

# Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM)

Grant M. Walker<sup>1</sup> · Gregory Hickok<sup>1</sup>

Published online: 30 July 2015  
© Psychonomic Society, Inc. 2015

**Abstract** Speech production is studied from both psycholinguistic and motor-control perspectives, with little interaction between the approaches. We assessed the explanatory value of integrating psycholinguistic and motor-control concepts for theories of speech production. By augmenting a popular psycholinguistic model of lexical retrieval with a motor-control-inspired architecture, we created a new computational model to explain speech errors in the context of aphasia. Comparing the model fits to picture-naming data from 255 aphasic patients, we found that our new model improves fits for a theoretically predictable subtype of aphasia: conduction. We discovered that the improved fits for this group were a result of strong auditory–lexical feedback activation, combined with weaker auditory–motor feedforward activation, leading to increased competition from phonologically related neighbors during lexical selection. We discuss the implications of our findings with respect to other extant models of lexical retrieval.

**Keywords** Speech production · Computational models · Neuropsychology

Speech production has been studied from several theoretical perspectives, including psycholinguistic, motor control, and neuroscience, often with little interaction between the approaches. Recent work, however, has suggested that integration

may be productive, particularly with respect to applying computational principles from motor control, such as the combined use of forward and inverse models, to higher-level linguistic processes (Hickok, 2012, 2014a, 2014b). Here we explore this possibility in more detail by modifying Foygel and Dell's (2000) highly successful psycholinguistic, computational model of speech production, using a motor-control-inspired architecture, and assess whether the new model provides a better fit to data and in a theoretically interpretable way.

We first present the theoretical foundations for this work by (1) describing the motivations behind Foygel and Dell's (2000) semantic–phonological model (SP), (2) briefly summarizing the motor-control approach, (3) highlighting some principles from our recent conceptual attempt to integrate the approaches, and (4) describing our modification of SP using a fundamental principle from motor-control theory to create our new semantic–lexical–auditory–motor model (SLAM). We then present the computational details of both the SP and SLAM models, along with simulations comparing SP with SLAM. To preview the outcome of these simulations, we found that SLAM outperforms SP, particularly with respect to a theoretically predictable subcategory of aphasic patients. We conclude with a discussion of how the new model relates to some other extant models of word production.

## The SP model

SP has its roots in Dell's (1986) theory of retrieval in sentence production, which was developed to account for the speech errors, or slips of the tongue, found in large collections of natural speech. To this end, the theory integrated psychological and linguistic concepts: From psychology it adopted the notion of computational simultaneity, in which multiple internal representations compete for selection prior to production,

---

**Electronic supplementary material** The online version of this article (doi:10.3758/s13423-015-0903-7) contains supplementary material, which is available to authorized users.

---

✉ Grant M. Walker  
grantw@uci.edu

<sup>1</sup> Department of Cognitive Sciences, University of California, Irvine, CA, USA

and from linguistics it incorporated hierarchical levels of representation, as well as the separation at each level between stored lexical knowledge and the applied generative rules.

Dell, Schwartz, Martin, Saffran, and Gagnon (1997) proposed a computational model that limited the focus to single-word production, but extended the theoretical scope to include explanations of speech errors in the context of aphasia. The basic idea was that the pattern of aphasic speech errors reflects the output of a damaged speech production system, which could be modeled by adjusting parameters in the normal model to fit aphasia data. The model's architecture consisted of a three-layer network with semantic, lexical, and phonological units, and the connections among the units were selected by the experimenters to approximate the structure of a typical lexical neighborhood (Fig. 1). Word production was modeled as a spreading-activation process, with noise and decay of activation over time. Damage was implemented by altering the parameters that control the flow of activation between representational levels. Simulations were then used to identify parameter values that generated frequencies of error types that were similar to those made by aphasic patients.

Due to the computationally intensive nature of the simulation method, however, comprehensive explorations were effectively limited to only two parameters at a time. Nevertheless, in a series of articles beginning with Foygel and Dell (2000), two free parameters in the model were identified that account for an impressive variety of the data derived from a picture-naming task, including clinical diagnostic information (Abel, Huber, & Dell, 2009), lexical frequency effects (Kittredge, Dell, Verkuilen, & Schwartz, 2008), characteristic error patterns associated with different types of aphasia (Schwartz, Dell, Martin, Gahl, & Sobel, 2006), characteristic patterns of recovery (Schwartz & Brecher, 2000), and interactive error effects (Foygel & Dell, 2000). These two free parameters were the connection strengths between semantic and lexical representations (the *s-weight*) and between lexical and phonological representations (the *p-weight*), an architecture known as *SP*. *SP* has been used to explain performance on

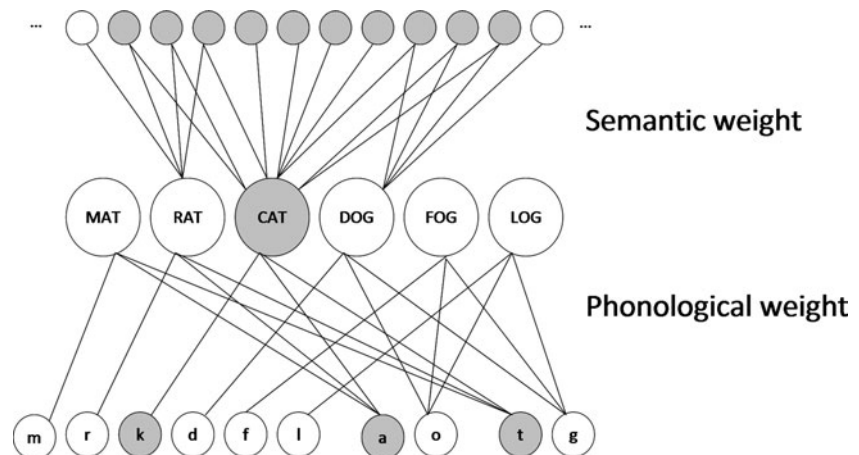
other tasks, as well, such as word repetition (Dell, Martin, & Schwartz, 2007), and to predict the location of neurological damage seen in clinical imaging (Dell, Schwartz, Nozari, Faseyitan, & Branch Coslett, 2013), although here we will focus primarily on its relevance to picture-naming errors.

*SP* pertains specifically to computations that occur between the semantic and phonological levels. It is assumed that the output of the model is a sequence of abstract phonemes that must then be converted into motor plans for controlling the vocal tract. We next turn to some fundamental constructs that have come out of research on how motor effectors are, in fact, controlled.

## Motor-control theory

At the broadest level, motor control requires sensory input to motor systems for initial planning and feedback control. It requires input for planning to define the targets of motor acts (e.g., a cup of a particular size and orientation and in a particular location relative to the body) and to provide information regarding the current state of the effectors (e.g., the position and velocity of the hand relative to the cup). Without sensory information, action is impossible, as natural (Cole & Sedgwick, 1992; Sanes, Mauritz, Evarts, Dalakas, & Chu, 1984) and experimental (Bossom, 1974) examples of sensory deafferentation have demonstrated. Sensory information has also been shown to provide critical *feedback* information during movement (Wolpert, 1997; Wolpert, Ghahramani, & Jordan, 1995), which provides a mechanism for error detection and correction (Kawato, 1999; Shadmehr, Smith, & Krakauer, 2010). When precise movements are performed rapidly, however, as in speech production, feedback mechanisms may be unreliable, due to feedback delay or a noisy environment. In this case, a state feedback control system can be supplemented with forward and inverse models (Jacobs, 1993), enabling the use of previously learned associations between motor commands and sensory consequences to guide the effectors toward sensory goals. This arrangement implies that the motor

**Fig. 1** The semantic–phonological (SP) model architecture



and sensory systems are tightly connected, even prior to on-line production or perception.

In the case of speech, the most critical sensory targets are auditory (Guenther, Hampson, & Johnson, 1998; Perkell, 2012), although somatosensory information also plays an important role (Tremblay, Shiller, & Ostry, 2003). Altered auditory feedback has been shown to dramatically affect speech production (Houde & Jordan, 1998; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Yates, 1963), and changes in a talker's speech environment can lead to “gestural drift”—that is, changes in his or her articulatory patterns (i.e., accent; Sancier & Fowler, 1997). Additionally, neuroimaging experiments investigating covert speech production have consistently reported increased activation in auditory-related cortices in the temporal lobe (Callan et al., 2006; Hickok & Buchsbaum, 2003; Okada & Hickok, 2006).

Some particularly relevant evidence for the role of the auditory system in speech production has come from neuropsychological investigations of language. Striking patterns of impaired and intact language-processing abilities resulting from neurological injury have led theorists to propose separate auditory and motor speech representations in the brain (Caramazza, 1991; Jacquemot, Dupoux, & Bachoud-Lévi, 2007; Pulvermüller, 1996; Wernicke, 1874/1969). Patients with conduction aphasia (Goodglass, 1992), for example, have fluent speech production, suggesting preserved motor representations. These patients also have good auditory comprehension and can recognize their own errors, suggesting spared auditory representations. Despite these abilities, they make many phonemic errors in production and have trouble with nonword repetition. This pattern is typically explained as resulting from damage to the interface between the separate auditory and motor systems (Anderson et al., 1999; Geschwind, 1965; Hickok, 2012; Hickok et al., 2000). This point regarding conduction aphasia has important theoretical implications, as we discuss below.

## Conceptual integration

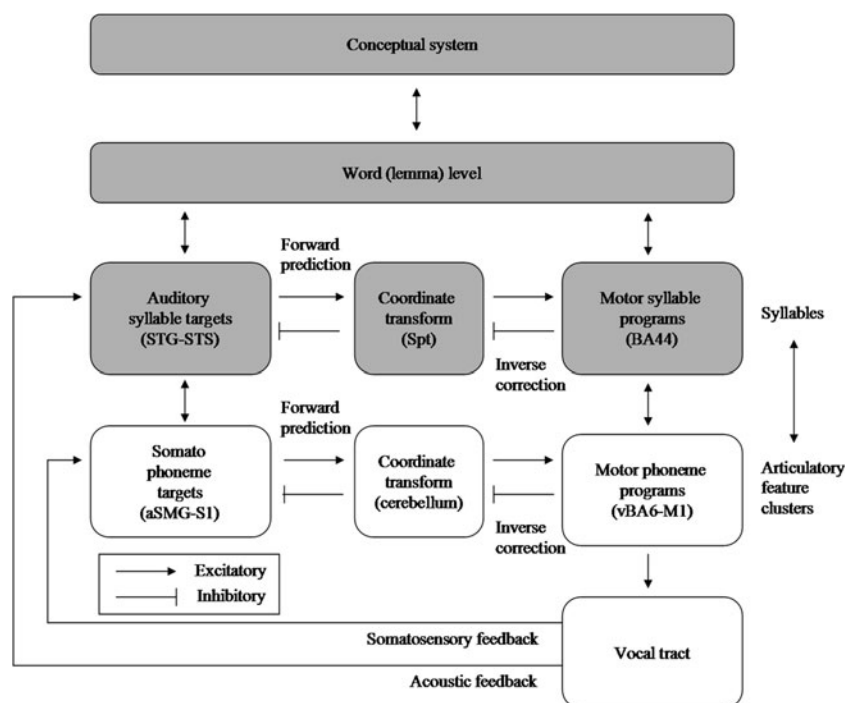
The hierarchical state feedback control (HSFC; Hickok, 2012) model provides a theoretical framework for the integration of psycholinguistic notions with concepts from biological motor-control theory. This conceptual framework is organized around three central principles. The first is that speech representations have complementary encodings in sensory and motor cortices that are activated in parallel during speech production, all the way up to the level of (at least) syllables. The second principle is that a particular pattern of excitatory and inhibitory connections between the sensory and motor cortices, mediated by a sensorimotor translation area, implements a type of forward/inverse model that can robustly guide motor representations toward sensory targets, despite the potential for errors in motor

program selection during early stages of motor planning/activation. The third principle is that the sensorimotor networks supporting speech production are hierarchically organized, with somatosensory cortex processing smaller units on the order of phonemes (or more accurately, phonetic-level targets such as bilabial closure, which can be coded as somatosensory states), and auditory cortex processing larger units on the order of syllables (i.e., acoustic targets). A schematic of the HSFC framework is presented in Fig. 2; it is clear that the top portion (darker colors) embodies the two steps of SP but breaks down the phonological component into two subcomponents, an auditory–phonological network and a motor–phonological network. This conceptual overlap has inspired our creation of a new computational model that is directly related to the first principle and is partially related to the other two principles. We reasoned that the architectural assumptions of the HSFC model can be evaluated, in part, by integrating them with an established and successful computational model of naming, SP; if the architectural changes led to improved modeling performance, this would provide support for the new framework.

## The SLAM model

SLAM is a computational model of lexical retrieval that divides phonological representations into auditory and motor components (Fig. 3). The dual representation of phonemes directly follows from the first HSFC principle. The choice to label the sensory units as auditory representations is motivated by the third principle—specifically, that this level of coding is larger than the phonetic feature. Neither SP nor SLAM includes inhibitory connections, and thus the second HSFC principle is not directly implemented; however, the pattern of connections in the SLAM model does implement a type of forward/inverse model that can reinforce potentially noisy motor commands. Our goal here was to modify the computational assumptions of SP as little as possible in order to assess the effects of the architectural assumption of separate motor and sensory phonological representations.

During picture-naming simulations, activation primarily flows from semantic to lexical to auditory to motor units—hence the model's acronym, SLAM. There is also a weaker, direct connection between lexical and motor units. The existence of this lexical–motor connection acknowledges that speech production may occur via direct information flow from lexical to motor units, an assumption dating back to Wernicke (1874/1969), which is needed to explain preserved fluency and spurts of error-free speech in conduction aphasia. However, the connection is always weaker than the lexical–auditory route (again, Wernicke's original idea), motivated by several points. First, the auditory–lexical route is presumed to develop earlier and to be used more frequently than the lexical–motor route. Longitudinal studies have shown that

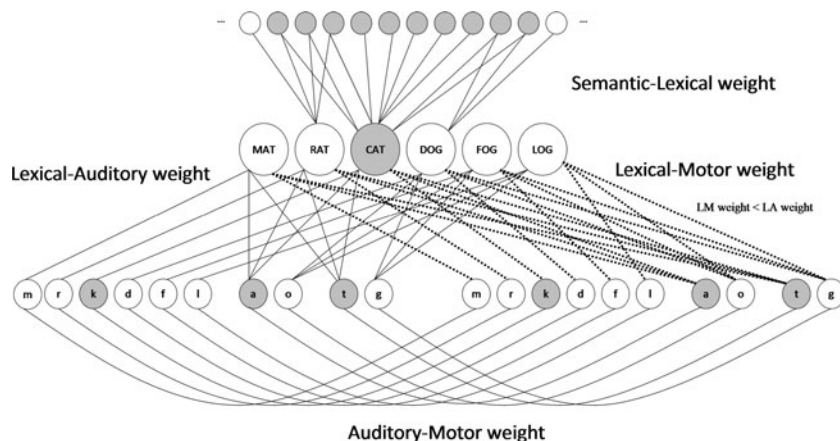


**Fig. 2** A schematic diagram of the hierarchical state feedback control (HSFC) framework (Hickok, 2012)

children begin to comprehend single words several months before they produce them, and they acquire newly comprehended words at nearly twice the rate of newly produced words (Benedict, 1979). Second, motor-control theory dictates that motor plans are driven by their sensory targets. During development, the learner must make reference to auditory targets, in order to learn the mapping between speech sounds and the motor gestures that reproduce those sounds (Hickok, 2012; Hickok, Houde, & Rong, 2011). Third, in the context of aphasia, comprehension deficits tend to recover more than production deficits (Lomas & Kertesz, 1978), suggesting a stronger association between lexical and auditory–phonological representations.

The assumption that the lexical–auditory mapping is always stronger than the lexical–motor mapping has an

important consequence: It means that the SLAM model is not merely the SP model with an extra part; in fact, there is effectively zero overlap in the parameter spaces covered by SP and SLAM. The reason for this is as follows. Given the SLAM architecture shown in Fig. 3, it is clear that one *could* implement SP simply by setting the connection weights in the lexical–auditory and auditory–motor mappings to zero and letting the lexical–motor weights vary freely. This would make SP a proper subset of SLAM, allowing SLAM to cover a parameter space (and therefore fits to data) identical to that of SP. However, this architectural possibility was explicitly excluded by implementing our assumption that lexical–auditory weights are always stronger than lexical–motor weights: If the lexical–auditory weights are zero, then the lexical–motor weights must also be zero and cannot vary freely—thus



**Fig. 3** The semantic–lexical–auditory–motor (SLAM) model architecture



effectively excluding the parameter subspace used by SP. This further allows us to test SLAM's assumption that the lexical–auditory route is the primary one used in naming. We can also examine model performance with the opposite constraint—namely, when the lexical–auditory weights are always less than the lexical–motor weights—a variant we might call “SLMA” to reflect the lexical–motor dominance and that would include SP parameter space as a subset. SLAM and SLMA have the same numbers of free parameters, both of which are more than that of SP, but with different assumptions regarding the connection strength patterns. If SLAM were to do better than SLMA, even though SLMA implements SP as a proper subset of its parameter space, it would demonstrate that the primacy of the lexical–auditory route is not only theoretically motivated, but also necessary for the observed improvements.

To summarize, we hypothesized that SLAM would characterize deficits in the general aphasia population at least as well as SP, and would primarily benefit the modeling of conduction aphasia. Recall that conduction aphasia is best explained as a dysfunction at the interface between auditory and motor speech representations that affects the phonological level, in particular (Hickok, 2012; Hickok et al., 2011). Thus, a naming model that incorporates a mapping between auditory–phonological and motor–phonological representations should provide a better fit for speech errors resulting from dysfunction in this mapping. To test this hypothesis, we compared the SP and SLAM model fits to a large set of aphasic picture-naming data.

