Integrating accounts of speech production: the devil is in the representational details

Brenda Rapp\textsuperscript{a}, Adam Buchwald\textsuperscript{b} & Matthew Goldrick\textsuperscript{c}

\textsuperscript{a} Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA
\textsuperscript{b} Department of Communicative Sciences and Disorders, New York University, New York, NY, USA
\textsuperscript{c} Department of Linguistics, Northwestern University, Evanston, IL, USA

Published online: 11 Oct 2013.
Integrating accounts of speech production: the devil is in the representational details

Brenda Rapp*, Adam Buchwaldb and Matthew Goldrickc

aDepartment of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA; bDepartment of Communicative Sciences and Disorders, New York University, New York, NY, USA; cDepartment of Linguistics, Northwestern University, Evanston, IL, USA

(Received 11 September 2013; accepted 18 September 2013)

This commentary focuses on two key aspects of Hickok’s proposal that distinguish it from other theories of speech production. Unlike many other accounts, auditory targets play a central and early role in speech production. This proposal also adopts a generally reductionist approach to the production of speech, relying almost exclusively on sensory and motor processes to represent sound structure. This eliminates various levels of phonological representation that play key roles in theories motivated by psycholinguistic and cognitive neuropsychological research. The general and specific issues raised by Hickok’s approach are examined within two specific areas: the structure of sound representations and patterns of performance in “conduction aphasia.”

Keywords: phonological representations; conduction aphasia; cognitive neuropsychology; psycholinguistics

The key innovation of the Hickok proposal is the central and prominent role of auditory targets in speech production. Based on the notion that an auditory target is the “ultimate goal of a speech act” (p. 14), and auditory targets are “elevated” from their traditionally more peripheral role within many motor, psycholinguistic and cognitive neuropsychological proposals. In these accounts, auditory representations are typically involved solely in the monitoring of speech output and, therefore, enter the speech production process only subsequent to the generation of some form of production target. Instead, Hickok proposes that auditory targets play a central and early role in speech production. Specifically, as soon as one moves beyond the semantic/lemma level of representation the word production process is driven by both auditory (and motor) information, regardless of whether the task itself involves auditory input. In other words, auditory targets play a similar role in spontaneous speech, picture naming and oral reading as they do in a task such as repetition where there is an explicit external auditory target.

As we discuss below, there are a number of representational consequences of Hickok’s proposal, which is a generally reductionist approach to speech production that relies almost exclusively on sensory (auditory and somatosensory) and motor processes to represent sound structure. Specifically, there is a significant “flattening” of the representational structures, eliminating the various levels of phonological representation that psycholinguistic and cognitive neuropsychological research has posited play key roles in moving from semantic representations to motor processes. Below we discuss some general and specific issues raised by such an approach, focusing on two areas: the structure of sound representations and patterns of performance in “conduction aphasia”.

Distinct levels of sound representations

Identifying the levels and content of neurocognitive representations used for production is one of the fundamental goals of psycholinguistic theories. With respect to sound structure, Hickok focuses on phonemes and syllables, which he characterises as differing in terms of their granularity. In the framework he presents, when words are selected at the semantic/lemma level, the phonological representations of the words are then retrieved as hierarchically organised auditory/sensory, motor and somatosensory representations of syllable- and phoneme-information that “may be distributed, in a partially overlapping fashion, across the two hierarchical levels” (Hickok, 2012, p. 139). However, a number of critical aspects of this proposal have not been specified, including the precise representational structure of hierarchical yet overlapping representations, their activation dynamics and the means for translating between the auditory–motor and auditory–somatosensory representations used for controlling the production of syllables and phonemes. Below, we discuss two sets of findings whose explanation within this framework will require further elaboration: first, there is clear evidence that segments and syllable structure are stored and retrieved independently; second, there is also strong evidence that there are levels
of sound structure representation that are abstract and do not encode sensory or motor information.

First, there is evidence that prior to their integration for production, syllable structures are stored and retrieved independently from the component segments of syllables. For example, syllable structure can be primed in the absence of segmental overlap (Sevald, Dell, & Cole, 1995). Segments can also exhibit effects that are independent of the syllable positions in which they occur. More than 20% of contextual phonological errors contain segments produced in the wrong syllable position (e.g., film → flim; Vousden, Brown, & Harley, 2000). In addition, individual segments can be primed even when produced in different syllable positions (Damian & Dumay, 2009). Given the proposed strong hierarchical organisation of syllabic- and phoneme-level information, it is unclear how Hickok’s framework could account for such results without further specification.

Second, there is substantial evidence that there are multiple levels of segment-sized representations that specify different types of information. One distinction is between position-specific vs. context-independent aspects of speech production. An example of evidence supporting this type of distinction comes from Buchwald and Miozzo (2011) who reported two impaired speakers who both made /s/-deletion errors in /s/-stop clusters (e.g., stable). One participant often produced a syllable-initial stop consonant (e.g., aspirated [th], as in table), reflecting an error that occurred before context-specific representations were specified (or retrieved). In contrast, the other participant often produced an unaspirated stop (as in dable), consistent with the motor plan that would have been produced in the specific context of the cluster, indicating an error that occurred after context-specific representations were specified. Hickok’s framework currently does not account for this distinction. The difference in the error patterns cannot be captured by a distinction between more or less “coarsely” or “broadly” specified representations; rather it requires a distinction between levels of representation at which context-specific adjustments have or have not been applied. Other findings suggest a distinction between phonological representations/processes that are highly sensitive to the phonetic (articulatory/acoustic) complexity of a structure (e.g., greater difficulty processing clusters vs. singleton consonants; coda vs. onset consonants) as opposed to processes that are strongly sensitive to lexical (word-level) properties of phonological structures (e.g., greater difficulties processing segments within low- vs. high-frequency words). This is supported by the observation of a number of individuals with acquired production deficits who exhibit contrasting performance patterns: those sensitive to lexical factors but not phonetic complexity vs. individuals sensitive to phonetic but not lexical factors (e.g., Goldrick & Rapp, 2007; Romani & Galluzzi, 2005; Romani, Galluzzi, Bureca, & Olson, 2011; Romani, Olson, Semenza, & Granà, 2002).

Finally, while speech motor control theories emphasise continuous, gradient representations, certain results from aphasia suggest that errors can arise at processing levels utilising discrete segmental representations. Buchwald, Rapp, and Stone (2007) presented articulatory and acoustic data showing that an aphasic individual’s vowel insertion errors (clone → “cologne”) did not reflect gradient articulatory mistiming but rather the insertion of a discrete vowel segment. The contrasting pattern in which insertion errors reflect gradient articulatory mistiming is commonly described as a frequent feature of apraxia of speech (McNeil et al., 2009) and can also be observed in certain contexts in second-language speakers (Davidson, 2005). Taken together, these findings reflect the existence of multiple levels of sound structure that differ not in “grain size” but in the nature of the sound structure information specified at each level.

“Conduction aphasia” as empirical evidence

In support of the prominent role of the auditory target in speech production, Hickok argues that “conduction aphasia is one of the strongest sources of empirical validation for the proposed model” (p. 10). This term is a clinical label that refers to a cluster of symptoms with some variability in the specific symptomatology. Conduction aphasia is largely considered a repetition disorder with the chief symptoms being impaired repetition of auditory targets in the face of good speech perception and comprehension. In addition, in speech tasks that do not involve auditory targets (e.g., picture naming, spontaneous speech and oral reading), phonemic errors are produced in the absence of the dysfluencies that are typically associated with motor difficulties. Hickok focuses on the co-occurrence of phonemic errors in both repetition and naming in the context of intact motor execution and auditory comprehension. Along lines pursued by Wernicke (1874/1969), Hickok attributes this pattern to a disruption of the process he refers to as auditory–motor coordinate transformation. Given that, in Hickok’s proposal, the auditory target is key to speech production and auditory–motor coordinate transformation is required for the spoken production of all but the very highest frequency words in all speech production tasks, a disruption to this process predicts sound errors across all speech tasks. This would, therefore, provide an account of the co-occurrence of phonological errors in repetition and naming despite intact perception, comprehension and motor functions. There are several issues associated with proposing conduction aphasia as one of the strongest empirical supports for Hickok’s proposal. First, there is no independent evidence that the errors and difficulties observed in these cases result specifically from a failure
in computing auditory–motor transformation. That is, no
evidence has been presented that the characteristics of
errors produced or the variables affecting performance in
these cases clearly and specifically originate in a failure in
translating between auditory and motor syllable-sized
targets.

Second, it is unclear how the proposal can account for
more detailed aspects of patterns of performance in
several relevant cases of acquired language impairment.
For example, Jacquemot, Dupoux, and Bachoud-Lévi
(2007) reported the case of an individual with good
speech comprehension and perception but low accuracy
in non-word repetition – the task that serves to index the
integrity of the auditory–motor coordinate transformation
process. In the Hickok proposal, the auditory target plays
a key role in producing infrequent forms when production
originates in semantic processes (in picture naming).
Presumably, the auditory target also plays a similar role
when production is based on visual-orthographic pro-
cesses (in oral reading). On this basis, we would expect
both non-word reading and picture naming of low
frequency words to be especially reliant on the auditory
target. Therefore, for individuals with damage to the
auditory–motor coordinate transformation process,
marked impairments in producing infrequent forms in
both picture naming and reading are predicted. However,
in this case, low frequency word naming and non-word
reading (77% and 83%, respectively) were both found to
be significantly superior to non-word repetition (35%; see
also Bachoud Lévi & Dupoux, 2003). Patterns such as
these indicate that the auditory–motor coordinate trans-
formation may not play the critical role that is proposed in
the Hickok account. The key features of this pattern have
been explained in alternative frameworks in which the
phonemic errors are more typically attributed to a deficit
in speech planning related to reduction in some type of
phonological buffering or working memory capacity
(Goodglass, 1992; see also Baldo, Klosterman, & Dron-
kers, 2008) or instability of phonemic representations.

Also seemingly problematic are the reports of indi-
viduals with good speech perception and comprehension
and good repetition but who nonetheless produce primar-
ily phonemic errors in naming and spontaneous speech
(e.g., Goldrick & Rapp, 2007; Kay & Ellis, 1987; Law,
2004; McCarthy & Warrington, 1984). Their good com-
prehension indicates that the link between word meaning
and auditory form is intact. Their strong repetition
performance indicates intact auditory targets, auditory–
motor coordinate transformation and motor execution.
The fact that they produce primarily phonemic errors and
relatively few semantic paraphasias in naming suggests that
their difficulties do not arise within access to lemmas. But,
under the Hickok proposal, subsequent to lemma selec-
tion, such individuals should be able to rely on their intact
auditory word form – motor link. Given the centrality of
the auditory word form for production, it is a challenge for
Hickok’s framework to explain why such individuals
make substantial numbers of phonemic errors in word
production.

Conclusions: the challenges of true integration
The integration of motor and psycholinguistic theories is
much needed; Hickok is most definitely to be commended
for taking on this challenge and attempting to place it
within our as yet limited understanding of the neural
substrates of language production processes. One of the
challenges in developing interdisciplinary accounts is
using vocabulary in a way that members of the relevant
disciplines can understand. There are many terms used in
Hickok’s proposal (e.g., phonological, phonemic and
amodal) that have distinct meanings in psycholinguistic
and linguistic traditions and should either be used in a
manner that is shared by these traditions or clearly
redefined – allowing readers to avoid confusion in
attempting to infer the intended meanings. A prime
example is the use of the term “phonological”. Despite
the pervasive use of the term in Hickok’s article, there is
little that is recognisable as phonological (as opposed to
acoustic, somatosensory or motoric) in this proposal.

More broadly, rather than an integration of psycholin-
guistic and motor theories, the proposal appears largely
to be a reduction of the former to the latter. Numerous lines of
psycholinguistic and cognitive neuropsychological research
have provided a rich evidential base for the multi-level
system of phonological representations and processes that
have been proposed to underlie the translation between
word meaning and speech production. Accounting for these
findings with the “representationally flatter” sensory-motor
approach that Hickok proposes is a key challenge that will
need to be met if the proposal is to prosper. One cannot
help but wondering if, in the course of doing so, the
auditory target may descend from the central position in
which Hickok has placed it to a lesser modular role,
somewhat comparable to what Hickok, Costanzo, Capasso,
and Miceli (2011; see also Rogalsky, Love, Driscoll,
Anderson, & Hickok, 2011) have so coherently argued is
played by motor representations in speech perception.

Funding
Preparation of this commentary was supported in part by
the NSF grant [BCS0846147] to MG.

References
syntactic and semantic variables on word form retrieval.
Cognitive Neuropsychology, 20, 163–188. doi:10.1080/
02643290242000907
either a cook or a baker: Patients with conduction aphasia get
the gist but lose the trace. *Brain and Language*, 105, 134–140. doi:10.1016/j.bandl.2007.12.007


