THE UNIVERSITY of EDINBURGH

# Edinburgh Research Explorer

# Prediction at all levels: forward model predictions can enhance comprehension

OPEN ACCESS

Prediction at all levels: forward model predictions can enhance comprehension

Simon Garrod

Institute for Neuroscience and Psychology

University of Glasgow

Chiara Gambi

Martin J. Pickering

Department of Psychology

University of Edinburgh


Address for correspondence:

Martin Pickering

Department of Psychology

University of Edinburgh

7 George Square Edinburgh EH8 9JZ

Tel.: 0131 650 3447 Fax: 0131 650 3461

Email:martin.pickering@ed.ac.uk

# Abstract

We discuss two limitations of Hickok's account. First, we propose that ideas from motor control and planning should be brought wholesale into psycholinguistics, so that processing at every level of the linguistic hierarchy (from concepts to sounds) should be recast in terms of forward model predictions and implementation. Second, we argue that motor involvement can sometimes enhance perception. We conclude that our account is consistent with a dual route model of comprehension in which different routes to prediction can interact.

Hickok (in press) is part of an interesting move in psycholinguistics to treat language production as a specific instance of motor control. In particular, his HSFC model offers a comprehensive neurocomputational account of articulatory (somatosensory) and auditory monitoring in speech production. Importantly, the state feedback control is hierarchically organized, so that he can begin to address psycholinguistic phenomena that involve linguistic representational levels such as phonemes and syllables. But why stop there? It should be possible to bring motor control and planning wholesale into psycholinguistics, so that the full range of representations in his Figure 2 (i.e., including the conceptual system and the lemma level) are included in such an account.

In fact, there is much evidence that different levels of representation in language production are integrated (e.g., Dell, 1986), just as for comprehension (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994). As a recent example, consider the evidence for phonological influences on word selection (e.g., Jaeger, Furth, & Hilliard, 2012), which suggests that processing at higher levels can make use of information from the phonological level. So if motor control and planning are involved at the level of the phoneme or syllable, we would expect them to also be involved in higher levels of processing, concerned, for instance, with syntax and semantics.

Thus, an alternative to Hickok's (in press) approach is to relate language processing as a whole to action, so that language production is treated as a form of action. For example, Pickering and Garrod (2013) argued that just as actors construct predictions about their forthcoming actions using forward models (see Wolpert, 1997), speakers predict what they are going to say using forward models. They then compare those predictions with the utterance as they produce it. Crucially, they do this at multiple linguistic levels, including semantics, syntax, and phonology. They compare their predicted semantics with the implemented semantics, then the predicted syntax with the implemented syntax, then the predicted phonology with the implemented phonology. This provides a form of production-based monitoring (cf. Laver, 1980) that contrasts with comprehension-based monitoring (Levelt, 1989) and more recent conflict-based accounts (Nozari, Dell, & Schwartz, 2011). By

contrast to Hickok (in press), in Pickering and Garrod's account forward predictions are computed from efference copies of production commands, and there are no inhibitory connections between production commands and sensory targets.

Pickering and Garrod (2013) also treat language comprehension as a form of action perception. In action perception, people can predict other people's actions using "prediction-by-simulation". When predicting another's hand movement, they determine what they would do if it were their hand. In other words, I covertly imitate your movements to determine the intention behind your movement and use that intention to predict (my percept of) your movement. The mapping from intention to prediction involves the same forward model as when I move my own hand, and this forward model constitutes part of the action system. Similarly, in comprehension, I covertly imitate your utterance and predict (my percept of) what you are about to say. This process applies to different aspects of the upcoming utterance; for example the listener might predict that the upcoming word will refer to something edible (Altmann & Kamide, 1999) or will begin with a vowel (DeLong, Urbach, & Kutas, 2005).

Prediction-by-simulation therefore suggests a close parallel between prediction of self and other. As speaking involves motor activity, it suggests that comprehension will also engage motor processes. In fact, there is much evidence for motor involvement during language comprehension and that such involvement is enhanced under conditions of difficulty (Adank, 2012; D'Ausilio, Bufalari, Salmas, & Fadiga, 2012; Scott, McGettigan, & Eisner, 2009). It is conceivable that such activation is response bias (Venezia et al., 2012) or a by-product of associative priming (see Hickok, 2013). There is good evidence that such activation is not necessary for comprehension (and in this respect we agree with Hickok). But we argue that such motor involvement is the result of prediction and that such prediction is more important under conditions of difficulty, specifically because the system has to rely on prediction rather that the unreliable bottom-up signal (e.g., in Kalman Filter terms; Pickering & Garrod, 2007). This explains behavioural findings such as the observation that overt imitation aids comprehension of accented speech (Adank et al., 2010).