## Computational implementation

### Patient data

All data were collected from the Moss Aphasia Psycholinguistic Project Database (Mirman et al., 2010; [www.mappd.org](http://www.mappd.org)). The database contains deidentified data from a large, representative group of aphasic patients, including responses on the Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996), a set of 175 line drawings of common nouns. All patients in the database had postacute aphasia subsequent to a left-hemisphere stroke, without any other diagnosed neurological comorbidities, and they were able to name at least one PNT item correctly. We analyzed the first PNT administration for all patients in the database with the demographic information available, including aphasia type and months postonset ( $N=255$ ). The cohort consisted of 103 anomic, 60 Broca's, 46 conduction, 35 Wernicke's, and 11 other aphasics with transcortical sensory, transcortical motor, postcerebral artery, or global

etiologies. The median months poststroke was 28 [1, 381], and the median PNT percent correct was 76.4 [1, 99].

### Computational models

As we mentioned above, SP was first presented by Foygel and Dell (2000). The model's approach to simulating picture naming instantiates an interactive, two-step, spreading-activation theory of lexical retrieval and consists of a three-layer network, with individual units representing semantic, lexical, and phonological symbols (Fig. 1). The number of units and the pattern of connections are intended to approximate the statistical probabilities of speech error types in English, by implementing the structure of a very small lexical neighborhood consisting of only six words, one of which is the target. The model includes six lexical units, with each connected to ten semantic units representing semantic features. Semantically related words share three semantic units, meaning that on a typical trial, with only one word that is semantically related to the target, the network has a total of 57 semantic units. Each lexical unit is also connected to three phonological units, corresponding to an onset, vowel, and coda. There are ten phonological units total: six onsets, two vowels, and two codas. Words that are phonologically related to the target differ only by their onset unit, and the network always consists of two such words. Finally, the remaining two words in the network are unrelated to the target, with no shared semantic or phonological units. On 20 % of the trials, one phonologically related word is also semantically related, creating a neighbor that has a “mixed” relation to the target.

Simulations of picture naming begin with a boost of activation delivered to the semantic units. Two parameters, S and P, specify the bidirectional weights of lexical–semantic and lexical–phonological connections, respectively. Activation spreads simultaneously between all layers, in both directions, for eight time steps according to a linear activation rule with noise and decay. Then, a second boost of activation is delivered to the most active lexical unit, and activation continues to spread for a further eight time steps. Finally, the most active phonological onset, vowel, and coda units are selected as output to be compared with the target. Production errors occur due to the influence of noise as activation levels decay, which can be mitigated by strong connections. Responses are classified as correct, semantic, formal, mixed, unrelated, or neologism. For a given parameter setting, a multinomial distribution over these six response types is estimated by generating many naming attempts with the model. These distributions may then be compared with those that

result from the naming responses produced by aphasic patients.

SLAM retains many of the details of SP, consistent with our aim to primarily assess the effects of the architectural modification. The semantic and lexical units remain unchanged, but there is an additional copy of the phonological units, with one group designated as auditory and the other as motor (Fig. 3). Four parameters specify the bidirectional weights of semantic–lexical (SL), lexical–auditory (LA), lexical–motor (LM), and auditory–motor (AM) connections. The LA and LM connections are identical to the P connections in the SP model, with each lexical unit connecting to three auditory and three motor units, whereas the AM connections are simply one-to-one. Simulations of picture naming are carried out in the same two-step fashion as with SP, with boosts delivered to the semantic and then the lexical units, and phonological selection occurring within motor units.

### Fitting data

In order to fit data, the model is evaluated with different sets of parameters that yield sufficiently different output distributions, creating a finite-element map from parameter space to data space, and vice versa. This process involves, first, selecting a set of parameter values (e.g., S and P weights), then generating many naming attempts with the model using that parameter set, in order to estimate the frequency of each of the six types of responses that occur with that particular model setup. Once those frequencies have been determined, that weight configuration becomes associated with the output distribution in a paired list called a *map*. Each point in the map represents a prediction about the type of error patterns that are possible when observing aphasic picture naming. One way to evaluate a model, then, is to measure how close its predictions come to observed aphasic error patterns. The distance between an observed distribution and the model's nearest simulated distribution is referred to as the model's *fit* for that data point. The root mean squared deviation (RMSD) is an arbitrary but commonly used measure of fit, which can be interpreted as the average deviation for each response type. For example, an RMSD of .02 indicates that the observed proportions deviate from the predicted proportions by .02, on average (e.g., predicted=[.50, .50]; observed=[.48, .52]). Thus, a lower RMSD value indicates a better model fit. Immediately, the question arises of how many points one should generate, and how to select the parameters to avoid generating redundant predictions.

In their Appendix, Foygel and Dell (2000) provided guiding principles for generating a variable-resolution map of parameter space, along with an example algorithm. They noted that the particular choice of mapping

algorithm likely would have little impact on the fit results, as long as it yielded a comprehensive search; however, given the inherently high computational cost of mapping, a particular algorithm may affect the map's maximum resolution in practice. A second algorithm for parameter space mapping was given by Dell, Lawler, Harris, and Gordon (2004), and these maps are considered to be the standard for SP, since they are available online and have been used in subsequent publications. This SP map has 3,782 points with 10,000 samples at each point and required several days of serial computation to generate. Clearly, the computational cost associated with the mapping procedure represents a considerable bottleneck for developing and testing models. Adding new points to the map improves the chances of a prediction lying closer to an observation, with diminishing returns as the model's set of novel predictions winnows. As Dell has suggested, because the goal is to find the best fit, adding more points to improve model performance is probably a worthy pursuit (G. Dell, personal communication, July 12, 2013). Moreover, because SLAM has two additional parameters, we needed to modify the mapping procedure to generate maps more efficiently.

We greatly improved efficiency by redesigning the mapping algorithm to take advantage of its inherent parallelism. There are two main iterative steps in the mapping algorithm: point selection and point evaluation. The coordinates of a point in parameter space are defined by a possible parameter setting for the model (point selection), and a corresponding point in data space is defined by the proportions of response types generated with that parameter setting (point evaluation). The point evaluation step is extremely amenable to parallelization, because the simulations involve computations across independent units, independent samples, and independent parameter sets. Point selection, however, required a new approach to foster parallelism: Delaunay mesh refinement.

The Delaunay triangulation is a graph connecting a set of points, such that the circumcircle of any simplex does not include any other points in the set. This graph has many favorable geometric properties, including the fact that edges provide adjacency relationships among the points. The new point selection algorithm takes advantage of these adjacency relationships. Beginning with the points lying at the parameter search range boundaries and their centroid, if the separation between any two adjacent points in parameter space exceeds a threshold distance (RMSD) in data space, their parameter space midpoint is selected for evaluation and is added to the map. These new points are then added to the Delaunay mesh, and the process reiterates until all edges are under threshold. Thus, on each iteration, the point selection algorithm yields multiple points to be evaluated in parallel across the entire parameter search range. Parallel processing was executed on

a graphics processing unit (GPU) to further improve efficiency.<sup>1</sup>

Before statistically comparing SP's and SLAM's performances, we studied the effects of map resolution on the model fits. First, we generated a very high-resolution map for each model using a low RMSD threshold of .01 to encourage continued exploration of the parameter space. Each map included 10,000 samples at each point, and the parameters varied independently in the range [.0001, .04]. The maximum parameter values were selected to be near the lowest values that yielded the highest frequency of correct responses, so that reduced values would lead to more errors. Due to the use of a low mapping threshold, the algorithm was halted before completion, after generating an arbitrarily large number of points. Early termination is not a great concern, because the algorithm efficiently selects points over the full search range. This fact also makes it a trivial matter to reduce the map resolution while still covering the full space.

The mapping procedure generated an initial 31,593 points for the SLAM model, with parameters freely varying; then, in accordance with the SLAM architecture, all points with  $LM \geq LA$  were removed, yielding a SLAM map with 17,786 points. The full SP map had 57,011 points. Next, we created 50 lower-resolution maps for each model by selecting subsets from the larger maps, with logarithmically spaced numbers of points from 5 to 17,000. For each map, we calculated the mean fit for the aphasic patients as a whole and for each of the diagnosis groups, excluding the heterogeneous diagnosis group. Figure 4 plots the fit curves. As we expected for both models, adding points improved the fits with diminishing returns. The relative fit patterns appeared to stabilize around 2,321 points, marked by vertical lines in the figure. We therefore chose to compare SP and SLAM at this map resolution; our findings should apply to any higher-resolution map comparisons, with trends favoring SLAM as resolution increases.

To compare the new parallel-generated maps with the standard serially generated maps, we also identified a parallel SP map resolution that yielded similar performance in terms of mean and maximum fit to the values reported in Schwartz et al. (2006). For this set of 94 patients, a parallel SP map with 189 points resulted in a mean and a maximum RMSD of .0238 and .0785, as compared with the reported values of .024 and .084, respectively. As expected, the parallel algorithm selected points much more efficiently than the serial algorithm,

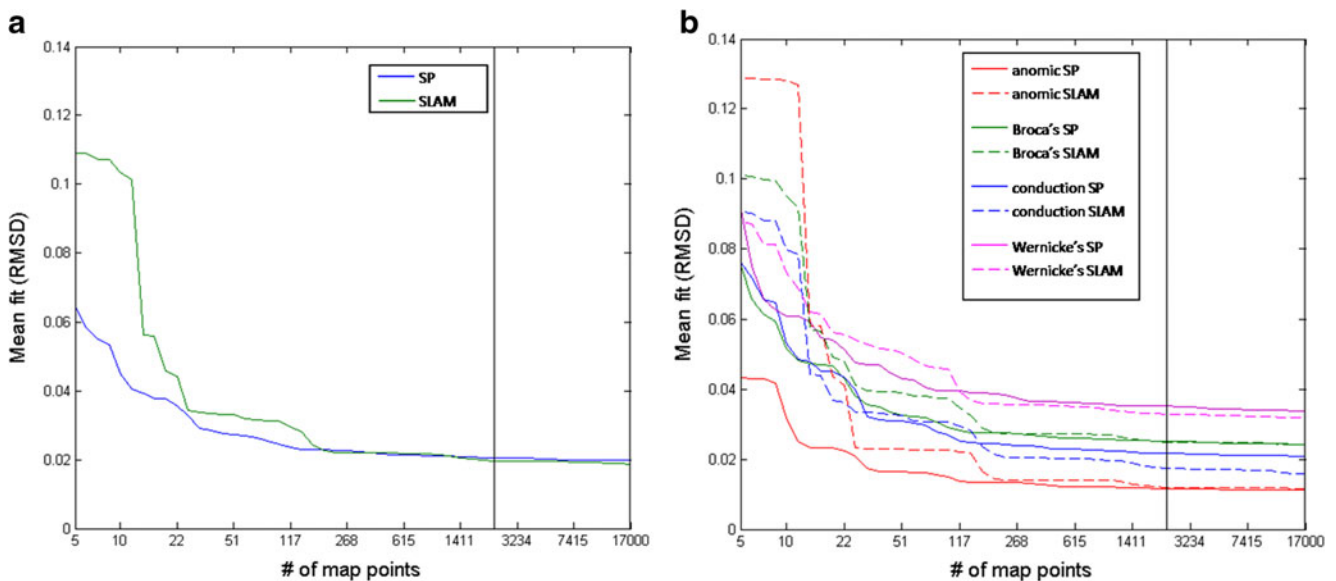
requiring many fewer predictions to achieve similar performance. We used this lower map resolution as a baseline, to compare the effects of adding points to the standard SP map with the effects of augmenting SP's structure. Because our fitting routine yielded better fits than the standard SP maps that have been available to researchers online (Dell et al., 2004), we have provided our fitting routine, with adjustable map resolutions, along with our new model, at the following Web address: <http://cogsci.uci.edu/~alns/webfit.html>

## Results

First, we examined our hypothesis that SLAM would fit the data at least as well as SP for the general aphasia population. All analyses were performed using the MATLAB software package. As we mentioned above, we chose to use RMSD as our measure of fit (where a lower value means a better fit). Table 1 provides descriptive statistics of the model fits for the entire sample of patients, as well as for the five subtypes of aphasia. Figure 5 shows a scatterplot comparing the SP and SLAM fits. The solid diagonal line represents the hypothesis that the models are equivalent, and the dotted lines indicate one standard deviation of fit difference in the sample. It is clear that both models do quite well overall, with the majority of patients clustering below .02 RMSD. Although the models tend to produce similar fits in general, it is also clear that a subgroup of patients falls well outside the 1-SD boundaries. The inset in Fig. 5 shows a bar graph comparing the numbers of patients who were better fit ( $>1$  SD) by SP or SLAM, demonstrating that SLAM provides better fits for a subgroup of patients without sacrificing fits in the general population.

Next, we examined our hypothesis that SLAM would improve the model fits specifically for conduction aphasia. Figure 6 displays the RMSD differences between the models for individual patients, grouped by aphasia type; positive difference values indicate improved fits for SLAM over SP. It is clear that the SLAM model provides the largest and most consistent fit improvements for the conduction group, and a majority of the fits for Wernicke's patients also benefit from the new model. The fact that Wernicke's aphasia was also better fit by SLAM is consistent with the HSFC theory. Wernicke's aphasia is associated with neuroanatomical damage very similar to that of conduction aphasia, and acute Wernicke's aphasia often recovers to be more like a conduction profile, suggesting a partially shared locus of impairment. For a statistical comparison of the fit improvements between the five aphasia subtypes, we performed a one-way analysis of variance (ANOVA) on the RMSD changes, which indicated at least one significant difference between the group means ( $p < .001$ ). A follow-up multiple comparison test indicated that the conduction group benefited more from SLAM than any other

<sup>1</sup> At the time of writing the manuscript, the authors were unaware of any freely available parallel algorithm to incrementally construct the Delaunay triangulation in arbitrary dimensions. We therefore implemented point evaluation and edge bisection using CUDA C and the Thrust library, executing these steps on a GPU, while the Delaunay triangulation was constructed on the computer's central processing unit (CPU) using the CGAL library. Performance tests comparing the parallel point evaluation step to a serial C++ implementation, running on an Nvidia Tesla K20Xm GPU and an Intel 1,200-MHz 64-bit CPU, respectively, demonstrated a speedup by a factor of 26.0.



**Fig. 4** Mean fit curves for (A) all patients and (B) diagnosis groups. The black vertical lines in the panels indicate the maps that were used for statistical comparisons

group, since the 95 % confidence interval for the mean fit improvement did not overlap with that of any other group, including Wernicke's.

To further validate these results, we tested whether fit improvements due to increasing the SP map resolution specifically favored any of the diagnosis groups. Unlike our theoretically motivated structural changes, this method of improving model fits was not expected to favor any particular group. We compared the model fits for an SP map with 189 points, which on average is equivalent to the standard SP map in the literature, to the higher-resolution SP map with 2,321 points. For the group of 255 patients, increasing the number of SP map points significantly improved the average fit from .0230 RMSD to .0206 RMSD ( $p < .001$ ). The improvement in fit was significant for all diagnosis groups (all  $ps < .001$ ); however, a one-way ANOVA with follow-up multiple comparison tests showed that no group had significantly greater improvement than every other group (no disjoint confidence intervals), unlike the result produced by our structural changes, which specifically favored the conduction group. Instead, the Wernicke's group improved most, whereas the anomic group improved least, consistent with the observation that these groups are already the worst and best fits for SP, respectively. The implication is that the improvements in fit caused by our theoretically motivated manipulation of the SP model's architecture are qualitatively different from the improvements gained by other methods.

