Published in final edited form as:

Lang Cogn Process. 2014 January 1; 29(1): 52–59. doi:10.1080/01690965.2013.852907.

Toward an Integrated Psycholinguistic, Neurolinguistic, Sensorimotor Framework for Speech Production

Gregory Hickok

University of California, Department of Cognitive Sciences, Irvine, CA 92697, USA,
gshickok@uci.edu

General comments

Cross-disciplinary integration is hard work. We are naturally comfortable with the theoretical paradigm of our training and have committed time and effort to its development, often with significant theoretical success. Integrating across fields requires us to tear down some of what we've built, re-think long-held assumptions, and submit to the fact that the early stages of an integrated research agenda will lack some of the theoretical crispness of the separate areas alone. Integration pushes us outside of our comfort zone. This is truth is manifest in the present proposal and is not overlooked by commenters, who point out several empirical and theoretical issues from within their traditions that need to be worked out. But while we may lose a little explanatory power within a given theoretical approach (hopefully temporarily), we gain explanatory power across domains. Integrative models need to be evaluated on the larger stage, not just from the perspective of one approach.

Some view my integration attempt as a bit lopsided. For example, Rapp et al. write, "rather than an integration of psycholinguistic and motor theories, the proposal appears largely to be a reduction of the former to the latter." Other authors may share this impression. Two general points are relevant. The first is that one will tend to notice deviations from the standard model in one's own tradition more so than deviations from the standard model in other traditions. So Rapp et al. may have noticed that the HSFC's use of motor control constructs like forward prediction is a deviation from traditional psycholinguistic models, but they don't comment on the model's lack of a true efference copy, or the involvement of forward prediction in the motor planning stage itself, or the existence of a lemma level of representation that provides the input to the "internal model", or the parallel input to sensory and motor systems, or the hierarchical arrangement of the circuits, aligned weakly to linguistic levels of representation – all deviations from standard motor speech control models. Speech motor control researchers may see a slightly reversed bias. Guenther notes that the abstract phoneme categories of linguists, and which I try to link to aspects of the model, aren't much good for the dynamics of motor control: "phonemes are not static vocal tract configurations but instead are inherently time-varying entities." To Guenther, the HSFC is not sufficiently sensitive to what the motor system needs for speech control. Both perspectives are valid, of course. The question is how are they related?

The second point is that the decision to incorporate a construct from one tradition over another was not driven by an *a priori* goal, such as to reduce linguistics to motor control or shoehorn linguistics into motor control. The goal is *integration*, to take the best features of each tradition and see if we end up with a model that is better as a result. This required a comparative evaluation of the various features. Consider, as an example, *feedback control/self-monitoring*, two independently developed constructs that appear to serve somewhat similar functions in motor control and psycholinguistic models respectively. They differ, however, in theoretical weight within the respective fields. Feedback control is a central feature of motor control models whereas in psycholinguistic models, it has a status

(traditionally anyway) that is closer to an appendage: talkers can, in fact, notice when they make an error, so there must be a mechanism to monitor speech. Given that feedback control has proven its worth in the motor control literature and has a computational grounding, I weighted the motor control version of this construct more heavily in constructing the integrated frame. In one sense, I've used research in motor control to *beef up* the theoretical centrality of the psycholinguistic construct of *self-monitoring*; there is no reductionist agenda here (see below for further comment).

More than one commenter expressed concerns that “motor” or “auditory” representations are insufficient to do the work of linguistic processes. For example, Indefrey writes, “There is, in fact, evidence for a *phonological* rather than a *motor* system level of internal feedback,” and Rapp et al. suggest that my proposal entails “a significant ‘flattening’ of the representational structures, eliminating the various levels of phonological representation....” These kinds of concerns reflect the persistence of the within-tradition barriers, which hold that auditory and motor systems are completely different from language systems and as such do not themselves involve rich representational structures that are capable of a high level of abstraction. A major point of my argument is that this is a *false distinction*. I do not distinguish, as Indefrey does, between “motor” and “phonological” systems, in principle, although I may distinguish between sensorimotor systems in the service of speech versus non-speech domains (an empirical question). And by integrating linguistic and sensorimotor systems I am not flattening or eliminating any of the representational structures. I am merely arguing that what we previously called “phonological” representations are in fact high-level, transformationally-linked sensory and motor representations, which are no less abstract than traditional linguistic categories or processes. The same mistake or bias is evident in some of the embodied cognition literature in which it is argued that abstract cognitive operations can be embodied or grounded in sensorimotor systems, thus supposedly eliminating traditional “cognitive” representations and computations (Barsalou, 2008; Pezzulo et al., 2011). At first blush, my proposal could be lumped in this tradition, but I don't see it that way. “Grounding” doesn't simplify the representation or computation, nor does it eliminate cognitive or linguistic operations, it just situates them in a different neurocomputational context. If anything, “grounding” complicates sensorimotor systems.

