex (Brodmann Areas 41, 42 and 22) lend themselves to an analogous parcellation (Uppenkamp *et al.*, 2006). **(B)** Network of 3 areas used in Experiment 1. Each colour-filled oval represents a pool of (excitatory) pyramidal cells; inhibitory mechanisms are not shown. Lines between areas represent random, sparse and patchy connections. **(C)** Network of all 6 perisylvian areas used in Experiment 2. [*Adapted from* Garagnani et al. (2008), their Fig. 3].

Figure 3. Contribution of local inhibition and adaptation to tone-elicited standard and deviant responses and to the MMN to frequency change. The 4 panels **(A-D)** show the results for the 4 networks used in Experiment 1. Simulated STD (dashed lines) and DEV (solid lines) responses are plotted on the left, resultant MMNs on the right insets. The *x*-axes give time (in simulation time-steps); *y*-axes give total network activation (averaged across trials and stimuli). Vertical bars in the MMN curves indicate standard errors (SE). Black segments on the *x* axis indicate stimulus onset-time and duration. Note the presence of an MMN (DEV > STD response) in all cases except for panel **(A)**, where the STD is larger than the DEV response.

Figure 4. **Simulation results on area-specific activations underlying the MMN.** Simulated activations in areas A1, AB and PB underlying STD **(A),** DEV **(B)** and MMN responses **(C)** in networks with inhibition-only are plotted against time. Note the strongly attenuated STD response in AB. **(D)** An illustration of how the differential A1 and AB activations contributing to the N1 and MMN responses can lead to different locations of the single estimated underlying current dipole (yellow circles; circle size indicates "source strength"). A1 and AB peak activations are depicted as striped (MMN response) and filled (N1 response) vertical bars.

Figure 5. **Simulated standard, deviant and MMN responses to familiar and unfamiliar stimuli.** Simulated responses to familiar "word" (W, red curves) and unfamiliar "pseudoword" patterns (PW, blue lines) presented in an oddball design for adaptation-only (first column from the left), inhibition-only (column 2) and combined adaptation-inhibition (columns 3-4) networks. STD **(A),** DEV **(B)** and MMN responses **(C)** are plotted against time. Note the overlap between W and PW MMNs (dashed red and dashed blue curves) for the adaptation-only network (not present in the others) and the larger MMN to words than to PW at time-steps 6 & 7 (note the different scale used for panel (C), column 4). Vertical bars indicate SE.

5. Figure
Click here to download high resolution image

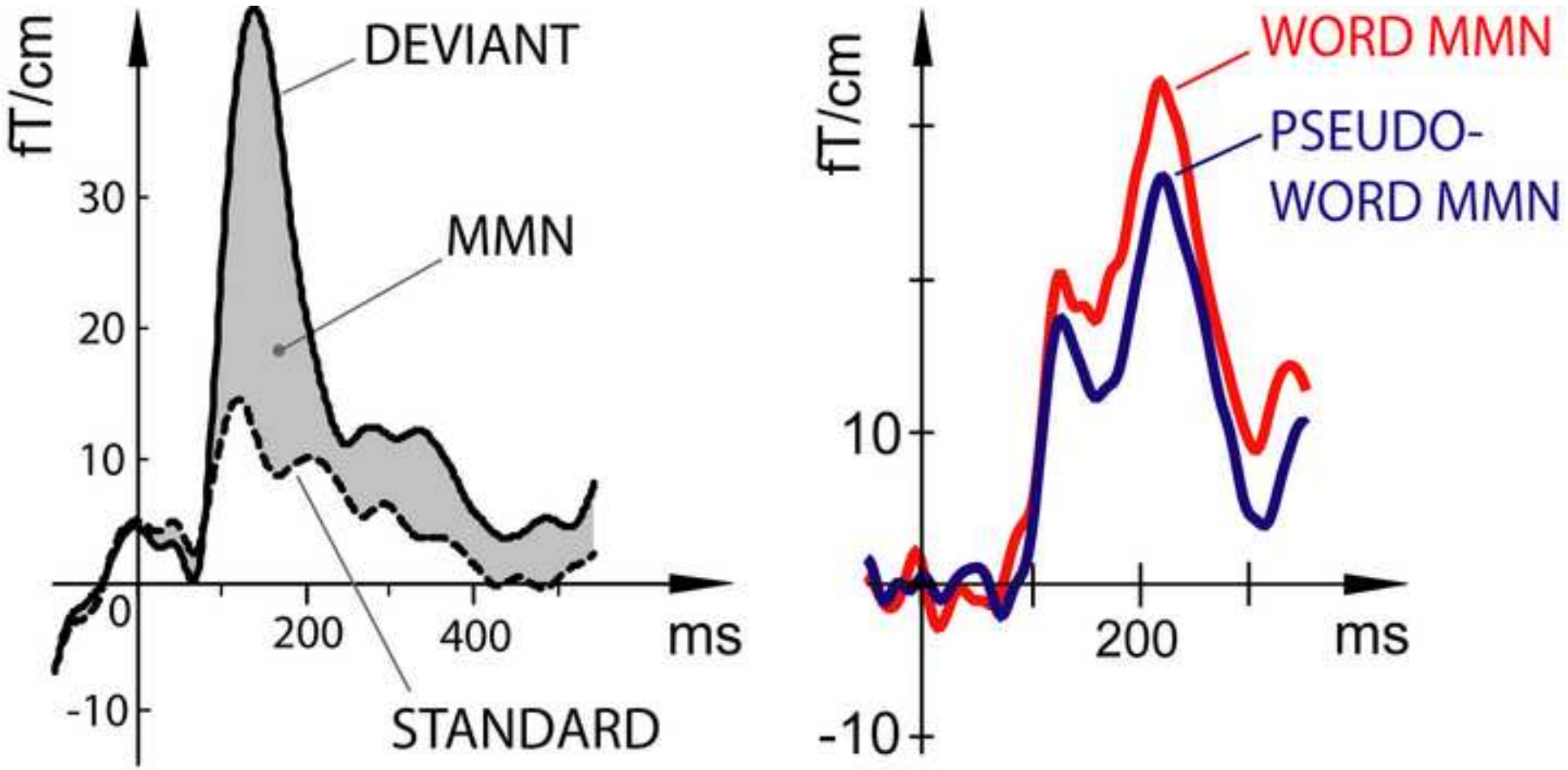

Figure.1

5. Figure

Click here to download high resolution image

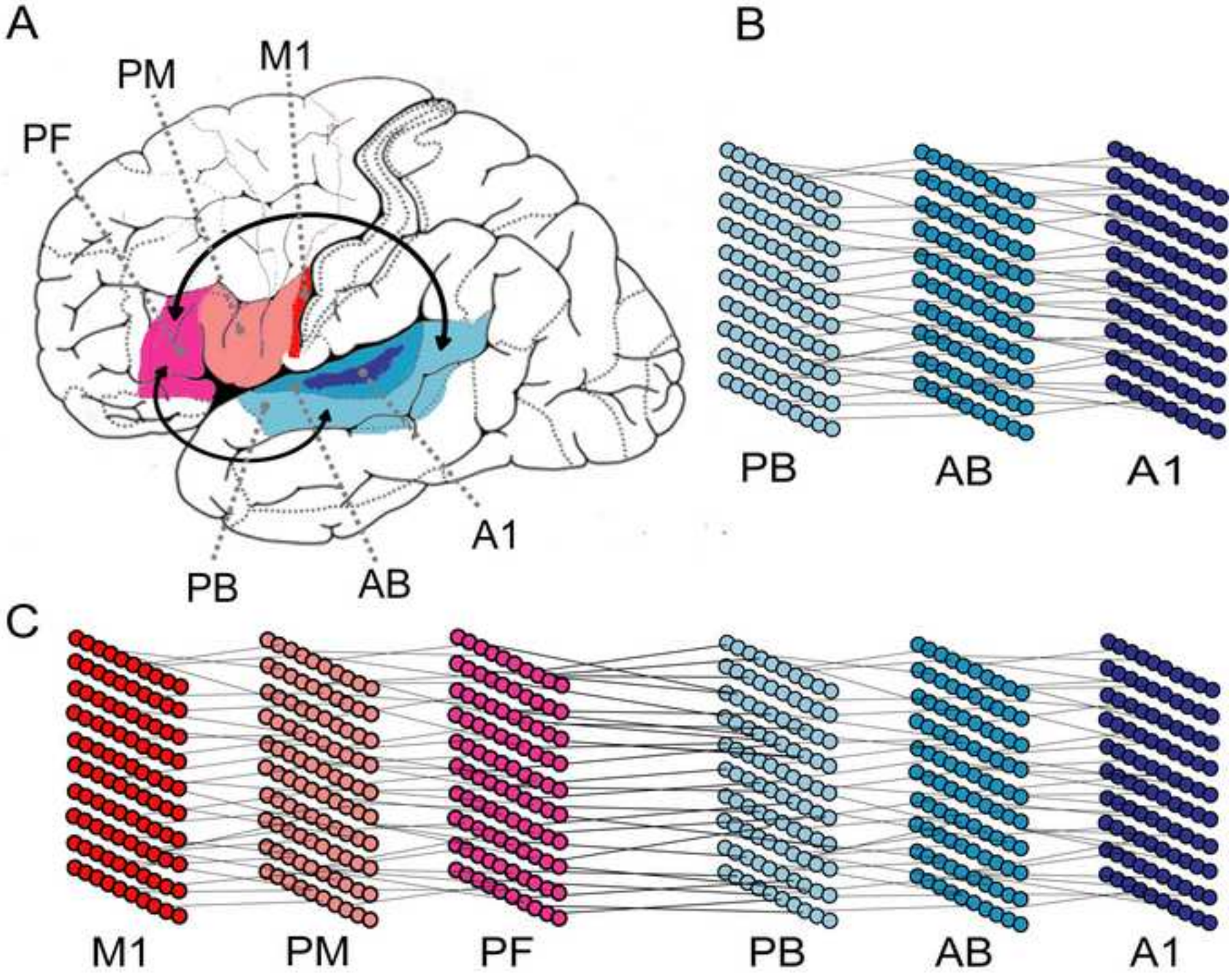

Figure.2

5. Figure

Click here to download high resolution image

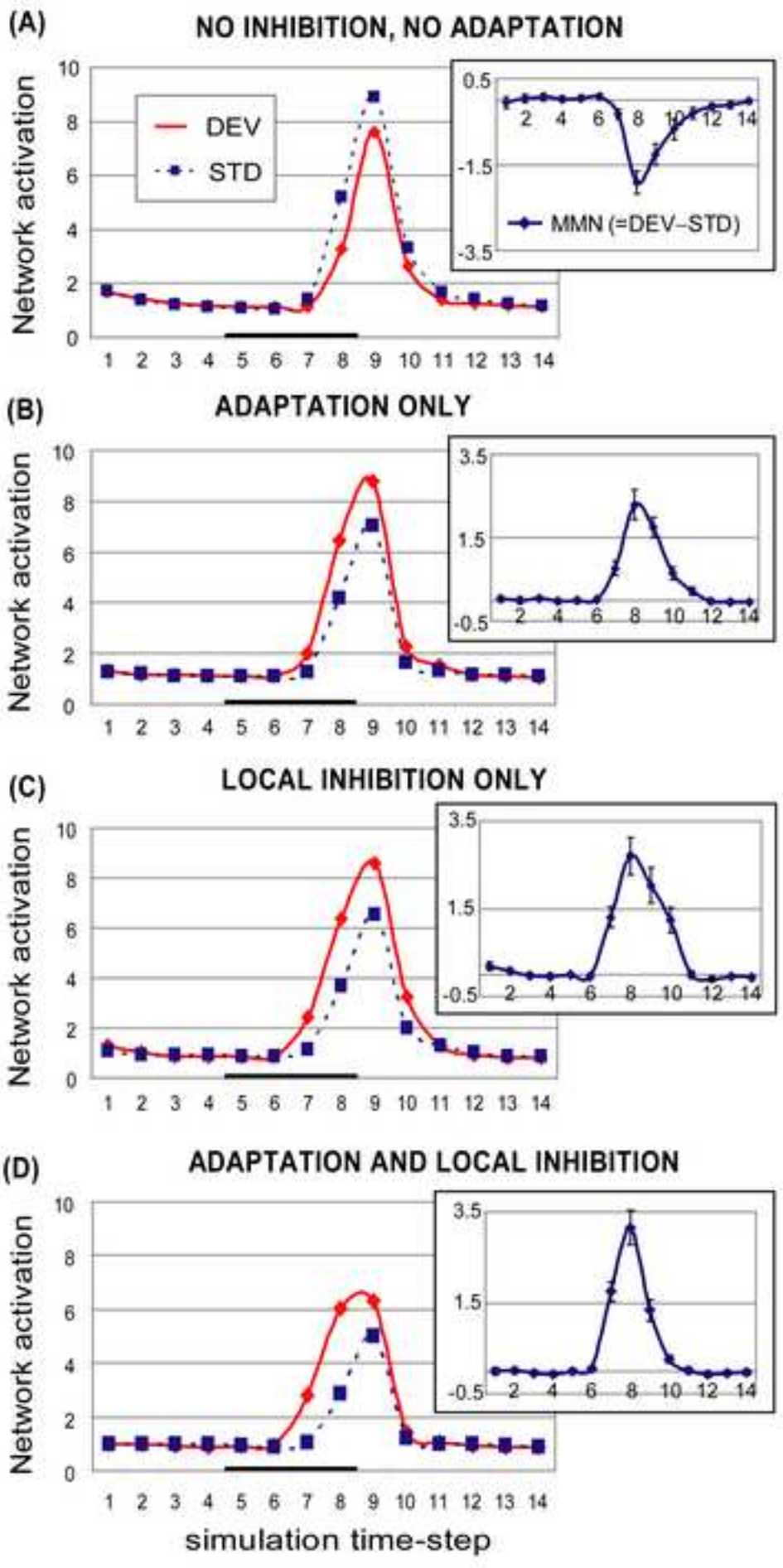

Figure 3

5. Figure
Click here to download high resolution image

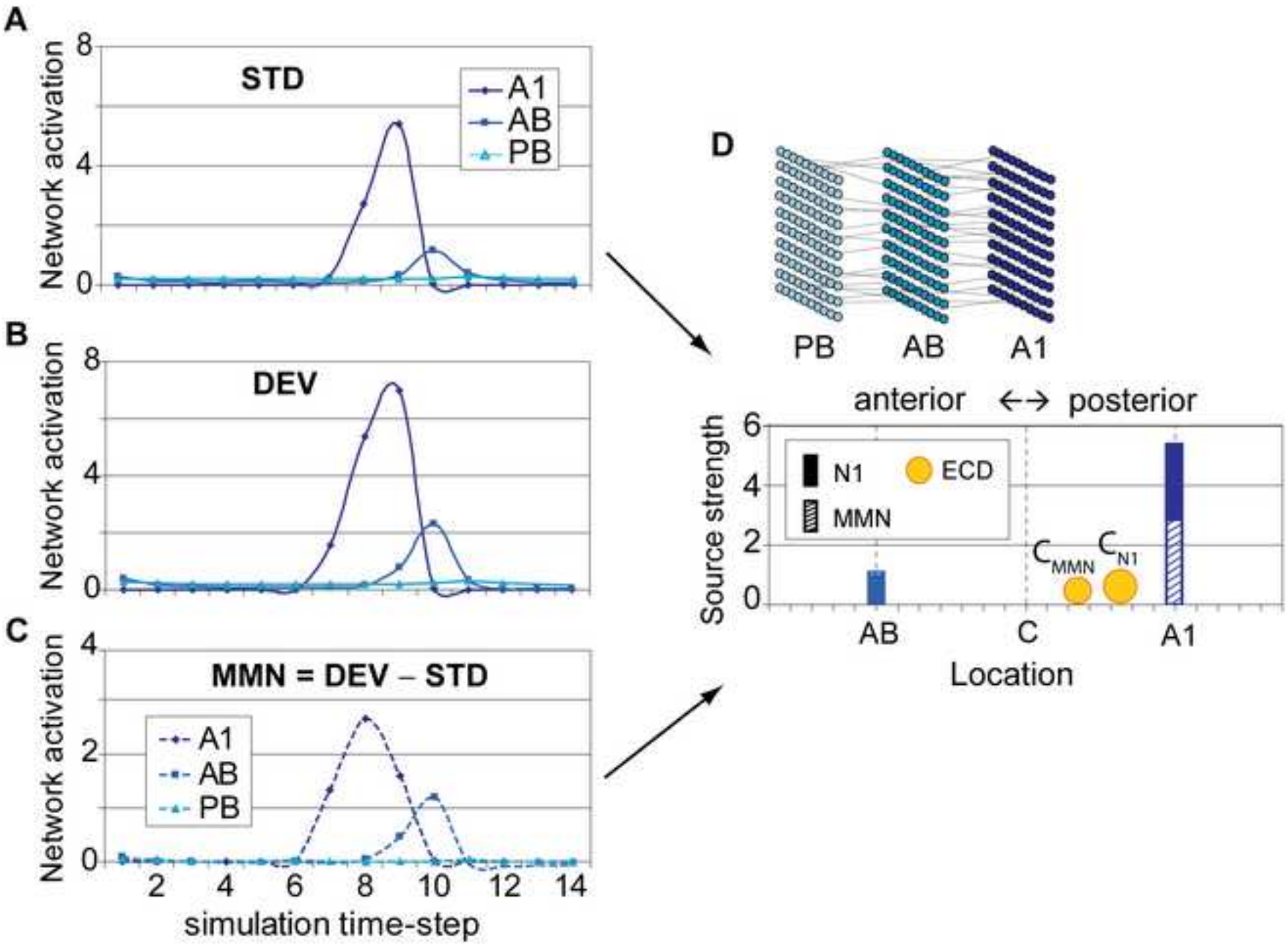

Figure.4

5. Figure

Click here to download high resolution image

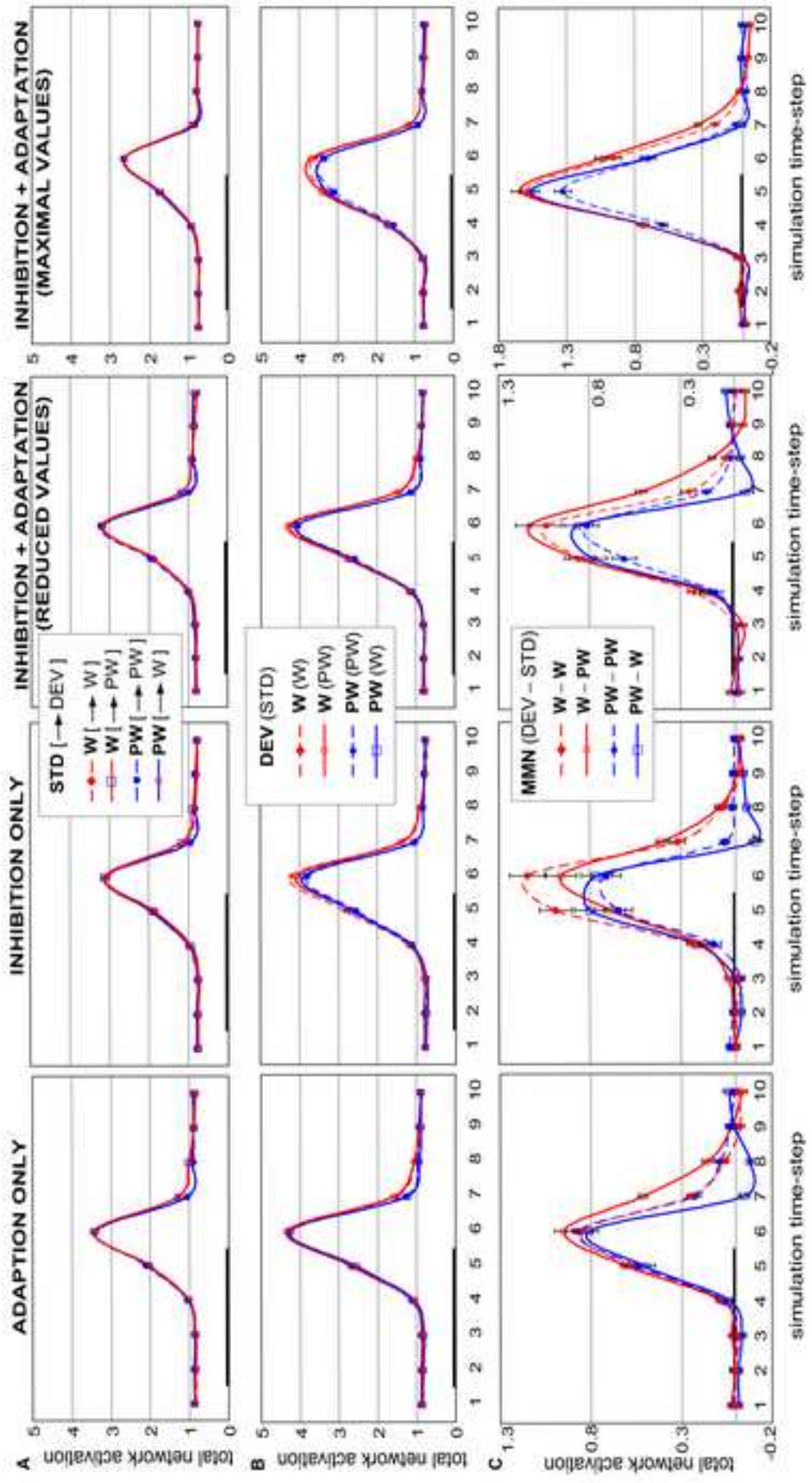

Figure.5