We also hypothesized that the conduction naming pattern should be fit by a particular SLAM configuration: strong LA and weak AM weights. For the patients who exhibited the greatest improvements in fit, this was indeed

the case. Figure 7 uses boxplots to display the SLAM weight configurations for the 20 patients (13 conduction, five Wernicke's, one anomic, one Broca's) who exhibited the greatest fit improvements ( $>2$  SDs). Figure 8 shows data from an example patient with conduction aphasia, along with the corresponding SP and SLAM model fits. The best-fitting weights in the SP model were .022 and .017, for S and P, respectively. The SLAM model for this patient yielded .023 and .013 for SL and LM, respectively, whereas the LA weight was maximized at .04, and the AM weight was minimized at .0001. For this patient, SLAM reduced the SP fit error by .0135 RMSD. This example also illustrates that SLAM's largest fit improvements over SP are accompanied by a consistent increase in the predicted frequency of formal errors, along with a consistent decrease in semantic (and in unrelated) errors. This trade-off in formal errors for semantic errors is most likely to occur at the first, lexical-selection step. The dual nature of formal errors, that they can occur during either lexical or phonological selection, is one of the hallmarks of the SP model. Foygel and Dell (2000) showed that formal errors during lexical selection increase when phonological feedback to lexical units outweighs the semantic feedforward activation. In conduction aphasia, large LA weights provide strong phonological feedback to lexical units, whereas small AM and LM weights provide weak phonological feedforward to the motor units. With LM greater than AM, more activation flows from the incorrect, phonologically related lexical items, thereby increasing formal errors at the expense of semantic errors. The implication, that strong auditory–phonological feedback influences lexical selection in conduction aphasia,



**Table 1** Descriptive statistics for SLAM and SP model fits

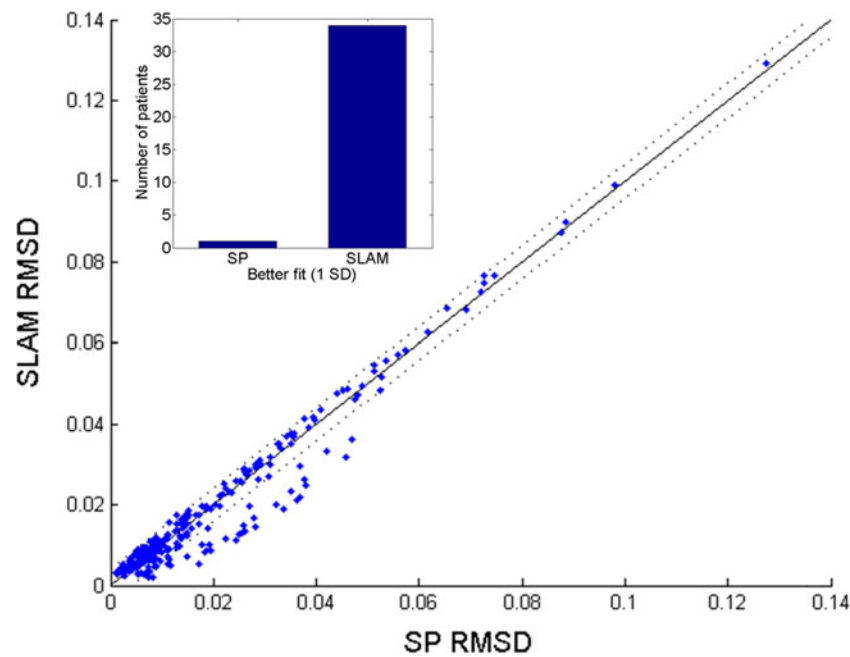
		SP			SLAM				
		S	P	RMSD	RMSD	SL	LM	LA	AM
<b>All</b> <i>N</i> = 255	Mean	.0241	.0232	.0194	.0187	.0242	.0152	.0300	.0254
	St. Dev.	.0118	.0075	.0189	.0189	.0121	.0059	.0085	.0135
	Median	.0256	.0224	.0127	.0111	.0250	.0151	.0300	.0300
	IQR	[.0170, .0341]	[.0179, .0281]	[.0071, .0264]	[.0067, .0246]	[.0176, .0347]	[.0113, .0200]	[.0225, .0388]	[.0151, .0388]
	Range	[.0001, .0400]	[.0062, .0400]	[.0010, .1273]	[.0019, .1977]	[.0001, .0400]	[.0026, .0300]	[.0010, .0400]	[.0001, .0400]
<b>Anomic</b> <i>N</i> = 103	Mean	.0299	.0274	.0110	.0115	.0299	.0181	.0308	.0296
	St. Dev.	.0081	.0070	.0095	.0095	.0085	.0054	.0080	.0113
	Median	.0296	.0266	.0082	.0085	.0300	.0176	.0325	.0350
	IQR	[.0241, .0376]	[.0219, .0318]	[.0049, .0137]	[.0056, .0141]	[.0250, .0400]	[.0151, .0200]	[.0250, .0400]	[.0204, .0400]
	Range	[.0054, .0400]	[.0106, .0400]	[.0010, .0654]	[.0019, .0685]	[.0063, .0400]	[.0038, .0300]	[.0101, .0400]	[.0001, .0400]
<b>Broca's</b> <i>N</i> = 60	Mean	.0215	.0218	.0238	.0240	.0217	.0143	.0267	.0266
	St. Dev.	.0125	.0071	.0243	.0250	.0128	.0053	.0088	.0144
	Median	.0205	.0202	.0145	.0149	.0200	.0126	.0250	.0325
	IQR	[.0139, .0334]	[.0174, .0262]	[.0076, .0312]	[.0075, .0300]	[.0144, .0313]	[.0101, .0176]	[.0200, .0350]	[.0188, .0400]
	Range	[.0001, .0400]	[.0075, .0400]	[.0012, .1273]	[.0019, .1292]	[.0001, .0400]	[.0026, .0275]	[.0101, .0400]	[.0001, .0400]
<b>Conduction</b> <i>N</i> = 46	Mean	.0245	.0182	.0203	.0157	.0250	.0120	.0323	.0163
	St. Dev.	.0110	.0053	.0153	.0137	.0110	.0048	.0089	.0134
	Median	.0259	.0177	.0175	.0110	.0275	.0126	.0375	.0144
	IQR	[.0020, .0331]	[.0145, .0219]	[.0078, .0282]	[.0063, .0217]	[.0188, .0338]	[.0088, .0138]	[.0250, .0400]	[.0038, .0238]
	Range	[.0001, .0400]	[.0062, .0300]	[.0019, .0720]	[.0028, .0727]	[.0001, .0400]	[.0038, .0275]	[.0101, .0400]	[.0001, .0400]
<b>Wernicke's</b> <i>N</i> = 35	Mean	.0126	.0195	.0332	.0318	.0123	.0115	.0305	.0233
	St. Dev.	.0095	.0059	.0209	.0225	.0096	.0051	.0080	.0130
	Median	.0133	.0193	.0294	.0275	.0126	.0101	.0325	.0250
	IQR	[.0039, .0187]	[.0152, .0248]	[.0155, .0448]	[.0139, .0472]	[.0032, .0185]	[.0076, .0151]	[.0232, .0388]	[.0123, .0350]
	Range	[.0002, .0400]	[.0070, .0327]	[.0042, .0979]	[.0038, .0989]	[.0001, .0400]	[.0038, .0225]	[.0163, .0400]	[.0001, .0400]
<b>Other</b> <i>N</i> = 11	Mean	.0180	.0255	.0275	.0283	.0178	.0173	.0293	.0246
	St. Dev.	.0148	.0073	.0222	.0221	.0148	.0057	.0080	.0123
	Median	.0173	.0257	.0143	.0125	.0176	.0176	.0275	.0275
	IQR	[.0019, .0265]	[.0220, .0301]	[.0087, .0488]	[.0106, .0488]	[.0023, .0250]	[.0151, .0200]	[.0250, .0372]	[.0200, .0319]
	Range	[.0003, .0400]	[.0133, .0400]	[.0049, .0617]	[.0057, .0628]	[.0001, .0400]	[.0076, .0288]	[.0151, .0400]	[.0001, .0400]

represents a novel prediction of our model that is supported by the data.

Finally, we tested the criticality of our assumption that LA weights must be greater than LM weights. We repeated our original analysis, this time comparing SP to SLMA, an alternative version of SLAM that has lexical–motor dominance instead of lexical–auditory dominance. SLMA was fit with a four-parameter map with 2,321 points, the same size as the SLAM map, culled from the 13,807 discarded SLAM points, ensuring that LM weights were always greater than or equal to the LA weights. Figure 9 is a scatterplot comparing the SP and SLMA model fits; the diagonal lines are the same as those in Fig. 5. When this alternative model architecture was used, there were no noticeable improvements over SP; the maximum change in fit was only .0038 RMSD. Thus, the mere presence of additional parameters in SLAM was not what

caused the observed fit improvements; their theoretically motivated arrangement was necessary, as well.

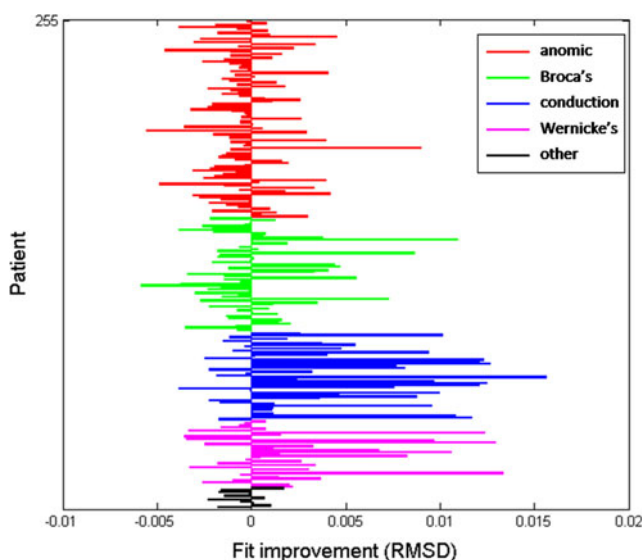
We also explored the necessity of the LM weights, testing the importance of our two routes. We fixed the LM weights at .0001 (effectively zero) by using 323 points from the full SLAM map to fit the data, thus yielding a three-parameter model, and we compared these fits with the fits from an SP map that had the same number of points. This three-parameter model that lacked direct LM connections did much worse than the two-parameter SP model, yielding an average fit of .10 RMSD. This catastrophic failure was due to the fact that not enough activation reached the motor units via the lexical–auditory–motor route. Recall that activation is multiplied by a fraction at each level, yielding lower activation after two steps through the lexical–auditory–motor route than after the one-step lexical–motor route. Without the combined inputs to



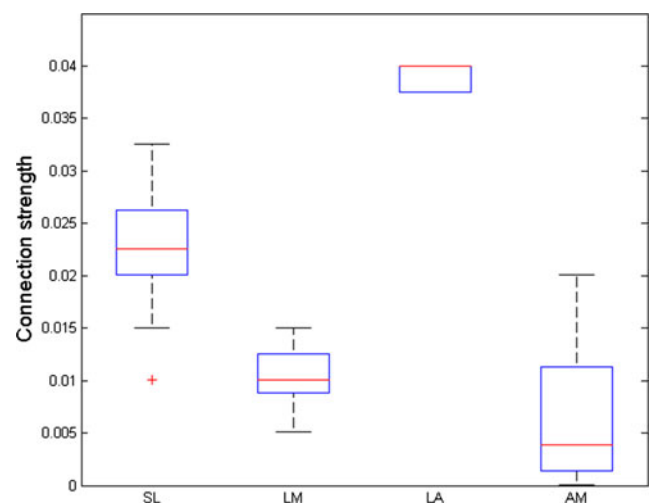
**Fig. 5** Scatterplot comparing model fits between SP and SLAM. The solid diagonal line represents equivalent fits; the dotted lines represent 1 *SD* of fit difference in the sample. The majority of patients are fit well by both models, and a subgroup of patients are fit notably better by SLAM (inset)

motor units from the two routes, the model could only produce a maximum estimate of 65 % correct responses. Although HSFC theory does predict that direct lexical–motor connections are required for normal levels of correctness, the weaker input to motor units from the auditory–motor route raises the concern that our initial choice of SLAM parameter constraints gave more prominence to the lexical–motor route than the HSFC theory warrants. We therefore explored the SLAM parameter space further, and we discovered alternative parameter constraints that yielded qualitatively similar results: In the

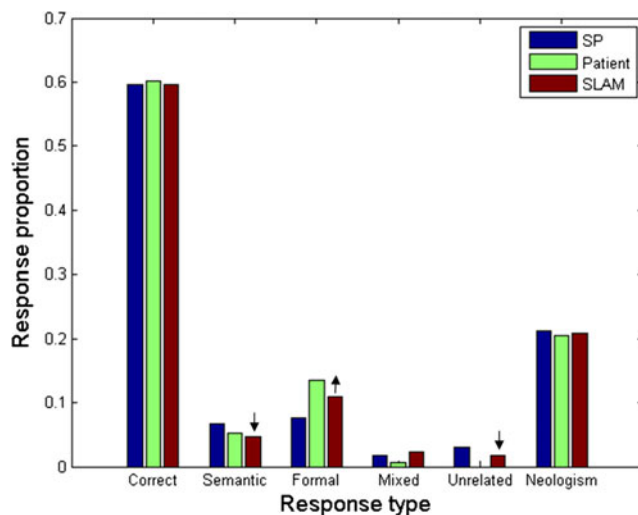
“healthy model,” the SL and LA weights have the usual maximum value of .04, whereas the LM weights have a maximum of .02, and AM weights have a maximum of .5; in aphasia, the parameters are free to vary below those values. This parameter arrangement ensures that the primary source of phonological feedback to the lexical layer is usually from auditory units, enables the auditory–motor route to provide strong activation to motor units during naming, and removes the previous constraint that in damaged states, the LM weights must always be lower than the LA weights. As with the original choice of



**Fig. 6** Individual fit changes between the SP and SLAM models. Positive values indicate better SLAM fits

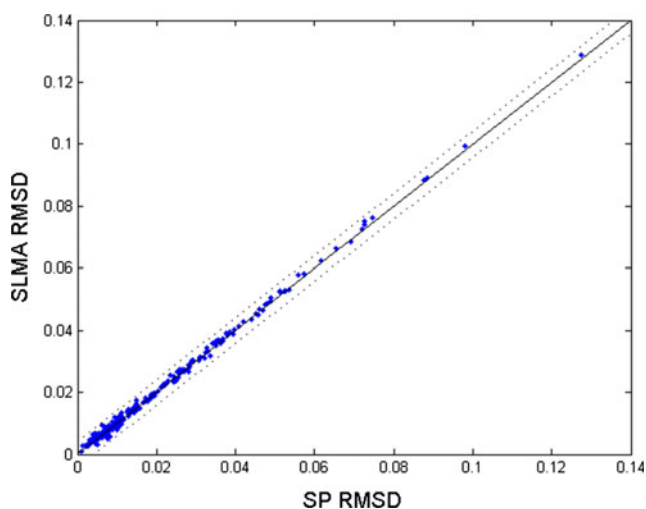


**Fig. 7** Boxplots showing the SLAM weights for the group of 20 patients with the greatest fit improvements. As expected, a model profile with high lexical–auditory and low auditory–motor weights leads to the greatest improvements over the SP model



**Fig. 8** Naming response distribution from an example patient with conduction aphasia, along with the corresponding SP and SLAM model fits. Arrows indicate how SLAM improves the fit to data, by increasing formal at the expense of semantic and unrelated errors. The SLAM model reduced the fit error for this patient by .0135 RMSD

SLAM parameter constraints, we observed fits similar to that of SP in the general population (Fig. S1), with noticeable improvements for the conduction naming pattern (Fig. S2), accompanied by high LA and low AM weights. With this alternative arrangement, a three-parameter model with LM weights fixed at .0001 still does not perform as well as the two-parameter SP model (Fig. S3), although the failure is no longer catastrophic, due to compensation by strong AM weights. To summarize, these investigations confirm our main finding that a second source of phonological feedback, predicted by HSFC theory to come from the auditory system, is the critical component for improving model fits.



**Fig. 9** Scatterplot comparing model fits between SP and the semantic–lexical–motor–auditory (SLMA) model, an alternative architecture with the same number of parameters as SLAM, but with lexical–motor dominance instead. The lines are the same as in Fig. 5. Unlike SLAM, SLMA provides no obvious fit improvements over SP

## Discussion

We put forward a new computational model of naming, SLAM, inspired by a recent conceptual model, HSFC, aimed at integrating psycholinguistic and motor-control models of speech production. SLAM implemented the HSFC claims that sublexical linguistic units have dual representations within auditory and motor cortices, and that the conversion of auditory targets to motor commands is a crucial computation for lexical retrieval, even prior to overt production.

We showed that augmenting the well-established SP model to incorporate auditory-to-motor conversion into the lexical-retrieval process allowed the model to explain general aphasic naming errors at least as well as the original SP model, while improving the model’s ability to account for conduction naming patterns in particular. The improvements in model fits were predicted to result from parameter settings with strong LA and weak AM weights. Examining the naming responses of 255 aphasic patients—the largest analysis of PNT responses to date—we confirmed our predictions, and additionally demonstrated that, unlike our theoretically motivated structural changes, improvements due to added map resolution were not specific to any aphasia type. We also discovered that the predicted weight configuration, which yielded the greatest fit improvements, did so by increasing formal errors at the expense of semantic errors. It is worth noting in this context that Schwartz et al. (2006) identified three anomalous subgroups whose naming patterns significantly deviated from SP’s predictions, one of which exhibited too many formal errors. Two of the patients in this subgroup had conduction aphasia, and the other had Wernicke’s aphasia. SLAM provides a plausible explanation for this subgroup. The increase of formal errors at the expense of semantic errors in conduction aphasia suggests that a significant proportion of their phonologically related errors were generated at the lexical-selection stage, rather than the phonological-selection stage, a novel prediction of our model. We also found that two separate phonological routes were required to produce the effect. Although the auditory–motor integration loop described by HSFC theory currently is not modeled in detail within SLAM, parallel inputs and feedback to separate auditory and motor systems are a prerequisite for state feedback control. The results of our modeling experiments thereby support the assumptions of the HSFC framework.

Although we pitted SP and SLAM against one another, they share many of their essential features. Thus, much of SLAM’s success can be attributed to the original SP model’s assumptions. The notions of computational simultaneity, hierarchical representation, interactivity among hierarchical layers, localized damage, and continuity between random and well-formed outputs are what enabled good predictions. The fact that we were able to successfully extend the model reinforces the utility of these ideas. Similarly, much of the criticism of SP applies equally to SLAM. For instance, the very small lexicon can only approximate the

structure of a real lexicon, and the semantic representations are arbitrarily defined. Although the model is interactive, it does not include lateral or inhibitory connections, which are essential features of real neurological systems. Also, the model does not deal directly with temporal information, which constitutes a large body of the psycholinguistic evidence regarding speech processing. Nevertheless, for examining the architectural assumptions of the HSFC, SP provided a useful test bed, in that it has been the best computational model available.

One further advantage of SLAM over SP (and over similar models that assume a unified phonological network) is that SLAM provides a built-in mechanism for repetition. Repetition is often used in addition to naming as a test of lexical-retrieval models, because repetition involves the same demands on the motor production system as naming, but lacks the semantic search component. In order to simulate repetition, however, some form of auditory representation is necessary, even if it is implicit. In Foygel and Dell (2000), the single-route SP model was used to simulate repetition, without explicitly modeling the auditory input, by assuming that perfect auditory recognition delivers a boost directly to lexical units, essentially just the second step of naming. Later, to account for patients with poor naming but spared repetition abilities, a direct input-to-output phonology route was added to the model (Hanley, Dell, Kay, & Baron, 2004). This dual-route model grafts the “nonlexical” route on to SP, leaving the architecture and simulations of naming unchanged; the two routes are used only during repetition. Although several studies have generated empirical support for the idea that the two routes are indeed used in repetition (Nozari, Kittredge, Dell, & Schwartz, 2010), our study suggests that both routes are used in naming as well, potentially providing a more cohesive account of the computations underlying these tasks. Given that SLAM already requires the auditory component for naming, we intend to develop it to simulate repetition as well, allowing for more direct comparisons to this alternative dual-route model in the future.

