



When brain regions talk to each other during speech processing, what are they talking about? Commentary on Gow and Olson (2015)

James M. McQueen, Frank Eisner & Dennis Norris

To cite this article: James M. McQueen, Frank Eisner & Dennis Norris (2016) When brain regions talk to each other during speech processing, what are they talking about? Commentary on Gow and Olson (2015), *Language, Cognition and Neuroscience*, 31:7, 860-863, DOI: [10.1080/23273798.2016.1154975](https://doi.org/10.1080/23273798.2016.1154975)

To link to this article: <http://dx.doi.org/10.1080/23273798.2016.1154975>



Published online: 25 Apr 2016.



Submit your article to this journal [↗](#)



Article views: 128



View related articles [↗](#)



View Crossmark data [↗](#)

COMMENTARY

When brain regions talk to each other during speech processing, what are they talking about? Commentary on Gow and Olson (2015)

James M. McQueen^{a,b}, Frank Eisner^a and Dennis Norris^c

^aDonders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; ^bMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; ^cMRC Cognition and Brain Sciences Unit, Cambridge, UK

ABSTRACT

This commentary on Gow and Olson questions in three ways their conclusion that speech perception is based on interactive processing. First, it is not clear that the data presented by Gow and Olson reflect normal speech recognition. Second, Gow and Olson's conclusion depends on still-debated assumptions about the functions performed by specific brain regions. Third, the results are compatible with feedforward models of speech perception and appear inconsistent with models in which there are online interactions about phonological content. We suggest that progress in the neuroscience of speech perception requires the generation of testable hypotheses about the function(s) performed by inter-regional connections.

ARTICLE HISTORY

Received 16 July 2015
Accepted 10 February 2016

KEYWORDS

Speech perception; sentential context effects; Granger causality; function of neural connections; feedback vs. feedforward models

Gow and Olson (2015) present evidence from a phonemic probe verification task in which listeners used sentential context to bias their interpretation of ambiguous phonemes. Processes in the left supramarginal gyrus (SMG) and the left posterior middle temporal gyrus (pMTG) appeared to “Granger cause” changes in neural activity in the posterior superior temporal gyrus (pSTG). These effects were observed specifically on trials in which the critical phonemes were identified in a manner that was consistent with the sentence context. Gow and Olson argue that this evidence supports an interactive account of speech perception. In this commentary, we present three arguments why this evidence does not allow inferences to be drawn about the presence or absence of interactive processes in speech recognition.

First, we question whether the magnetic resonance imaging-constrained magnetoencephalographic/electroencephalographic activation patterns used for the Granger causality analyses can plausibly be said to reflect normal speech processing. Gow and Olson (2015) had their participants listen to the sentences in the knowledge that they would see one of four probe letters (P, B, T, D) after each sentence, and that they would have to indicate whether they had heard the corresponding phoneme. This task is likely to engage additional processes beyond those normally involved in online comprehension. How might participants

perform this task? Perhaps the participants attempted to remember the sentences and then consulted that memory when the probes were presented. Perhaps the participants performed phoneme monitoring for all four possible probes as they heard the sentences. Perhaps the participants retroactively judged whether the two interpretations of the final word fitted with the sentence context. On any of these accounts, however, the task would engage additional mnemonic processes that are not part of normal speech comprehension. One might counter that all laboratory tasks used to study speech perception involve some additional processes. For example, in classical speech perception tasks such as phoneme monitoring, behaviour in those tasks may reflect task-specific (meta-linguistic) processes (as discussed, for example, by Norris, McQueen, & Cutler, 2000). However, at least in the case of phoneme monitoring (Norris et al., 2000) and lexical decision (Norris, 2006; Norris & McQueen, 2008), there are well-developed computational models of how those tasks are performed. The fundamental problem with Gow and Olson's task is that we have little idea how it might be performed. It requires the same meta-linguistic judgments as phoneme monitoring compounded with extra memory demands, and considerable uncertainty as to the exact strategy that participants might adopt. With this much complexity it is hard to have any confidence that the Granger causality effects reflect processes that are engaged in normal

listening rather than the way the listeners performed the probe verification task.