So if this is just a re-naming exercise – “phonological” is now “high-level sensorimotor” -- what's the point? The point is that once we recognize a parallel between “linguistic sensorimotor” and “non-linguistic sensorimotor” processes, we can draw on a wider empirical and theoretical base, a kind of comparative research program. We can use motor control research as a model for language science *and vice versa*. In the end, it doesn't matter whether we call the system linguistic or sensorimotor, cognitive or embodied. What matters is the nature of the representations, the architecture, and the computations. What excites me about the prospects of an integrated research program is that there appears to be overlap between traditionally linguistic domains and non-linguistic sensorimotor domains in terms of the architectures and computations. To a first approximation, what may primarily distinguish between domains then -- what distinguishes a linguistic system from a manual control system -- is the representational bits that are plugged into those computational architectures. If we hold to our within tradition biases and treat the domains as fundamentally different, then we will not be in a position to profit from cross-disciplinary exploration.

In what follows I address the more specific issues that arose in the various commentaries, organized somewhat thematically.

Phonemes and linguistic structure

Bakovi makes some important clarifications regarding generative phonology, particularly regarding the notion of phonemes and segments. I don't disagree with any of his comments, which will be helpful as the work progresses. I will comment that my terminological imprecision does not reflect a misunderstanding so much as it does an intentional vagueness. Outside of linguistics, "phoneme" is used generically or non-technically to refer to segment-sized units. We are barely at a stage where we can map (or at least consider hypotheses for mapping) *vague* linguistic concepts to neural circuits, let alone precise theoretical objects. Therefore I used phoneme in its non-technical sense. Admittedly, it would have been helpful to clarify this upfront. Also, my focus on the Chomsky and Halle model of articulatory feature bundles is a reflection of its dominance in the research circles outside of generative phonology. More broadly, perhaps the disagreement in phonological theory – in particular, whether the theory should be cast in terms of articulatory or auditory features (or both) can be informed by thinking about the problem in the context of a hierarchical sensorimotor control circuit. For example, I don't think it is an accident that linguists (Prince & Smolensky, 2004) and motor control scientists (Todorov & Jordan, 2002) independently hit on the notion of optimality -- not that optimality is necessarily the correct approach, but that the problem in the two domains has led to similar hypothesized solutions, which suggests some overlap in the process under investigation. There may be an opportunity for cross-fertilization.

Rapp et al. view my integrative approach as "reductionist" and highlight several pieces of evidence demonstrating that speech processing involves abstract linguistic structures that cannot be explained within a representationally "flattened" sensorimotor framework. I've pointed out above that this is a misunderstanding (my own fault, to be sure) of what I'm trying to do and what a sensorimotor framework is capable of doing from a representational standpoint. I will comment on some of their specific points to illustrate further. Rapp et al. note that "prior to their integration for production, syllable structures are stored and retrieved independently from the component segments of syllables." They then cite several sources of evidence for the separability of segments and syllable structure *in production tasks* (see Dell's comment) and claim that this is difficult for a hierarchical model such as mine to accommodate. Again, if I'm reading between the lines accurately, the authors are buying into a (false) distinction between levels of representation on one hand and "motor production" on the other. They seem to conceptualize the motor programs involved in production as low-level, "flattened", "continuous" and "gradient" codes. But this is not the case. The motor system (indeed the sensorimotor system) has long been recognized to be massively hierarchical, from simple reflex circuits to cortically mediated sensorimotor transformations (Andersen, 1997; Wolpert, 1997). The hierarchical organization continues within cortical circuits such that some regions and even single neurons have been argued to code abstract motor goals (e.g. "grasp") rather than specific movements (Grafton & Hamilton, 2007; Haruno, Wolpert, & Kawato, 2003; Rozzi, Ferrari, Bonini, Rizzolatti, & Fogassi, 2008). There is no reason, then, that the sensorimotor system can't code speech acts at high levels of abstraction, such as syllabic frames, and *during production* integrate these with lower-level segment-sized representations. The existence of hierarchical structure in speech fits nicely with what is known about the motor system. Rapp et al. cite other evidence to make similar points but again there is nothing incompatible in principle with a sensorimotor account and the existence of structure in linguistic processes. It is true, as they point out, that the present outline of the model does not include all the necessary processes, but this is precisely what an integrative research *program* should work on filling in.