Hickok (in press; see also Hickok, 2012a, 2012b) suggests that such involvement of

motor areas in speech perception is likely to be limited and unlikely to support recognition and comprehension. In his state feedback control model, predictions correspond to inhibitory signals. Activation of motor areas therefore leads to suppression of activity in auditory areas. This mechanism is useful during speech production because it enhances the detection of deviations from the intended auditory target (and also because it guarantees that auditory targets are deactivated as soon as appropriate motor commands have been selected; see p. 12 of the target article). However, according to Hickok, if a similar mechanism were implemented to support recognition of others' speech (i.e., in perception as opposed to production), it would be of little use, as it would reduce perceptual sensitivity (Hickok, 2012b, p. 399).

However, the proposal that production and comprehension both make use of forward models does not necessarily imply that prediction of self and other should bring about precisely the same effects. Rather, sensory suppression might occur during production because we can accurately predict the timing of speech sounds when they are self-generated. The timing of prediction during comprehension of other people is likely to be less accurate, and it is possible that the predictions only lead to suppression when they precisely match the timing (or indeed the form) of the utterance. Indeed, a recent MEG study by Tian and Poeppel (2013) suggests that forward model predictions might have an enhancing effect on sensory areas when stimulation of those areas follows the computation of predictions. They found that M200 responses in auditory temporal areas were enhanced when participants had just articulated or imagined themselves articulating the same auditory stimulus (a syllable), compared to a different stimulus. They argued that overt and covert articulation uses forward models to compute somatosensory predictions that then fine-tune auditory perception, hence leading to enhancement of the M200 in auditory areas. In contrast, when participants listened to or imagined themselves listening to the stimulus this led to Repetition Suppression in the M200, hence suggesting the involvement of a different neural pathway.

Hickok (in press) assumes a dual-route framework in which the ventral stream is concerned with the relationship between phonological words and conceptual representations,

and the dorsal stream is concerned with the relationship between speech gestures and the sounds they make (with his paper being concerned with the dorsal stream). Our account may be compatible with a dorsal-ventral split with respect to prediction. Pickering and Garrod (2013) proposed that comprehenders can use what they termed prediction-by-association instead of prediction-by-simulation. Prediction-by-association does not recruit the language production system, and instead relies on recurrent patterns of co-occurence (e.g., in the linguistic input) to generate predictions. We suggest that prediction-by-simulation is used in the dorsal route and is involved in, for example, turn taking (Scott et al., 2009). Prediction-by-association is used in the ventral route and facilitates understanding.

However, we propose that the ventral stream can assist the dorsal stream (e.g., using understanding to determine turn ending; de Ruiter et al., 2006). In addition, the dorsal stream can assist the ventral stream in order to facilitate understanding under conditions of difficulty (e.g., noise, degradation) or when time and resources allow (e.g., when speech is sufficiently slow) – and under such conditions, comprehenders can sometimes covertly or overtly complete what they hear. Importantly, this proposal means that motor involvement occurs during comprehension but is not necessary for comprehension. It also makes the clear prediction that people with compromised use of their motor system will be less able than other people to benefit from the dorsal stream under conditions of difficulty.

References

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood (ALE) meta-analyses. *Brain and Language, 122(1),* 42-54.

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological science, 21(12),* 1903-1909.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73(3),* 247-264.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex, 48(7),* 882-887.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review, 93(3),* 283.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8(8),* 1117-1121.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82(3),* 515-535.

Hickok, G. (2012a). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13,* 135-145.

Hickok, G. (2012b). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual stream model. *Journal of Communication Disorders, 45(6),* 393-402.

Hickok, G. (2013). Predictive coding? Yes, but from what source? *Behavioral and Brain Sciences, 36(4),* 358.

Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5),* 1439.

Laver, J.D.M. (1980). Monitoring systems in the neurolinguistic control of speech production.

In VA. Fromkin (Ed.), *Errors in linguistic performance*. New York, Academic Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MIT Press, Cambridge, MA.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101(4),* 676.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63(1),* 1-33.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in cognitive sciences, 11(3),* 105-110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36(4),* 329-392. doi: http://dx.doi.org/10.1017/S0140525X12001495

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action: Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10,* 295-302.

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience, 25(7),* 1020-1036.

Venezia, J. H., Saberi, K., Chubb, C., & Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Frontiers in psychology, 3,* doi: 10.3389/fpsyg.2012.00157

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences, 1(6),* 209-216.