Second, Gow and Olson's (2015) conclusion hinges on a form of "reverse inference" (Poldrack, 2006). Inferences about the relationship between different cognitive processes are based on assumptions about the functions performed by particular brain regions. But it is not yet clear what the functions of pSTG, SMG and pMTG are. There is agreement that the pSTG, for example, is involved in speech processing as part of a postero-dorsal processing stream (see, e.g. Hickok & Poeppel, 2007; Scott & Johnsrude, 2003), but there is still disagreement about what exactly it does. This stream, which connects auditory cortex, inferior parietal lobe, and prefrontal cortex, has been associated not only with acoustic-phonetic processing (Gow, 2012) but also with a variety of other speech-related functions, including sensorimotor integration (Warren, Wise, & Warren, 2005), auditory imagery (Price, 2012), and verbal working memory (Buchsbaum & D'Esposito, 2009; McGettigan et al., 2011). Region-to-function inferences are even more difficult in the case of SMG. While Gow and Olson interpret the SMG activity as reflecting access to lexical representations, this region has also been linked to sublexical processes such as mapping between phonology and articulation (Rauschecker & Scott, 2009), phonological short-term memory (Jacquemot & Scott, 2006), and meta-linguistic judgments including phoneme categorisation and discrimination (Obleser & Eisner, 2009; Turkeltaub & Coslett, 2010). In addition, there is no consensus that speech-specific auditory analysis and the computations leading up to lexical access are in fact necessarily part of the postero-dorsal stream, as these functions are also commonly attributed to a processing stream in the temporal lobe which extends anteriorly (e.g. DeWitt & Rauschecker, 2012; Ueno, Saito, Rogers, & Lambon Ralph, 2011) or ventrally (Hickok & Poeppel, 2007) from auditory cortex.

Given our limited knowledge about the mapping between brain regions and (sub-)lexical processing, there are other reverse inferences one could plausibly make about the observed interactions between pSTG and SMG. For example, they might reflect the verbal working memory component that is engaged in the probe verification task. The argument about interactive processing rests on the assumptions that activity in the pSTG specifically reflects the pre-lexical acoustic-phonetic processing that leads to the identification of phonemes, and that SMG activity specifically reflects retrieval of lexical representations, but these assumptions are questionable.

Third, and perhaps most fundamentally, it is not clear what Gow and Olson (2015) mean by interactive processing. Even if we assume, for the sake of argument, that

the reported Granger causality effects do reflect normal speech processing and that the regions of interest do what Gow and Olson claim, it would still not be clear what type of "interaction" those effects would reflect. There are in fact a number of possibilities. Consider the influence of SMG on STG. Does this reflect online feedback about phonological content from word-form representations to phonemic processing, as in the TRACE model (McClelland & Elman, 1986), offline perceptual learning processes (Kilian-Hütten, Valente, Vroomen, & Formisano, 2011; Norris, McQueen, & Cutler, 2003), attentional processes (Obleser & Kotz, 2010), or binding of different representational components into a coherent whole (as suggested by Gow and Olson)?

The fact that this question has not yet been answered means that the current results are compatible with feed-forward models of speech perception such as the Merge model (Norris et al., 2000). Merge is consistent with the idea that perceptual processing can be influenced by higher levels to serve functions such as perceptual learning, attentional control, and binding because these functions do not require online feedback of phonological information (as e.g. in TRACE). Merge (and other feed-forward models) require further development to include these functions, but would be challenged by the current data only if it could be shown that the Granger causality effects reflect online feedback about phonological content.

There are three reasons to suspect that this is not the case. First, this kind of online feedback serves no useful purpose in speech recognition. Note that this point was at the core of the Norris et al. (2000) argument against this form of feedback. In contrast, learning, attentional control, and binding are useful processes. The second reason is that, while most behavioural and neuroscientific evidence is not diagnostic about whether or not there is online feedback (as discussed by Gow & Olson, 2015), some behavioural evidence appears to be inconsistent with online feedback. In particular, within-task dissociations between lexical involvement in phonemic decision-making and the lack of lexical involvement in perceptual compensation for coarticulation are inconsistent with online feedback (McQueen, Jesse, & Norris, 2009). Finally, the data presented by Gow and Olson indicate that, to a large degree, the causality effects are too late to plausibly reflect online interactions about phonetic content. The critical words were 300 ms long, and the critical ambiguous stop consonants were 20 ms long. As shown in their Figure 4, most effects (of SMG on STG, and especially of MTG on SMG) occur 300 ms after the onset of the critical word (i.e. after word offset, and 280 ms after the critical sound). It is possible

that the lateness of these effects may reflect task-related post-perceptual filtering based on the fit of the two interpretations of the critical word to the sentential context. Rather than being an indication of online feedback, the effects may instead be due to offline feedback about the outcome of this filtering process to the phonetic analysis stage that could support learning, or binding, or attentional control. Note that while one might expect binding and attentional control processes to operate from word onset, this does not exclude the possibility that aspects of such processes operate in the time window of the present effects. The timing pattern suggests that whatever processes these effects may reflect, they are not online, contrary to the view that they are markers of interactive feedback.