Laganaro's comments are similar in spirit to Rapp et al.'s in that she argues for abstract segmental representations and views this observation as incompatible with a sensorimotor-

inspired architecture. Laganaro reminds us, appropriately so, that phoneme-sized units are slippery in speech production, resulting in classic speech error patterns, which, it is suggested, do not have an explanation within the HSFC framework. It is true that the proposed model does not explicitly deal with how these errors occur, and it should. However, there is nothing about the framework that precludes the existence of segments and the fact that they can slip during production. (I question featural-based segment representations in the auditory system as part of the ventral stream, not as part of the dorsal, sensorimotor stream.) For example, a higher-level motor plan in the auditory-Spt-BA44/6 circuit could frame a sequence of syllable-sized units, which are then further specified by slotting in a sequence of somatosensory targets/lower-level motor control plans associated with vocal tract cycle endpoint states (~segments); errors in this hierarchical mapping could lead to exchanges and other errors. Laganaro suggests that the HSFC architecture should correct such errors because these states will result in mismatches between motor plans and sensory targets. This is true, and I would argue this is what happens the vast majority of the time when the production mechanism succeeds in hitting the intended targets. But just like psycholinguistic computations are prone to the occasional slip, so too are sensorimotor implementations of these psycholinguistic computations.

Dell makes two important points. One is that slips of the tongue and slips of the ear exhibit different patterns, with the former clearly implicating the phoneme as a planning unit and the latter failing to make a strong case for the role of the phoneme in perception. *Dell* does a better job than I did in making this point convincingly. But then he questions whether the phoneme may be a prominent planning unit in all languages. Phonemes clearly play a major role in English and Dutch, but in other languages with simpler syllable structures this level of representation may play less of a role. This possibility is entirely consistent with a sensorimotor framework for speech production: the system forms categories and fine-tunes action to an extent necessary to hit sensory targets.

Psycholinguistic architecture

Indefrey argues against the proposal of parallel inputs from the lemma level to auditory- and motor-phonological networks, citing evidence from MEG and TMS, which shows that motor-related speech areas activate 200 msec after temporal lobe areas in production tasks. He suggests that this observation is more consistent with a serial architecture. While these data are valuable, they may not reveal the complete picture. Electrocorticographic (ECoG) recordings reveal a more variable picture regarding the timing of motor and auditory-related circuits, which may depend on the production task (Cervenka et al., 2013; Edwards et al., 2010; Korzeniewska, Franaszczuk, Crainiceanu, Kus, & Crone, 2011). Moreover, parallel input to motor and auditory systems, does not necessarily imply simultaneous or equal magnitude inputs. For example, I consider the lemma-to-auditory associations more robust than the lemma-to-motor associations, which could translate into an activation timing difference in population-based recordings. Note that even by Indefrey's estimates, both motor- and auditory-related speech systems are activated well before speech articulation, consistent with the present proposal.

Roelofs' comment pushes the integrative approach forward. He first points out that the HSFC model does not have an account of syllabification facts (true) and then he suggests a modification of model in which input to the auditory- and motor "syllable" level of representation (see above qualifications) is not direct from the lemma level, but is mediated by a phoneme (segment) layer of representation. A simulation is provided showing that the same excitation-inhibition patterns hold in his modified model as in the original HSFC model simulation. This is exactly the kind of approach I believe we need. Rather than simply stating that an integrative approach doesn't work for this or that psycholinguistic (or motor

control) fact, Roelofs made the effort to see what would happen if some additional psycholinguistic machinery were built in to the framework. Maybe it works, maybe it doesn't, but we're not going to know unless we try. That said, I'm not sure Roelofs' solution to dealing with syllabification is the only one. For example, it may be that the syllable targets coded in the auditory system are less rigid than Roelofs is assuming (e.g., *get* and *it*) or maybe they are but when combined at another level, the syllabification patterns can be modified (see Rapp et al.'s arguments for abstract syllable frame representations). It might be relevant that patients with cerebellar dysarthria, presumably a lower-level speech motor control deficit, have a characteristic reduction in syllable duration variation.

