

The Neurocognitive Mechanisms of Speech Production

Rhonda McClain¹ and Matthew Goldrick²

PENNSYLVANIA STATE UNIVERSITY¹

NORTHWESTERN UNIVERSITY²

Abstract

The study of how a speaker produces meaningful utterances is a vital research topic in psychology and cognitive neuroscience, offering a window into not only language itself, but also cognition more broadly. We review neurocognitive and behavioral evidence that supports a common set of principles for theories of the processes underlying speech production: independent levels of representation/processing; parallel activation and (limited) interaction; and structure-sensitive selection and sequencing. We then examine the interaction between speech production and other domains of cognition, reviewing research on monitoring, executive control and attention, and the intention to communicate. This work reveals how production is highly integrated with other domains of cognition, suggesting a number of interesting avenues for future research.

Key Terms: speech production, levels of representation, interaction, selection, monitoring, control, intention

Introduction

Contemporary investigations of the psychological processes underlying language production have their roots in the investigation of spontaneous speech errors in native English speakers (Fromkin, 1971; Garrett, 1975). Over the following four decades, researchers have used an ever-widening range of data to constrain theories of language production. This includes a diverse array of behavioral data, including speech errors that are experimentally induced; errors that arise as a result of neuropsychological impairment; and chronometric studies of the timing of production behaviors. These behavioral data are increasingly integrated with neurophysiological measures, including electrophysiological measures, indices of disruptions to neural processing, and changes in metabolic

activity in brain regions. This methodological advances have been coupled with a consideration of a more diverse array of populations, including speakers of a variety of different native languages as well as multilingual speakers.

The recent contributions to Goldrick, Ferreira, and Miozzo (2014) provide a detailed review of this work at levels of processing ranging from meaning to articulation. Rather than summarize all of these findings, the first section of this chapter aims to provide a high-level overview of the current state of research into speech production, focusing on four points of relative theoretical consensus concerning the internal structure of the spoken production system. We then turn to more recent work that explores how speech production processes are integrated with other aspects of the cognitive system.

Basic Principles of Production

While many aspects of the organization of speech production are clearly contentious, across research methods and populations there is a substantial amount of agreement concerning its basic organizational principles. In the following sections, we briefly review a few key pieces of evidence that have motivated these principles, as well as areas of disagreement among current theoretical proposals.

Independent levels of representation/processing

The systematic observations of spontaneous speech errors by pioneering investigators like Fromkin (1971) and Garrett (1975) revealed that rather than being random distortions of processing the distribution and types of errors showed regular patterns. One key observation from this work is that different types of errors are governed by distinct principles. For example, Garrett (1975) examined the distribution of errors involving the exchange of words (*I'm writing a mother to my letter*) versus individual sounds (*beast of burden* → *burst of beadan*). Word exchange errors in this corpus were strongly influenced by their grammatical properties (*mother* and *letter* are both nouns) and could involve elements that were separated by several lexical items (e.g., the exchanging words could occur in different syntactic phrases). In contrast, sound exchanges tended to occur over shorter distances (within a syntactic phrase, one or two words apart) and were influenced by sound similarity (e.g., the preceding /b/ sound in the exchange above). These distinct patterns suggest some degree of fractionation of the process of producing speech—such that some errors have their origin within a syntactically-driven processes operating over word-sized representations, whereas others arise in sound structure processes manipulating individual components of word form. In the intervening decades, data from other methodologies and

populations has bolstered these initial observations; this has led current theories of speech production (e.g., Caramazza, 1997; Dell, 1986; Levelt, Roelofs, & Meyer, 1999) to converge on the assumption that there are independent levels of representation/processing that encode word meaning (semantics), word form (phonology), and grammatical structure (syntax).

Data from acquired language impairments has provided clear evidence for independent processing of word meaning and form. If these are two independent aspects of processing, it should be possible for the link between them to be disrupted. Such an individual would have intact comprehension as well as the ability to process word forms for production yet still exhibit difficulties in producing speech. Consistent with this, multiple individuals have been reported in the literature who, subsequent to neurological impairment, produce semantically (but not phonologically) related errors specifically within spoken production (Caramazza & Hillis, 1990; Rapp, Benzing, & Caramazza, 1997). These individuals can successfully comprehend speech and avoid producing semantic errors in writing (e.g., naming a picture of a brush “comb” while writing B-R-U-S-H); this suggests that their processing of word meaning is intact. The fact that they produce semantically but not phonologically related errors in speaking suggests that their difficulty does not arise in processing of word form—if it did, we would expect the individuals to produce errors that are similar in sound to the target. Further evidence of intact form processing can be seen in their ability to correctly repeat words. This suggests that their disruption is specific to the processes that allow meaning to access word forms—processes that are only necessary in an architecture that assumes separate processing of these two types of information.

Further evidence for distinct processes underlying the processing of word form and meaning comes from studies showing that difficulties in processing word meaning vs. form in production are associated with damage to distinct brain regions (see Race & Hillis, 2014, for a review). Mirman et al. (2015) recently presented new evidence in favor of this type of distinction. Rather than define *a priori* critical behavioral profiles, they used factor analysis of 99 individuals with chronic post-stroke aphasia to identify four major components to variation in memory, speech perception, speech production, and comprehension performance. Consistent with earlier work, this revealed two major divisions in patterns of performance—semantic vs. phonological processing—each of which could then be further (roughly) subdivided into perception and production. These different dimensions of variation in impaired performance also mapped onto damage to distinct brain regions (consistent with earlier neuroimaging and aphasia studies). In production, variation in semantic processing was associated with damage to the anterior temporal lobe, whereas variation in word form processing was associated with damage to regions superior to the Sylvian fissure. The distinct neural substrates underlying processing of word form and meaning suggests they are both independent processing stages at both the cognitive and neural level.

Data from unimpaired language processing has provided clear evidence that grammatical structure is represented independently from word meaning and form. Bock (1986) showed the use of a particular grammatical structure could be primed by the structure of previously encountered sentences, independent of overlap in form and meaning of content words. In her study, English-speaking participants repeated prime sentences and then described semantically unrelated pictures. The pictures depicted events that could be described using a prepositional object sentence construction (*The woman is showing the dress to the man*) or a double object (*The woman is showing the man the dress*). The results showed that the probability of using one of these alternative structure was influenced by the previously repeated sentence. Repeating a semantically unrelated sentence with the prepositional object structure (e.g., *A rock climber sold some cocaine to an undercover agent*) increased the likelihood that speakers would subsequently use a prepositional object description; the complementary pattern was observed with a double object prime. Subsequent work has confirmed that such *syntactic* or *structural priming* occurs even when function words like *to* are not shared across sentences and even across languages that use distinct function and content words (see Pickering & Ferreira, 2008, for a detailed review). Such a pattern is best understood by assuming that within the process of producing speech there is a representation of grammatical structure, independent of meaning and form, which can be independently primed.

Within each of these levels of representation, multilingual speakers maintain some degree of distinct representations of each language—while capitalizing on shared linguistic structure (see Kroll & Tokowicz, 2005, for review and discussion). The fact that speakers can successfully acquire languages with distinct grammatical and sound patterns provides *prima facie* evidence for such distinctions. For example, English-Spanish bilinguals can learn to correctly describe a picture of a white house using an adjective-noun sequence in English (*the white house*) and a noun-adjective sequence in Spanish (*la casa blanca*). This independence is tempered by shared representations/processes for overlapping structures. As noted above, syntactic priming can occur across languages. For example, in Spanish-English bilinguals, the use of a passive vs. active construction (*The boy is startled by the alarm* vs. *The alarm startles the boy*) can be primed by repetition of a Spanish sentence with parallel structure (e.g., ‘The truck is being chased by the taxi’: *El camión persigue el taxi*; Hartsuiker, Pickering, & Veltkamp, 2004). In many cases, this priming is roughly equivalent across languages, consistent with shared structural representations (see Pickering & Ferreira, 2008, for a review).