In summary, we suggest that the observed effects do not necessarily reflect pre-lexical and lexical processing both because the task employed may engage additional mnemonic processes, and because this conclusion rests on still-debated assumptions about the functions of the brain regions that are involved. Furthermore, to the extent that these effects do reflect pre-lexical and lexical processing, the data are fully consistent with models in which, during online processing, phonological content is fed forward from the pre-lexical processor to the lexicon but not back again. Indeed, the data appear to be inconsistent with models in which there are online interactions about phonological content.

Gow and Olson (2015) conclude that their results “support an interactive view of sentence context effects on speech perception”. We suggest that one reason why this conclusion can be questioned is that Gow and Olson paint too simplistic a view of the theoretical landscape. It is no longer the case that a binary choice has to be made between interactive and non-interactive models. Instead, there are a range of different kinds of “interaction”, and it is possible that higher level processes can influence lower levels of processing for functions such as learning, attention, and binding without those functions necessarily being combined with online processing interactions about the phonological content of the speech being heard. This means that, in studies of speech perception, care needs to be taken in specifying what kind of “top-down” effect is being investigated. In neuroscientific studies, testable hypotheses about the functions performed by inter-regional connections need to be generated. Gow and Olson have advanced the field not only by investigating what the relevant regions and their connections are, but also by establishing directionality and causality in the information flow between them. Their method thus makes it possible to analyse in detail how the different brain regions involved in speech comprehension talk to each

other. The next step is to specify what those regions are talking about.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

FE is supported by the NWO gravitation consortium Language in Interaction.

References

- Buchsbaum, B. R., & D'Esposito, M. (2009). Repetition suppression and reactivation in auditory-verbal short-term recognition memory. *Cerebral Cortex*, *19*(6), 1474–1485. doi:10.1093/cercor/bhn186
- DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, *109*, E505–E514. doi:10.1073/pnas.1113427109
- Gow, D. W., Jr. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language*, *121*(3), 273–288. doi:10.1016/j.bandl.2012.03.005
- Gow, D. W., & Olson, B. (2015). Sentential influences on acoustic-phonetic processing: A Granger causality analysis of multimodal imaging data. *Language, Cognition and Neuroscience*. Advance online publication. doi:10.1080/23273798.2015.1029498
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402. doi:10.1038/nrn2113
- Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, *10*(11), 480–486. doi:10.1016/j.tics.2006.09.002
- Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*, *31*(5), 1715–1720. doi:10.1523/JNEUROSCI.4572-10.2011
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86. doi:10.1016/0010-0285(86)90015-0
- McGettigan, C., Warren, J. E., Eisner, F., Marshall, C. R., Shanmugalingam, P., & Scott, S. K. (2011). Neural correlates of sublexical processing in phonological working memory. *Journal of Cognitive Neuroscience*, *23*(4), 961–977. doi:10.1162/jocn.2010.21491
- McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical-prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language*, *61*, 1–18. doi:10.1016/j.jml.2009.03.002
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357. doi:10.1037/0033-295X.113.2.327

- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. doi:10.1037/0033-295X.115.2.357
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–325. doi:10.1017/S0140525X00003241
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238. doi:10.1016/S0010-0285(03)00006-9
- Oblaser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, 13(1), 14–19. doi:10.1016/j.tics.2008.09.005
- Oblaser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20, 633–640. doi:10.1093/cercor/bhp128
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2), 59–63. doi:10.1016/j.tics.2005.12.004
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62, 816–847. doi:10.1016/j.neuroimage.2012.04.062
- Rauschecker, J., & Scott, S. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724. doi:10.1038/nn.2331
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100–107. doi:10.1016/S0166-2236(02)00037-1
- Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, 114(1), 1–15. doi:10.1016/j.bandl.2010.03.008
- Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72, 385–396. doi:10.1016/j.neuron.2011.09.013
- Warren, J. E., Wise, R. J. S., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences*, 28, 636–643. doi:10.1016/j.tins.2005.09.010