

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

Grant M. Walker¹ · Gregory Hickok¹

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

¹ 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

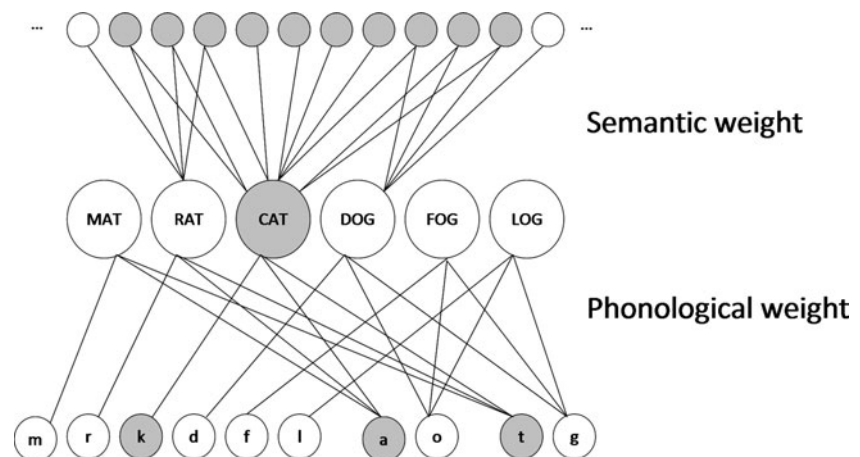
other tasks, as well, such as word repetition (Dell, Martin, & Schwartz, 2007), and to predict the location of neurological damage seen in clinical imaging (Dell, Schwartz, Nozari, Faseyitan, & Branch Coslett, 2013), although here we will focus primarily on its relevance to picture-naming errors.

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

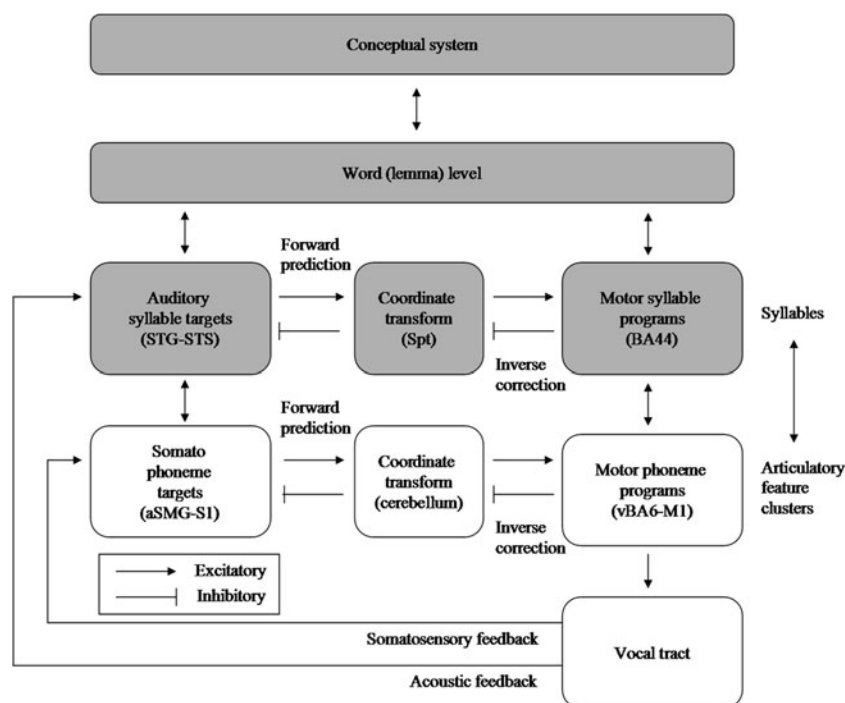


Fig. 2 A schematic diagram of the hierarchical state feedback control (HSFC) framework (Hickok, 2012)

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

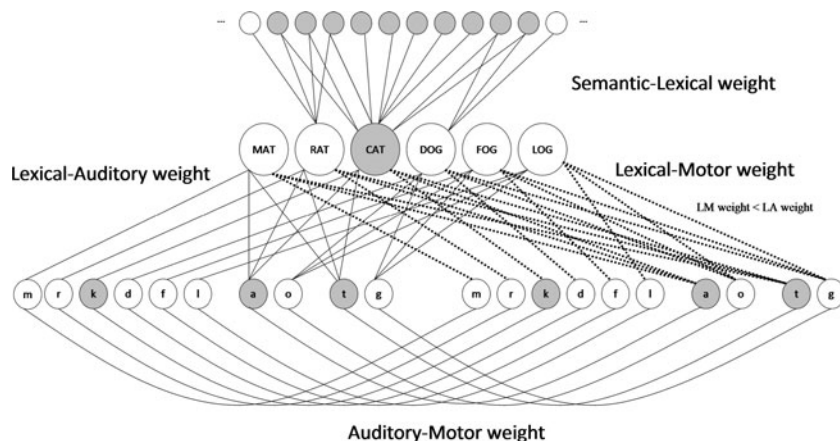


Fig. 3 The semantic–lexical–auditory–motor (SLAM) model architecture

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). 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.¹

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

¹ 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.

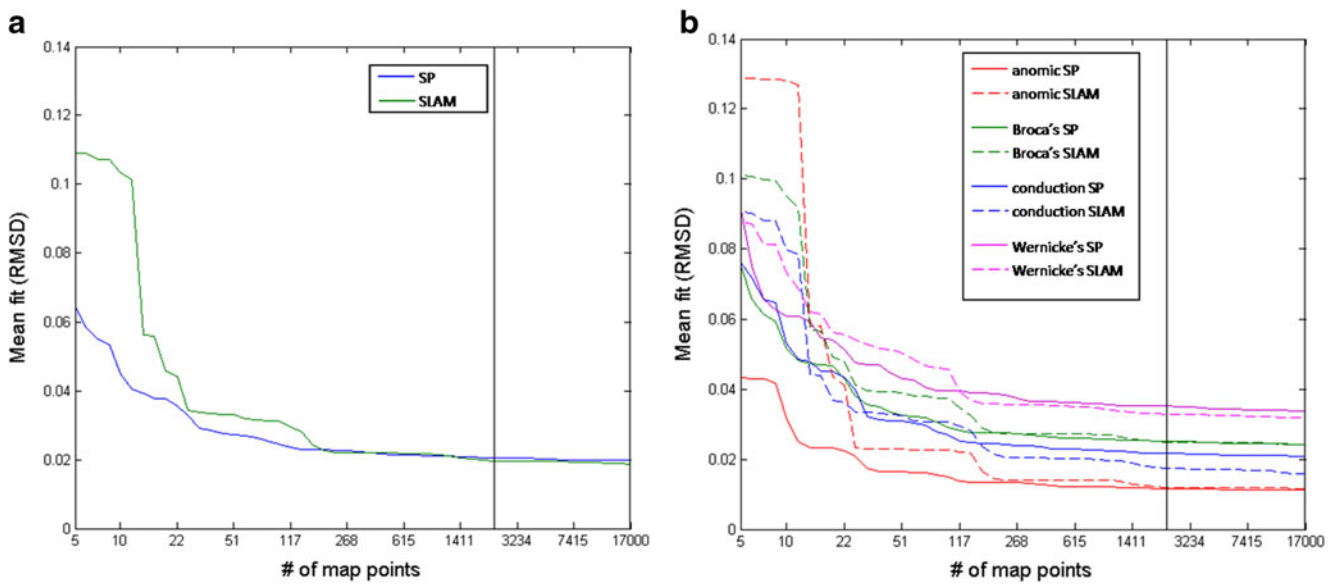


Fig. 4 Mean fit curves for (A) all patients and (B) diagnosis groups. The black vertical lines in the panels indicate the maps that were used for statistical comparisons

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
All <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]
Anomic <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]
Broca's <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]
Conduction <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]
Wernicke's <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]
Other <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