Although SLAM does not employ learning or time-varying representations, another lexical retrieval model that does implement these features has also adopted a similar separation of auditory and motor speech representations. Ueno, Saito, Rogers, and Ralph (2011) presented Lichtheim 2, a “neurocomputational” model, which simulates naming, repetition, and comprehension for healthy and aphasic speech processing, using a network architecture in which each layer of units corresponds to a brain region. Lichtheim 2 does not categorize speech error types according to SP’s more detailed taxonomy, however, making it hard to compare directly with SLAM. Furthermore, since our goal with SLAM was to investigate the effects of the separate phonological representations, and Lichtheim 2 shares this architectural assumption, we did not compare the models directly. In Lichtheim 2, the phonology of the input and the output is represented by a pattern of phonemic features presented one cluster at a time, and semantic representations are temporally static and statistically independent of their

corresponding phonological representations. The model is simultaneously trained on all three tasks, and hidden representations are allowed to form in a largely unconstrained manner. The trained network can then be “lesioned” in specific regions to simulate aphasic performance. We see much in common between our approaches in terms of their theoretical motivations, proposing psycholinguistic representations grounded in neuroanatomical evidence. Furthermore, the use of a single network to perform multiple tasks is very much in line with our plans to develop the SLAM model. One major difference between SLAM and Lichtheim 2 is that SLAM maintains an explicit hierarchical separation between lexical units and phonological units, allowing for selection errors at either stage. This hierarchical separation was essential for making our successful predictions regarding conduction naming patterns. It remains to be seen how our proposed architecture will cope with multiple tasks simultaneously.

Another model of lexical production, *WEAVER++/ARC* (Roelofs, 2014), has been proposed as an alternative to Lichtheim 2. Although this model uses spreading activation through small, fixed networks, as SP does, it also employs condition-action rules to mediate task-relevant selection of the network’s representations, thereby implementing a separation of declarative and procedural knowledge. Like Lichtheim 2, this model does not apply the detailed error taxonomy examined by SLAM, and so we did not compare them directly. Importantly, though, *WEAVER++/ARC* and Lichtheim 2 largely agree on most cognitive and computational issues, especially the primary one investigated by SLAM: the participation of separate auditory and motor–phonological networks in speech production. Additionally, like SLAM and Lichtheim 2, *WEAVER++/ARC* simulates the conduction aphasia pattern by reducing weights between the input and output phonemes. The primary disagreement between *WEAVER++/ARC* and Lichtheim 2 is an anatomical one: Should the lexical–motor connections for speech production be associated with the (dorsal) arcuate fasciculus or the (ventral) uncinate fasciculus? At present, the SLAM model is compatible with either position.<sup>2</sup> *WEAVER++/ARC* does differ from SLAM with respect to one important theoretical point, however. In *WEAVER++/ARC*, the input and output lexical units are separated, and in naming, activation primarily flows from lexical output units to motor units. Auditory units then provide stabilizing activation to motor units through an auditory feedback loop (i.e., motor to auditory to motor), rather than being activated by a single lexical layer in parallel with motor units to serve as sensory targets. This runs contrary to our finding that

<sup>2</sup> One might wonder whether the lexical–motor and auditory–motor connection weights were generally correlated in our sample. They were not ( $r = .10$ ,  $p = .09$ ). This seems to indicate that these mappings are functionally and anatomically distinct; however, *WEAVER++/ARC* also allows these routes to be independently lesioned, so this is not necessarily a strong point of disagreement.



strong lexical–auditory feedback influenced lexical selection for conduction aphasia. Again, it remains to be seen whether our assumption of a single lexical layer can account for multiple tasks as Lichtheim 2 and WEAVER++/ARC do, which we intend to explore in future work.

The SLAM model falls into a broad class of models that can be described as “dual-route” models—that is, models that posit separate but interacting processing streams controlling behavior. Much of this work relates directly to Hickok and Poeppel’s (2000, 2004, 2007) neuroanatomical dual-stream framework for speech processing, in that the mapping between auditory and motor speech systems corresponds to the dorsal stream, whereas the mapping between auditory and lexical–semantic levels corresponds to the ventral stream. Although Hickok and Poeppel discussed this cortical network from the perspective of the auditory speech system, which diverges into the two streams, picture naming traverses both streams, going from conceptual to lexical to auditory (ventral stream) and from auditory to motor (dorsal stream). One difference between the SLAM model and the Hickok and Poeppel framework is that explicit connectivity is assumed between the lexical and motor–phonological networks. Hickok and Poeppel assumed (but didn’t discuss) connectivity between conceptual and motor systems, but did not specifically entertain the possibility of lexical-to-motor speech networks. The present model, along with the HSFC, thus refines the Hickok and Poeppel dual-stream framework.

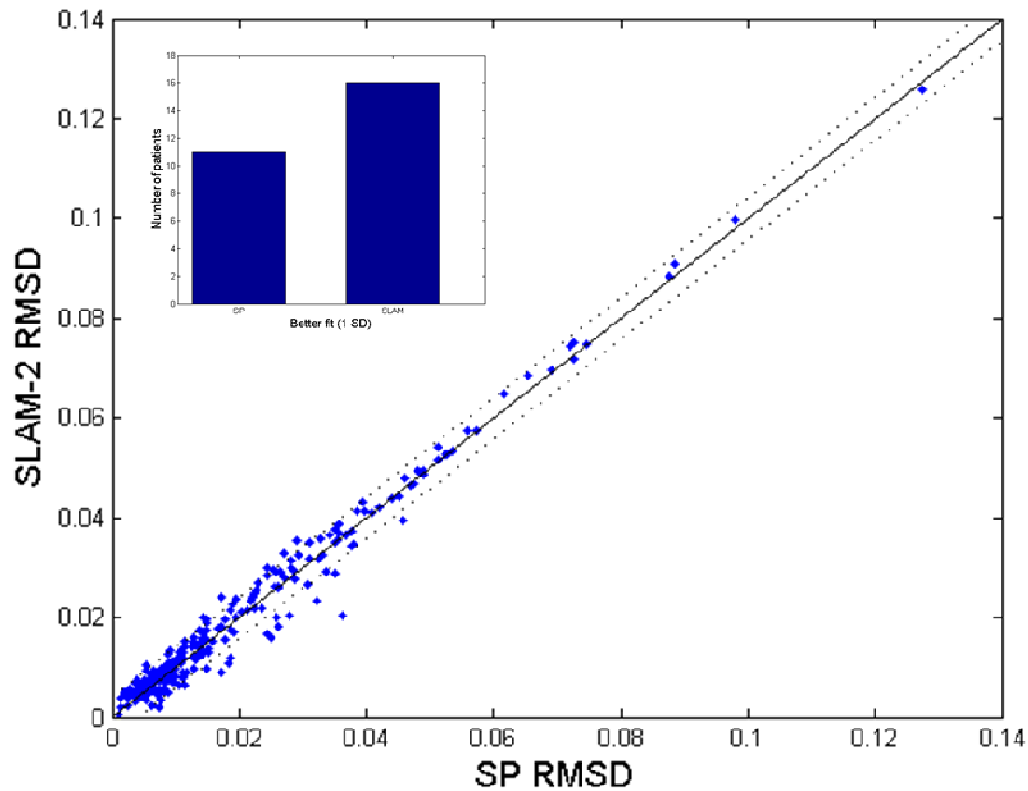
## References

- Abel, S., Huber, W., & Dell, G. S. (2009). Connectionist diagnosis of lexical disorders in aphasia. *Aphasiology*, 23, 1353–1378.
- Anderson, J. M., Gilmore, R., Roper, S., Crosson, B., Bauer, R. M., Nadeau, S., . . . Heilman, K. M. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke–Geschwind model. *Brain and Language*, 70, 1–12.
- Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, 6, 183–200.
- Bossom, J. (1974). Movement without proprioception. *Brain Research*, 71, 285–296.
- Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage*, 31, 1327–1342. doi:10.1016/j.neuroimage.2006.01.036
- Caramazza, A. (1991). Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. In A. Caramazza (Ed.), *Issues in reading, writing and speaking: A neuropsychological perspective* (pp. 15–44). Amsterdam: Springer.
- Cole, J. D., & Sedgwick, E. M. (1992). The perceptions of force and of movement in a man without large myelinated sensory afferents below the neck. *Journal of Physiology*, 449, 503–515.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321. doi:10.1037/0033-295X.93.3.283
- Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, 21, 125–145. doi:10.1080/02643290342000320
- Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56, 490–520.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 801–838. doi:10.1037/0033-295X.104.4.801
- Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Branch Coslett, H. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, 128, 380–396. doi:10.1016/j.cognition.2013.05.007
- Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, 43, 182–216. doi:10.1006/jmla.2000.2716
- Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain*, 88(237–294), 585–644.
- Goodglass, H. (1992). Diagnosis of conduction aphasia. In S. E. Kohn (Ed.), *Conduction aphasia* (pp. 39–49). Hillsdale: Erlbaum.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611–633. doi:10.1037/0033-295X.105.4.611-633
- Hanley, J. R., Dell, G. S., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology*, 21, 147–158.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13, 135–145. doi:10.1038/nrn3158
- Hickok, G. (2014a). The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes*, 29, 2–20. doi:10.1080/01690965.2013.834370
- Hickok, G. (2014b). Toward an integrated psycholinguistic, neurolinguistic, sensorimotor framework for speech production. *Language and Cognitive Processes*, 29, 52–59. doi:10.1080/01690965.2013.852907
- Hickok, G., & Buchsbaum, B. (2003). Temporal lobe speech perception systems are part of the verbal working memory circuit: Evidence from two recent fMRI studies. *Behavioral and Brain Sciences*, 26, 740–741.
- Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A. K., Naeve-Velguth, S., Strupp, J. P., . . . Ugurbil, K. (2000). A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: Implications for the explanation of conduction aphasia. *Neuroscience Letters*, 287, 156–160.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69, 407–422. doi:10.1016/j.neuron.2011.01.019
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402. doi:10.1038/nrn2113
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216.
- Jacobs, O. L. R. (1993). *Introduction to control theory*. Oxford: Oxford University Press.
- Jacquemot, C., Dupoux, E., & Bachoud-Lévi, A. C. (2007). Breaking the mirror: Asymmetrical disconnection between the phonological input

- and output codes. *Cognitive Neuropsychology*, 24, 3–22. doi:10.1080/02643290600683342
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9, 718–727.
- Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, 25, 463–492. doi:10.1080/02643290701674851
- Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *Journal of the Acoustical Society of America*, 110, 2845–2848.
- Lomas, J., & Kertesz, A. (1978). Patterns of spontaneous recovery in aphasic groups: A study of adult stroke patients. *Brain and Language*, 5, 388–401.
- Mirman, D., Strauss, T. J., Brecher, A., Walker, G. M., Sobel, P., Dell, G. S., & Schwartz, M. F. (2010). A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive function. *Cognitive Neuropsychology*, 27, 495–504. doi:10.1080/02643294.2011.574112
- Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, 63, 541–559.
- Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, 98, 112–117. doi:10.1016/j.bandl.2006.04.006
- Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25, 382–407.
- Pulvermüller, F. (1996). Hebb's concept of cell assemblies and the psychophysiology of word processing. *Psychophysiology*, 33, 317–333.
- Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, 24, 121–134.
- Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33–48. doi:10.1016/j.cortex.2014.07.001
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421–436.
- Sanes, J. N., Mauritz, K. H., Evarts, E. V., Dalakas, M. C., & Chu, A. (1984). Motor deficits in patients with large-fiber sensory neuropathy. *Proceedings of the National Academy of Sciences*, 81, 979–982.
- Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasic picture naming. *Brain and Language*, 73, 62–91.
- Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, 54, 228–264. doi:10.1016/j.jml.2005.10.001
- Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33, 89–108. doi:10.1146/annurev-neuro-060909-153135
- Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423, 866–869.
- 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
- Wernicke, C. (1969). The symptom complex of aphasia: A psychological study on an anatomical basis. In R. S. Cohen & M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science* (pp. 34–97). Dordrecht: Reidel. **Original work published 1874.**
- Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1, 209–216.
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269, 1880–1882.
- Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60, 213–232. doi:10.1037/h0044155

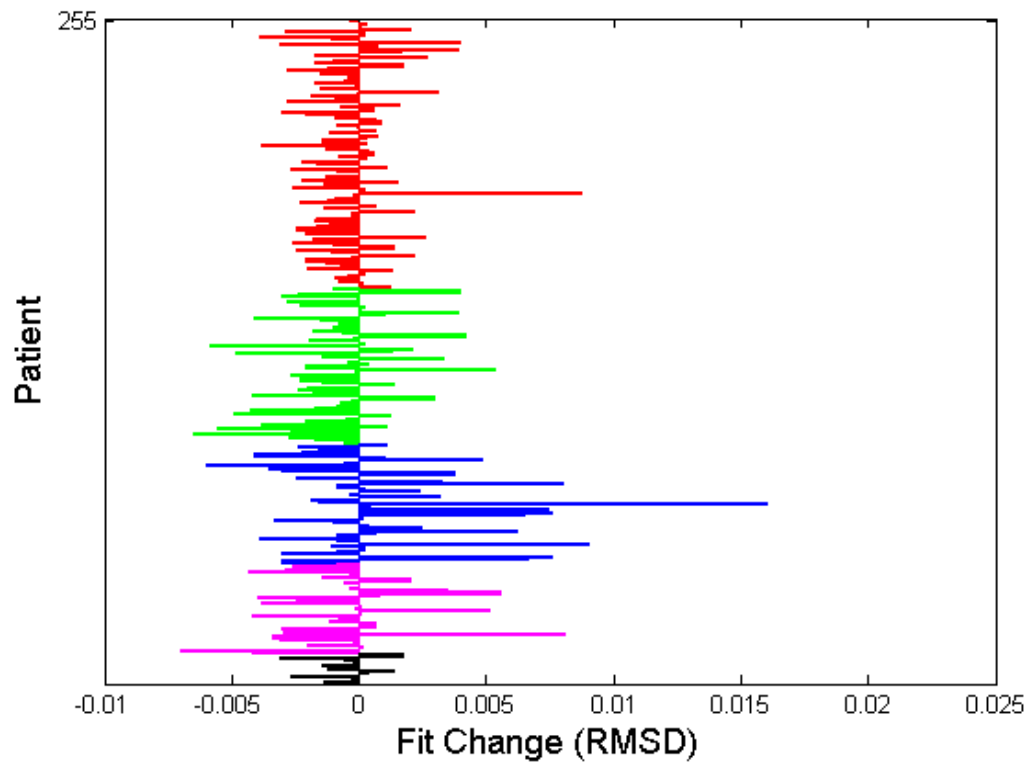
## Supplementary Material

We repeated our analyses with alternative parameter constraints for SLAM, called SLAM-2, investigating whether our improved fits relied too much on the lexical-motor route. The new parameters were constrained such that SL and LA max = .04, LM max = .02<sup>1</sup>, AM max = .5, and they were free to vary below those values (min = .0001). We compared a SLAM-2 map with 2,321 points against the SP map (Figure S1). The results were qualitatively similar to our initial SLAM model results, with good fits in general, and notable improvements specifically for the Conduction group (Figure S2), accompanied by high LA weights and low AM weights. Removing the LM route (i.e., fixing LM = .0001) created a 3-parameter model with a 745-point map, which still failed to outperform the 2-parameter SP model (Figure S3). Thus, the critical component for the observed fit improvements is the separate phonological feedback to the lexical layer from auditory units.



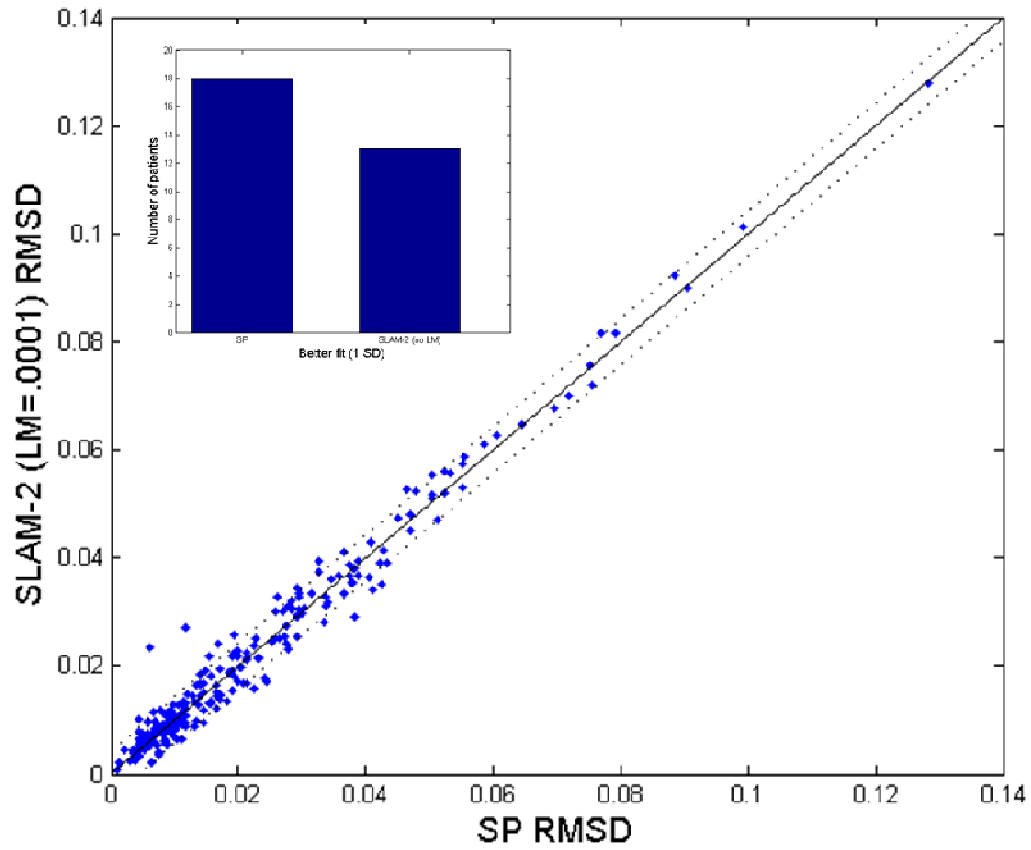
**Figure S1.** Scatterplot comparing model fits for SP and SLAM-2, testing whether our improved fits relied too much on LM weights. Diagonal lines are the same as in Figure 5. Once again, results show good fits overall, and SLAM outperforms SP for a subgroup of patients.