Concerns about conduction aphasia

I consider conduction aphasia to be a critical piece of evidence in support of the SFC architecture. As such it has attracted some commentary and deserves a fairly detailed response.

Rapp et al. assert that there is no independent evidence that conduction aphasia is a disorder of auditory-motor transformation. Let me clarify what we know in this respect:

1. Conduction aphasics have an elevated level of phonemic-type errors without deficits in perceiving those same speech sounds (they readily detect their overt errors) and without a general articulatory dysfluency (they are capable of very fluent strings of syllables; see (Goodglass, Quadfasel, & Timberlake, 1964) for an important comment on fluency in aphasia). It is therefore quite difficult to explain the deficit in either auditory- or motor-phonological terms at any level.
2. The manner in which conduction aphasics frequently "grope" for a correct word form, so called *conduite d'approche* (Goodglass, 1992; Kohn, 1984), is phenomenologically much like the kind of manual groping observed optic ataxia, a visuomotor integration disorder (Perenin & Vighetto, 1988). In both cases, deviation from the target does not seem to provide the relevant information for making a rapid, accurate correction, which is explainable in terms of a sensorimotor integration deficit.
3. The two dominant competing hypotheses for conduction aphasia are a deficit in auditory-motor transformation (the present claim, a variant on Wernicke's older claim) and a deficit involving the phonological short-term memory (STM) buffer (Baldo, Klostermann, & Dronkers, 2008). The STM buffer hypothesis provides no explanation for why the errors show up not only in verbatim repetition but also in single word naming and spontaneous speech, which minimizes STM demands; it provides no explanation for why patients can't immediately correct the errors they detect (see #2); and has nothing to say about why delayed auditory feedback is not as disruptive to conduction aphasics' fluency as it is to controls (Boller, Vrtunski, Kim, & Mack, 1978). An auditory-motor account has ready explanations for all these features.
4. Within the lesion distribution of conduction aphasia lies area Spt (Buchsbaum et al., 2011), which exhibits features found in dorsal stream visuo-motor integration areas such as AIP and LIP (Andersen, 1997; Colby & Goldberg, 1999). Namely, it shows sensory-motor response properties (Hickok, Buchsbaum, Humphries, & Muftuler, 2003), contains both auditory- and motor-weighted units (evident in voxel preferences) (Hickok, Okada, & Serences, 2009), and is motor-effector (vocal-tract) selective (Pa & Hickok, 2008). This same region is activated during altered auditory feedback, that is when a mismatch occurs between the expected

sensory consequence of a motor speech and the actual sensory consequence (Tourville, Reilly, & Guenther, 2008).

5. Spt shows a similar sensorimotor response pattern to speech and tonal melodies suggesting that the circuit is not phonologically specific (at the resolution of fMRI). Conduction aphasics have been found to also have deficits not just in speech but also in repeating back a sequence of tones (Gordon, 1983; Strub & Gardner, 1974).

Taken together I submit that there is substantial independent evidence that conduction aphasia is a disorder of auditory-motor transformation.

Rapp et al. also wonder about how the performance of individual case studies might be explained in within the HSFC framework. One example is a patient with dramatically worse performance of auditory non-word repetition than either low-frequency word repetition or non-word reading. Auditory non-words represent the stimulus type that taxes the auditory-motor network most substantially; there is simply no other way to go from acoustic stimulus to motor speech codes than through the auditory-motor transformation network. Low-frequency words have a semantic route available, thus explaining the difference, and if work in phonological short-term memory is correct, written information has another route also, via mappings between graphemes and motor articulatory codes (A. Baddeley, Thomson, & Buchanan, 1975; A. D. Baddeley, 1992). Another pattern claimed to be problematic is patients with good auditory comprehension and repetition but with relatively frequent phonemic paraphasias. This pattern is readily explained in the terms of the one aspect of these tasks that differ dramatically, whether the acoustic trace of the word is provided overtly and therefore presumably very robustly (auditory comprehension and repetition) or must be internally generated/accessed (volitional production). If there is a disruption of processes involving in accessing auditory-phonological forms internally or if activation of the forms themselves are compromised, the attested pattern could result because of the difference in activation strength. There is no serious challenge to the HSFC framework here.