In addition to representations and processes encoding word meaning, word form, and grammatical structure, most theories include a level of lexical representation which serves to bind together these distinct aspects of structure—e.g., for *cat*, a representation <CAT> linking {furry} {feline}, NOUN, and /k/ /ae/

/t/ (e.g., Caramazza, 1997; Dell, 1986; Levelt, et al., 1999). In bilinguals, it is typically assumed that there are distinct lexical representations for each lexical item in each language (e.g., <CAT> as well as <GATO> for Spanish-English bilinguals; Kroll & Tokowicz, 2005). Some theories eschew such explicit representations, relying instead on learned, distributed representations to mediate the link between form, meaning, and grammatical structure (see Woollams, 2015, for a recent review).

Theories also incorporate mechanisms for representing morphological patterns (the structure and organization of the smallest meaning-bearing units in language; e.g., expressing {cat} and {plural} through the combination of *cat* and *s*). As such patterns hold at the interface of other types of knowledge (meaning + form, syntax + form), many theories assume that such information is reflected in the structure of lexical representations—but may have influences distributed across multiple processes (e.g., lexical and phonological; see Cohen-Goldberg, 2013, for a recent review).

Within these broad structural categories, there is considerable disagreement regarding specific assumptions. For example, many theories incorporate multiple levels of lexical representation; for example, distinguishing lexical-syntactic *lemma* representations from lexical-form *lexeme* representations (Levelt et al., 1999; c.f. Caramazza, 1997). Theories contrast not only in the number of distinct levels of representation but the content of each level. At the level of form, most theories distinguish relatively abstract, context-independent representations from relatively more specified, context-dependent representations. For example, the pronunciation of the sound /t/ differs in its acoustic and articulatory realization across different sounds structure contexts (its syllable positions, surrounding segments, stress, etc.). In the word *note* /t/ is pronounced with a full closure; the tongue completely stops the flow of air forming an obstruction at the roof of the mouth. In contrast, in *notable* the corresponding sound involves a very rapid closure and release (a “flap” sound). There is empirical evidence suggesting that speakers utilize two levels of form representation to represent this pattern. At one level, the /t/ is *note* and *notable* is identical; at a subsequent level of planning, the two sounds are represented as distinct form*s (i.e., flap vs. full closure; Buchwald & Miozzo, 2011; Goldrick & Rapp, 2007). However, debate continues regarding the precise nature of this representational distinction and whether the processing of context-dependent and independent information is interactive or more serial in nature (see Buchwald, 2014, for a review). The contributions to Goldrick et al. (2014) provide detailed discussion of these theoretical contrasts as well as the relevant empirical data supporting various positions.

Parallel activation and interaction

In addition to noting cases where elements such as words and sounds move within phrases, Fromkin (1971) and Garrett (1975) observed cases where multiple words fused into novel combinations of sounds (e.g., *clarinola* blends *clarinet* and *viola*). Similar phenomena occur over larger units such as phrases, where elements from multiple syntactic constructions fuse to form a phrase with a distinct meaning (e.g., *The blind have a missing sensory deficit* blends *The blind have a missing sense* and *The blind have a sensory deficit*; see Bock, 2011, for a review). Blends of individual sounds have also been observed; the acoustic and articulatory properties of speech errors reflect properties of both the intended and intruding sound (e.g., in *pig* → *big*, the resulting /b/ has properties intermediate between correctly produced /b/ and /p/; Goldrick & Blumstein, 2006). Such context-dependent blending of multiple intended sounds may also influence non-native accents; due to the difficulties in production processing, the acoustic and articulatory properties of speech sounds may reflect varying degrees of blends between the native and non-native language (Goldrick, Runnqvist, & Costa, 2014).

Current theories have modeled such phenomena by assuming that: one, at each level of processing, multiple representations are simultaneously activated in parallel (see Melinger, Branigan, & Pickering, 2014, for a review); and two, there is interaction between different stage of processing. Specifically, activation from multiple representations at one stage is allowed to influence processing at subsequent stages. For example, at the lexical level, <CLARINET> and <VIOLA> may both be activated in the context of discussing an orchestra; the observed blend *clarinola* results because activation from both lexical representations is allowed to *cascade*, influencing the activation of their corresponding word form representations.

The consequences of parallel and cascading activation can be observed not only in speech errors but also in non-errorful responses. For example, Peterson and Savoy (1998) asked participants to prepare to name a picture; after a delay, a cue indicate whether the picture name should be produced or a visually presented word should be read aloud. The critical pictures had a dominant label (e.g., *couch*) with a near-synonym (*sofa*). When asked to name the picture, participants produced the dominant label. However, response times in word reading suggested there was activation of the word form of the synonym; words that shared sounds with the synonym (e.g., *soda*) showed faster naming latencies than unrelated words. Similar phenomena have been observed in bilingual language production, where the coactivation of translation equivalents has been argued to lead to the facilitation of retrieval of words sharing form across the two languages (cognates; Costa, Caramazza, & Sebastián-Gallés, 2000). Given that coactivation poses a challenge for selecting words in a single language, there has been a great deal of

research dedicated to the potential constraints on cross-language activation during speech production (see Kroll & Navarro-Torres, Chapter 8, this volume).

After over a decade of extended debate concerning the extent of interaction between lexical and phonological levels (for opening salvos, see Dell & O'Seaghdha, 1991; Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991), there is now widespread agreement that there is some degree of interaction between semantic, lexical, and phonological levels in production (see e.g., Roelofs, 2008, for discussion from the perspective of those previously favoring less interaction). Most theories assume that parallel activation of semantic representations leads to the activation of multiple lexical representations; they also agree that these multiple lexical representations lead to the simultaneous activation of multiple word form representations. While such interactions are a functional part of production processing, there is also widespread agreement that there are critical limitations on their strength and extent (Dell & O'Seaghdha, 1992; Rapp & Goldrick, 2000; Roelofs, 2008; see Goldrick, 2006, for a review). What remains contentious is the presence and extent of reciprocal interactions between these levels of representation. Some theories assume that word form processing can exert a strong influence on lexical representations (Dell, 1986); others assume that such influences are present but limited (Rapp & Goldrick, 2000); while others maintain such direct interactions are absent, occurring only through perceptual monitoring processes (Roelofs, 2004). Similar debates arise with respect to the degree to which grammatical structure processes exhibit bidirectional vs. unidirectional interaction with processes involved in form and meaning (see Vigliocco & Hartsuiker, 2002, for review and discussion).

While the presence and strength of connections linking distinct levels of processing constrains interaction, it is important to note there are inherent asymmetries present in the production process. In the context of normal communication, speech production begins with the intention to express a message; meaning—as the first actor on the stage—therefore holds an inherent priority over other levels of processing. Coupled with restrictions on connectivity, this inherent asymmetry can serve to restrict interactions between subsequent levels of processing. For example, in the absence of direct connections from meaning to form, sound structure representations are activated only after passing through lexical and/or grammatical processes—reducing the degree to which word form processing can dominate production. This ‘staged’ view of production (with processing proceeding from meaning, to grammatical and lexical processing, and only then to form) is a unifying assumption of the models discussed above. However, recent electrophysiological evidence has suggested that there is early and near-simultaneous activation of semantic and phonological information during naming in both monolingual (Miozzo, Pulvermüller, & Hauk, 2015) and bilingual speakers (Strijkers, Baus, Runnqvist, Fitzpatrick, & Costa, 2013; Strijkers, Costa, & Thierry, 2010). The divergence between these empirical results and the

assumptions of current theoretical approaches to speech production is an area that clearly warrants detailed investigation.