<sup>1</sup> The initial mapping procedure set LM max at .04, and then points with LM greater than or equal to .02 were removed, yielding an actual max LM of .0188.



**Figure S2.** Individual fit changes between the SP and SLAM-2 models. Positive values indicate better SLAM-2 fits. Anomic = red, Broca's = green, Conduction = blue, Wernicke's = magenta, Other = black.





**Figure S3.** Scatterplot comparing model fits for SP and SLAM-2 with the LM route removed. Diagonal lines are the same as in Figure 5. This 3-parameter model does not perform as well as the 2-parameter SP model, and thus it does not yield the improvements seen with the 4-parameter model.

- author: Grant M. Walker and Gregory Hickok
- year: 2016
- journal: Psychonomic Bulletin Review
- DOI: 10.3758/s13423-015-0903-7
- filename: 2016Walker\_Hickok\_slam\_ja.md

## 発話産出への計算機的アプローチの橋渡し：意味(S)-語彙(L)-聴覚(A)-運動(M) モデル (SLAM)

- Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM)

### 要旨:

音声生成は、心理言語学的な視点と運動制御の視点の両方から研究されているが、両アプローチ間の相互作用はほとんどない。本研究では、心理言語学と運動制御の概念を統合して音声生成の理論を構築することの説明力を評価した。一般的な心理言語学的な語彙検索モデルに、運動制御にヒントを得たアーキテクチャを追加することで、失語症の文脈における発話エラーを説明する新しい計算モデルを作成した。255人の失語症患者から得られた絵画命名データに対するモデルの適合度を比較したところ、理論的に予測可能な失語症のサブタイプである「伝導」の適合度が向上することがわかった。これは、聴覚-語彙フィードバックの活性化が強く、聴覚-運動フィードフォワードの活性化が弱いために、語彙選択の際に音素的に関連する隣人との競争が激化した結果であることを発見した。今回の発見は、他の既存の語彙検索モデルにも影響を与えるものと思われる。

キーワード: Speech production; Computational models; Neuropsychology

音声生成は、心理言語学、運動制御、神経科学など、いくつかの理論的観点から研究されてきたが、これらのアプローチ間の相互作用はほとんどなかった。しかし、最近の研究では、統合が生産的である可能性が示唆されている。特に、順方向モデルと逆方向モデルを組み合わせるなど、運動制御からの計算原理をより高レベルの言語プロセスに適用することに関してである (Hickok, 2012, 2014a, 2014b)。ここでは、Foygel and Dell (2000) の音声生成の心理言語学的計算モデルを、運動制御にインスパイアされたアーキテクチャを用いて修正することで、この可能性をより詳細に探り、新しいモデルがデータによりよくフィットし、理論的に解釈可能な方法で提供されるかどうかを評価する。

まず、本研究の理論的基礎を示すために、(1) Foygel and Dell (2000) の semantic-phonological model (SP) の背後にある動機を説明し、(2) 運動制御アプローチを簡単に要約し、(3) 両アプローチを統合する我々の最近の概念的試みからいくつかの原理を強調し、(4) 運動制御理論からの基本原理を用いて SP を修正し、新しい semantic-lexical-auditory-motor model (SLAM) を作成した。次に、SP モデルと SLAM モデルの計算方法の詳細と、SP モデルと SLAM モデルの比較シミュレーションを行う。シミュレーションの結果、特に理論的に予測可能な失語症患者のサブカテゴリに関しては、SLAM が SP よりも優れていることがわかった。最後に、この新しいモデルが他の既存の単語生成モデルとどのように関連しているかを議論して締めくくる。

## 1. SP モデル

SP は、Dell (1986) の文生成における検索の理論に端を発している。この理論は、大量の自然音声に見られるスピーチエラー (舌の滑落) を説明するために開発された。この理論は、心理学と言語学の概念を統合したものである。心理学からは「計算上の同時性」という概念を取り入れ、複数の内部表現が生成前に選択されるようにした。また、言語学からは、階層的な表現レベルと、保存された語彙知識と適用された生成規則を各レベルで分離することを取り入れた。

Dell, Schwartz, Martin, Saffran, and Gagnon (1997) は、計算モデルを提案した。このモデルでは、対象を単一の単語生成に限定しているが、理論的範囲は失語症の文脈における音声エラーの説明にまで拡大している。基本的な考え方は、失語症の音声エラーのパターンは、損傷を受けた音声生成システムの出力を反映しており、失語症のデータに合うように通常のモデルのパラメータを調整することでモデル化できるというものである。モデルの構造は、意味、語彙、音韻の各ユニットからなる3層のネットワークで構成されており、ユニット間の接続は、典型的な語彙領域の構造に近似するように実験者が選択した (図1)。単語の生成は、時間とともに活性化のノイズと減衰を伴う、拡散活性化過程としてモデル化された。損傷を与えるには、表現レベル間の活性化の流れを制御するパラメータを変更する必要があった。その後、シミュレーションを行い、失語症患者に類似したエラータイプの頻度を生み出すパラメータ値を特定した。

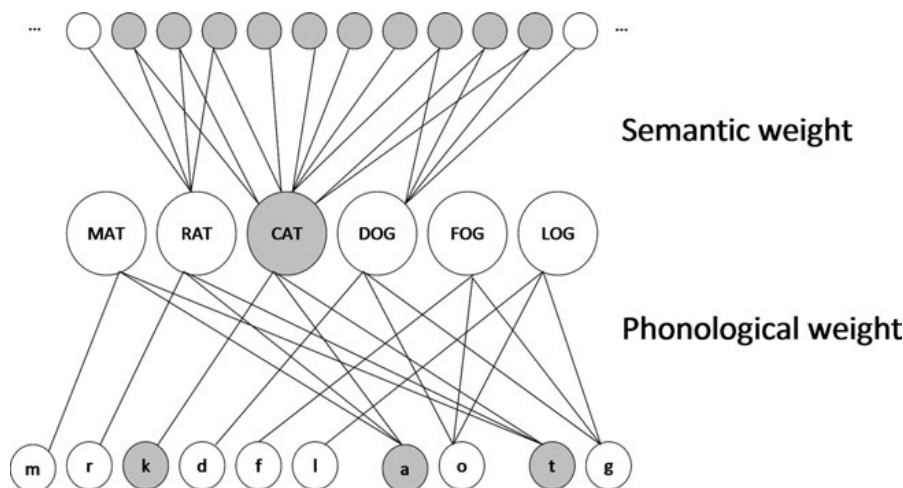


図1 SP (Semantic-Phonological) モデルのアーキテクチャ

しかし、シミュレーション手法は計算量が多いため、包括的な検討は事実上、一度に2つのパラメータに限られていた。しかし、Foygel and Dell (2000) に始まる一連の論文では、モデル内の2つの自由パラメータが特定され、臨床診断情報を含む絵画命名課題から得られた驚くほど多様なデータを説明することができた (Abel, Huber, & Dell, 2009)。臨床診断情報 (Abel, Huber, & Dell, 2009)、語彙頻度効果 (Kittredge, Dell, Verkuilen, & Schwartz, 2008)、異なるタイプの失語症に関連する特徴的なエラーパターン (Schwartz, Dell, Martin, Gahl, & Sobel, 2006)、回復の特徴的なパターン (Schwartz & Brecher, 2000)、および相互的なエラー効果 (Foygel & Dell, 2000) を含む、絵画命名課題から得られた非常に多様なデータを説明するモデルが特定された。この2つの自由パラメータは、意味的表現と語彙的表現の間の接続強度 (s-重み) と、語彙的表現と音韻的表現の間の接続強度 (p-重み) であり、SP と呼ばれるアーキテクチャであった。SP は、単語の繰り返し (Dell, Martin, & Schwartz, 2007) や、臨床画像で見られる神経損傷の位置の予測 (Dell, Schwartz, Nozari, Faseyitan, & Branch Coslett, 2013) など、他の課題のパフォーマンスを説明するのにも用いられているが、ここでは主に絵画命名課題でのエラーとの関連性に焦点を当てる。

SP は、特に意味レベルと音韻レベルの間で行われる計算に関係している。モデルの出力は、抽象的な音素の列であり、それを声道を制御するための運動計画に変換しなければならないと想定されている。次に、運動効果器が実際にどのように制御されるかについての研究から生まれた、いくつかの基本的な構成要素について説明する。

## 2. 運動制御理論 motor-control theory

最も広いレベルで、運動制御は、最初の計画とフィードバック制御のために、運動システムへの感覚入力が必要とする。計画には、運動行為の対象 (例えば、特定の大きさや方向のカップを、身体に対して特定の位置に置く) を定義するための入力と、効果器の現在の状態 (例えば、カップに対する手の位置と速度) に関する情報を提供するための入力が必要である。自然界 (Cole & Sedgwick, 1992; Sanes, Mauritz, Evarts, Dalakas, & Chu, 1984) や実験 (Bossom, 1974) での感覚脱失の例が示すように、感覚情報がなければ行動は不可能である。また、感覚情報は、運動中に重要なフィードバック情報を提供し (Wolpert, 1997; Wolpert, Ghahramani, & Jordan, 1995), エラー検出と修正のメカニズムを提供することが示されている (Kawato, 1999; Shadmehr, Smith, & Krakauer, 2010)。しかし、音声生成のように精密な動作を高速に行う場合には、フィードバックの遅れやノイズの多い環境などにより、フィードバック機構が信頼できない場合がある。このような場合には、状態フィードバック制御システムにフォワードモデルと逆モデルを追加することで (Jacobs, 1993)、運動命令と感覚結果の間の過去に学習した関連性を利用して、効果器を感覚目標に導くことができる。このようにして、運動系と感覚系は、オンラインでの生成や知覚に先立って、密接に結びついていることが示唆される。

スピーチの場合、最も重要な感覚ターゲットは聴覚だが (Guenther, Hampson, & Johnson, 1998; Perkell, 2012)、体性感覚情報も重要な役割を果たす (Tremblay, Shiller, & Ostry, 2003)。聴覚フィードバックの変化は、音声生成に大きな影響を与えることがわかっており (Houde & Jordan, 1998; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Yates, 1963)、発話者の発話環境が変化すると、「ジェスチャードリフト」つまり発話パターン (アクセント) が変化することがある (Sancier & Fowler, 1997)。さらに、隠語の発話を調査した神経画像実験では、一貫して側頭葉の聴覚関連皮質の活性化の増加が報告されている (Callan et al., 2006; Hickok & Buchsbaum, 2003; Okada & Hickok, 2006)。

音声生成における聴覚システムの役割については、言語に関する神経心理学的な研究から、特に重要な証拠が得られている。神経系の損傷によって言語処理能力が損なわれるパターンと損なわれないパターンが顕著に見られることから、理論家たちは脳内の音声表現を聴覚と運動に分けて考えるようになった (Caramazza, 1991; Jacquemot, Dupoux, & Bachoud-Lévi, 2007; Pulvermüller, 1996; Wernicke, 1874/1969)。例えば、伝導性失語症 (Goodglass, 1992) の患者は、流暢な発話をしており、運動表象が保たれていることを示唆している。また、これらの患者は、聴覚的な理解力が高く、自分の誤りを認識できることから、聴覚的な表象が保たれていることが示唆される。このような能力があるにもかかわらず、発話時に多くの音素性の誤りを犯し、非単語の繰り返しに問題がある。このパターンは一般的に、聴覚系と運動系の間のインターフェースの損傷によるものと説明されている (Anderson et al., 1999; Geschwind, 1965; Hickok, 2012; Hickok et al., 2000)。伝導性失語症に関するこの点は、以下に述べるように、理論的にも重要な意味を持っている。

## 3. 概念的統合 Conceptual integration

階層的状態フィードバック制御 (HSFC; Hickok, 2012) モデルは、心理言語学的な概念と生物学的な運動制御理論の概念を統合するための理論的枠組みを提供する。この概念的な枠組みは、3つの中心的な原理を中心に構成されている。1つ目の原理は、音声表現は感覚皮質と運動皮質で補完的に符号化されており、音声生成時には (少なくとも) 音節のレベルまで並行して活性化されるというものである。第2の原理は、感覚運動翻訳領域を介した、感覚皮質と運動皮質の間の特定のパターンの興奮性および抑制性の接続により、運動の計画/活性化の初期段階で運動プログラムの選択にエラーが生じる可能性があるにもかかわらず、運動表象を感覚ターゲットに向けて強固に導くことができる、一種のフォワード/インバースモデルが実装されているというものである。3つ目の原理は、音声生成を支える感覚運動ネットワークは階層的に構成されており、体性感覚皮質は音素 (正確には、体性感覚の状態としてコード化できる両唇閉鎖などの音素レベルのターゲット) の順に小さなユニットを処理し、聴覚皮質は音節 (すなわち音響的ターゲット) の順に大きなユニットを処理するというものである。HSFC のフレームワークの概略を図2に示す。上部 (濃い色) は SP の2つのステップを体現しているが、音韻成分を聴覚・音韻ネットワークと運動・音韻ネットワークの2つのサブコンポーネントに分解していることが明らかである。このような概念の重なりから、私たちは、第1原理に直接関連し、他の2つの原理に部分的に関連する新しい計算モデルを作成した。我々は、HSFC モデルのアーキテクチャの仮定を、既に成功しているネーミングの計算モデルである SP と統合することで、部分的に評価することができると考えた。

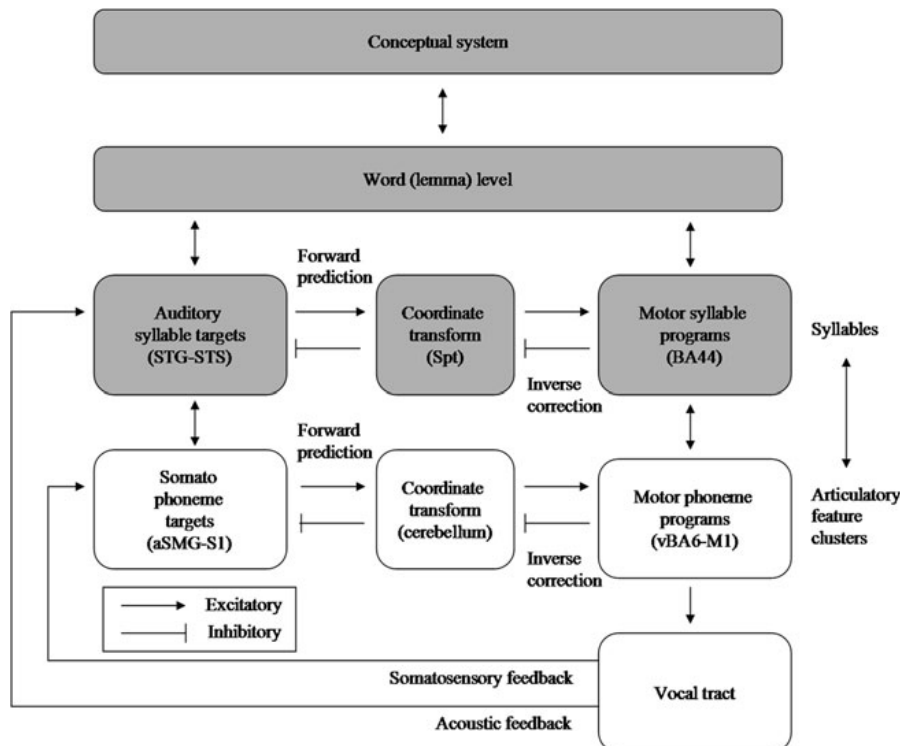


図2 hierarchical state feedback control (HSFC) フレームワークの概略図 (Hickok, 2012)

## 4. SLAM モデル

SLAM は、音素表現を聴覚成分と運動成分に分けて表現する語彙検索の計算モデルである (図3)。音素の二重表現は、HSFC の第一原理から直接導かれる。感覚ユニットを聴覚表現としたのは、第3の原理、すなわち、このレベルの符号化は音素よりも大きいという原則に基づくものである。SP も SLAM も抑制性の接続を含んでいないため、第2のHSFC 原理は直接実装されていないが、SLAM モデルの接続パターンは、潜在的にノイズの多い運動命令を強化することができる一種のフォワード/インバースモデルを実装している。私たちの目的は、SP の計算上の仮定をできる限り変更せずに、運動と感覚の音韻表現が別々であるというアーキテクチャ上の仮定の効果を評価することであった。

絵画命名シミュレーションでは、主に意味的ユニット、語彙的ユニット、聴覚的ユニット、運動的ユニットの順に活性化されるため、このモデルの頭文字をとって SLAM と呼ばれる。また、語彙ユニットと運動ユニットの間には、より弱い直接的な接続がある。この語彙と運動のつながりの存在は、語彙ユニットから運動ユニットへの直接的な情報の流れによって音声生成が行われている可能性を認めている。これはWernicke (1874/1969) にまで遡る仮定であり、流暢さが保たれていることや、ミスのない発話の瞬発力を説明するのに必要なものである。伝導性失語症では、流暢性が保たれ、誤りのない発話が急増する。しかし、この接続は、語彙-聴覚ルート (これも Wernicke のオリジナルアイデア) よりも常に弱いものである。まず、聴覚-語彙ルートは語彙-運動ルートよりも早く発達し、より頻繁に使用されると考えられている。縦断的な研究によると、子供は一つの単語を理解し始めるのは、その単語を発する数ヶ月前であり、新たに理解した単語を獲得する割合は、新たに発した単語を獲得する割合の約 2 倍であることがわかっている (Benedict, 1979)。第二に、運動制御理論では、運動計画は感覚的なターゲットによって動かされるとされている。発達段階において、学習者は、音声とその音声を再現する運動ジェスチャーの間のマッピングを学ぶために、聴覚的なターゲットを参照する必要がある (Hickok, 2012; Hickok, Houde, & Rong, 2011)。第三に、失語症の場合、理解力の障害は生産力の障害よりも回復する傾向があり (Lomas & Kertesz, 1978)、これは語彙と聴覚音韻の表現の間に強い関連性があることを示唆している。

語彙-聴覚マッピングが語彙-運動マッピングよりも常に強いという仮定は、重要な結果をもたらす。それは、SLAM モデルが単に SP モデルに余分な部分を加えたものではなく、実際には SP と SLAM がカバーするパラメータ空間には実質的にゼロのオーバーラップがあるということである。その理由は以下の通りである。図 3 に示した SLAM のアーキテクチャを考えると、語彙-聴覚、聴覚-運動のマッピングの接続重みをゼロにし、語彙-運動の重みを自由に变化させるだけで、SP を実装できることは明らかである。そうすれば、SP は SLAM の適切なサブセットとなり、SLAM は SP と同じパラメータ空間をカバーすることができる(したがって、データへの適合性も)。しかし、このようなアーキテクチャの可能性は、「語彙聴覚重みは語彙運動重みよりも常に強い」という我々の仮定を実装することで、明示的に除外された。聴覚-語彙重みがゼロであれば、運動-語彙重みもゼロでなければならず、自由に变化させることができないため、SP で使われていたパラメータ部分空間を効果的に除外することができた。これにより、SLAM の「命名時には語彙聴覚ルートが主に使われる」という仮定をさらに検証することができる。また、逆の制約、すなわち、語彙的聴覚の重みが語彙的運動の重みよりも常に小さい場合のモデルの性能も調べることができる。これは、語彙-運動の優位性を反映して「SLMA」と呼ばれ、SP のパラメータ空間をサブセットとして含む変形である。SLAM と SLMA は、自由パラメータの数は同じで、どちらも SP よりも多いが、接続強度のパターンに関する仮定が異なる。もし、SLMA が SP をパラメータ空間の適切な部分集合として実装しているにもかかわらず、SLAM が SLMA よりも優れた結果を出したとしたら、語彙-聴覚ルートの優位性が理論的に動機付けられているだけでなく、観測された改善に必要であることを示すことになる。

要約すると、我々は、SLAM が SP と同様に一般の失語症患者の障害を特徴づけることができ、主に伝導性失語症のモデル化に役立つという仮説を立てた。伝導性失語症は、聴覚的音声表現と運動的音声表現の接点における機能障害であり、特に音韻レベルに影響を及ぼすと説明されている (Hickok, 2012; Hickok et al., 2011)。したがって、聴覚-音韻表現と運動-音韻表現の間のマッピングを組み込んだ命名モデルは、このマッピングの機能障害に起因する音声エラーに、よりよく適合するはずである。この仮説を検証するために、大規模な失語症の絵画命名データに対して、SP モデルと SLAM モデルの適合性を比較した。

## 5. 計算機による実装

### 5.1. 患者データ

すべてのデータは、Moss Aphasia Psycholinguistic Project Database (Mirman et al., 2010; [www.mappd.org](http://www.mappd.org)) から収集した。このデータベースには、代表的な失語症患者のデータが含まれている。その中には、Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) の回答が含まれている。データベースに登録されたすべての患者は、左半球に脳卒中を発症した後の急性失語症で、他に神経学的合併症の診断を受けておらず、少なくとも 1 つの PNT 項目を正しく挙げることができた。データベースに登録されている患者のうち、失語症のタイプや発症後の月数などの人口統計学的な情報が得られたすべての患者の最初の PNT を分析した (N=255)。対象となったのは、超皮質感覚性失語症、超皮質運動性失語症、脳動脈後遺症、全身性失語症のうち、無表情失語症 103 例、ブローカ失語症 60 例、伝導性失語症 46 例、ウェルニック失語症 35 例、その他の失語症 11 例であった。脳卒中後の月数中央値は 28 (1, 381), PNT 正答率の中央値は 76.4 (1, 99) であった。

### 5.2 計算モデル

先に述べたように、SP は Foygel and Dell (2000) によって初めて発表された。このモデルでは、絵画命名課題をシミュレートするために、語彙検索の相互 2 段階拡散活性化理論を具体化しており、意味記号、語彙記号、音韻記号を表す個々のユニットを持つ 3 層のネットワークで構成されていた (図1)。ユニットの数と接続のパターンは、英語の発話誤りタイプの統計的確率を近似することを目的としており、6 つの単語(そのうち 1 つはターゲット) からなる非常に小さな語彙領域の構造を実装している。このモデルには 6 つの語彙ユニットが含まれ、それぞれが意味的特徴を表す 10 の意味ユニットに接続されている。意味的に関連のある単語は 3 つの意味的ユニットを共有しており、ターゲットと意味的に関連のある単語が 1 つしかない典型的な試行では、ネットワークには合計 57 の意味的ユニットがあることになる。また、各語彙ユニットは、オンセット、母音、コーダに対応する 3 つの音韻ユニットに接続されている。音学的ユニットは、6 つのオンセット、2 つの母音、2 つのコーダの合計 10 個ある。ターゲットに音韻的に関連する単語は、そのオンセットユニットだけが異なり、ネットワークは常にそのような 2 つの単語で構成される。最後に、ネットワーク内の残りの 2 つの単語は、ターゲットとは無関係で、意味的にも音韻的にも共有されるユニットはない。20 % の試行では、音韻的に関連のある単語が意味的にも関連しており、ターゲットと「混合」関係にある隣接語が生まれる。

絵画命名シミュレーションは、意味ユニットに活性化のブーストがかかることから始まる。S と P という 2 つのパラメータは、それぞれ語彙-意味、語彙-音韻の接続の双方向の重みを指定する。活性化は、ノイズと減衰を伴う線形活性化規則に従って、8 つの時間ステップの間、すべての層の間で同時に両方向に広がる。その後、最も活発な語彙単位に 2 度目の活性化が与えられ、さらに 8 時間ステップにわたって活性化が広がり続ける。最後に、最も活性化された音韻のオンセット、母音、コーダの各ユニットがターゲットと比較される出力として選択される。活性化レベルが低下すると、ノイズの影響で生成エラーが発生するが、これは強い接続によって緩和される。応答は、正解、意味錯語、形式錯語、混合錯語、無関連語、新造語に分類される。所与のパラメータ設定に対して、モデルを用いて多くの命名試行を行うことで、これら 6 つの応答タイプの多項分布が推定される。これらの分布は、失語症患者の命名反応から得られる分布と比較することができる。

SLAM は、SP の細部の多くを残しており、アーキテクチャの変更による効果を主に評価するという我々の目的に合致している。意味ユニットと語彙ユニットはそのままだが、音韻的ユニットのコピーが追加されており、一方を聴覚、もう一方を運動としている (図3)。4 つのパラメータは、意味-語彙(SL)、語彙-聴覚(LA)、語彙-運動(LM)、聴覚-運動(AM) の各接続の双方向の重みを指定する。LA と LM の接続は、SP モデルの P 接続と同じで、1 つの語彙ユニットが 3 つの聴覚ユニットと 3 つの運動ユニットに接続されるのに対し、AM の接続は単純に 1 対 1 である。絵画命名課題のシミュレーションは、SP モデルと同様に、意味ユニット、語彙ユニットの順にブーストをかけ、運動ユニット内で音韻選択を行うという 2 段階の方法で行われた。



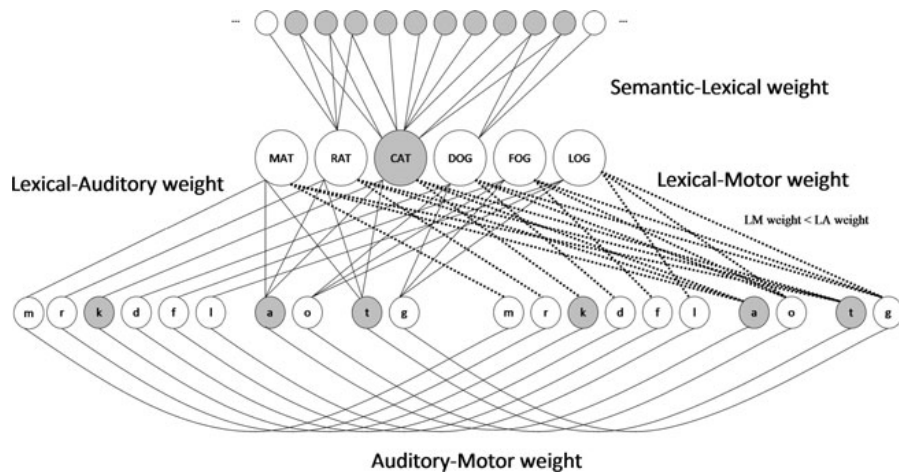


図 3 SLAM (Semantic-lexical-auditory-motor) モデルのアーキテクチャ

### 5.3 データの適合

データに適合させるためには、十分に異なる出力分布をもたらす異なるパラメータセットを用いてモデルを評価し、パラメータ空間からデータ空間へ、またはその逆の有限要素マップを作成する。この処理過程では、まずパラメータ値の集合 (S および P の重みなど) を選択し、次にそのパラメータセットを使用してモデルで多くの命名を試み、特定のモデル設定で発生する 6 種類の反応のそれぞれの頻度を推定する。これらの頻度が決定されると、その重みの設定は、マップと呼ばれるベアリストの出力分布と関連付けられる。マップの各点は、失語症の絵画命名を観察したときに起こりうるエラーパターンの種類についての予測を表している。モデルを評価する 1 つの方法は、その予測が、観察された失語症のエラーパターンにどれだけ近いかを測定することである。観察された分布と、モデルがシミュレートした最も近い分布との距離は、そのデータポイントに対するモデルの適合度と呼ばれる。RMSD (平均自乗偏差 root mean squared deviation) は任意の尺度だが、一般的に使用される適合度の尺度であり、各反応タイプの平均的な偏差と解釈できる。例えば RMSD が 0.02 の場合、観察された比率が予測された比率から平均して 0.02 だけ逸脱していることを示す (例えば、予測 = [50, 50]; 観察 = [48, 52])。このように、RMSD の値が低いほど、モデルの適合性が高いことを示している。すぐに、いくつかのポイントを生成すべきか、また、冗長な予測を生成しないためにどのようにパラメータを選択すべきかという疑問が生じる。

Foygel and Dell (2000) は付録において、パラメータ空間の可変解像度マップを生成するための指針を、アルゴリズムの例とともに示している。彼らは、マッピングアルゴリズムの選択は、包括的な検索を行う限り、適合結果にほとんど影響を与えないだろうと述べている。しかし、マッピングには本質的に高い計算コストがかかるため、特定のアルゴリズムが実際のマップの最大解像度に影響を与える可能性がある。Dell, Lawler, Harris, and Gordon (2004) により、パラメータ空間マッピングの 2 つ目のアルゴリズムが与えられた。これらのマップは、オンラインで入手可能であり、その後の出版物でも使用されているため、SP の標準的なマップと考えられる。この SP マップは、3,782 点、各点に 10,000 個のサンプルがあり、生成には数日間の逐次計算が必要であった。このマップの作成には、数日間の連続計算が必要であった。このように、モデルの開発・検証には、マップ作成のための計算コストが大きな障害となっている。マップに新しいポイントを追加すると、予測が観測に近づく可能性が高くなるが、モデルの新しい予測のセットが絞られていくにつれて、リターンは減少していく。Dell が示唆しているように、目的はベストフィットを見つけることなので、モデルの成績を向上させるためにポイントを増やすことは、おそらく価値のある追求である (G. Dell, personal communication, July 12, 2013)。さらに、SLAM には 2 つのパラメータが追加されているため、より効率的にマップを生成するためには、マッピング手順を変更する必要がある。

我々はマッピングアルゴリズムを再設計して、その固有の並列性を利用することで、効率を大幅に改善した。マッピングアルゴリズムには、主に 2 つの反復ステップがある。点選択と点評価である。パラメータ空間の点の座標は、モデルの可能なパラメータ設定によって定義され (点選択)、データ空間の対応する点は、そのパラメータ設定で生成された応答タイプの割合によって定義される (点の評価)。シミュレーションでは、独立したユニット、独立したサンプル、独立したパラメータセットを用いて計算を行うため、ポイント評価のステップは非常に並列化しやすいものとなっている。しかし、点選択には、並列化のための新しいアプローチ、ドロネーメッシュ精密化 (Delaunay mesh refinement) が必要であった。

ドロネー三角錐とは、点の集合を結ぶグラフで、任意のシンプレックスの円周がその集合の他の点を含まないようなものである。このグラフには、点同士の隣接関係を示す辺など、幾何学的に有利な性質が多くある。新しい点選択アルゴリズムは、この隣接関係を利用している。パラメータ探索範囲の境界に位置する点とそのセントロイドから始まり、パラメータ空間で隣接する 2 点間の距離がデータ空間でのしきい値 (RMSD) を超えた場合、そのパラメータ空間の中点が評価対象として選択される。空間の中点が評価対象として選択され、マップに追加されます。その後、これらの新しい点がドロネーメッシュに追加され、すべてのエッジがしきい値以下になるまで、この処理が繰り返される。このように、点選択アルゴリズムは、各反復において、パラメータ探索範囲全体にわたって複数の点を並列に評価することになる。並列処理には GPU (Graphics Processing Unit) を使用し、さらに効率を高めている (脚注1)。

脚注 1. 筆者らは原稿執筆時点で、任意の次元のドロネー三角錐を段階的に構築するための自由に利用できる並列アルゴリズムを知らなかった。そこで、CUDA C と Thrust ライブラリを用いて点の評価と辺の二等分を実装し、GPU 上で実行するとともに、CGAL ライブラリを用いてコンピュータの中央演算処理装置 (CPU) 上でドロネー三角法を構築した。Nvidia Tesla K20Xm GPU と Intel 1,200 MHz 64-bit CPU を用いて、点の評価を C++ で並列処理した場合と比較したところ、26.0 倍の高速化を実現した。

SP と SLAM の性能を統計的に比較する前に、地図の解像度がモデルの適合性に与える影響を調べた。まず、パラメータ空間の継続的な探索を促すために、0.01 という低い RMSD 閾値を用いて、各モデルに対して非常に高解像度のマップを作成した。各マップには、各ポイントで 10,000 個のサンプルが含まれており、パラメータは [0.0001, 0.04] の範囲で独立して変化していた。パラメータの最大値は、最も高い頻度で正解が得られる最小値に近い値を選択し、値を小さくすると誤りが多くなるようにした。低いマッピング閾値を使用したため、任意の数のポイントを生成した後、アルゴリズムが完了する前に停止した。このアルゴリズムでは、探索範囲全体にわたって効率的にポイントを選択しているため、早期終了はあまり気にならない。この事実により、マップの解像度を下げても、全空間をカバーすることは些細な問題となる。

マッピング手順では、SLAM モデルのパラメータを自由に变化させながら、最初に 31,593 点を生成した後、SLAM アーキテクチャに従って、LM=LA の点をすべて削除し、17,786 点の SLAM マップを生成した。SP マップは 57,011 点となった。次に、大きなマップからサブセットを選択して、各モデルについて 50 個の低解像度マップを作成した。マップのポイント数は 5 から 17,000 まで対数で区切られている。各マップについて、失語症患者全体と、異質な診断群を除いた各診断群の平均適合度を算出した。図 4 にはそのフィット曲線を示した。両方のモデルで予想されたように、ポイントを追加することでフィット感が向上したが、リターンは減少した。相対的なフィットパターンは、図中に縦線で示した 2,321 点付近で安定するようである。この結果は、より高解像度の地図を比較した場合にも当てはまり、解像度が上がるにつれて SLAM が有利になる傾向が見られる。

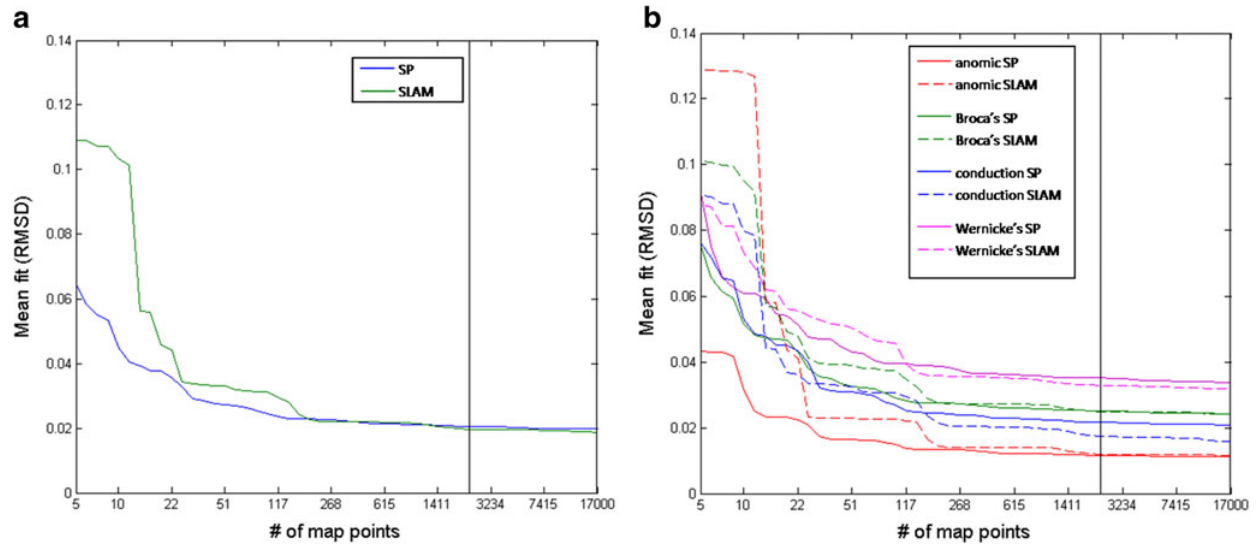


図4 (A) 全患者と (B) 診断群の平均フィット曲線。パネル内の黒い縦線は、統計的比較に使用したマップを示す

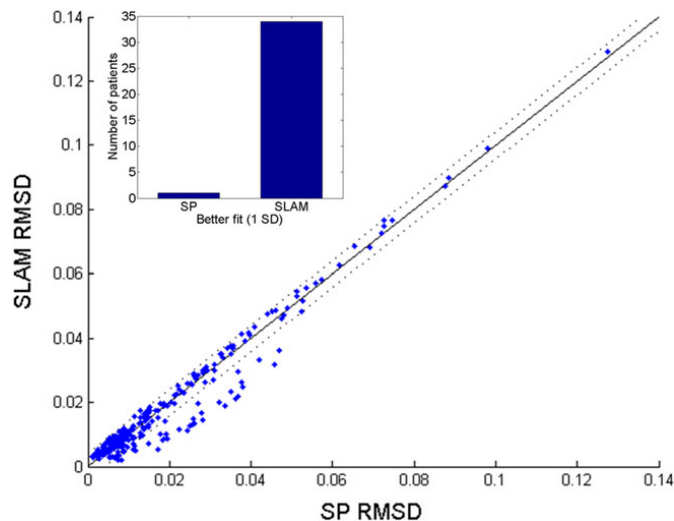
新しい並列生成マップを標準的な 逐次生成マップと比較するために、Schwartz ら (2006) の報告値と平均および最大適合度の点で同様の性能を得られる並列 SP マップの解像度も特定した。この 94 人の患者について、189 点の並列 SP マップを作成したところ、平均 RMSD は 0.0238、最大 RMSD は 0.0785 となり、報告値の 0.024、0.084 と比較しても遜色ない結果となった。予想通り、並列アルゴリズムは直列アルゴリズムよりもはるかに効率的にポイントを選択し、同程度の性能を得るために必要な予測数は少なくて済む。この低解像度のマップをベースラインとして、標準的な SP マップにポイントを追加した場合と、SP の構造を強化した場合の効果を比較した。我々のフィッティング・ルーチンは、これまで研究者がオンラインで利用してきた標準的な SP マップよりも優れたフィット感を得られたため (Dell et al.2004), マップの解像度を調整できるフィッティングルーチンと新しいモデルを次の Web アドレスで提供している <http://cogsci.uci.edu/~alns/webfit.html>

## 6. 結果

まず、一般的な失語症患者の場合、SLAM は少なくとも SP と同様にデータに適合するだろうという仮説を検証した。すべての分析は、MATLAB ソフトウェアパッケージを用いて行った。前述したように、我々は適合度の指標として RMSD を使用することにした (値が低いほど適合度が高いことを意味する)。表 1 は、失語症の 5 つの下位タイプだけでなく、患者の全サンプルのモデルフィットの記述統計を示している。図 5 は、SP と SLAM の適合度を比較した散布図である。実線の斜線は両モデルが同等であるという仮説を表し、点線はサンプルにおけるフィットの 1 標準偏差の違いを表している。どちらのモデルも全体的に非常によくできており、大部分の患者が 0.02 RMSD 以下でクラスタリングされていることがわかる。両モデルは全体的に似たようなフィット感を示す傾向があるが、一部の患者は 1-SD 境界から大きく外れていることも明らかである。図 5 の挿入図は、SP と SLAM のどちらがフィットするか (1 SD 以上) を比較した棒グラフであり、SLAM が一般集団でのフィットを犠牲にすることなく、患者のサブグループに対してより良いフィットを提供することを示している。

**Table 1** Descriptive statistics for SLAM and SP model fits

		SP			SLAM				
		S	P	RMSD	RMSD	SL	LM	LA	AM
<b>All</b> <i>N</i> = 255	Mean	.0241	.0232	.0194	.0187	.0242	.0152	.0300	.0254
	St. Dev.	.0118	.0075	.0189	.0189	.0121	.0059	.0085	.0135
	Median	.0256	.0224	.0127	.0111	.0250	.0151	.0300	.0300
	IQR	[.0170, .0341]	[.0179, .0281]	[.0071, .0264]	[.0067, .0246]	[.0176, .0347]	[.0113, .0200]	[.0225, .0388]	[.0151, .0388]
	Range	[.0001, .0400]	[.0062, .0400]	[.0010, .1273]	[.0019, .1977]	[.0001, .0400]	[.0026, .0300]	[.0010, .0400]	[.0001, .0400]
<b>Anomic</b> <i>N</i> = 103	Mean	.0299	.0274	.0110	.0115	.0299	.0181	.0308	.0296
	St. Dev.	.0081	.0070	.0095	.0095	.0085	.0054	.0080	.0113
	Median	.0296	.0266	.0082	.0085	.0300	.0176	.0325	.0350
	IQR	[.0241, .0376]	[.0219, .0318]	[.0049, .0137]	[.0056, .0141]	[.0250, .0400]	[.0151, .0200]	[.0250, .0400]	[.0204, .0400]
	Range	[.0054, .0400]	[.0106, .0400]	[.0010, .0654]	[.0019, .0685]	[.0063, .0400]	[.0038, .0300]	[.0101, .0400]	[.0001, .0400]
<b>Broca's</b> <i>N</i> = 60	Mean	.0215	.0218	.0238	.0240	.0217	.0143	.0267	.0266
	St. Dev.	.0125	.0071	.0243	.0250	.0128	.0053	.0088	.0144
	Median	.0205	.0202	.0145	.0149	.0200	.0126	.0250	.0325
	IQR	[.0139, .0334]	[.0174, .0262]	[.0076, .0312]	[.0075, .0300]	[.0144, .0313]	[.0101, .0176]	[.0200, .0350]	[.0188, .0400]
	Range	[.0001, .0400]	[.0075, .0400]	[.0012, .1273]	[.0019, .1292]	[.0001, .0400]	[.0026, .0275]	[.0101, .0400]	[.0001, .0400]
<b>Conduction</b> <i>N</i> = 46	Mean	.0245	.0182	.0203	.0157	.0250	.0120	.0323	.0163
	St. Dev.	.0110	.0053	.0153	.0137	.0110	.0048	.0089	.0134
	Median	.0259	.0177	.0175	.0110	.0275	.0126	.0375	.0144
	IQR	[.0020, .0331]	[.0145, .0219]	[.0078, .0282]	[.0063, .0217]	[.0188, .0338]	[.0088, .0138]	[.0250, .0400]	[.0038, .0238]
	Range	[.0001, .0400]	[.0062, .0300]	[.0019, .0720]	[.0028, .0727]	[.0001, .0400]	[.0038, .0275]	[.0101, .0400]	[.0001, .0400]
<b>Wernicke's</b> <i>N</i> = 35	Mean	.0126	.0195	.0332	.0318	.0123	.0115	.0305	.0233
	St. Dev.	.0095	.0059	.0209	.0225	.0096	.0051	.0080	.0130
	Median	.0133	.0193	.0294	.0275	.0126	.0101	.0325	.0250
	IQR	[.0039, .0187]	[.0152, .0248]	[.0155, .0448]	[.0139, .0472]	[.0032, .0185]	[.0076, .0151]	[.0232, .0388]	[.0123, .0350]
	Range	[.0002, .0400]	[.0070, .0327]	[.0042, .0979]	[.0038, .0989]	[.0001, .0400]	[.0038, .0225]	[.0163, .0400]	[.0001, .0400]
<b>Other</b> <i>N</i> = 11	Mean	.0180	.0255	.0275	.0283	.0178	.0173	.0293	.0246
	St. Dev.	.0148	.0073	.0222	.0221	.0148	.0057	.0080	.0123
	Median	.0173	.0257	.0143	.0125	.0176	.0176	.0275	.0275
	IQR	[.0019, .0265]	[.0220, .0301]	[.0087, .0488]	[.0106, .0488]	[.0023, .0250]	[.0151, .0200]	[.0250, .0372]	[.0200, .0319]
	Range	[.0003, .0400]	[.0133, .0400]	[.0049, .0617]	[.0057, .0628]	[.0001, .0400]	[.0076, .0288]	[.0151, .0400]	[.0001, .0400]



**図 5** SPとSLAMのモデルフィットを比較した散布図。斜めの実線は同等のフィット感を表し、点線はサンプルにおけるフィット感の差が1 SDであることを表している。大多数の患者は両モデルともよくフィットしており、一部の患者はSLAMの方が顕著にフィットしている (挿入図)。

次に、SLAMによってモデルの適合性が向上するという仮説を、伝導性失語症に限定して検証した。図6は、個々の患者のモデル間のRMSD差を失語症のタイプ別に表示したもので、差の値が正であれば、SLAMがSPよりも適合度を向上させていることを示している。伝導性失語症では、SLAMモデルが最も大きく一貫した適合性の向上をもたらした。ウェルニッケ失語症でも大部分の適合性がこの新しいモデルの恩恵を受けていることが明らかになった。ウェルニッケ失語症もSLAMによって適合度が向上したことは、HSFC理論と一致する。ウェルニッケ失語症は伝導性失語症と非常に似た神経解剖学的損傷を伴い、急性ウェルニッケ失語症はしばしば伝導性プロファイルに近い状態に回復することから、障害の局所が部分的に共有されていることが示唆される。5つの失語症のサブタイプ間の適合性の改善を統計的に比較するために、RMSDの変化について一元配置分散分析(ANOVA)を行ったところ、グループ平均の間に少なくとも1つの有意差があることが示された( $p < 0.001$ )。フォローアップの多重比較検定では、平均適合度改善の95%信頼区間がウェルニッケ失語症を含む他のグループのそれと重ならなかったことから、伝導グループは他のグループよりもSLAMの恩恵を受けたことが示された。

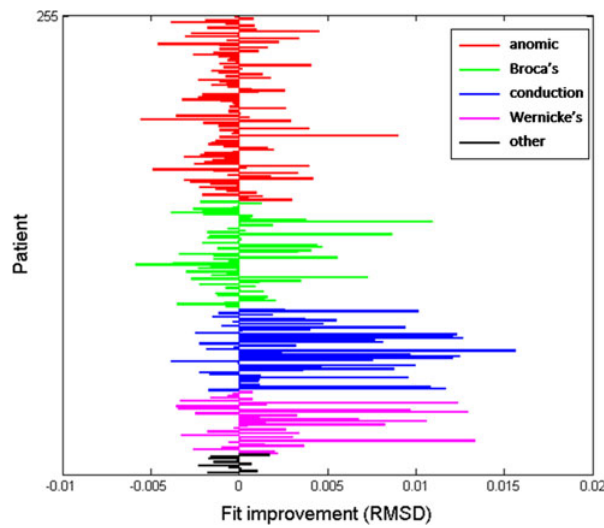


図 6 SP モデルと SLAM モデルの個々のフィット感の変化。正の値は SLAM のフィット感が高いことを示す

これらの結果をさらに検証するために、SP マップの解像度を上げることによる適合度の改善が、特定の診断グループに有利に働くかどうかを検証した。理論的な構造変化とは異なり、このモデルフィットの改善方法は、特定のグループに有利に働くとは考えられなかった。そこで、文献にある標準的な SP マップに相当する 189 点の SP マップと、より高解像度の 2,321 点の SP マップのモデルフィットを比較した。255 人の患者のグループでは、SP マップのポイント数を増やすことで、平均適合度が 0.0230 RMSD から 0.0206 RMSD へと有意に改善した ( $p < 0.001$ )。適合度の改善は、すべての診断群で有意であった (すべての  $ps < 0.001$ )。しかし、追跡的多重比較検定を用いた一元配置分散分析では、伝導群に特に有利な構造変更によって生じた結果とは異なり、どの群も他のすべての群よりも有意に大きな改善を示さなかった (信頼区間が不一致ではなかった)。これは、これらのグループがすでに SP に最も適合していないグループと最も適合しているグループであるという観察結果と一致している。つまり、理論的に SP モデルの構造を操作することで得られる適合度の改善は、他の方法で得られる向上とは質的に異なるということである。

また、伝導失語の命名パターンは、(語彙-聴覚) LA 重み係数が強く、(聴覚-運動) AM 重み係数が弱いという特定の SLAM 構成によって適合するはずだという仮説を立てた。適合度の改善が最も大きかった患者では、実際にそのようになった。図 7 は、最大の適合性改善 (2 SD 以上) を示した 20 人の患者 (伝導失語 13 人、ウェルニッケ失語 5 人、失名辞失語 1 人、ブローカ失語 1 人) の SLAM ウェイト構成を箱ひげ図で示したものである。図 8 は、伝導性失語症の患者のデータと、それに対応する SP モデルと SLAM モデルの適合を示している。SP モデルで最もフィットした重みは、S と P でそれぞれ 0.022 と 0.017 であった。この患者の SLAM モデルでは、(意味-語彙) SL 重み係数と (語彙-運動) LM 重み係数がそれぞれ 0.023 と 0.013 となり、(語彙-聴覚) LA 重み係数は 0.04 で最大となり、(聴覚-運動) AM 重み係数は 0.0001 で最小となった。この患者では、SLAM によって SP の適合誤差が 0.0135 RMSD に減少した。この例では、SLAM の SP に対する最大の適合性の向上は、形式的エラーの予測頻度の一貫した増加と、意味的エラー (および無関連エラー) の一貫した減少を伴うことも示されている。この形式的エラーと意味的エラーのトレードオフは、最初の語彙選択の段階で起こる可能性が高いと考えられる。形式的エラーは、語彙選択時にも音韻選択時にも発生するという二重の性質を持っており、SP モデルの特徴の一つとなっている。Foygel & Dell (2000) は、語彙選択時の形式的エラーは、語彙適合への音韻的フィードバックが意味フィードフォワード活性化を上回ると増加することを示した。伝導失語症では、(語彙-聴覚) LA 重み係数が大きいと語彙ユニットへの音韻フィードバックが強くなり、(聴覚-運動) AM 重み係数と (語彙-運動) LM 重み係数が小さいと運動ユニットへの音韻的フィードフォワードが弱くなる。(語彙-運動) LM 重み係数が (聴覚-運動) AM 重み係数より大きいと、音韻的に関連のある誤った語彙項目からより多くの活性化が流れ、意味的な誤りを犠牲にして形式的エラーが増加する。伝導性失語症では、聴覚-音韻の強いフィードバックが語彙選択に影響を与えるという意味は、我々のモデルの新しい予測であり、データによって支持されている。

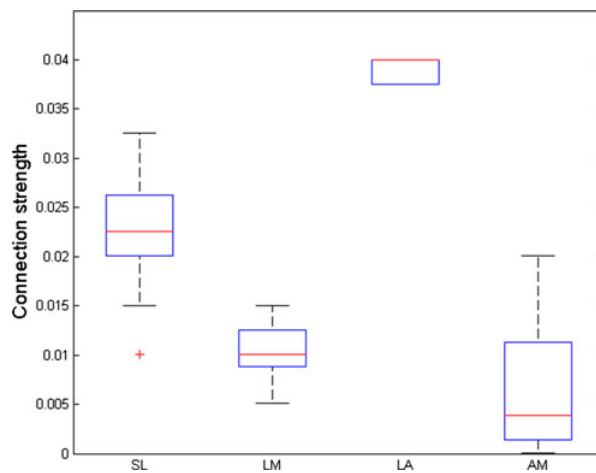


図 7 最も適合度が向上した 20 名の患者グループの SLAM 重みを示す箱ひげ図。予想通り、語彙-聴覚の重みが高く、聴覚-運動の重みが低いモデルプロファイルが、SP モデルに対する最大の改善につながった



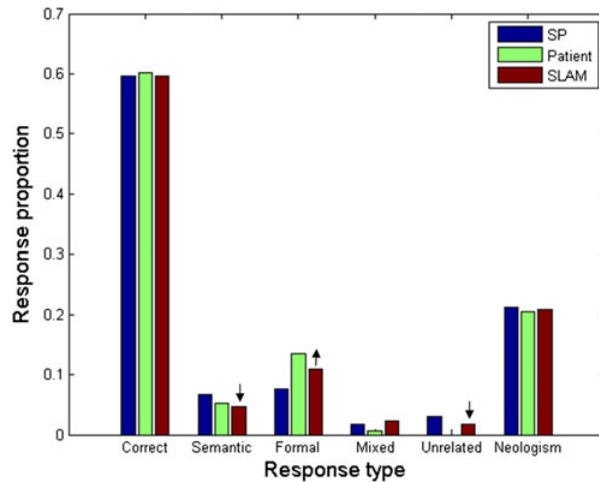


図 8 伝導性失語症の患者の命名反応分布と、それに対応する SP モデルおよび SLAM モデルの適合度。矢印は、SLAM が、意味エラーと無関連エラーを犠牲にして形式エラーを増加させることにより、データへの適合度を改善していることを示している。SLAM モデルは、この患者の適合度エラーを 0.0135 RMSD 減少させた。

最後に、「(語彙-聴覚) LA 重み係数が(語彙-運動) LM 重み係数よりも大きくなければならない」という仮定の重要性を検証した。今回は、SP と、SLAM の代替バージョンである SLMA (語彙・聴覚優位ではなく、語彙・運動優位) を比較した。SLMA は、SLAM マップと同じ大きさの 2,321 点の 4 パラメータマップにフィットさせ、SLAM で廃棄された 13,807 点から切り出し、LM 重み係数が常に LA 重み係数以上になるようにした。図 9 は、SP モデルと SLMA モデルの適合性を比較した散布図で、対角線は図 5 と同じである。このモデルアーキテクチャを使用した場合、SP に比べて顕著な改善は見られず、適合度の最大変化はわずか 0.0038 RMSD であった。このように SLAM では単にパラメータを追加しただけでは適合度の改善は見られず、理論的な根拠に基づいたパラメータの配置が必要であった。

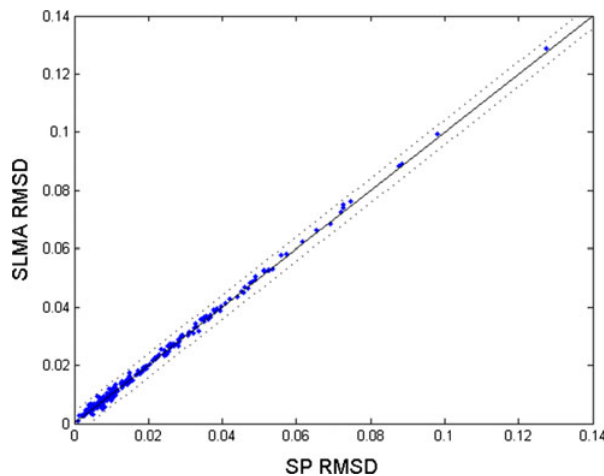


図 9 SP と、SLAM と同じ数のパラメータを持ち、代わりに語彙運動優位のアーキテクチャである SLMA (semantic-lexical-motor-auditory) モデルとのモデル適合性を比較した散布図。線は図 5 と同じ。SLAM とは異なり、SLMA では SP に対する明らかなフィット感の向上は見られない

また、(語彙-運動) LM 重み係数の必要性を探り、我々の 2 経路の重要性を検証した。SLAM マップの 323 点を用いて LM 重み係数を 0.0001 (実質的にゼロ) に固定してデータを適合させ、3 パラメータモデルを作成し、同じ点数の SP マップの適合度と比較した。直接的な LM 結合を持たないこの 3 パラメータモデルは、2 パラメータの SP モデルよりもはるかに悪い結果となり、平均適合度は 0.10 RMSD となった。この壊滅的な失敗は、語彙-聴覚-運動のルートを経由して運動ユニットに十分な活性化が到達しなかったことに起因する。活性化は各レベルで分数を乗じて行われるため、語彙-聴覚-運動のルートを経過した後の活性化は、語彙-運動のルートを経過した後の活性化よりも低くなることを思い出してほしい。2 つの経路からの運動ユニットへの入力組み合わせられていない場合、モデルは最大で 65 % の正答率の推定値しか得ることができなかった。HSFC 理論では、通常の正解率を得るためには語彙-運動間の直接的な接続が必要であるとされているが、聴覚-運動経路からの運動ユニットへの入力弱いことから、我々が最初に選択した SLAM パラメータ制約は、HSFC 理論が保証する以上に語彙-運動経路を重視しているのではないかと懸念が生じた。そこで我々は、SLAM のパラメータ空間をさらに調査し、質的に似た結果をもたらす別のパラメータ制約を発見した。健常者モデルでは、SL と LA の重みは通常の最大値である 0.04、LM の重みは最大値である 0.02、AM の重みは最大値である 0.5 であるが、失語症ではパラメータはこれらの値よりも自由に变化する。このようなパラメータ配置により、語彙層への音韻フィードバックの主な供給源が通常は聴覚ユニットからのものであることを保証し、聴覚-運動経路が命名時に運動ユニットに強い活性化を与えることを可能にし、損傷状態では LM 係数が LA 係数よりも常に低くなければならないという以前の制約を取り除くことができた。当初の SLAM パラメータ制約の選択と同様に、一般人の SP と同様の適合性が観察され (図 S1)、LA の重み係数が高く、AM の重み係数が低いことに伴い、伝導失語の命名パターンに顕著な改善が見られた (図 S2)。このように、LM 重み係数を 0.0001 に固定した 3 パラメータモデルは、2 パラメータ SP モデルほどの性能は得られなかったが (図 S3)、(聴覚-運動) AM 重み係数を強くすることで補正され、致命的な失敗ではなくなった。以上の結果から、HSFC 理論で予測されている聴覚系からの第 2 の音韻フィードバック源が、モデルの適合性を向上させるための重要な要素であるという我々の主な発見が確認された。

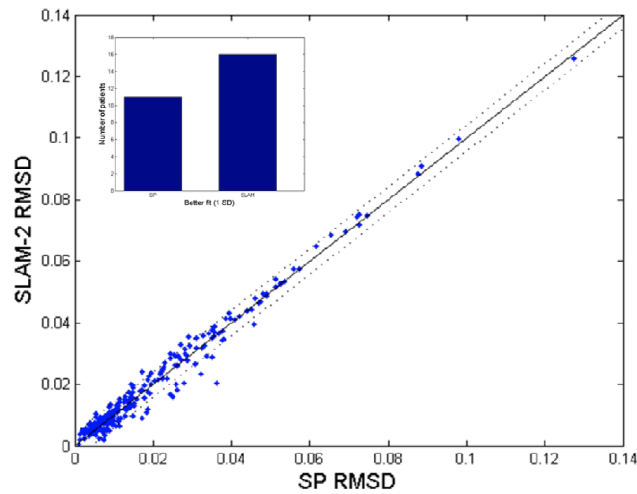


図 S1 SP と SLAM-2 のモデルフィットを比較した散布図。我々が改善した適合度が LM 係数に依存しすぎていないかどうかを検証している。対角線は図 5 と同じ。今回も全体的に適合しており、一部の患者では SLAM の方が SP よりも優れていることがわかる。

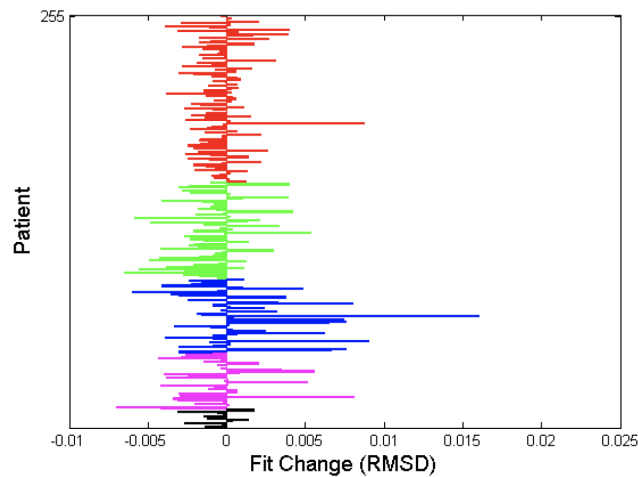


図 S2 SP モデルと SLAM-2 モデル間の個々の適合度の変化。正の値は SLAM-2 の適合度が良いことを示す。失名辞失語=赤、ブローカ失語=緑、伝導失語=青、ウィルニッケ失語=マゼンタ、他=黒

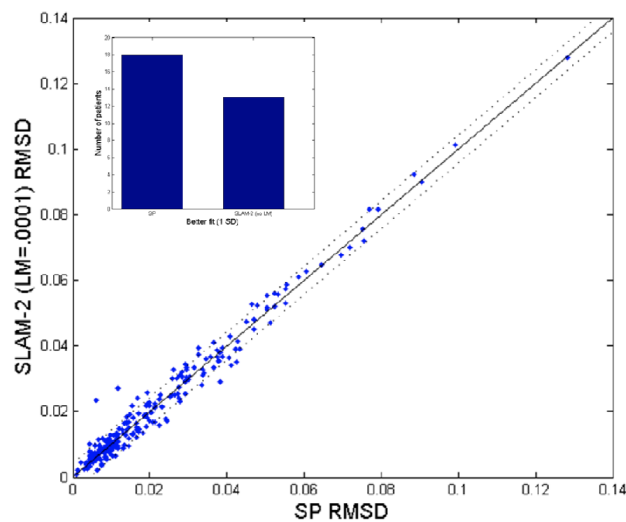


図 S3 LM 経路を削除した SP と SLAM-2 のモデル適合度を比較した散布図。対角線は図5と同じ。この 3 パラメータモデルは、2 パラメータの SP モデルよりも性能が劣るため、4 パラメータモデルのような改善は見られない。

## 議論

本研究では、言語心理学的モデルと運動制御モデルの統合を目的とした最近の概念モデル HSFC にヒントを得て、命名課題の新しい計算モデル SLAM を提唱した。SLAM は、HSFC の主張を実装したものであり、下位語彙の言語単位が聴覚皮質と運動皮質で二重に表現されていること、また、聴覚ターゲットを運動コマンドに変換することが、語彙の検索に重要な計算であることを、オバートな表出前であっても説明している。

我々は、定評のある SP モデルに聴覚から運動への変換を語彙検索処理過程に組み込むことで、このモデルが一般的な失語症の命名ミスを元の SP モデルと同等に説明できることを示した。特に伝導失語症の命名パターンを説明する能力が向上した。このようなモデルの適合性の向上は、(語彙-聴覚) LA 重み係数を強く、(聴覚-運動) AM 重み係数を弱くしたパラメータ設定によってもたらされると予測された。255 名の失語症患者の命名反応を調べたところ、これまでで最大規模の PNT 反応の分析結果となり、我々の予測を確

認することができた。さらに、理論的に導出された構造的変化とは異なり、マップの解像度の追加による改善は失語症のタイプに特異的なものではないことも明らかになった。また、予測された重み係数設定は、意味エラーを犠牲にして形式エラーを増加させることにより、最大の適合性の向上をもたらすことがわかった。Schwartz ら (2006) は、命名パターンが SP の予測から著しく逸脱した 3 つの異常な下位グループを特定した。この下位グループの患者のうち 2 人は伝導性失語症、もう 1 人はウェルニッケ失語であった。SLAM は、このサブグループのもっともらしい説明となる。伝導性失語症では、意味性エラーを犠牲にして形式性エラーが増加していることから、音韻に関連した誤りのかかなりの割合が、音韻選択段階ではなく語彙選択段階で発生していると考えられた。また、この効果を生み出すためには、2 つの異なる音韻経路が必要であることもわかった。SLAM では、HSFC 理論で説明されている聴覚運動統合ループは詳細にモデル化されていないが、状態フィードバック制御を行うためには、別々の聴覚運動系への並列入力とフィードバックが必要である。今回のモデリング実験の結果は、HSFC 理論の前提を裏付けるものである。

SP と SLAM を対決させてみたが、両者には共通の特徴がある。したがって、SLAM の成功の多くは、オリジナルの SP モデルの前提条件に起因している。計算の同時性、階層的な表現、階層間の相互作用、局所的な損傷、ランダムな出力と整形された出力の連続性などの概念が、優れた予測を可能にした。今回、モデルの拡張に成功したことで、これらのアイデアの有用性が確認できた。同様に、SP に対する批判の多くは SLAM にも同様に当てはまる。例えば、非常に小さい語彙は実際の語彙の構造に近似しているに過ぎず、意味表現は任意に定義されている。また、このモデルは相互的ではあるが、実際の神経系で必須の機能である横方向の接続や抑制性の接続が含まれていない。また、このモデルでは、音声処理に関する心理言語学的証拠の大部分を占める時間情報を直接扱っていない。とはいえ、HSFC のアーキテクチャ上の前提条件を検証するためには、SP は最良の計算モデルであるという意味で、有用な検証のための温床となった。

SLAM が SP (および統一された音韻ネットワークを想定した類似モデル) よりも優れている点は、SLAM には復唱のための機構が組み込まれていることである。復唱は、命名に加えて、語彙検索モデルの検証としてよく用いられる。復唱は、命名と同じように運動産出システムを必要とするが、意味検索の要素を持たないからである。しかし、復唱をシミュレートするためには、たとえ暗黙的であっても、何らかの聴覚表現が必要である。Foygel and Dell (2000) では、単一経路 SP モデルを用いて、聴覚入力を明示的にモデル化することなく、完全な聴覚認識が語彙ユニットに直接ブーストを与えると仮定して、基本的には命名の第 2 段階に過ぎない復唱をシミュレートした。その後、命名は不得手だが復唱能力には問題がない患者を説明するために、入力から出力への直接音韻ルートをモデルに追加した (Hanley, Dell, Kay, & Baron, 2004)。この二重ルートモデルは「非語彙」ルートを SP に移植したもので、命名のアーキテクチャやシミュレーションは変更せず、2 つの経路は反復時にのみ使用される。しかし、今回の研究では、2 つの経路が命名時にも使われていることが示唆されたため、これらの課題の基礎となる計算をより包括的に説明できる可能性がある。SLAM ではすでに命名に聴覚要素が必要とされているため、将来的には復唱をシミュレートできるように開発し、この代替双経路モデルとの直接的な比較を可能にしたいと考えている。

SLAM は学習や時間変化のある表現を採用していないが、これらの特徴を実装した別の語彙検索モデルでも、聴覚と運動の音声表現を同様に分離する方法を採用している。Ueno, Saito, Rogers, and Ralph (2011) は、健常者と失語症者の音声処理における命名、復唱、理解を、各層のユニットが脳領域に対応するネットワークアーキテクチャを用いてシミュレートする「神経計算論」モデルである Lichtheim 2 を発表した。しかし、Lichtheim 2 は、SP のより詳細な分類法に基づいて発話エラーのタイプを分類していないため、SLAM と直接比較することはできない。また、SLAM は音韻表現の分離による効果を調べることを目的としており、Lichtheim 2 もこのアーキテクチャの前提を共有しているため、直接比較することはできなかった。Lichtheim 2 では、入力と出力の音韻は、1 クラスタずつ提示される音韻特徴のパターンで表現され、意味表現は時間的に静的で、対応する音韻表現から統計的に独立している。このモデルは、3 つの課題で同時に学習され、中間層表現はほとんど制約を受けずに形成される。学習したネットワークを特定の領域を「病変」させることで、失語症の成績をシミュレートすることができる。神経解剖学的な証拠に基づいて心理言語表現を提案するという理論的動機の点で、我々のアプローチには多くの共通点がある。さらに、1 つのネットワークで複数課題を実行するという点では、SLAM モデルの開発計画と一致している。SLAM と Lichtheim 2 の大きな違いは、SLAM では語彙ユニットと音韻ユニットを明確に階層的に分離し、どちらの段階でも選択ミスを許容していることである。この階層的な分離は、伝導失語の命名パターンに関する予測を成功させるために不可欠であった。今回提案したアーキテクチャが、複数課題を同時に処理できるかどうかは、今後の課題である。

Lichtheim 2 の代替モデルとして、WEAVER++/ARC (Roelofs, 2014) という語彙生成モデルも提案されている。このモデルでは、SP と同様に小さな固定されたネットワークを介した拡散活性化を用いているが、ネットワークの表現のうち課題に関連した選択を仲介するために条件作用規則を採用し、宣言的知識と手続き的知識の分離を実現している。このモデルは Lichtheim 2 と同様、SLAM で検討された詳細なエラー分類法を適用していないため、直接比較することはできなかった。しかし、WEAVER++/ARC と Lichtheim 2 は、ほとんどの認知的・計算的問題、特に SLAM が調査した主要な問題である、音声生成に聴覚と運動音韻の別々のネットワークが参加していることについて、ほぼ一致している。また、WEAVER++/ARC は、SLAM や Lichtheim 2 と同様に、入力音素と出力音素の間の重みを減らすことで、伝導失語症パターンをシミュレートしている。WEAVER++/ARC と Lichtheim 2 の主な相違点は、解剖学的なものである。発音のための語彙と運動の接続は、(背側) 弧状筋節と(腹側) 無縁筋節のどちらに関連すべきか？現在のところ、SLAM モデルはどちらの位置にも対応しています (脚注2)。しかし、WEAVER++/ARC は、1 つの重要な理論的に関して SLAM とは異なる。WEAVER++/ARC では、入力と出力の語彙ユニットが分離されており、命名時、活性化は主に語彙出力ユニットから運動ユニットへと流れる。そして聴覚ユニットは、運動ユニットと並行して単一の語彙層が感覚ターゲットとして活性化されるのではなく、聴覚のフィードバックループ (すなわち、運動→聴覚→運動) を通じて運動ユニットに安定した活性化を与えるのである。このことは、伝導性失語症において、語彙-聴覚の強いフィードバックが語彙選択に影響を与えているという我々の発見とは相反する。繰り返しになるが、単一の語彙層という仮定が、Lichtheim 2 や WEAVER++/ARC のように複数課題を説明できるかどうかは、今後の研究で明らかにしたい。

SLAM モデルは、「二重経路」モデルと呼ばれる幅広いクラスのモデルに分類される。つまり、行動を制御する処理の流れは別々だが相互に作用するとするモデルである。この研究の多くは、Hickok と Poeppel (2000, 2004, 2007) の音声処理に関する神経解剖学的な二重経路の枠組みに直接関連している。Hickok と Poeppel は、この皮質ネットワークを、2 つの流れに分岐する聴覚音声システムの観点から議論したが、絵画命名は、概念→語彙→聴覚 (腹側経路)、聴覚→運動 (背側経路) と、両経路を横断する。SLAM モデルと Hickok and Poeppel の枠組みとの違いは、語彙ネットワークと運動音韻ネットワークの間に明示的な接続性を仮定していることである。Hickok と Poeppel は、概念系と運動系の間の接続性を想定していたが (議論はしていない)、語彙系から運動系への音声ネットワークの可能性については特に考慮していなかった。今回のモデルは、HSFC とともに、Hickok と Poeppel の二重経路の枠組みを改良したものである。