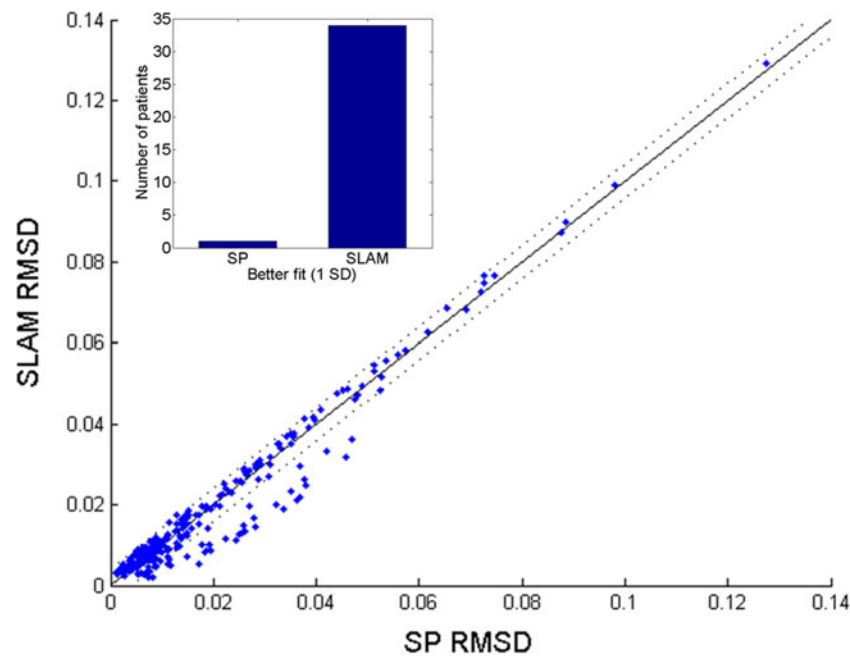


Fig. 5 Scatterplot comparing model fits between SP and SLAM. The solid diagonal line represents equivalent fits; the dotted lines represent 1 *SD* of fit difference in the sample. The majority of patients are fit well by both models, and a subgroup of patients are fit notably better by SLAM (inset)

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

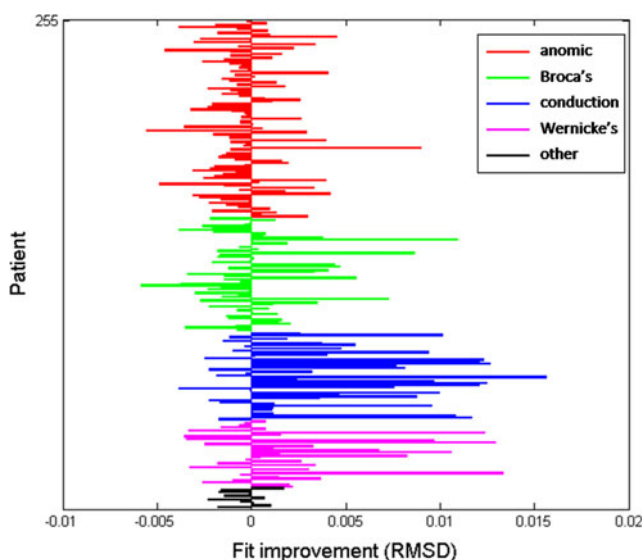


Fig. 6 Individual fit changes between the SP and SLAM models. Positive values indicate better SLAM fits

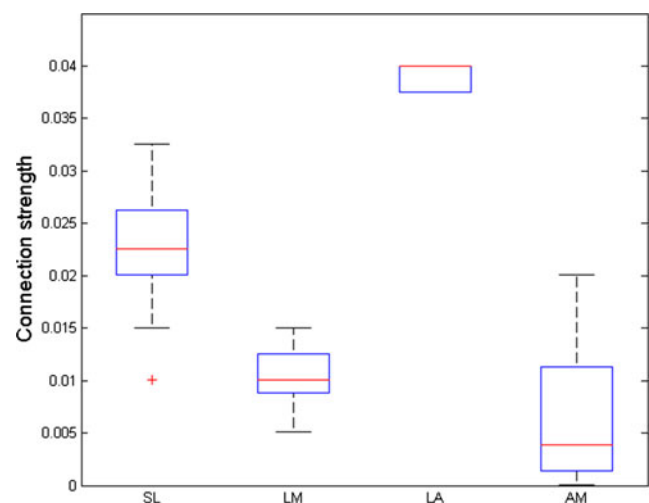


Fig. 7 Boxplots showing the SLAM weights for the group of 20 patients with the greatest fit improvements. As expected, a model profile with high lexical–auditory and low auditory–motor weights leads to the greatest improvements over the SP model

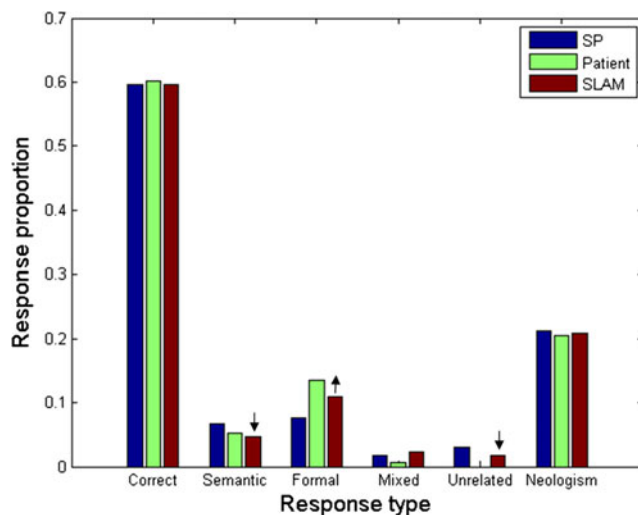


Fig. 8 Naming response distribution from an example patient with conduction aphasia, along with the corresponding SP and SLAM model fits. Arrows indicate how SLAM improves the fit to data, by increasing formal at the expense of semantic and unrelated errors. The SLAM model reduced the fit error for this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

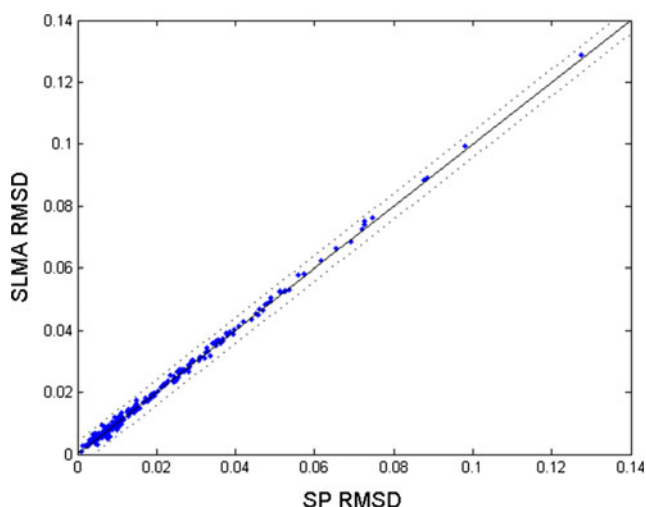


Fig. 9 Scatterplot comparing model fits between SP and the semantic–lexical–motor–auditory (SLMA) model, an alternative architecture with the same number of parameters as SLAM, but with lexical–motor dominance instead. The lines are the same as in Fig. 5. Unlike SLAM, SLMA provides no obvious fit improvements 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.² *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

² 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 = .02¹, 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.

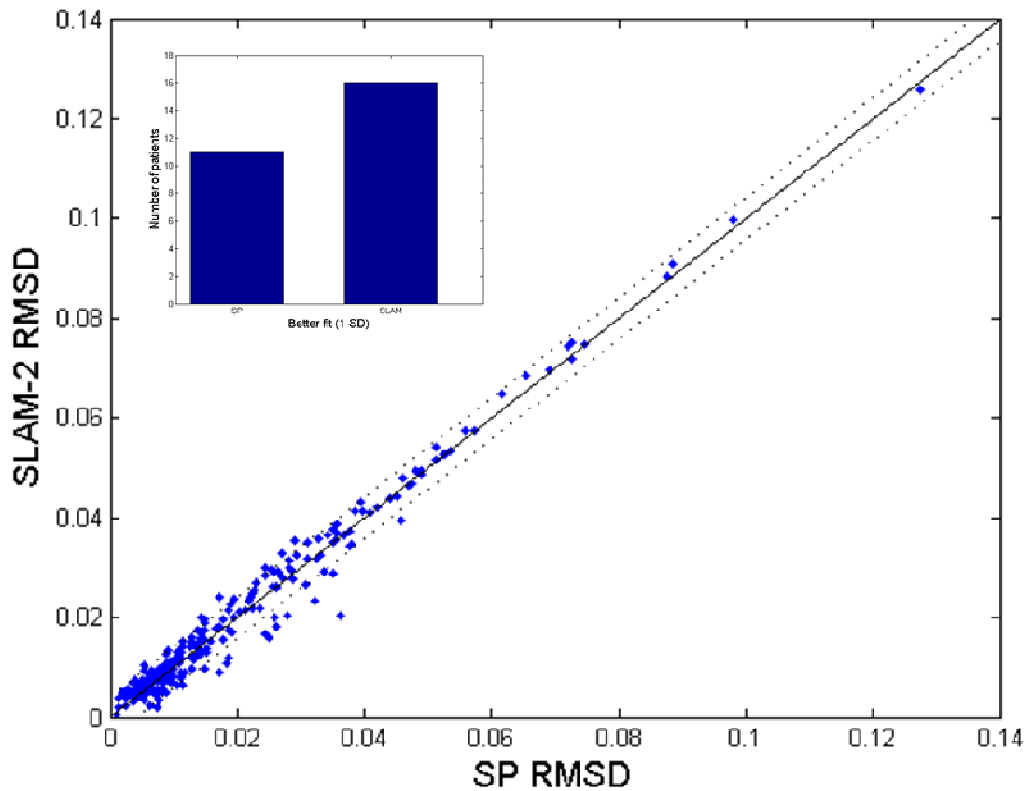


Figure S1. Scatterplot comparing model fits for SP and SLAM-2, testing whether our improved fits relied too much on LM weights. Diagonal lines are the same as in Figure 5. Once again, results show good fits overall, and SLAM outperforms SP for a subgroup of patients.

¹ 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.

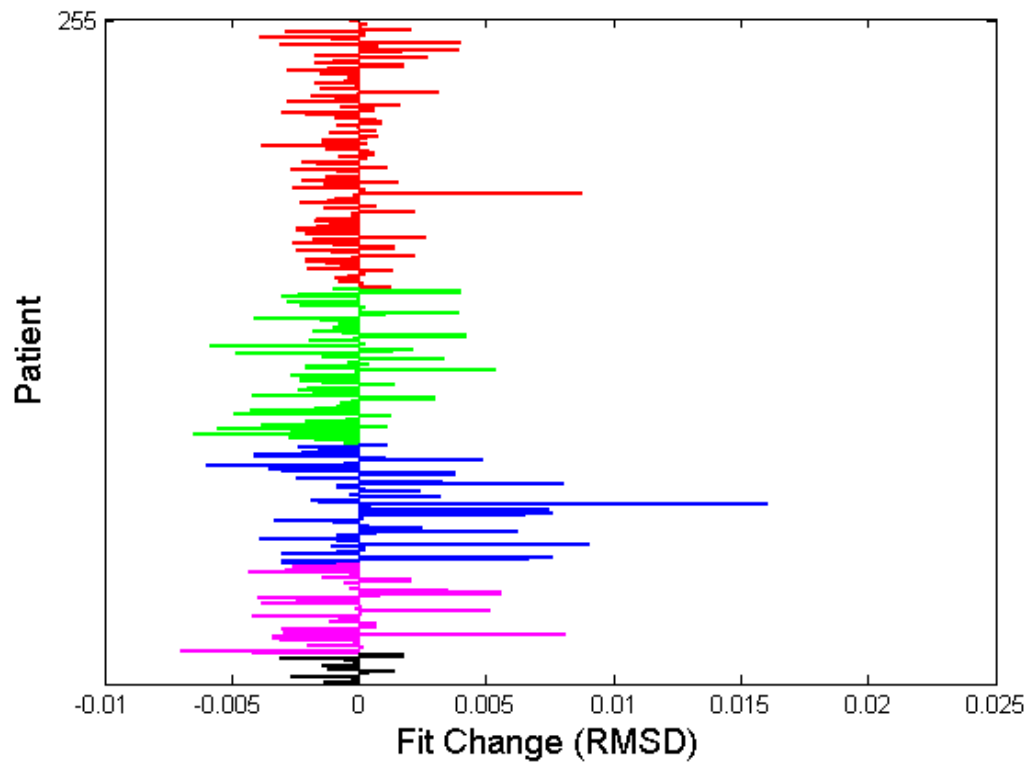


Figure S2. Individual fit changes between the SP and SLAM-2 models. Positive values indicate better SLAM-2 fits. Anomic = red, Broca's = green, Conduction = blue, Wernicke's = magenta, Other = black.

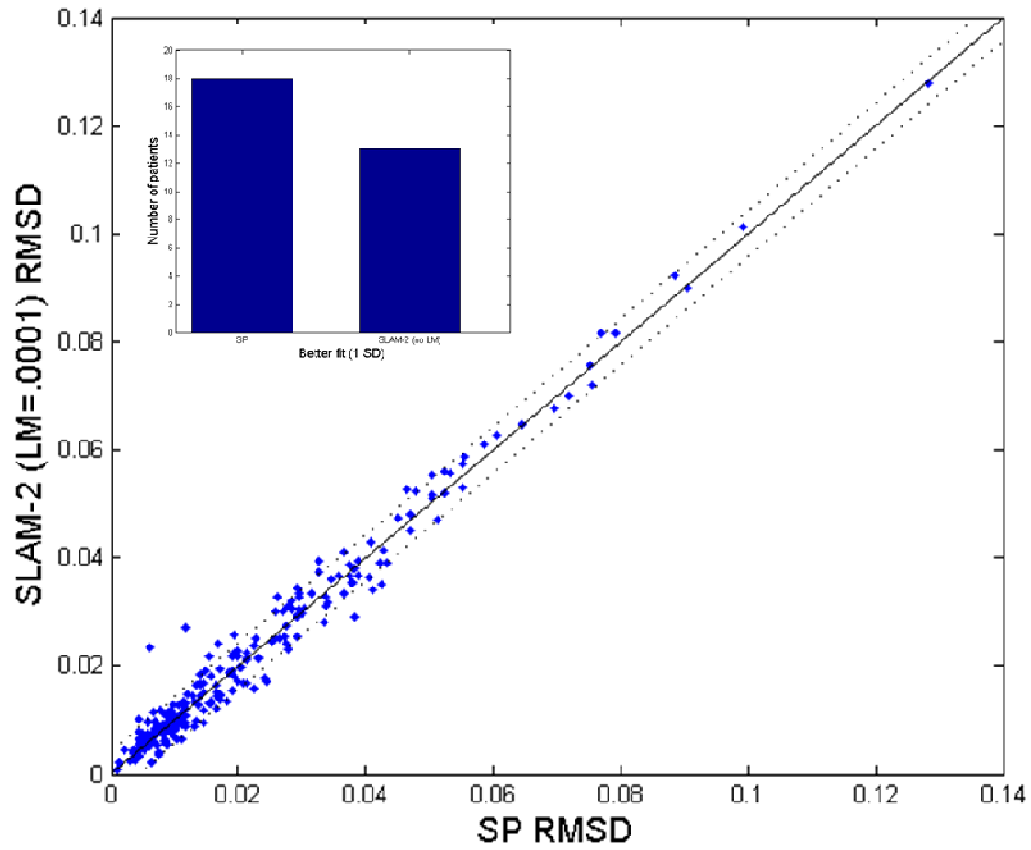


Figure S3. Scatterplot comparing model fits for SP and SLAM-2 with the LM route removed. Diagonal lines are the same as in Figure 5. This 3-parameter model does not perform as well as the 2-parameter SP model, and thus it does not yield the improvements seen with the 4-parameter model.

author: Grant M. Walker and Gregory Hickok year: 2015 (2016) journal: Psychonomic Bulletin Review volume: 23 pages: 339–352 title: "音声生成への計算機的アプローチの架け橋: 意味・語彙・聴覚・発話・運動モデル (SLAM)"

音声生成への計算機的アプローチの架け橋: 意味・語彙・聴覚・発話・運動モデル (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。

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て使用するなど、運動制御からの計算原理をより高レベルの言語処理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) の意味-音韻モデル (SP) の動機を説明し、
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、新しい意味-語彙-聴覚-運動モデル (SLAM) を作成することを説明する。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下位カテゴリーに関しては、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 (The SP model)

SP は [Dell, 1986] の文生成における検索理論にルーツを持つ。SP は自然な音声の大規模なコレクションに見られる発話誤りや舌足らず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、複数の内部表現が生成前の選択を競う計算的同時性の概念を採用し、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との間の各レベルでの分離を取り入れた。

[Dell, 1997] は、単一語の発話に限定した計算モデルを提案した。

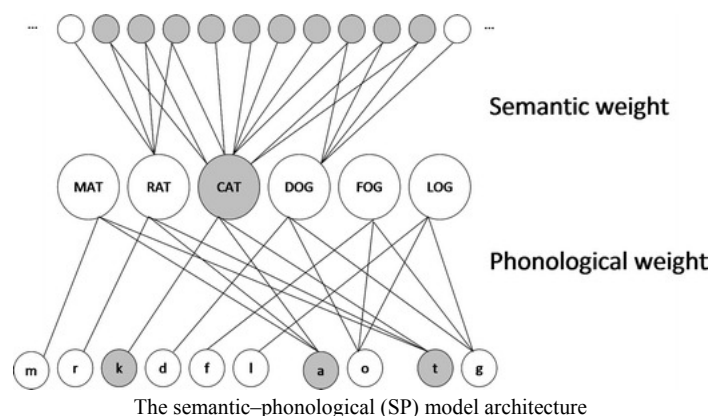
だが、理論的な範囲を拡大し、失語症の文脈における発話エラーの説明を含めるようにした。

基本的な考え方は、失語症の音声エラーのパターンは、損傷した音声生産システムの出力を反映しているもので、失語症のデータに適合するよう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だった。このモデルのアーキテクチャは、意味、語彙、音韻の3つのユニットからなる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、典型的な語彙近傍の構造に近似するように実験者によって選択された (図 1)。

単語の生成は、ノイズと時間の経過に伴う活性化の減衰を伴う拡散活性化過程としてモデル化した。

損傷は、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ことによって実装された。

次に、シミュレーションを用いて、失語症患者が行うエラータイプと類似したエラータイプの頻度を発生させるパラメータ値を特定した。



しかし、シミュレーション手法の計算量が多いという性質上、包括的な探索は事実上、一度に2つのパラメータのみに制限されていた。それにもかかわらず、[@2000Foygell_Dell_SP]から始まる一連の論文では、臨床診断情報を含む絵画命名課題から得られるデータの驚くほど多様性を説明するモデル内の2つの自由パラメータが特定された(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[@2006Schwartz_Dell_SP], 回復の特徴的なパターン(Schwartz & Brecher, 2000), および相互的エラー効果[@2000Foygell_Dell_SP]である。これら2つの自由パラメータは、意味表現と語彙表現の間の接続強度(s-weight)と、語彙表現と音韻律表現の間の接続強度(p-weight)であり、SPとして知られているアーキテクチャである。

SPは、単語反復(Dell, Martin, & Schwartz, 2007)や、臨床画像で見られる神経学的損傷の位置を予測するためにも使用されてきたが(Dell, Schwartz, Nozari, Faseyitan, & Branch Coslett, 2013), ここでは主に絵画命名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発生する計算に関連する。このモデルの出力は抽象的な音素系列であり、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仮定される。

次に、運動効果器が実際にどのように制御されているかについての研究から生まれた基本的な構成に目を向ける。

2. 運動制御理論 (Motor-control theory)

最も広いレベルでは、運動制御は初期計画およびフィードバック制御のための運動系への感覚的な入力が必要とする。それは、運動行為の目標(例えば、特定の大きさと向きのカップ、身体に相対する特定の位置にあるカップ)を定義し、効果器の現在の状態に関する情報(例えば、カップに対する手の位置と速度)を提供するための計画のための入力が必要とする。感覚情報がなければ、行動は不可能であることは、自然界(Cole & Sedgwick, 1992; Sanes, Mauritz, Evarts, Dalakas, & Chu, 1984)や実験(Bossom, 1974)での感覚脱失の例で実証されている。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、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系に順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。

音声の場合、最も重要な感覚目標は聴覚である(Guenther, Hampson, & Johnson, 1998; Perkell, 2012)が、体性感覚情報も重要な役割を果たしている(Tremblay, Shiller, & Ostry, 2003)。聴覚フィードバックの変化は発話に劇的な影響を与えることが示されており(Houde & Jordan, 1998; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Yates, 1963), 発話者の発話環境の変化は「ジェスチャー・ドリフト」、つまり発声パターンの変化(アクセント; Sancier & Fowler, 1997)を引き起こす可能性がある。さらに、秘密の発話を調査した神経イメージング実験では、側頭葉の聴覚関連皮質の活性化が増加していることが一貫して報告されている(Callan et al., 2006; Hickok & Buchsbaum, 2003; Okada & Hickok, 2006)。

発話生成における聴覚系の役割については、言語についての神経心理学的研究から特に関連性の高い証拠が得られている。神経損傷によって言語処理能力が損なわれた場合と損なわれていない場合の顕著なパターンがあることから、理論家は脳内の音声表現を聴覚系と運動系に分けて考えるようになった(Caramazza, 1991; Jacquemot, Dupoux, & Bachoud-Levi, 2007; Pulvermuller, 1996; Wernicke, 1874/1969)。例えば、伝導失語患者(Goodglass, 1992)は、流暢な発話をしており、運動表現が保存されていることを示唆している。また、これらの患者は聴覚理解力に優れており、自分の誤りを認識することができ、聴覚表現が保存されていることを示唆している。このような能力にもかかわらず、多くの音素性の誤りがあり、非単語の繰り返しに問題がある。このパターンは、一般に、分離した聴覚系と運動系の界面の損傷の結果として説明されている(Andersonら 1999; Geschwind, 1965; Hickok 2012; Hickok ら 2000)。伝導失語症に関するこの点は、以下に述べるように、重要な理論的意味合いを持つ。

3. 概念的統合 (Conceptional integration)

階層的状態フィードバック制御(HSFC[@2012Hickok_HSFC])モデルは心理言語学的概念と生物学的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。この概念的枠組みは3つの中心原理を中心に構成されている。

1. 第1に、音声表現には感覚皮質と運動皮質の相補的なエンコーディングが損税する。それらは音声生成中に(少なくとも)音節のレベルに至るまで並行して活性化される
2. 第2は、感覚野と運動野の間の興奮性と抑制性の接続の特定のパターンが、感覚運動翻訳領域によって媒介され、運動計画/活性化の初期段階で運動プログラムの選択に誤りが生じる可能性があるにもかかわらず、運動表現を感覚目標に向かって頑健に導くことができる一種の前方/後方モデルを実装しているということである。
3. 第3の原則は、音声生成をサポートする感覚運動ネットワークが階層的に組織化されていることであり、体性感覚野は音素(より正確には、体性感覚状態として符号化される二唇閉鎖などの音韻レベルの目標)の順に小さな単位を処理し、聴覚野は音節の順に大きな単位(すなわち音響目標)を処理する。

HSFC フレームワークの模式図を図ref{fig2}に示す。一番上の部分(色が濃い)はSPの2つのステップを体現している。音韻論的な要素が聴覚-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されていることがわかる。この概念的な重複により第一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することになった。我々はHSFCモデルのアーキテクチャ上の仮定を、確立された成功した命名の計算モデルであるSPと統合することで、部分的に評価できると推論した。

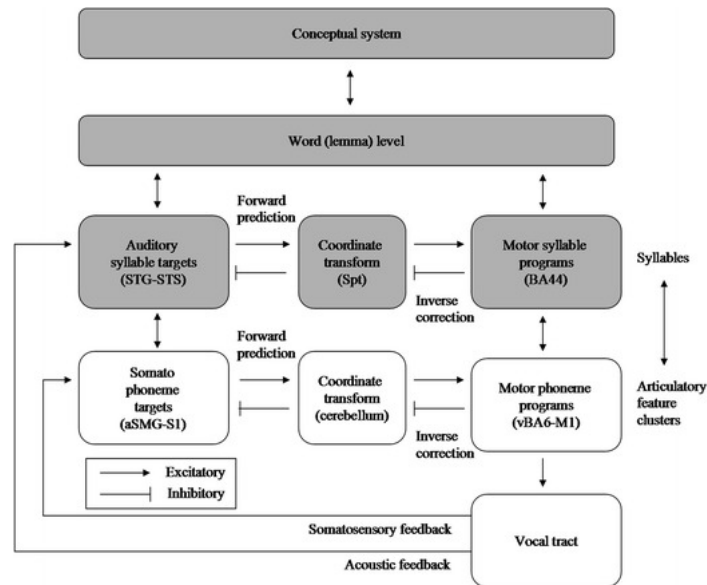


Fig. 2 A schematic diagram of the hierarchical state feedback control (HSFC) framework (Hickok, 2012)

4. The SLAM model

SLAM は 音素表現を聴覚要素と運動要素に分割した語彙検索計算モデルである (図 3)。音素の 2 重表現は第 1 の HSFC 原理に直接従っている。聴覚的表現とすることは第 3 原理に基づくものであり特にこのレベルの符号化は音素性よりも大きいことを意味する。しかし SLAM モデルの接続パターンはノイズの多い運動命令を強化することができる一種の前方/後方モデルを実装している。我々の目標は運動と感覚の音韻表現を分離するというアーキテクチャ上の仮定の効果を評価するために SP の計算上の仮定を可能な限り修正することであった。

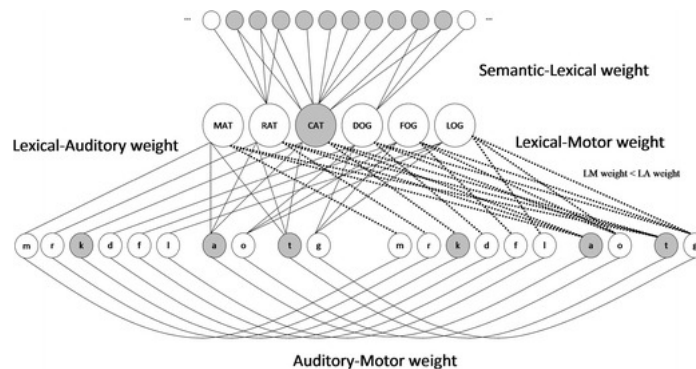


Fig. 3 The semantic-lexical-auditory-motor (SLAM) model architecture

絵画命名課題のシミュレーションでは、活性化は主に意味ユニットから語彙ユニット、聴覚ユニット、運動ユニットへと流れる。以降、このモデルを SLAM と表記する。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も存在する。これは Wernicke (1874/1969) の仮定にまで遡るものである。伝導失語症において流暢さが維持され、エラーのない発話をすることを説明するために必要なものである。しかし、この接続は語彙-聴覚経路 (これも Wernicke の当初の考え) よりも常に弱く、いくつかの点が動機となっている。

- 第 1 に聴覚-語彙経路は、語彙-運動経路よりも早期に発達し、より頻繁に使用されると想定される。縦断的研究によるば、子どもは単語を生成する数ヶ月前に単一の単語を理解し始め、新たに理解した単語は、新たに生成した単語の約 2 倍の速度で獲得することが示されている (Benedict, 1979)。
- 第 2 に運動制御理論は運動計画は感覚的な目標によって動かされていることを示している。発達の過程で、学習者は聴覚的目標を参照し、発話音とその音を再現する運動ジェスチャーの間のマッピングを学習しなければならない (Hickok, 2012; Hickok, Houde, & Rong, 2011)。
- 第 3 に失語症の文脈では、理解障害は生産障害よりも回復する傾向があり (Lomas & Kertesz, 1978) 語彙表現と聴覚音韻表現の間にはより強い関連性があることを示唆している。

語彙-聴覚マッピングは、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: すなわち、SLAM モデルが単に余分な部分を持つ SP モデルではないことを意味する。実際には SP と SLAM がカバーするパラメータ空間の重複は事実上ゼロである。その理由は以下の通りである。図 3 の SLAM アーキテクチャを考えると、SP モデルは、語彙-聴覚写像の結合係数と語彙-運動写像の結合係数、聴覚-運動写像の結合係数をゼロに設定し、語彙-運動写像の結合係数を自由に变化させることで SP を実装できることは明らかである。これにより、SP は SLAM の適切なサブセットになり、SLAM は SP と同じパラメータ空間 (したがってデータへのフィット) をカバーすることができるようになる。しかし、このアーキテクチャの可能性は、語彙-聴覚結合係数が語彙-運動結合係数よりも常に強いという我々の仮定を実装することによって明示的に排除された。語彙-聴覚結合係数が 0 であれば語彙-運動結合係数も 0 でなければならず、自由に变化することはできない。つまり SP によって使用されるパラメータ部分空間を事実上除外する。これにより、語彙-聴覚経路が命名に使用される主要な経路であるという SLAM の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-聴覚写像の結合係数が常に語彙-運動写像の結合係数よりも小さい場合にモデルの性能を調べることもできる。語彙-運動優位性を反映するために我々は「SLMA」と呼ぶこともできる。SLAM と SLMA は同数の自由パラメータを持ち、どちらも SP のそれよりも多い。だが、接続強度パターンに関しては異なる仮定を持つ。もし SLAM が SLMA よりも良い結果を出すとすれ

ば SLMA は SP をそのパラメータ空間の適切な下位集合として実装しているにもかかわらず、語彙-聴覚経路の優位性が理論的に動機づけられているだけでなく、観察された改善にも必要であることを実証することになるだろう。

要約すると SLAM は少なくとも SP と同様一般的な失語症の集団における障害を特徴づけ、主に伝導性失語症のモデル化に役立つであろうという仮説を立てた。伝導失語症は、特に音韻レベルに影響を与える聴覚と運動性音声表現の間のインターフェースの機能障害として説明されるのが最適である[@2011Hickok_state_feedback_control;@2012Hickok_HSFC]。したがって、聴覚-音韻表現と運動-音韻表現との間の写像を組み込んだ命名モデルは、この写像の機能不全に起因する発話エラーに対して、より良いフィット感を提供するはずである。この仮説を検証するために、失語性の絵画命名データの大規模なセットに対して SP モデルと SLAM モデルのフィットを比較した。

5. コンピュータ実装 (Computational implementation)

5.1 患者データ (Patient data)

すべてのデータは [Moss Aphasia Psycholinguistic Project Database \(Mirman et al., 2010\)](#) から収集された。このデータベースには、フィラデルフィア絵画命名検査 (PNT;[@1996Roach_Philadelphia_Naming_Test]) の回答を含む、失語症患者の大規模で代表的なグループからの同定されていないデータが含まれており、一般名詞 175 枚の線画セットが含まれている。データベースに登録されているすべての患者は、左半球脳卒中後の急性期失語症で、他の神経学的併存疾患の診断を受けておらず、少なくとも 1 つの PNT 項目を正確に命名することができた。我々は、失語症のタイプおよび発症後の月数 ($N = 255$) を含む、利用可能な人口統計学的情報を持つデータベース内のすべての患者について、最初に実施した PNT 検査結果を分析した。コホートは、アノミー型 103 人、プロカ型 60 人、伝導型 46 人、ウェルニッケ型 35 人、その他の失語症 11 人で、超皮質感覚型、超皮質運動型、脳後動脈型、または全球性の病因を有するものであった。脳卒中後の月数中央値は 28 [1,381], PNT の正答率中央値は 76.4 [1,99] であった。

5.2 計算論モデル (Computational models)

前述のように SP は [@2000Foygell_Dell_SP] によって最初に発表された。このモデルの絵画命名シミュレーションのアプローチは、2 段階相互拡散活性化理論を用いた語彙検索を実装したもので、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る(図 1)。ユニット数と接続パターンは、英語の音声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 つの単語のうち 1 つが対象となる単語で構成された非常に小さな語彙近傍構造を実装している。このモデルは 6 つの語彙ユニットを含み、それぞれが意味的特徴を表す 10 個の意味ユニットに接続されている。意味的関連する単語は 3 つの意味ユニットを共有しており、意味的にターゲットに関連する単語が 1 つだけの典型的な試行では、ネットワークは合計 57 の意味ユニットを持っていることを意味する。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。音韻ユニットは、6 つのオンセット、2 つの母音、2 つのコーダの合計 10 個であった。音韻的に関連するターゲット語は、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 2 つの単語で構成された。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あり、意味的にも音韻的にも共有されていない。20% の試行では、音韻的に関連する単語が意味的にも関連しており、ターゲットとの「混合」関係を持つ隣接語を作成している。

絵画命名課題のシミュレーションは、意味ユニットへの活性化を高めることから始まる。2 つのパラメータ S と P とは、それぞれ語彙-意味接続と語彙-音韻接続の双方向の重みを指定する。活性化は、ノイズと減衰を含む線形活性化則に従って 8 時刻ステップで、双方向に全層間で同時に広がる。その後、最も活性化されている語彙ユニットに 2 回目の活性化のブーストが送られる。活性化は、さらに 8 時刻ステップに渡って広がり続ける。最後に、最も活性化された音韻層のオンセット、母音、コーダユニットが出力として選択され、ターゲットと比較される。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。だが、これは強い接続によって緩和される。応答は、正しいもの、意味的、形式的、混合、無関連、新語に分類される。与えられた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 つの応答タイプにわたる多項分布が推定される。これらの分布を、失語症患者が作成した命名反応の結果と比較することができる。

SLAM は SP の詳細を多く保持しており、主にアーキテクチャ変更の効果を評価することを目的としている。意味ユニットと語彙ユニットは変更されていない。だが、音韻ユニットのコピーが追加されている。片方は聴覚ユニット、もう片方は運動ユニットである(図 3)。意味-語彙接続 (SL)、語彙-聴覚接続 (LA)、語彙-運動接続 (LM)、聴覚-運動接続 (AM) の双方向の重みを 4 つのパラメータで指定する。LA と LM の接続は SP モデルの P 接続と同じであり、各語彙ユニットは 3 つの聴覚と 3 つの運動ユニットに接続している。だが、AM の接続は単純に 1 対 1 である。絵画命名シミュレーションは SP モデルと同じ 2 段階の方法で行われ、意味ユニットと語彙ユニットにブーストが与えられ、運動ユニット内で音韻選択が行われる。