Selection and structure-sensitive sequencing

Given that multiple representations are simultaneously activated at different levels of representation/processing—each of which exerts some degree of mutual influence on the others—a clear challenge to theories is how to regulate such processes to allow for efficient production of a sequence of linguistic elements. Theories have incorporated two types of interrelated mechanisms to accomplish these goals. *Selection* mechanisms serve to regulate parallel activation and its attendant consequences for processing at other levels. *Sequencing* mechanisms regulate processing of multiple elements within an utterance to enable ordered behavior.

In general, the function of selection mechanisms is to enhance the processing of one representation relative to others activated in parallel. While all theories agree on this necessity of achieving this goal, there is considerable disagreement as to the structure of the mechanisms that accomplish it. One proposal is that representations inhibit one another, with the selected representation suppressing the activation of alternatives (e.g., Cutting & Ferreira, 1999). A second type of proposed mechanism enhances the activation of a selected representation, such that its relative activation is much greater than that of alternative structures (e.g., Oppenheim, Dell, & Schwartz, 2010). Finally, rather than directly altering the activation of representations, other mechanisms restrict activation flow, allowing only selected representations to influence other stage of processing (e.g., Levelt et al., 1999). The distinction between these various proposals is currently the focus of considerable debate in the field, particularly at the level of lexical processing (e.g., Spalek, Damian, & Bólte, 2013).

As suggested by the exchange errors reviewed above, a challenge for spoken production processing is producing linguistic elements in the appropriate sequence. It is clear that the mechanisms used to meet this challenge are sensitive to linguistic structure. For example, as noted above, word exchange errors are sensitive to grammatical category. Similar patterns are observed at the form level; the sounds participating in errors tend to occur at similar positions within the phonological structure (e.g., stressed vowels exchange with stressed, not unstressed vowels; sounds at the beginning of the syllable exchange with sounds at the beginning, not end, of other syllables; see Goldrick, 2011, for a recent cross-linguistic review). The nature of the mechanisms that allow for such structure-sensitive sequencing is a matter of considerable debate. Some models (e.g., Dell, Burger, & Svec, 1990) incorporate an explicit structural frame, the slots of which specify the types of elements that can occupy a position in the structure (e.g., a noun phrase frame would contain a slot that could be bound to any noun). Serial

order then reflects the activation of elements within the frame (e.g., ensuring that in an English noun phrase the determiner is activated before the noun). In control signal models (e.g., Vousden, Brown, & Harley, 2000), serial order reflects explicitly dynamic control representations instead of frames (e.g., oscillators which regularly cycle through a sequence of activation states). The changing states of these representations lead to the activation of the appropriate element in the sequence. Finally, sequence learning models (e.g., Chang & Fitz, 2014) develop internal representations to drive the production of sequences. These models are trained on the task of predicting the next element in a sequence based on a representation of what has been produced thus far. This allows for the emergence of distinct control structures cross-linguistically (based on the different statistical properties of each language).

Basic organizational principles of production

In summary, four basic organizational principles form the foundation for current theories of speech production:

- Independent levels of representation/processing: The meaning, form, and structural properties of utterances are reflected by distinct levels of representation and processing. Within each level, multilingual speakers maintain some degree of independence, encoding distinct structures in each language.
- Parallel activation of multiple representations: During the course of production, at each level of processing multiple representations are simultaneously activated in parallel.
- Interaction between levels of representation/processing: While there are distinct and independent processing levels, these overlap in time and exert some degree of mutual influence on one another.
- Selection and structure-sensitive sequencing: To support the production of specific linguistic elements, parallel activation is regulated by selection mechanisms that allow specific representations to dominate processing; to allow these elements to be produced in the proper order, interaction and parallel activation are constrained by structure-sensitive sequencing mechanisms.

How Does Production Interface with Other Cognitive Systems?

Most research in this domain has focused on the internal structure of speech production processes. More recent work has turned to the relationship between production processes and other aspects of cognition. The following sections examine three major issues that have been explored in previous work. A long-standing issue in production is how speakers monitor their speech; the first section examines production-based theories of this process, as well as more recent proposals that incorporate mechanisms external to the production system. We then examine how domain-general mechanisms contribute to selection processes within production. Finally, we turn to a small but critical body of recent work examining the role of intention in spoken production processing.

Monitoring

As mentioned previously, the study of speech errors has been fundamental to characterizing the architecture and dynamics of the speech production system. An important influence on the production of speech errors is monitoring: processes that inspect the appropriateness of speech on multiple levels of structure, helping to intercept errors before they are overtly produced (see Slevc & Ferreira, 2006, for a review). A key unanswered question is the extent to which monitoring reflects components of the language system or other components of cognition outside of language itself (e.g., Riès, Xie, Haaland, Dronkers, & Knight, 2013).

In the sections to follow, we begin by describing a set of research findings consistent with the idea that speech monitoring is generally characterized by the presence of internal and external monitors. Then we examine the support for three distinct theories of internal speech monitoring that include the Perceptual Loop Account (Levelt, 1983; Levelt, 1989), production-based accounts, and the Conflict-based Account (Nozari, Dell, & Schwartz, 2011). We suggest that the evidence favors the idea that monitoring reflects processes of speech production, comprehension, as well as cognitive control mechanisms external to the language system proper.

Evidence for internal and external monitoring components

Most theories of monitoring during speech production include two general components. The *internal monitor* serves to correct speech before any motor plan is executed. The *external monitor* relies on the results of the motor command, including proprioceptive and auditory information. If we think of the monitoring system as being hierarchically organized, like a tree with branches, then the

internal and external components form the top level of the tree. Beyond this level, the monitoring system has been hypothesized to branch into many specialized components. These specialized components are often referred to as *loops*. This refers to the fact that monitors serve to provide feedback about the output of production process, looping this information back to production to enable correction and re-formulation of utterances. Figure 1 summarizes Potsma's (2000) model and the hierarchy within the monitoring system, illustrating the internal and external loops, with descriptions of some specific components taken across both the internal and external levels. The figure also shows that the activity that takes place at each independent level of word representation also acts as its own loop (i.e., conceptual loop), but is guided by more peripheral activity (i.e., proprioception, tactile feedback, auditory feedback etc.).

[Insert Figure 1 about here.]

There is a window of time between speech planning and articulation lasting approximately 100 ms during which errors can be intercepted. In a highly referenced example, a speaker who was asked to orally describe the route of different colored circles as they moved in different paths demonstrated an interesting pattern in the timing of speech repairs. When the speaker had to describe the circles moving in a horizontal path, the speaker said that the objects were moving in the horizontal direction but only after a segment of the alternative word had intruded into overt speech. In other words, an error was spoken as “v-horizontal” (Levelt, 1989). What is interesting about this example is that the speaker only produced the phoneme /v/ before correcting their response. The very small amount of time that had passed between the error and the correction is not consistent with a mechanism that first monitors speech through comprehension, halts speech, and then repairs it. Internal monitors accomplish this by parsing inner (unarticulated) speech plans, inspecting them for potential errors and interrupting speech. The fact that we can anticipate errors prior to articulation supports the idea that such a mechanism is used to detect speech errors (Garnsey & Dell, 1984; Postma & Kolk, 1993). Additional evidence for such internal monitor mechanisms is that slips of the tongue are detected in the absence of auditory feedback (Dell & Repka, 1992; Lackner & Tuller, 1979; Postma & Kolk, 1992a, b; Postma & Noordanus, 1996).