Laganaro also questions whether conduction aphasia provides strong evidence for the HSFC model. If I understand correctly, the refutation is grounded in the idea that a sensorimotor transformation deficit can only account for errors in repetition (pseudoword repetition in particular) and can't explain exchange errors, often seen in conduction aphasia. This is incorrect. As in motor control generally, sensorimotor speech circuits provide a mechanism for detecting and correcting motor planning errors. I suggest further that this process operates internally *as part of the motor planning process* (very much in line with Levelt's notion of internal monitoring), and not just externally on overtly produced speech (Levelt's external monitoring). Given this, if conduction aphasics have damage to the region that computes the relation between sensory and motor representations of speech, then this internal error detection and correction mechanism will fail to catch errors in motor planning when they occur, leading to an increase in errors in all speech tasks, not just repetition. But the error rate will vary as a function of how much load a task places on the sensorimotor transformation system: pseudoword repetition is particularly problematic simply because it places the highest demands; production of high-frequency words or phrases will be easiest (although noise in the system can be expected to cause errors occasionally); and low-frequency words or sequences will fall somewhere in between with respect to their error rate. On the assumption that slips of the tongue in healthy individuals as well as conduction aphasics stem from failures in this sensorimotor circuit (one induced by random noise, the other by lesion), we should expect similar types of errors, reflecting the computational organization of motor planning, which *must* reflect linguistic organization.

The view from sensorimotor processing

Ziegler is understandably disappointed with my emphasis on the Chomsky & Halle definition of the phoneme and points out other important motor control-related features, such as sub-segmental temporal dynamics and supra-segmental rhythm. He suggests that my theory “stops halfway in its endeavor to establish an integrated model...” I take issue with this assertion and would argue instead that my theory stops *well short* of halfway! In an effort to integrate, I have dismantled the theoretical framework of *some* of the many approaches to speech production and reconstructed an impoverished frame of *some* of the pieces lying around. I have no doubts that the empirical and theoretical omissions in the HSFC model far exceed the bits it employs or attempts to deal with. Building in these missing details will be an important future task. So, I don’t disagree with Ziegler; we need to worry about sub-segmental temporal dynamics and suprasegmental features like rhythm. His suggestion that phonological primitives are more temporally dynamic is interesting (see also Guenther’s comments) and may mesh nicely with the proposed HSFC framework. Ziegler’s summary description of the function of the hierarchical circuits, for example, is very much in the spirit of what I have in mind: “the auditory channel specifies the ultimate goal of how a word must sound in order to be understood, whereas the somatosensory channel codes for the mechanical and aerodynamic events mediating this goal.”

Sato et al. and *Guenther* are primarily concerned with my proposed alignment between linguistic levels of representation and sensorimotor circuits (auditory=syllables, somatosensory=phonemes) and argue that there is good evidence for *auditory* phonemic representations. Guenther also specifically challenges the idea of a hierarchical arrangement of auditory and somatosensory circuits. I share the concern of these authors regarding the alignment of linguistic units and sensorimotor circuits. This is why I called it an oversimplification and a convenient shorthand. It is also why I clarified my claim: “With respect to auditory targets, I believe that the representational granularity can vary from individual speech sounds (defined in acoustic space), ranging from vowels and consonants that make noise to syllables to sequences of syllables.” So I don’t think I disagree at all with what Sato et al. and Guenther point out in this respect. I do want to be careful, though, about disentangling two points that I was trying to make in the target article. One is that some proposed linguistic representations, such as articulatory feature-defined segments, may line up more neatly with a somatosensory circuit. My oft-critiqued remark about voiceless stops is a good example. In one theory, the phoneme /t/ is defined by a cluster of articulatory features. *Assuming that theory*, it makes little sense to talk about an auditory representation of /t/, but makes some sense to think about a possible somatosensory representation of it. Several commentary authors (and one reviewer) have responded by saying something like “it’s a silly point because that’s the wrong phonological theory; clearly the representations are more dynamic, have acoustic features, and so on.” An alternative to this kind of response is this: maybe we are all talking about different, functionally comingled representations of basically the same bit of speech. One representation is static, another more dynamic and temporally elongate, one auditory, another somatosensory. This view suggests a different research approach. Rather than trying to uncover the “true”, unique representation of speech units and then try to localize its neural correlate, we should perhaps consider the many ways that speech might be represented in motor control or perceptual systems and for what purpose, and then we can map the neural correlates of these. My discussion of /t/ was intended to be more along the lines of this alternative approach: given that it seems possible to represent /t/ as a collection of articulatory features, how might that representation be coded. It doesn’t mean that is the only possible representation.

Regarding the more general point, emphasized by Guenther, about auditory and somatosensory circuits being hierarchically organized, I’ll say that I sympathize with

Guenther's perspective. My own comments about differences in the granularity of coding between auditory (coarser) and somatosensory (finer) circuits, seem to be at odds with my qualification, quoted above, that the auditory system can code anything from a noisy consonant to a sequence of syllables. The self-contradiction reflects my uncertainty, and highlights an important area of future research. I nonetheless remain convinced that a hierarchical organization remains a viable hypothesis. The reason behind my conviction is not the representational points that Guenther highlights, but rather the anatomical hierarchical organization of, and effects of damage to the auditory-Spt-premotor circuit compared to the somatosensory-(cerebellar)-motor circuit. From an anatomical hierarchy standpoint, somatosensory cortex is densely connected to primary motor cortex (perhaps with cerebellar modulation), whereas auditory cortex is more densely connected to premotor areas (BA6/44/45); this strongly suggests a functional hierarchy. Damage to the auditory-Spt-premotor circuit is associated with higher level phonological deficits such as those observed in conduction aphasia and apraxia of speech, whereas damage to sensorimotor cortex or the cerebellum result in lower-level dysarthric speech; again strongly suggestive of a functional hierarchy. If there is no hierarchical distinction between auditory and somatosensory circuits, we will need a different explanation for these facts.

Sensorimotor processes in language generally

Garrod and Pickering take up two issues. One concerns whether motor-based architectures might be useful for thinking about higher-levels of language processing and the other deals with the possibility that motor-based prediction might be involved in perception/comprehension. I think it is important to think deeply about these ideas and test them thoroughly. Although some authors have aligned syntactic processes with aspects of motor function (e.g., Michael Ullman's declarative/procedural model bears at least a family resemblance to this view (Ullman et al., 1997)), much work is needed before we can evaluate the theoretical and empirical benefits of framing, say, syntactic processes in sensorimotor terms. It is nonetheless an excellent topic for future research, and I would be lying if I said I hadn't thought about it in some depth.

Regarding the role of motor prediction in perception/comprehension: having proposed motor prediction as a possible mechanism for (mildly) enhancing perception (Hickok, Houde, & Rong, 2011), I am sympathetic to the idea. But after thinking about it a bit more deeply I am now rather skeptical for two reasons. One is that motor prediction is a computational mechanism designed to detect deviation from prediction *and to use that error signal to correct motor plans for action*. What makes these predictions useful is that they can quickly confirm or disconfirm the accuracy of motor plan, the thing that generated the prediction, and provide information on the magnitude of the correction needed. It is quite unclear, computationally, whether a prediction generated from a motor plan benefits access to a semantic representation, since motor prediction isn't designed to do this, since the semantic representation did not generate the prediction, and since the magnitude of deviation from prediction in auditory-motor space (*bat* vs. *pat* are similar acoustically and motorically) does not map neatly onto the magnitude of deviation in semantic space (*bat* vs. *pat* are semantically miles apart). One might argue that the motor system can be co-opted to assist in perception/comprehension. I agree this is a possibility. One can also use a screwdriver as a hammer, but a hammer will work better so why bother with the screwdriver except under extreme conditions (e.g., no hammer available)? This may be what the motor system contributes to perception: a last ditch effort to make sense of an ambiguous perceptual event, consistent with Garrod and Pickering's observation that it seems to get involved under very demanding conditions. The second reason I'm skeptical, is related to this last point. While there are plenty of claimed examples of motor prediction assisting perception, the effect sizes are extremely small, roughly 10% modulation under near-threshold conditions (e.g.,

(Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007)). And these examples may reflect decision bias more than perceptual discriminability (Venezia, Saberi, Chubb, & Hickok, 2012). Garrod and Pickering hit on the right answer, however, that prediction comes from two sources, the dorsal stream and the ventral stream, as I've argued previously (Hickok, 2012b). My working hypothesis is that the vast majority of the predictive influence on perception/comprehension comes from the ventral stream; that is, prediction based on semantic context, priming, and the like. It's a bit of a mundane hypothesis—the computational stream that is involved in understanding is the source of prediction for understanding, while the dorsal stream, involved in action, is the source of prediction for motor control—but I think it is correct.

Concerns about neural computation

Strijkers raises interesting questions regarding the interaction of Hebbian learning mechanisms, inhibitory motor-to-sensory feedback, and the dual stream architecture of the speech system. He is concerned that inhibitory feedback will result in a dis-association of auditory and motor phonological units: when inhibited, the auditory units no longer fire together with the motor units and therefore will “un-wire” from one another. Although this may be a serious concern in the toy models I've discussed (Hickok, 2012a), in a more realistic, fleshed out neurocomputational model is not a serious worry. Hebbian learning is likely mediated neurally by spike timing-dependent plasticity (STDP), which refers to the observation that synaptic strength modification is dependent on the *order* of firing between the pre- and postsynaptic cells within a *restricted* time window (Caporale & Dan, 2008). Further, synapses between two excitatory units and between inhibitory and excitatory units can have reverse polarities in terms of whether the synaptic link is potentiated or depressed. Thus, synaptic modification in the sort of sensorimotor circuit I've proposed (which is not unique to speech) can vary according to a range of spike timing, spike ordering, and cell type parameters; it's not as simple as Strijkers' quite reasonable logic implies. In addition, it is also possible that the motor-to-sensory inhibitory feedback is not direct, but mediated by an inhibitory interneuron such that motor units excite inhibitory interneurons (associated via standard fire-together, wire together STDP mechanisms), which then inhibit the excitatory auditory units to which they are paired. In short, there are biologically plausible ways around the sorts of problems Strijkers raises. More generally, it is well established that excitatory-inhibitory balance is critical to brain function and it is also well established that the brain nonetheless is capable of forming associations, likely via a Hebbian (STDP) mechanism. It is therefore a virtual given that there are solutions to Strijkers concerns. We are currently working on detailed spiking neural network models that aim to find these solutions.

Meta comments

Proposing an integrated model is much like proposing a multi-method, multi-disciplinary grant: everyone thinks it is a great idea and is desperately needed but they don't like *this* particular integration. This is a predictable outcome for the reasons outlined in my general comments. All commenters have accurately pointed out shortcomings of my attempts at integration in the domain of speech production, largely concerning details coming out of their own traditional approach. It's instructive, I think, to note the differences between the critics. Rapp et al., Laganaro, and Bakovi are concerned about a loss of linguistic richness, whereas Ziegler and Guenther see the model as lacking in terms of dynamic features necessary for proper motor control. Indefrey and Roelofs are concerned the network's architecture and processing steps, and Sato et al. and Guenther lament the alignment between linguistic constructs and neural circuits. It's a microcosm of the problem we face in working toward a true integration. And it's the beginning of a solution. In coming at the

proposal and the scientific problem from their own perspective each author notices a different set of issues to the exclusion of others. Why isn't Ziegler or Guenther or Strijkers worried about the fact that syllable structures can be primed and segments slip? Why aren't Rapp et al. puzzling over the role of rhythmic features or dynamic motor control structures in speech production? They should be! They are all relevant observations concerning the same overall process and maybe by pulling together the whole set of observations along with the lot of representational features and computational approaches we will be able to make even more progress than we have separately.

References

- Andersen R. Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London B Biological Sciences*. 1997; 352:1421–1428.
- Baddeley A, Thomson N, Buchanan M. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:575–589.
- Baddeley AD. Working memory. *Science*. 1992; 255:556–559. [PubMed: 1736359]
- Baldo JV, Klostermann EC, Dronkers NF. It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. *Brain and Language*. 2008; 105(2):134–140. [PubMed: 18243294]
- Barsalou LW. Grounded cognition. *Annual Review of Psychology*. 2008; 59:617–645.
- Boller F, Vrtunski PB, Kim Y, Mack JL. Delayed auditory feedback and aphasia. *Cortex*. 1978; 14(2): 212–226. [PubMed: 679702]
- Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, et al. Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain and Language*. 2011; 119(3):119–128. [PubMed: 21256582]
- Caporale N, Dan Y. Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*. 2008; 31:25–46.
- Cervenka MC, Corines J, Boatman-Reich DF, Eloyan A, Sheng X, Franaszczuk PJ, et al. Electrocorticographic functional mapping identifies human cortex critical for auditory and visual naming. *Neuroimage*. 2013; 69:267–276. [PubMed: 23274183]
- Colby CL, Goldberg ME. Space and attention in parietal cortex. *Annual Review of Neuroscience*. 1999; 22:319–349.
- Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, et al. Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage*. 2010; 50(1):291–301. [PubMed: 20026224]
- Goodglass, H. Diagnosis of conduction aphasia. In: Kohn, SE., editor. *Conduction aphasia*. Hillsdale, N.J.: Lawrence Erlbaum Associates; 1992. p. 39–49.
- Goodglass H, Quadfasel F, Timberlake W. Phrase length and the type and severity of aphasia. *Cortex*. 1964; 1:133–153.
- Gordon WP. Memory disorders in aphasia -- I. Auditory immediate recall. *Neuropsychologia*. 1983; 21:325–339. [PubMed: 6621861]
- Grafton ST, Hamilton AF. Evidence for a distributed hierarchy of action representation in the brain. *Hum Mov Sci*. 2007; 26(4):590–616. [PubMed: 17706312]
- Haruno M, Wolpert DM, Kawato M. Hierarchical MOSAIC for movement generation. *International Congress Series*. 2003; 1250:575–590.
- Hickok G. Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*. 2012a; 13(2):135–145.
- Hickok G. The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. *J Commun Disord*. 2012b; 45(6):393–402. [PubMed: 22766458]
- Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*. 2003; 15:673–682. [PubMed: 12965041]

- Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*. 2011; 69(3):407–422. [PubMed: 21315253]
- Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*. 2009; 101(5):2725–2732. [PubMed: 19225172]
- Kohn SE. The nature of the phonological disorder in conduction aphasia. *Brain and Language*. 1984; 23:97–115. [PubMed: 6478196]
- Korzeniewska A, Franaszczuk PJ, Crainiceanu CM, Kus R, Crone NE. Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (ERC) analysis of human electrocorticography (ECoG). *Neuroimage*. 2011; 56(4):2218–2237. [PubMed: 21419227]
- Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. *Current Biology*. 2007; 17(19):1692–1696. [PubMed: 17900904]
- Pa J, Hickok G. A parietal-temporal sensory-motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia*. 2008; 46:362–368. [PubMed: 17709121]
- Perenin MT, Vighetto A. Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain*. 1988; 111(Pt 3):643–674. [PubMed: 3382915]
- Pezzulo G, Barsalou LW, Cangelosi A, Fischer MH, McRae K, Spivey MJ. The mechanics of embodiment: a dialog on embodiment and computational modeling. *Front Psychol*. 2011; 2:5. [PubMed: 21713184]
- Prince, A.; Smolensky, P. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell Publishing; 2004.
- Rozzi S, Ferrari PF, Bonini L, Rizzolatti G, Fogassi L. Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur J Neurosci*. 2008; 28(8):1569–1588. [PubMed: 18691325]
- Strub RL, Gardner H. The repetition defect in conduction aphasia: Mnestic or linguistic? *Brain and Language*. 1974; 1:241–255.
- Todorov E, Jordan MI. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*. 2002; 5(11):1226–1235.
- Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*. 2008; 39(3):1429–1443. [PubMed: 18035557]
- Ullman MT, Corkin S, Coppola M, Hickok G, Growdon JH, Koroshetz WJ, et al. A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedure system. *Journal of Cognitive Neuroscience*. 1997; 9:266–276. [PubMed: 23962016]
- Venezia JH, Saberi K, Chubb C, Hickok G. Response bias modulates the speech motor system during syllable discrimination. *Frontiers in Psychology*. 2012; 3
- Wolpert DM. Computational approaches to motor control. *Trends Cogn Sci*. 1997; 1(6):209–216. [PubMed: 21223909]