5.3 データフィッティング (Fitting data)

データに適合させるため、モデルは十分に異なる出力分布をもたらす異なるパラメータのセットで評価され、パラメータ空間からデータ空間への有限要素地図が作成され、またその逆も行われる。この過程では、まず、パラメータ値のセット (例えば、S と P の重み) を選択し、次に、そのパラメータセットを使用して、その特定のモデルのセットアップで発生する 6 種類の応答のそれぞれの頻度を推定するために、そのパラメータセットを使用してモデルの多くの命名試行を生成する。これらの頻度が決定されると、その重み構成は、地図と呼ばれる対リストの出力分布に関連付けられる。この地図の各点は、絵画命名を観察する際に起こりうる可能なエラーパターンの型についての予測を表している。モデルを評価する一つの方法は、その予測が観測された無相性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。観測された分布とモデルの最も近いシミュレーションされた分布との間の距離を、そのデータ点に対するモデルの適合度という。開平方平均二乗偏差 (RMSD) は、任意だが、一般的に使用される適合度の尺度で、各応答タイプの平均偏差として解釈できる。たとえば RMSD が 0.02 の場合、観測された割合が予測された割合から平均で 0.02 だけ逸脱していることを示す (たとえば、予測 = [.50, .50]; 観測 = [.48, .52])。このように RMSD 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、どのくらいのポイントを生成すべきか、冗長な予測値を生成しないようにパラメータをどのように選択するか、という問題が生じる。

[@2000Foygell_Dell_SP] は付録として、パラメータ空間の可変解像度地図を生成するための指針となる原則をアルゴリズム例とともに提供している。彼らは、マッピングアルゴリズムの特定の選択は、それが包括的な検索をもたらす限り、適合結果にほとんど影響を与えない可能性が高いと指摘している。パラメータ空間マッピングのための第二のアルゴリズムは Dell, Lawler, Harris, and Gordon (2004) によって提供された。これらの地図はオンラインで利用可能であり、その後の出版物でも使用されているため、これらの地図が SP の標準と考えられている。この SP 地図は各点で 10,000 サンプルの 3,782 ポイントを持ち、生成に数日の連続計算を必要とした。明らかに、マッピング手順に関連する計算コストは、モデルの開発とテストのためのかなりのボトルネックとなっています。地図に新しい点を追加することで、予測が観測に近い位置にある可能性が向上する。だが、モデルの新しい予測値のセットが細くなるにつれてリターンは減少する。Dell が提案しているように、目標は最適な適合を見つけることなので、モデルの性能を向上させるために地図点を追加することは、おそらく追求する価値のあることである (G. Dell, 私信 July 12, 2013)。さらに SLAM には 2 つの追加パラメータがあるため地図をより効率的に生成するためにマッピング手順を修正する必要がある。

我々は、マッピングアルゴリズムを再設計し、その固有の並列性を利用することで、効率を大幅に向上させた。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。パラメータ空間内の点の座標は、モデルのための可能なパラメータ設定によって定義され (点選択)、データ空間内の対応する点は、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 (点評価)。点評価ステップは、シミュレーションが独立した単位、独立したサンプル、独立したパラメータセットにまたがって計算されるため、並列化に非常に適している。しかし、点選択には、並列化を促進するための新しいアプローチが必要であった。ドローネメッシュの精密化である。

デローネ三角測量は、点の集合を結ぶグラフであり、任意の単項の円周が集合の他の点を含まないグラフである。このグラフは、辺が点の間の隣接関係を提供するという事実を含む、多くの有利な幾何学的特性を持っている。新しい点選択アルゴリズムは、これら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、パラメータ空間内の隣接する 2 つの点の間の距離がデータ空間内のしきい値距離 (RMSD) を超える場合、それらのパラメータ空間の中点が評価のために選択され、マップに追加される。次に、これらの新しい点がデローネメッシュに追加され、すべての辺がしきい値以下になるまで処理が繰り返される。このように、各反復において、点選択アルゴリズムは、パラメータ探索範囲全体にわたって並列に評価される複数の点をもたらす。並列処理は、効率をさらに向上させるために、グラフィックス・プロセッシング・ユニット (GPU) 上で実行された。

SP と SLAM の性能を統計的に比較する前に、モデルフィットに対する地図解像度の効果を調査した。最初に、パラメータ空間の継続的な探索を奨励するため、小さな RMSD しきい値 0.01 を使用して、各モデルについて超高解像度の地図を作成した。各地図は各点で 10,000 サンプルを含み、パラメータは独立して $[.0001, 0.04]$ の範囲で変化した。パラメータの最大値は、値を小さくするとより多くのエラーが発生するように、正しい応答の頻度が最も高くなる最低値の近くになるように選択された。低いマッピングしきい値を使用したため、アルゴリズムは、任意に大きな数の点を生成した後、完了前に停止された。アルゴリズムは完全な探索範囲にわたって効率的に点を選択するので、早期終了は大きな懸念ではない。この事実はまた、完全な空間をカバーしながら地図の解像度を下げることの些細な問題にしてい。

このマッピング手順で SLAM モデルの初期の 31,593 点を生成し、パラメータを自由に变化させた。その後 SLAM アーキテクチャに従って $LM \geq LA$ を持つすべての点を削除し 17,786 点の SLAM 地図を得た。完全な SP 地図は 57,011 点であった。次に大きな地図の中から、対数的に間隔をおいて 5 点から 17,000 点までの部分集合を選択して、各モデルについて 50 個の低解像度地図を作成した。各地図について、失語症患者全体と診断群ごとに、異種診断群を除いた平均フィットを計算した。図 4 はフィット曲線をプロットしたものである。両方のモデルで予想されたように、点を追加することでフィットはリターンが減少しながら改善された。図中の縦線でマークされた 2,321 点付近で相対的な適合パターンは安定しているように見えた。したがって、我々は、この地図解像度で SP と SLAM を比較することを選択した。

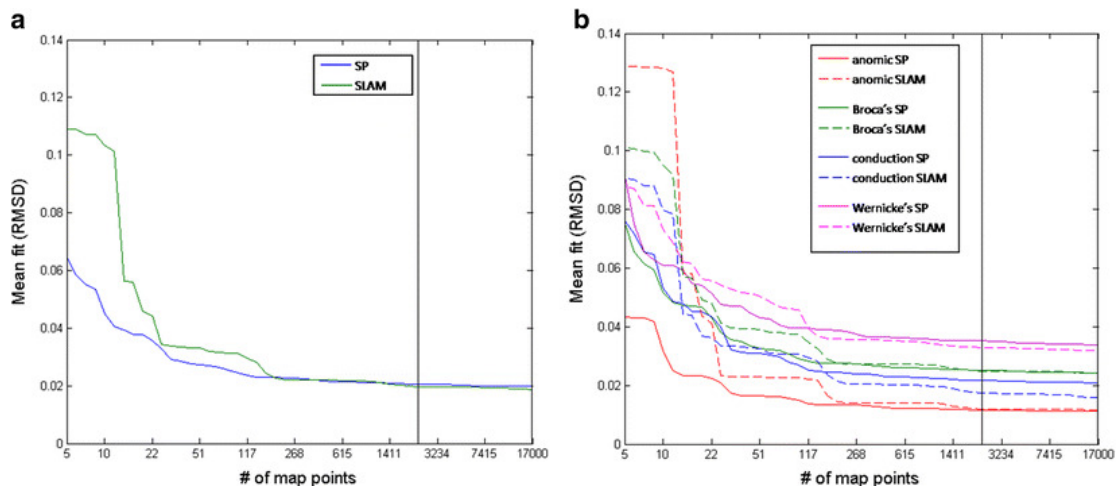


図 4 (A) 全患者と (B) 診断群の平均フィット曲線。パネル内の黒い縦線は、統計的比較に使用したマップを示す

新しい並列生成された地図と標準的な連続生成された地図とを比較するため、我々はまた、平均および最大フィットの点で、Schwartz ら (2006) で報告された値と同様の性能をもたらす並列 SP 地図の解像度を同定した。この 94 人の患者について、189 点の並列 SP 地図では、報告されている値である 0.024 と 0.084 と比較して、平均値と最大 RMSD はそれぞれ 0.0238 と 0.0785 であった。予想通り、並列アルゴリズムは逐次アルゴリズムよりもはるかに効率的に点を選択し、同様の性能を達成するために必要な予測値が少なくなった。我々は、標準の SP 地図に点を追加する効果と SP の構造を増強する効果を比較するために、この低解像度の地図をベースラインとして使用した。我々のフィッティングルーチンは、研究者がオンラインで利用可能な標準的な SP 地図よりも良いフィッティングが得られたので (Dell et al., 2004) 我々は、我々の新しいモデルと共に、地図解像度を調整可能にした我々のフィッティングルーチンを、以下の Web アドレスで提供した。

<http://cogsci.uci.edu/~alns/webfit.html>

6. 結果 (Results)

最初に SLAM が一般的な失語症集団の SP と同様に少なくともデータに適合するという仮説を検証した。すべての分析は MATLAB ソフトウェア・パッケージを使用して実行された。上述したように我々は適合尺度として RMSD を使用することを選択した(ここでは、値が低いほど適合が良いことを意味する)。表 1 は患者の全サンプルおよび失語症の 5 つの下位タイプについてのモデル適合の記述統計量を示している。図 5 は SP と SLAM の適合を比較した散布図である。実線の対角線はモデルが等価であるという仮説を表し、点線はサンプルの適合差の 1 標準偏差を示している。両方のモデルが全体的に非常によく機能し、大多数の患者が 0.02 RMSD 以下にクラスタリングしていることが明らかである。両モデルは一般的に似たような適合をもたらす傾向があるが、患者のサブグループが 1-SD の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 のどちらの方が適合が良かった (> 1 SD) 患者の数を比較した棒グラフで SLAM が一般集団での適合を犠牲にすることなく、より良い適合を患者のサブグループに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
All N = 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]
Anomic N = 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]
Broca's N = 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]
Conduction N = 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]
Wernicke's N = 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]
Other N = 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]

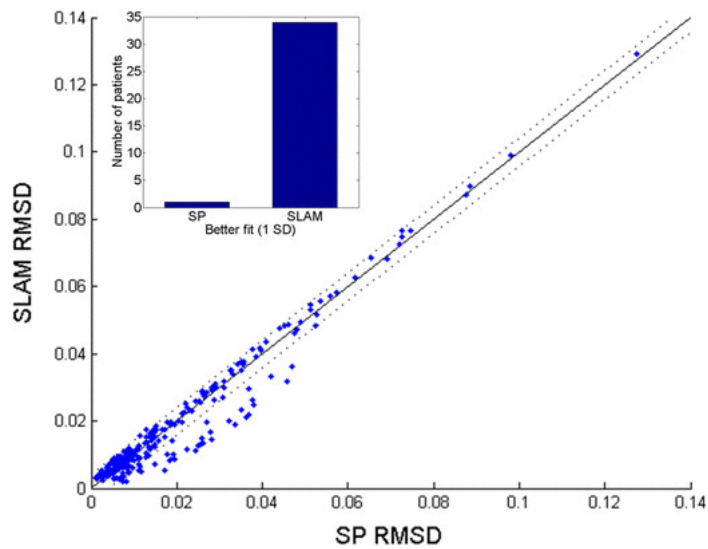


図 5 SP と SLAM のモデルフィットを比較した散布図。斜めの実線は同等のフィット感を表し、点線はサンプルにおけるフィット感の差が 1 SDであることを表している。大多数の患者は両モデルともよくフィットしており、一部の患者は SLAM の方が顕著にフィットしている (挿入図)。

次に SLAM が伝導失語症に特化したモデル適合を改善するという仮説を検証した。図 6 は失語症のタイプ別にグループ化した個々の患者のモデル間の RMSD の差を示している。正の差の値は SP よりも SLAM の方がフィットが改善されたことを示している。SLAM モデルは伝導失語グループに対して最大かつ最も一貫した適合改善を提供しており Wernicke 患者の適合の大部分も新しいモデルの恩恵を受けていることは明らかである。Wernicke の失語症も SLAM によって適合が改善されたという事実は HSFC 理論と一致している。ウェルニッケ失語症は、伝導失語症と非常によく似た神経解剖学的損傷と関連しており、急性ウェルニッケ失語症では、伝導性プロファイルに近い状態に回復することが多く、障害の部位が部分的に共有されていることを示唆している。5 つの失語症サブタイプ間の適合改善を統計的に比較するために、我々は RMSD の変化について一要因分散分析 (ANOVA を行い、グループ間の平均値の間に少なくとも 1 つの有意差 ($p < 0.001$) を示した。フォローアップ多重比較試

験では、平均フィット改善の 95% 信頼区間が Wernicke を含む他のどの群のそれとも重ならなかった。なので、導通群は他のどの群よりも SLAM からより多くの恩恵を受けたことが示された。

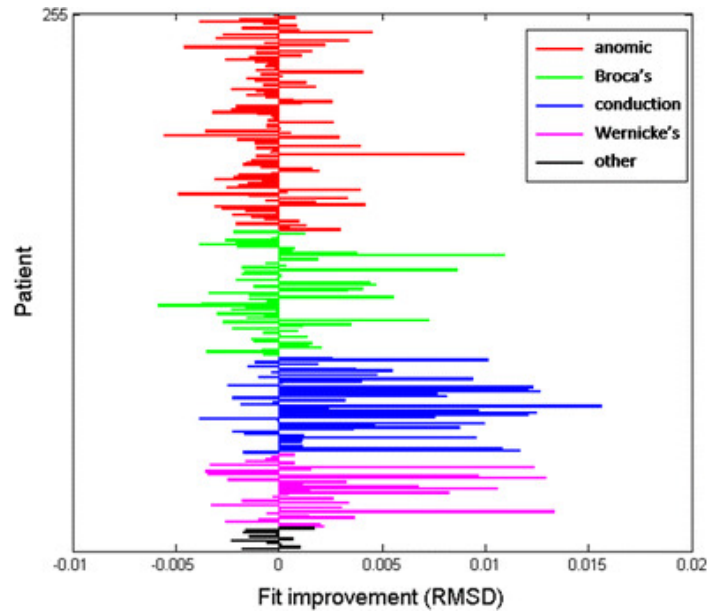


図 6. SP モデルと SLAM モデルの個々のフィットの変化。正の値は SLAM のフィットが高いことを示している

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フィット改善が、どの診断群にも特に有利に働くかどうかを検証した。理論的に動機づけられた構造的な変更とは異なり、モデルフィットを改善する方法では、特定の群に有利になるとは予想されなかった。我々は、平均して文献の標準的な SP マップと同等の 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 人の患者群では SP マップのポイント数を増やすことで、平均フィットが 0.0230 RMSD から 0.0206 RMSD に有意に改善した ($p < .001$)。フィットの改善は、すべての診断グループに対して有意であった (すべての $p < 0.001$)；しかし、フォローアップ多重比較検定を用いた一要因 ANOVA では、どの群も他のすべての群よりも有意に大きな改善を持っていなかったことが示された (不連続信頼区間がない)、特に伝導グループを有利にした我々の構造変更によって生成された結果とは異なる。代わりに Wernicke 群が最も改善したのに対し、失名辞群は最も改善しなかった。だがこれは、これらの群がそれぞれ SP のための最悪のフィットと最良のフィットであるという観察と一致する。このことは、理論的に動機づけられた SP モデルのアーキテクチャの操作によってもたらされた適合性の改善は、他の手法によって得られた改善とは質的に異なるということを示している。

我々はまた、伝導失語型命名パターンは特定の SLAM 構成によって適合されるべきであるという仮説を立てた: 強い LA と弱い AM の重み。フィット感の最大の改善を示した患者では、これは実際にそうであった。図 7 は最大の適合改善 (2 SD 以上) を示した 20 人の患者 (伝導 13 人、ウィルニッケ型 5 人、失名辞型 1 人、プロカ型 1 人) の SLAM ウェイト構成を表示するためにボックスプロットを使用している。図 8 は伝導性失語症の例の患者からのデータを、対応する SP と SLAM モデルの適合とともに示している。SP モデルの最適な重みは S と P について、それぞれ 0.022 と 0.017 であった。この患者の SLAM モデルでは SL と LM がそれぞれ 0.023 と 0.013 であったのに対し LA の重みは 0.04 で最大化され AM の重みは 0.0001 で最小化された。この患者では SLAM は SP の適合誤差を 0.0135 RMSD で減少させた。この例は SLAM の SP に対する最大の適合改善が、意味的 (および無関連) エラーの一貫した減少とともに、形式エラーの予測頻度の一貫した増加を伴っていることも示している。形式エラーと意味エラーのトレードオフは最初の語彙選択ステップで発生する可能性が高い。形式エラーの二重の性質つまり語彙選択と音韻選択のいずれかの段階で発生する可能性があることは SP モデルの特徴の一つである。[@2000Foygell_Dell_SP] は語彙選択時の形式的誤りは、語彙ユニットへの音韻的フィードバックが意味的フィードフォワードの活性化を上回る場合に増加することを示した。伝導失語症では LA の重みが大きいと語彙ユニットへの強い音素フィードバックが得られるのに対し、AM と LM の重みが小さいと運動ユニットへの弱い音素フィードフォワードが得られる。AM よりも LM の方が大きい場合、不正確で音韻論的に関連した語彙項目からより多くの活性化が流れ、それによって意味的な誤りを犠牲にして形式的な誤りが増加する。伝導失語症では、聴覚-音韻フィードバックの強さが語彙選択に影響を与えていることが示唆された。

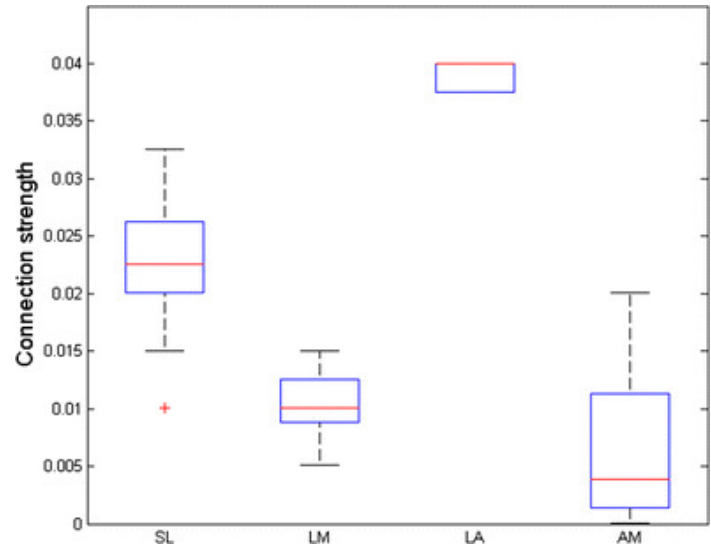


図 7 最も適合度が向上した 20 名の患者群の SLAM 重みを示す箱ひげ図。予想通り， 語彙-聴覚の結合係数が大きく， 聴覚-運動の結合係数が小さいモデルプロファイルが， SP 値の改善に大きく貢献している。

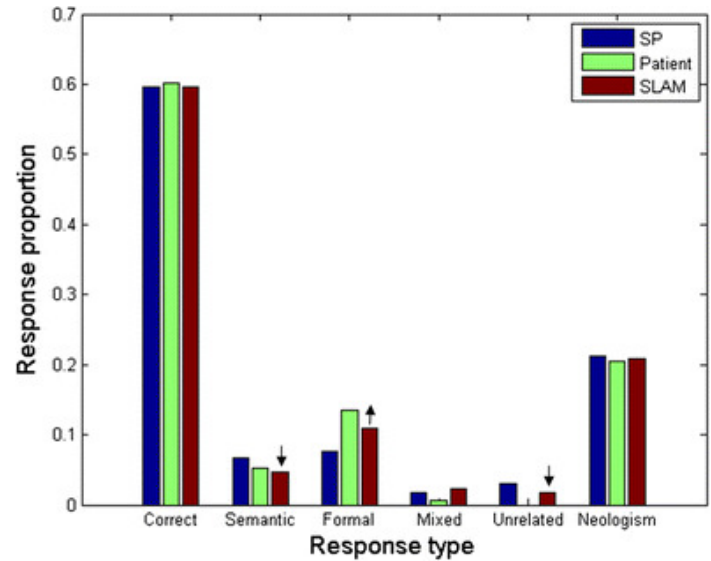


図 8 伝導失語症の患者の命名反応分布と， それに対応する SP モデルおよびSLAM モデルの適合性。矢印は， SLAM が， 意味エラーと無関連エラーを犠牲にして形態エラーを増加させることにより， データへの適合性を改善していることを示している。SLAM モデルは， この患者のフィットエラーを 0.0135 RMSD 減少させた。

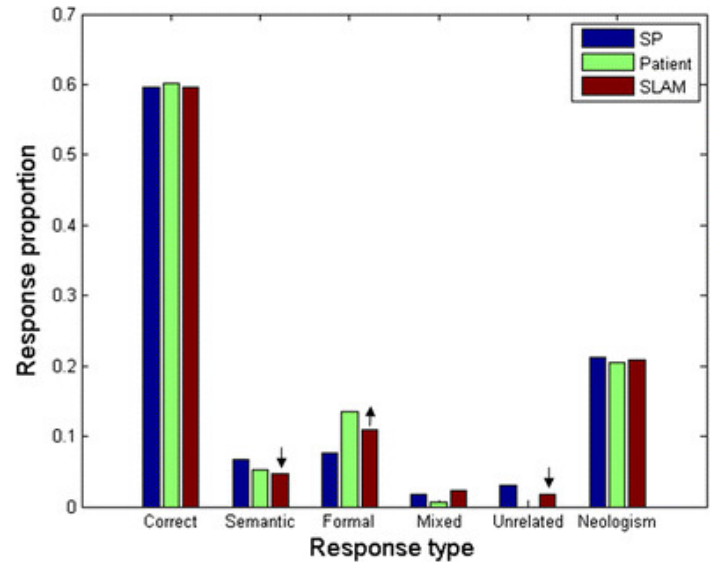


図 9 SP と SLAM と同じ数のパラメータを持ち， 代わりに語彙-運動優位のアーキテクチャである SLMA (semantic-lexical-motor-auditory) モデルとのモデル適合性を比較した散布図。線は図 5

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の向上は見られない

7. 議論 (Discussion)

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、準語彙 (sublexical) 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、聴覚的標的を運動指令に変換することが、言語表出に先行する場合であっても、語彙検索に重要な計算であることを説明している。

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プロセスに組み込むことで、このモデルが一般的な失語症の命名エラーを元の SP モデルと同等に説明できることを示した。特に伝導失語患者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、LA の重みを強く、AM の重み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、理論的に動機づけられた構造的変化とは異なり、地図の解像度の追加による改善は、失語症の型に依存しないことも実証された。また、最も高い適合性を得ることができると予測された結合係数重みの設定は、意味エラーを犠牲にして形態エラーを増加させることによってもたらされることがわかった。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 つの異常な下位群を特定した。この下位群患者のうち 2 人は伝導失語であり、もう 1 人はウェルニック失語であった。SLAM は、この下位群のもっともらしい説明となる。伝導失語症では、意味エラーを犠牲にして形式エラーが増加していることから、音韻関連エラー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 つの別々の音韻ルート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を詳細にモデル化していないが、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。計算同時性、階層的表現、階層間相互作用、局所的損傷、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、非常に小さな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、有用なテストベッドとなった。

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査としてよく用いられる。復唱は、命名と同様に運動産出系を必要とするが、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、名前を覚えるのが苦手が反復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経路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 つの経路が名前を覚えるとき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、この代替二重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、反復、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」モデルである Lichtheim2 を発表した。しかし、Lichtheim2 は、SP の詳細な分類法に基づいて発話エラーの型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、Lichtheim2 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。Lichtheim2 では、入力と出力の音韻は、1 クラス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、隠れた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の特定の領域を「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な動機の点で、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 Lichtheim2 の大きな違いは、SLAM では語彙単位と音韻単位を明確に階層的に分離し、どちらの段階でも選択エラーを許容していることである。この階層的な分離は、伝導失語症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。

Lichtheim2 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課題に関連した選択を媒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施している。このモデルは Lichtheim2 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 Lichtheim2 は、ほとんどの認知的・計算的問題、特に SLAM が調べ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。さらに、SLAM や Lichtheim2 と同様、WEAVER++/ARC は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のパターンをシミュレートしている。WEAVER++/ARC と Lichtheim2 の間の主な不一致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、(背側の) 弧状筋節と (腹側の) 無縁筋節のどちらに関連すべきか？しかし、WEAVER++/ARC は 1 つの重要な理論的ポイントに関して 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、命名時に、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語彙層が感覚標的として活性化されるのではなく、聴覚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。このことは、伝導失語において、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、Lichtheim2 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」モデルと呼ばれる広いクラスのモデルに分類される。すなわち、行動を制御する処理の流れは別々だが、相互に作用すると考えるモデルである。この研究の多くは、Hickok and 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。つまり、聴覚と運動音声系との間の写像は背側経路に対応し、聴覚と語彙・意味レベルの間の写像は腹側経路に対応する。Hickok and Poeppel は、この皮質ネットワークを、2つの経路に分岐する聴覚音声システムの観点から議論したが、絵画命名課題は、概念→語彙→聴覚(腹側経路)、聴覚→運動(背側経路)と、両方の経路を横断する。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との間に明示的な接続性を仮定していることである。Hickok and Poeppel は、概念系と運動系間の接続性を想定していたが(議論はしなかった)、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 and Poeppel の二重経路の枠組みを改良したものである。