In addition to internal monitors, speakers also use external speech monitors that rely on the overt speech that speakers generate (see Figure 1). Because speakers can hear themselves as they produce speech, speakers should be able to detect errors via the auditory stream (Hartsuiker, 2014). Consistent with the use of such a mechanism for monitoring, speakers are worse at detecting errors when they cannot hear themselves speak (Lackner & Tuller, 1979). Furthermore,

speakers who have acquired hearing loss later in life show declines in their ability to properly monitor the phonetic properties of their speech. They experience difficulty in compensating for articulatory disruptions (e.g., speaking while clenching an object with your teeth). However, after being fitted with cochlear implants the speakers improve in their control ability (Jones & Munhall, 2000), suggesting auditory feedback is used to control and correct aspects of online speech production.

Considering the findings demonstrating reliance on both the internal and external monitors, one plausible explanation of speakers' ability to monitor their own speech is that they recruit both internal and external monitors in parallel. This predicts that while we are speaking, we are generating relevant information about the quality of our speech from various monitors. Among major theories of monitoring, there is disagreement about which monitoring channels provide the crucial information for acting upon errors (and which play secondary roles). We start our review first with the Perceptual Loop theory (Levelt, 1983, 1989), a seminal account of monitoring during speech production.

Perceptual Loop Theory

The Perceptual Loop theory (Levelt, 1983, 1989) is a comprehension-based monitor. The main assumption of this theory is that the speaker exploits the same processes used to comprehend other speakers' utterances when they engage in error detection. The monitor checks whether information in the utterance matches the speakers' intentions on two levels, inner speech and overt speech. At the level of inner speech, the phonetic plan corresponding to the output is temporarily held in a pre-articulatory buffer, at which point the monitor can survey the material (Hartsuiker & Kolk, 2001). This mechanism is referred to as the inner loop (see Figure 1). At the same time, speech is articulated, and information available in the speech stream is checked for appropriateness. This is referred to as the auditory feedback loop.

Behavioral Evidence

As discussed above, there is evidence that both internal and external mechanisms are involved in monitoring speech—broadly consistent with the Perceptual Loop theory's claim that there are two streams for monitoring. However, the claim that speakers detect errors through comprehension has been a matter of controversy. If comprehension is a key component in the ability to detect errors in speech production, then comprehension ability and error detection should correlate. Speakers who have poor comprehension should also be poor at detecting errors in their own speech. Research directly testing the correlations between comprehension and monitoring has failed to find such a relationship (Nickels & Howard, 1995; Nozari et al., 2011). Dissociations that manifest bi-directionally, with patients who have poor auditory comprehension having intact monitoring abilities (Marshall, Rappaport, & Garcia-Bunuel, 1985) and vice versa

(Butterworth & Howard, 1987; Liss, 1998; McNamara, Opler, Au, Durso and Albert, 1992; Marshall, Robson, Pring, & Chiat, 1998) provide further evidence challenging the role of a comprehension-based mechanism in production monitoring.

Neural Evidence

In evaluating the Perceptual Loop theory, there has been limited consideration of the neural correlates of the inner speech loop, auditory feedback loop, and comprehension at large. Neural data might be especially relevant for evaluating the validity of the Perceptual Loop theory, since it states that comprehension of one's self as feedback is supported by the same system that comprehends others' speech. Christoffels, Forminsano, and Schiller (2007) tested this prediction by having participants name pictures aloud under normal hearing conditions and in the presence of pink noise. Pink noise served to disrupt monitoring via auditory feedback. Regions that are more active under normal conditions were taken to be involved in monitoring via auditory feedback. Only a small region of areas involved in comprehending others' speech (the superior temporal gyrus; STG) were activated more strongly in the normal hearing vs. noise condition. In contrast, greater activation was observed in brain areas (the anterior cingulate, ACC) that have been implicated in cognitive control across many domains. Although previous research on the functional role of the ACC has implicated it in conflict-monitoring and error detection, the results of the study were interpreted under as favoring a more general monitoring function: continuous monitoring of output performance, regardless of the particular performance domain. This suggests that such domain-general mechanisms may be more critical for monitoring during speech production than the comprehension system.

Production-based theories

Production-based monitors function by measuring the flow of activation within the production system (Laver, 1973, 1980; MacKay, 1987, 1992a, b; Schlenk, Huber, & Wilmes, 1987). For example, Node Structure theory (MacKay, 1987, 1992a, b) utilizes a monitor that analyzes patterns of activation between interconnected representational units or nodes. This is based on the idea that errors arise when there is prolonged activation between nodes within the production system that do not already share connections (MacKay, 1992). Suppose that during word form planning speakers co-activate nodes corresponding to particular speech sounds. During correct production, such coactivation patterns will reflect the structural patterns of a language; for example, initial consonant sequences might include /st/, /sp/, /pl/ etc. When coactivation violates these patterns—e.g., activating the initial consonant sequence /sr/—the monitor may respond by halting further production activity.

Behavioral Evidence

Because error detection is based on patterns of activation, Node Structure theory predicts that if coactivation patterns are more rapidly instantiated they will be detected more quickly. Faster articulation therefore entails more rapid error detection. Consistent with this, Oomen and Potsma (2001) found that when participants described the path of objects moving at varying speeds, the rate of error detection and repair scales with the rate of articulation. The Perceptual Loop theory predicts that the error-to-cutoff timing will be extended when speech is produced at a rapid rate. Speaking at a fast rate means that the number of syllables produced per interval of time will increase. The consequence of more syllables being produced is that the auditory loop processes greater information before speech is halted and the error can be corrected. However, contrary to the Perceptual Loop account, Oomen and Potsma (2001) found faster error to cutoff latencies in rapid production conditions than speech produced at a normal pace. Therefore, data focusing on the repair of errors is more consistent with the Node Structure theory than the Perceptual Loop theory.

Neural Evidence

If the production system alone is responsible for error detection, then one would predict that errors elicit neural activity associated with language-related processes, but do not engage more domain-general error related processes. Abel et al. (2009) tested this prediction by imaging brain activity in healthy adult speakers as they named pictures. Brain areas associated with error monitoring (and error generation more generally) were indexed by comparing activation patterns associated with correct vs. error responses. Critically, error responses elicited greater activation in areas implicated in domain-general error processes (bilateral ACC, prefrontal, and premotor regions). Parallel to the imaging results reviewed above, this suggests that the language system interacts with other cognitive systems to perform error monitoring.

Conflict-based monitoring theory

Recently, a production-based monitoring theory has been proposed that focuses on the interaction between domain-general conflict detection mechanisms and the speech production system (Nozari et al., 2011). This Conflict-based monitoring theory states that the domain-general control system serves as the central monitor and the speech production is the subordinate component, relaying information to the core error detection device. The speech production system generates information about conflict between simultaneously activated response options. Then, the domain-general system uses that information to conduct error detection.

Neural Evidence

ERPs arguably provide some of the strongest evidence for the involvement of a conflict-detection mechanism in speech production. The error-related negativity (ERN) is a frontally-distributed, negative-going brain-wave component that peaks within 80-100 ms of a speech error (Masaki et al., 2001; Möller, Jansma, Rodríguez-Fornells, & Münte, 2007). This same component is observed during error-free processing when there is a high degree of competition for selection (Gaschunak & Schiller, 2008).

Critically, the ERN appears to index the engagement of brain regions involved in domain-general monitoring. This response appears to arise in the anterior cingulate cortex (ACC) and the basal ganglia (Hermann, Römmler, Ehlis, Heidrich, & Fallgatter, 2004; Kiehl, Liddle, & Hopfinger, 2000; Miltner, Lemke, Weiss, Holroyd, Scheffers, & Coles, 2003; Ullsperger & von Cramon, 2001; van Veen & Carter, 2002). The ACC is engaged during conflict arising in many tasks beyond speech production. For example, the Erikson Flanker task is a visuo-motor task that does not require the use of language. Participants indicate the direction of central target arrow while ignoring flanking arrows. Trials that induce conflict (with flanking arrows pointing in reverse direction from the target) increase activity in the ACC relative to congruent trials (where flanking arrows match the direction of the target). Because the recruitment of ACC (as well as the ERN) appears in both linguistic and non-linguistic tasks, it suggests that is a general central error-processing system.

Summary: Monitoring and Interaction with Other Cognitive Systems

In this section, we have shown not only the importance of the speech monitoring system but also provided an example of how speech production interacts with other cognitive domains. Whereas early studies of the monitoring system for speech production were concerned with language-based processes, newer neural investigations have shown that there is substantial contribution from domain-general control processes. This suggests that monitoring is not entirely dependent on production or comprehension. A key issue for future work is to articulate the division of labor between domain-general control mechanisms, production, and comprehension-based monitoring systems.

Executive Control

The domain general mechanisms briefly discussed above have often been referred to as executive control mechanisms (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). In this section, we consider the structure of such mechanisms in more detail, examining how they contribute to processes that regulate the effects on parallel activation and interaction in production.

A prerequisite to exploring how control can aid speech production is to have a sufficiently concrete definition of control. In other words, what are the relevant forms of control? A number of observations highlight the idea that speech production may depend on top-down control and importantly, speak to the forms of control that could be relevant for speech planning. One such set of observations comes from studies of bilingual language control. That is, there is increasing evidence that the complexities of producing speech as a bilingual speaker draw on domain-general executive functions (see Kroll & Navarro-Torres, Chapter 8 for a review of the most compelling results on engagement of top-down control mechanisms during bilingual production). When proficient bilinguals speak, they rarely make intrusions into an unintended language. This is essential since bilinguals often interact with speakers who are not proficient speakers of the two languages. To prevent intrusions in these contexts, bilinguals may maintain the task goal of speaking only the mutually known language (Green & Abutalebi, 2013). Convincing evidence for the idea that bilinguals engage top-down control during speech production has been revealed in studies of pathological language switching. Pathological language switching occurs when bilinguals alternate between languages across different utterances without the ability to willfully to control switching. For example, a case study of pathological language switching revealed that a Frulian-Italian bilingual alternated into Italian about 40% of the time that Frulian was to be spoken and switched into Frulian about 43% of the time that Italian was required (Fabbro, Skrap, Aglioti, 2000). Critically, the patient had lesions to the left frontal cortex and the right anterior cingulate—regions shown to be active in domain-general executive control tasks. This suggests that control of speech production may draw on a domain-general executive system.

Executive control has taken center stage in investigations of top-down control during speech production, but, in fact, top-down control may reflect the dynamics and coordination of different types of control mechanisms. In addition to executive control, researchers have begun to explore the role of attention in the control of internal representations relevant for speech production. Attentional mechanisms have been implicated in speech production, specifically in facilitating retrieval of words for objects that are in focus and intended for production while filtering objects that are out of focus from intentional word retrieval (for a review, see Schotter, Jia, Ferreira, & Rayner, 2014). This new research permits evaluation of whether control of speech production is achieved through a range of cognitive processes.

In the sections that follow, we review evidence for the role of executive control and attention in monolingual speech, bilingual speech, and speech production by aphasic speakers. We begin with a theoretical overview of executive control and attention. The potential roles of these processes in speech production are examined in patient data, neural data, and behavioral data. Finally, we address a larger question about the degree to which top-down control affects production.

As most instances of speech production are highly practiced, control mechanisms might not be required to execute such highly automatic tasks. Does the evidence suggest that top-down control is engaged during speech all of the time or under particular circumstances?

Executive Control

Miyake et al. (2000) proposed that three types of independent yet interrelated processes underlie executive control:

- *Shifting* between mental sets allows us to engage in different tasks (Monsell, 1996; Norman & Shallice, 1986). One context in which shifting skills may be relevant for speech production is when bilinguals switch between languages (e.g., saying *casa* instead of *house*). By analogy, the lexicons in the first and second language serve as the mental sets. Then, the requirement to speak in a particular language is much like applying different categorization rules over different mental sets, as in the domain-general test of shifting known as the Wisconsin Card Sorting task (Grant & Berg, 1948). One specific mechanism that could uniquely tie shifting to language switching may be unbinding the mental sets from the stimuli on which the categorization rule is just applied. In tasks where mental sets must be alternated between, research demonstrates that shifting costs can be reduced in the presence of an external cue that signals the upcoming shift. The benefit of being able to exploit an external cue is even larger when there is a long onset between the cues and the stimulus that must be responded to (Meiran, 1996). The fact that increasing the interval between cue and stimulus is additionally beneficial to performance suggests that individuals exploit this interval to unbind mental sets. Similar effects of lengthening cue-stimulus intervals have been observed in bilingual language switching (e.g., Guo, Luo, Chen, & Li, 2013)
- *Updating* involves revising items held in working memory by replacing old information with new information (Jonides & Smith, 1997; Morris & Jones, 1990). These items may literally be words previously spoken or untended for upcoming utterances, or the items may be the goals that are currently relevant for the task. Bilingual language production also illustrates how updating may function in single word production. Updating may be engaged by bilinguals at the level of the language schema allowing them to more efficiently activate the relevant language.
- *Inhibition* supports selection among competing alternatives, such as translation equivalents in a bilingual's two languages. As discussed in the introduction, the lexical representations <PERRO> and <DOG> are likely to be highly active when a Spanish-English bilingual sees the object of the

dog. Inhibition is not merely driven by the presence of coactivation. Given that for most bilinguals, one language is typically more dominant language than the other, a prepotent response is activated whenever the weaker language must be spoken. To prevent the prepotent response in the dominant language from being produced, it is hypothesized that one mechanism of control involves deliberate response suppression (Green, 1998). More specifically, both motoric stopping and conflict monitoring have been suggested to contribute to the ability to suppress a response (Chatham, Claus, Kim, Curran, Banich, & Munakata, 2012).

Inhibition

Behavioral evidence for inhibition of semantic relatives. During lexical access, words semantically related to the target (e.g., for target *dog*, *cat*, *mouse*, *horse*, etc.) become activated and influence production processing (Crowther & Martin, 2014, Howard et al. 2006; Oppenheim, Dell, & Schwartz, 2010). The introduction to this chapter identified production-internal mechanisms that might allow the target to dominate processing over these semantically related alternatives. In recent research, there has been much interest in whether domain-general inhibition serves to resolve competition between co-activated representations in speech production.

For example, Crowther and Martin (2014) used an individual difference approach to examine the role of executive control on semantic interference in production. Several measures of executive control were measured, including the Stroop task which indexed the ability to inhibit distracting responses. They examined how these measures influenced performance in the semantic blocking task, where groups of pictures are presented in large sets of trials that are semantically related (dog, cat, rat) or unrelated (dog, chair, plane). As illustrated in the example here, a small set of exemplars are shown as items within the semantically related and unrelated sets. Numerous cycles occur, such that all of the pictures serving as items in either the related or unrelated set are shown within every cycle. A series of cycles will be presented successively to form a larger semantically related or unrelated block. Slower reaction times and greater errors are typically observed in the semantically related blocks relative to the semantically unrelated blocks, consistent with the idea that semantic relatives compete for selection with target words. The results demonstrated that the size of this semantic interference effect was enhanced in individuals that had more difficulty in the Stroop task. Shao, Roelofs, Martin and Meyer (2015) draw similar conclusions based on distributional measures of the engagement of inhibitory processes (Ridderinkhof, 2002). However, Fink (2016) failed to find evidence that prior use of inhibition in a non-verbal task increased semantic interference effects; the lack of transfer effects suggests there may be limits on the degree to which semantic processes require the use of inhibitory processes.

Neural evidence of inhibition of semantic relatives. In the semantic blocking task, reaction times undoubtedly reflect a mixture of many factors, both facilitatory and inhibitory, at multiple levels of processing. For example, in addition to inhibition during lexical selection, facilitation may occur during conceptual processing because of repetition of visual/conceptual features across sets of pictures. Potentially, this might obscure the role of top-down control processes during speech production, especially inhibition. One means of addressing this issue is using neural techniques to isolate facilitatory from inhibitory processes (Aristei, Melinger, & Rahman, 2011) and separate early conceptual versus later selection-based effects (Janssen, Carreiras, & Barber, 2011).

One line of research adopting this approach has used the left inferior frontal gyrus (LIFG) to index the use of inhibition. This frontal lobe region is commonly activated in response inhibition paradigms (Garavan et al., 2002; Konishi et al., 1999; Konishi et al., 1998; Menon, Adelman, White, Glover, & Reiss., 2001; Rubia, Smith, Brammer, & Taylor, 2003), suggesting it plays a role in domain-general inhibition. Several studies have suggested that disruptions to this region enhance semantic blocking effects. Pisoni, Papagno, and Cattaneo (2012) used transcranial direct current stimulation (tDCS) to disrupt processing in a region implicated in domain-general inhibition—the left inferior frontal gyrus (LIFG). Relative to a sham stimulation control condition, they found that disruption of LIFG increased the semantic interference effect. Other research suggests that chronic disruption of the LIFG as a result of brain damage increases semantic interference effects. Individuals with lesions to the LIFG demonstrate larger interference effects than non-brain damaged speakers and individuals with lesions that are not localized to the frontal lobe (McCarthy and Kartsounis, 2000; Schnur, Schwartz, Brecher, & Hodgson, 2006). Taken together, these results suggest that there is an underlying relationship between inhibition and the semantic blocking effect.

Bilingualism and inhibition. As discussed in the introduction, bilinguals naturally experience competition between words in the two languages. The joint activation of two languages during speech planning occurs despite the bilingual speaker's intention to speak a single language (for a review, see Kroll, Bobb, & Wodniecka, 2006). With respect to speech production, the majority of the evidence shows that competition persists to late stages of production, including lexical selection and even phonetic encoding. In the face of persistent competition between words in the two languages, the evidence favors inhibition as one of the solutions to the problem of parallel activation of both languages.

If bilinguals engage inhibition to resolve competition between words, then there should be traces of this inhibition on subsequent naming trials. Inhibition of the L1 during L2 production should persist into subsequent L1 naming trials.

Misra, Guo, Bobb, and Kroll (2012) found evidence in support of this prediction by having participants name pictures in two single-language blocks, either L1 (native language) first followed by L2 (second language) or the reverse. When bilinguals produced the names of pictures in the L1 after speaking the L2, response times were slower relative to cases where L1 was named first. This behavioral result has been replicated in a within-participant design (Branzi, Martin, Abutalebi, & Costa, 2014) and has been shown to be quite long-lasting, persisting over a substantial number of L1 naming blocks (McClain, Rossi, & Kroll, in preparation). Electrophysiological data from this paradigm also supports a role for inhibition. Misra et al. (2012) found that naming L1 after L2 evoked an N200 component in ERPs (which has been argued to index domain-general inhibitory processes; but, see also Branzi, et al., 2014).

Shifting

As reviewed, the evidence from blocked picture naming suggests that in the purely L2 context, there is sustained control engaged over the L1. In contrast, contexts in which bilinguals are required to speak both languages in rapid alternation may impose a different set of demands, requiring the engagement of domain-general shifting mechanisms. Previous studies provided some initial support for the idea that greater experience switching languages during bilingual speech production trains domain-general abilities in shifting. Prior and Gollan (2011) found that Spanish-English bilinguals showed smaller switching costs on a domain-general switching task relative to monolinguals and Chinese-English bilinguals, presumably because they engage in language switching more than the other groups. However, more recent work suggests that the ability to switch languages as a bilingual may not reflect a one-to-one correspondence between shifting and language switching. Gollan and Goldrick (2016) find that in younger and older adults the ability to switch languages in connected speech is not correlated with individual differences in domain-general switching. Multiple studies have demonstrated that older bilinguals experience declines in domain-general switching, but not in language switching in single word tasks (Calabria, Branzi, Marne, Hernández, & Costa, 2015; Weissberger, Wierenga, Bondi, & Gollan, 2012; but see Gollan & Goldrick, 2016, for data from connected speech). Gollan and Goldrick (2016) find while (compared to younger adults) older adults have greater difficulty in domain-general switching, aging does not increase the degree of difficulty in producing language switches that violate vs. respect regular patterns of code switching. This suggests that domain-general deficits fail to interact with language-specific control of language switching.

These mixed results may reflect the fact that language switching is not the result of a single component process within executive control. It is increasingly clear that bilingual differences in cognition reflect multiple aspects of bilingual experience (Kroll & Bialystok, 2013). Therefore, it might not be necessary to

completely abandon the idea that shifting reflects some portion of bilinguals' ability to perform task switching and vice versa. To reconcile discrepancies in the literature, the rationale of Miyake et al.'s (2000) theory of executive control may offer insight. Shifting may partially underlie language switching. In addition, greater clarity on the role of executive components and language switching might be gained by seeking evidence for systematic, but not one-to-one relationships between executive control and performance. Resolving this issue embodied is an ongoing goal of bilingualism research (see Kroll & Navarro-Torres, Chapter 8).

Updating

Few studies have examined the potential contributions of updating to speech production. Piai and Roelofs (2013) hypothesized that updating could be required when individuals perform two tasks that overlap temporally. In this context, participants may actively choose among the two tasks for which task will be responded to, holding the other task in working memory while deciding when to resume it. Consistent with this, Piai and Roelofs found that picture naming speed in a dual task context (naming plus tone discrimination) correlated with a domain-general measure of updating ability. This evidence implicates a role for executive control in a common situation we find ourselves every day, talking while performing a concurrent task.

Attention

Domain-general processes outside of these three basic elements of executive function may also contribute to control during production. Attention can be understood as processes that occur to ensure that the massive amount of information available during processing is reduced to only the information relevant to our current goals. Central to most theoretical views of attention is the idea that we have limited capacity for processing information (Broadbent, 1958). Out of this traditional interpretation has come the metaphor of attention as a "spotlight" (Derryberry & Tucker, 1994). Stimuli that are not focused "under the spotlight" do not reach awareness or only receive superficial processing (Posner, 1994; Treisman, 1960). One important difference about how selective attention is utilized in speech production and other domains (e.g., visual or auditory perception) is that in speech production, selection occurs for internal conceptual representations and not information from the environment (Kan & Thompson-Schill, 2004). Nevertheless, in the speech domain attention may similarly function to maintain relevant information, disregarding irrelevant information (Bodenhausen & Hugenberg, 2009). Jongman, Roelofs and Meyer (2015) find evidence that individual differences in domain-general measures of the ability to sustain attention to a task correlates with picture naming speed, both under high demand (in the context of a dual task) and under low demand (when picture naming is the only task).

Summary: Executive Control and Speech Production

Existing research supports the idea that top-down control is generally recruited during speech production. The weaker view, that executive control and attention is critical only when speakers encounter high production demands, was not supported. Evidence demonstrates that top-down control generalizes beyond high load contexts (Jongman et al., 2015). In addition, executive functions were implicated in speech production for multiple speaker groups (Prior & Gollan, 2011; Shao et al., 2015). Therefore, a diverse set of findings all converge upon the idea that there is an interaction between top-down control and speech production. A key issue for future work is better understanding the balance between these domain-internal and domain-general mechanisms in processing.

Intentions

The processes we have discussed so far in this chapter underlie our ability to act on the intention to produce speech. Intention leads to the formulation of a message the speaker wishes to communicate (see Konopka & Brown-Schmidt, 2014, for a review of these processes) and then to the processes that specify a form to communicate this message (the focus of the discussion above). What is less well-understood is the structure of processes that initiate this process; the formation of the intention to speak. The goal of this section is to discuss recent advances in the state of knowledge concerning intention as a processing stage in speech production. Central questions include how and when does intention exert its influence on the speech production system at the neural level.

Previous work suggests the intention system is primarily engaged through medial frontal lobe structures. Though there has been discussion about how these structures are implicated in executive control and speech production, traditionally these structures have not been discussed with respect to their particular role in intention. A second discovery regarding the intention system is that, at the neural level, the speech production system undergoes rapid restructuring in the face of the intention to actively produce speech. Specifically, we discuss evidence that intention processes may be engaged by relying on the medial frontal lobe structures, shifts in hemispheric lateralization towards a left dominant network, and pre-activation of the lexical system.

Neural substrates of the intention network

It is now well-recognized that the core speech production processes, though widely distributed throughout the brain, are heavily reliant on left-hemisphere structures. For example, many neuroimaging studies investigating the role of different regions in performance of production tasks (e.g., word repetition, picture naming, reading aloud) have revealed that the core processes of word production

(i.e., propositional/conceptual processes and word-form encoding processes) are supported by a left-lateralized production network (Indefrey & Levelt, 2000; Simonyan & Fuertinger, 2015). In contrast, as we review below, functions supporting the formation of intentions are situated in the *medial* frontal regions of the brain, including the cingulate cortex, paracingulate cortex, presupplemental motor area (pre-SMA), supplementary motor area (SMA), and supplementary eye fields (Passingham, Bengtsson, & Lau, 2010).

Recent research has focused on the extent and the nature of interaction between the medial-frontal intention network and the left lateral production regions, such as Broca's area. The existence of the frontal aslant white matter tract, directly connecting medial-frontal regions and the left fronto-lateral speech areas, suggests that speaking requires coordination of processing across these different brain regions. If the coordination between processing within the two areas were critical for speech production, then the presence of frontal aslant abnormalities should lead to a high degree of impairment during speech production. There is some evidence that patients diagnosed with the grammatical/nonfluent variety of primary progressive aphasia (PPGA) demonstrate reduced speech output because of reduced integrity in this white matter tract. Catani et al. (2013) observed that behaviorally, agrammatic/nonfluent aphasic speakers produce fewer mean words per minute than other subgroups of speakers diagnosed with PPGA. Moreover, the extent of damage to the frontal aslant predicted this group's fluency during speech production. This finding supports the idea that the ability to engage the intention network is critical for speech production.

If the medial frontal lobe is critically involved in intentional production, it should be less engaged when speech production tasks are externally rather than internally cued. Crosson et al. (2001) asked healthy participants to generate exemplars in a semantic fluency task (e.g., name as many animals as you can think of). Participants performed this under their own pacing (internally-cued), when they saw the word "next" appear on the screen (weak externally-cued), or when given the category label and descriptive cues that would prompt them to retrieve a specific target exemplar (strong externally-cued). Results demonstrated that activity in the pre-SMA and inferior frontal sulcus increased as tasks progressed from internally to externally-cued, suggesting that the intentional nature of the task has a great role in the degree to which the medial frontal lobe is activated.

Intention modulates the production network

The processes reviewed in the first section of this chapter support automatic retrieval of production representations. Strijkers, Na Yum, Grainger, and Holcomb (2011) examined whether or not intentions serve as an additional top-down process that facilitates such processes, speeding early conceptual, lexical, and

post-lexical processes that may be influenced by the intention to speak. In effect, they hypothesized that intention to speak speeds up lexical access. To test their hypothesis, overt picture naming was compared to a task that does not induce intention to articulate, semantic categorization of pictures. In line with their predictions, previous evidence suggests that a key difference between overt naming and picture classification is the level of detail encoded. The picture classification task involves conceptual, lexical, and morphophonological access, but not articulatory processes (e.g., Abdel Rahman & Aristei, 2010). ERPs revealed early divergences between picture naming and semantic categorization. The ERP components that were differentially modulated included the N170, a marker of early visual processing, and the P200, which is sensitive to lexical access. Critically, the very early time course of the effects was interpreted as evidence that the production system increases baseline levels of the lexical system to perform more efficient visual/semantic processing and lexical access. As noted in the introduction, this early time course of modulation poses a challenge for the widely accepted staged view of lexical production. A related idea is that intention facilitates production by re-configuring the functional relationships between brain regions from a ‘resting’ to an ‘active’ state. Similar to Strijkers et al. (2011), this idea challenges the view that the production system is always prepared to react to speech-relevant stimuli. Instead, the intention to speak alters the way the speech production system is configured. Simonyan and Fuertinger (2015) explored this idea by contrasting the interrelationships in activity across brain regions at rest vs. during active, intentional speech production (self-initiated sentence repetition). Roughly, their analysis allows us to examine what brain regions are added to the networks already active at rest specifically for the sake of production. Their findings indicate that engaging in intentional speech is associated with changes in hemispheric lateralization and enhances specific regional connectivity (e.g., cerebellum connectivity). This provides further evidence that there is rapid modification of the existing speech production network when the intention to speak is present.

Conclusion and Future Directions

In this chapter, we focused on multiple aspects of speech production, bridging long-standing ideas with more recent trends in the study of speech production (see Figure 2). First, we identified fundamental properties of the speech production system that have been successful in explaining a wide range of empirical phenomena. We argued that the research evidence is consistent with the notion that there are independent levels of representation/processing, multiple

representations activated in parallel during the course of production, interactions between levels of representation/processing, and requirements for selection and structure-sensitive sequencing. In the next section, we emphasized research suggesting possible links between speech production and other mental capacities, including domain-general monitoring, executive control and attention, and intentional processes. These research devoted to these topics is growing at a rapid pace.

[Insert Figure 2 about here.]

Considering the contribution of other cognitive domains opens up the possibility of new explanations of speech production phenomena. One might conclude that speech errors are rare because there is limited interactivity and cascade within the system, as well as production-internal mechanisms of selection. A large body of research on monitoring during speech production suggests that the picture is more complex. Most critical to the present discussion is the idea that monitoring during speech production relies on brain networks implicated in monitoring outside of the language domain. One important step in future research on the role of domain-general monitoring in speech production will be to generate more fine-grained information about the neural substrates of monitoring mechanisms. For example, precisely which substrates overlap during speech production and non-speech domains. This will lead to greater understanding of the brain regions that support speech production.

The historical focus on monolinguals in speech production research may also have obscured the role of domain-general processes. Research conducted with bilinguals conducted with aphasic patients has provided clearer evidence for contributions of top-down control to speech production. Recent work has shown that executive control and attention contribute in monolingual speech production as well, even extending to cases in which speech is executed under low cognitive demands. We feel that a promising avenue of future research will involve incorporating architectural components representing specific control mechanisms into formal models of monolingual speech production. Though we caution that to make strong claims, data should be modeled along specifications that do and do not assume a role for top-down control mechanisms. This approach, as illustrated in Crowther and Martin (2014), will be needed to explain the range of circumstances in which different forms of control are engaged during speech production.

We also examined research on the role of intention in speech production, which is the least understood area suggesting intersection between speech production and other aspects of cognition. We reviewed evidence which suggests

that brain regions supporting the ability to act on intention during speech production comprise a network distinct from that underlying speech production more generally. However, these networks are functionally integrated, with intention facilitating the formulation of messages and lexicalization of speech. A potentially fruitful area for investigation concerns the role of intention in the selection of words in the bilingual lexicon, especially as it pertains to the ability to flexibly control activation of the two languages. This is particularly important since, bilingual language control, has been a topic of great debate in bilingualism research. According to some theories of bilingual language production, intention can be used to activate a restricted pool of words in only the target language (Costa & Caramazza, 1999; Costa, Miozzo, & Caramazza, 1999; Finkbeiner, Gollan, & Caramazza, 2006; Roelofs, 1998). An alternative theory claims that bilinguals control selection of words by inhibiting the unintended language (e.g., Green, 1998). The behavioral and neural evidence reviewed in the section on executive control converge to suggest that inhibition plays at least some role in selection. However, it is possible that intention to speak a particular language may also contribute. Of interest is whether the patterns that have been already demonstrated in intention research, suggesting pre-activation of the system will capture differences between the intention to speak the weaker and dominant language, and vary in unilingual and code-switching language contexts.

The past several decades have seen an ever-broadening scope of speech production research, moving beyond its origins in domain-specific processes motivated by speech error patterns in English to an integrated, neurocognitive view of processing that draws on a wide set of paradigms, populations, and mechanisms. Emerging data, while confirming some long-held assumptions, challenges others. The next several decades hold the potential for tremendous progress in advancing our understanding of the production of spoken language.

Author Note

Preparation of this manuscript was supported by National Institutes of Health Grant HD077140. Thanks to Emily Cibelli for comments on the manuscript.

References

- Abel, S., Dressel, K., Kümmerer, D., Saur, D., Mader, I., Weiller, C., & Huber, W. (2009). Correct and erroneous picture naming responses in healthy subjects. *Neuroscience Letters*, *463*, 167-171.

- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Bock, J. K. (2011). ‘How much correction of syntactic errors *are* there, anyway?’ *Language and Linguistics Compass*, 5/6, 322-335.
- Bodenhausen, G. V., & Hugenberg, K. (2009). *Attention, perception, and social cognition*. In F. Strack and J. Förster (Eds.), *Social cognition: The basis of human interaction* (pp. 1-22). New York: Psychology Press.
- Branzi, F. M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia*, 52, 102-116.
- Broadbent, D. E. (1958). *The effects of noise on behaviour*. In C.M. Harris (Ed.), *Handbook of Noise Control* (pp. 10-34). New York: McGraw-Hill.
- Buchwald, A. (2014). Phonetic processing. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.). *The Oxford handbook of language production* (pp. 245-258). Oxford: Oxford University Press.
- Butterworth B, Howard D. (1987). Paragrammatism. *Cognition*, 26, 1–37.
- Calabria, M., Branzi, F. M., Marne, P., Hernández, M., & Costa, A. (2015). Age-related effects over bilingual language control and executive control. *Bilingualism: Language and Cognition*, 18, 65-78.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208.
- Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex*, 26, 95–122.
- Chatham, C. H., Claus, E. D., Kim, A., Curran, T., Banich, M. T., & Munakata, Y. (2012). Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. *PloS one*, 7:e31546.
- Cohen-Goldberg, A. M. (2013). Towards a theory of multimorphemic word production: The heterogeneity of processing hypothesis. *Language and Cognitive Processes*, 28, 1036-1064.
- Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, 2, 231-244.
- Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1283–1296.
- Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection?. *Journal of Memory and Language*, 41, 365-397.
- Crosson, B., Sadek, J. R., Maron, L., Gökçay, D., Mohr, C. M., Auerbach, E. J., & Briggs, R. W. (2001). Relative shift in activity from medial to lateral frontal cortex during internally versus externally guided word generation. *Journal of Cognitive Neuroscience*, 13, 272-283.
- Crowther, J. E., & Martin, R. C. (2014). Lexical selection in the semantically blocked

- cyclic naming task: the role of cognitive control and learning. *Frontiers in Human Neuroscience*, 8, 9.
- Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 318–344.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S., Burger, L. K., & Svec, W. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104, 123–147.
- Dell, G. S. & O'Sheaghda, P. G. (1991) Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, 98, 604–614.
- Dell, G. S., & O'Seaghda, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42, 287–314.
- Derryberry, D., & Tucker, D. M. (1994). Motivating the focus of attention. In P. M. Niedenthal & S. Kitayama (Eds.), *The heart's eye: Emotional influences in perceptions and attention* (pp. 167-196), San Diego: Academic Press.
- Fabbro, F., Skrap, M., & Aglioti, S. (2000). Pathological switching between languages after frontal lesions in a bilingual patient. *Journal of Neurology, Neurosurgery & Psychiatry*, 68, 650-652.
- Fink, A. (2016). *The role of domain-general executive functions, articulation, and conceptualization in spoken word production*. Doctoral dissertation, Northwestern University, Evanston, IL. Available at http://www.linguistics.northwestern.edu/documents/dissertations/2016-05-02_Angela_Fink_PhD_Dissertation-submitted.pdf
- Finkbeiner, M., Gollan, T., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem?. *Bilingualism: Language and Cognition*, 9, 153-166.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27–52.
- Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *Neuroimage*, 17, 1820-1829.
- Garnsey, S. M., & Dell, G. S. (1984). Some neurolinguistic implications of prearticulatory editing in production. *Brain and Language*, 23, 64-73.
- Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.) *The psychology of learning and motivation* (vol. 9, pp. 133–177). New York: Academic Press.
- Goldberg, G. (1985). Supplementary motor area structure and function: review and hypotheses. *Behavioral and Brain Sciences*, 8, 567-588.
- Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, 21, 817–855.
- Goldrick, M. (2011). Linking speech errors and generative phonological theory. *Language and Linguistics Compass*, 5, 397–412.

- Goldrick, M., Ferreira, V., & Miozzo, M. (2014). *The Oxford handbook of language production*. Oxford: Oxford University Press.
- Goldrick, M., Runnqvist, E., Costa, A. (2014). Language switching makes pronunciation less native-like. *Psychological Science*, 25, 1031 – 1036.
- Gollan, T., & Goldrick, M. (2016). Grammatical constraints on language switching: Language control is not just executive control. *Journal of Memory and Language*, 90, 177-199.
- Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, 38, 404-411.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, 1, 67-81.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25, 515-530.
- Guo, T., Liu, F., Chen, B., & