The GluA2 phenomenon can be explained by the loss of coupling between Ser831 phosphorylation and channel gating (**Fig. 1c**), although the molecular mechanism of such inhibition is still unknown. Nevertheless, the dominance of the GluA2 subunit presents a paradox: the channel conductance of synaptic AMPARs is increased during LTP[9], despite the presence of GluA2 subunits in the receptors[11,12]. The enigma was partially resolved by the observation that synaptic AMPARs can temporally change their subunit composition toward GluA2-lacking forms during the early phase of LTP[13]. However, the study by Kristensen and colleagues[1] goes deeper into AMPAR regulation by looking at properties of the AMPAR when it is in the complex with transmembrane AMPAR regulatory proteins (TARPs), which are known to regulate both AMPAR synaptic trafficking and receptor function[14]. These findings offer a new explanation for the underlying regulation of synaptic AMPARs: TARPs (TARP γ2 and γ8) have another, unexpected role in preserving CaMKII-mediated potentiation of channel conductance despite the presence of GluA2, perhaps because the gate coupling efficiency can still be increased (**Fig. 1d**). The question of how exactly TARPs can exert this function will likely await crystallographic studies of AMPAR-TARP complexes, which have recently expanded our understanding of the structural basis for activation of homomeric AMPARs[15].

Of particular importance, the authors also looked at the regulation of AMPARs in hippocampal neurons expressing different GluA1 mutants. This allowed them to evaluate the functional roles of different key receptor residues in the much more relevant neuronal environment. The analysis of resulting changes in AMPAR properties explicitly confirmed that the Ser831 phosphorylation is specific for the increase in single-channel conductance and that other receptor residues or phosphorylation of receptor-associated proteins were not involved. Thus, Kristensen *et al.*[1] offer insights into intrinsic molecular events that underlie the regulation of neuronal AMPARs by CaMKII, and likely the regulation of synaptic strength during hippocampal LTP. This impressive feat of structure-function analysis advances our understanding of the complexity of molecular mechanisms behind the form of synaptic plasticity that is responsible for memory.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Kristensen, A.S. *et al. Nat. Neurosci.* **14**, 727–735 (2011).
2. Whitlock, J.R., Heynen, A.J., Shuler, M.G. & Bear, M.F. *Science* **313**, 1093–1097 (2006).
3. Citri, A. & Malenka, R.C. *Neuropsychopharmacology* **33**, 18–41 (2008).
4. Lisman, J., Schulman, H. & Cline, H. *Nat. Rev. Neurosci.* **3**, 175–190 (2002).
5. Barria, A., Muller, D., Derkach, V., Griffith, L.C. & Soderling, T.R. *Science* **276**, 2042–2045 (1997).
6. Lee, H.K., Barbarosie, M., Kameyama, K., Bear, M.F. & Huganir, R.L. *Nature* **405**, 955–959 (2000).
7. Derkach, V., Barria, A. & Soderling, T.R. *Proc. Natl. Acad. Sci. USA* **96**, 3269–3274 (1999).
8. Derkach, V.A. *Biophys. J.* **84**, 1701–1708 (2003).
9. Benke, T.A., Lüthi, A., Isaac, J.T. & Collingridge, G.L. *Nature* **393**, 793–797 (1998).
10. Jin, R., Banke, T.G., Mayer, M.L., Traynelis, S.F. & Gouaux, E. *Nat. Neurosci.* **6**, 803–810 (2003).
11. Lu, W. *et al. Neuron* **62**, 254–268 (2009).
12. Oh, M.C. & Derkach, V.A. *Nat. Neurosci.* **8**, 853–854 (2005).
13. Plant, K. *et al. Nat. Neurosci.* **9**, 602–604 (2006).
14. Milstein, A.D. & Nicoll, R.A. *Trends Pharmacol. Sci.* **29**, 333–339 (2008).
15. Sobolevsky, A.I., Rosconi, M.P. & Gouaux, E. *Nature* **462**, 745–756 (2009).

When what you see is not what you hear

Chandramouli Chandrasekaran & Asif A Ghazanfar

Different facial movements are better predictors of speech sounds than others. A new study investigates the neural processes that use this predictive signal and what happens to brain rhythms when it goes awry.

A recent game of counting with an infant named Aaron revealed something both adorable and unnerving. With mounting excitement, Aaron slowly counted "one, two," and then, after what seemed a promising start, he began articulating "ffff" instead of "three." The adults were silent, certain he was going to say four. Just when all hope was lost, he ended up saying "ffff-three", to the delight of all. This common speech error that infants make as they learn how to produce various consonant sounds reveals two levels of predictions that the adults made as they listened to Aaron. The first prediction was that three would follow two when counting and the second, more subtle prediction was that watching someone articulate the /f/ fricative should lead to hearing a word that begins with /f/ (four, rather than ffff-three). These predictions are built through our vast experience with audiovisual speech. One working model for why we are unnerved when we expect someone to say one thing and instead hear another ("Oh, I thought you were going to say…") is that our brains build representations of speech and its contexts (who is saying what and where) whose goal is to make accurate predictions by reducing errors. What does the brain do when faced with such an error?

In this issue, Arnal *et al.*[1] leveraged the natural structure of audiovisual speech in two ways to get at this question. First, when someone speaks, there is a consistent natural time lag between the facial movements and the sound produced by the vocal folds: vision precedes audition and is used as a predictive signal[2]. Second, certain facial movements are better predictors of subsequently voiced speech than others. For example, movements involving the lips are readily visible and consonant-vowel pairs beginning with /p/ or /m/ are therefore easier to predict than ones beginning with consonants that are produced at the back of the mouth and are therefore not as visible, such as /k/ or /g/. Arnal *et al.*[1] generated prediction errors by manipulating the validity (or congruence) between these visual and auditory components of speech. For example, in some conditions they presented subjects with a visual /pa/ followed by a mismatched /ga/ sound. Furthermore, because certain facial movements are more predictive than others, the experimenters also varied the magnitude of the prediction error. For example, the visual /pa/ will generate a bigger prediction error than the less visible /ka/ if mismatched with a /ga/ sound. The authors therefore mismatched auditory speech with highly predictive visual cues that would then generate large prediction error signals, as well as less predictive visual cues that would generate smaller error signals. Combining this manipulation of two natural features of audiovisual speech with magnetoencephalography (MEG) allowed them to study the role of prediction errors on the spatiotemporal dynamics of cortical networks.

Arnal *et al.*[1] found prediction error effects in MEG responses over presumptive auditory

The authors are in the Neuroscience Institute and Department of Psychology, Princeton University, Princeton, New Jersey, USA.
e-mail: chandram@princeton.edu or asifg@princeton.edu

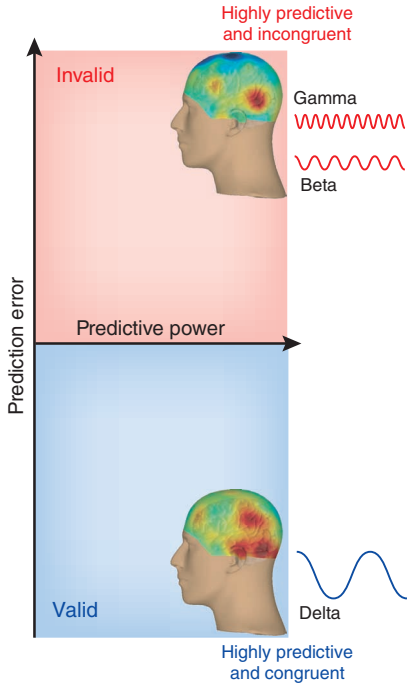

Figure 1 Low and high prediction errors in audiovisual speech generate different spatiotemporal dynamics. Prediction errors vary as a function of the predictability and congruence of audiovisual input. When visual input is highly predictive and is followed by its congruent auditory input, it leads to activation of a large higher order integrative network linked by low-frequency rhythms. In contrast, when visual prediction is strongly violated, it leads instead to a localized network around the superior temporal sulcus and dynamics mediated by higher frequency brain rhythms. Adapted from Arnal *et al.*[1].

cortical regions. Responses to speech sounds in auditory cortex are expected to be modulated by incoming, predictive visual signals. They found that the largest effect of a prediction error (that is, an invalid visual signal) occurred late in the response, approximately 350–500 ms after the presentation of the auditory component of the speech. This late timing is consistent with predictive coding theories that posit that the brain continuously uses available information to update an internal estimate of the world to generate predictions about upcoming sensory signals[3,4]. More specifically, the idea is that at every stage of the cortical hierarchy, a higher order area provides a prediction about an upcoming signal to a lower order area through feedback connections. This lower order area then relays the actual signal up to the higher order area and the mismatch between the prediction and the feedforward input is estimated by the circuit. This means that the temporal profile of responses in a given area should reflect the dynamics of this feedforward/feedback interaction, with later responses presumably reflecting the effect of feedback from higher order areas. In the case of the late responses in and around the auditory cortex, these presumably reflect the mismatch between a prediction based on the visual speech signal and the actual auditory speech signal. A likely higher order source for the visual speech prediction input to auditory cortex is the superior temporal sulcus[5–7].

By breaking down the MEG responses over auditory cortex into brain rhythms using time-frequency analyses and then correlating their phase-locking strength to the MEG sensors, Arnal *et al.*[1] revealed qualitative differences between the brain rhythms elicited by valid versus invalid visual speech, as well as differences in the cortical network engaged by the two different processes (**Fig. 1**). For the predictive and valid pairings (low prediction errors), a large network involving higher order language areas was linked by a low-frequency rhythm (a 3–4-Hz delta rhythm; **Fig. 1**). This suggests that when the predictions are valid, there is nothing more to account for at lower levels (such as the phoneme level) and the brain can therefore proceed with higher order computations (such as semantics and syntax) related to speech. In contrast, for the highly predictive and invalid pairing (when the prediction error was large), higher frequency brain rhythms (14–15-Hz beta and 70–90-Hz gamma rhythms) were present (**Fig. 1**). In this error context, these rhythms linked to a more circumscribed region centered on the superior temporal sulcus and primary sensory areas. These more local spatiotemporal dynamics presumably reflect an attempt by these circuits to resolve this prediction error. To recap, when visual predictions are validated by auditory input, it leads to the smooth activation of a higher order integrative network at a low temporal frequency. However, when visual predictions are violated, faster local processing is engaged to resolve the error.

As with any good study, these findings generate more questions. One is the functional role of these different brain rhythms in the predictive coding scheme. The authors suggest that an increase in beta rhythms is the signature of a prediction error[1], but others have suggested the opposite: that novel unexpected events (ones that generate very large prediction errors) lead to a decrease in beta rhythms[8]. Furthermore, a vexing concern regarding brain rhythms in general is that they seem fickle in their behavior and location. For example, in an attention task, beta band activity increases in early visual cortex, decreases in frontal cortex and is essentially unchanged in parietal cortex[9]. Similarly, for alpha band (~8–14 Hz) rhythms, both the source (cortical layer) and its relationship to behavior seem to differ across cortical areas[10]. Thus, it is unclear what exactly the cortical rhythms related to audiovisual speech prediction errors are doing. They are likely to be doing different things in different areas, although they may be phase-locked with each other or with some common node (for example, auditory cortex).

As the data of Arnal *et al.*[1] show, the general idea of predictive coding needs further elaboration from a neurophysiological perspective. Most pressing is, what is the timescale of prediction errors? The example above with Aaron shows that multiple predictions can exist simultaneously: one at the level of words (three should follow two) and another at the level of phonemes ("ffff" predicts four rather than three). This suggests that different prediction errors can evolve on the order of milliseconds, seconds and minutes, all of which need their own mechanisms. As some brain rhythms can be very slow relative to the typical electroencephalographic bands[11], is each timescale of prediction error related to the periods of particular brain rhythms? By linking different brain rhythms to different neural networks in an audiovisual speech context, Arnal *et al.*[1] provide the foundation for generating these and related questions as well as our first insights into how such processes might work.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Arnal, L., Wyart, V. & Giraud, A. *Nat. Neurosci.* **14**, 797–801 (2011).
2. Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A. & Ghazanfar, A.A. *PLOS Comput. Biol.* **5**, e1000436 (2009).
3. Rao, R.P. & Ballard, D.H. *Nat. Neurosci.* **2**, 79–87 (1999).
4. Friston, K. *Phil. Trans. R. Soc. Lond. B* **360**, 815–836 (2005).
5. Ghazanfar, A.A., Chandrasekaran, C. & Logothetis, N.K. *J. Neurosci.* **28**, 4457–4469 (2008).
6. Kayser, C. & Logothetis, N.K. *Front. Integr. Neurosci.* **3**, 7 (2009).
7. Noesselt, T. *et al. J. Neurosci.* **27**, 11431–11441 (2007).
8. Engel, A.K. & Fries, P. *Curr. Opin. Neurobiol.* **20**, 156–165 (2010).
9. Siegel, M., Donner, T.H., Oostenveld, R., Fries, P. & Engel, A.K. *Neuron* **60**, 709–719 (2008).
10. Bollimunta, A., Chen, Y., Schroeder, C.E. & Ding, M. *J. Neurosci.* **28**, 9976–9988 (2008).
11. Leopold, D.A., Murayama, Y. & Logothetis, N.K. *Cereb. Cortex* **13**, 422–433 (2003).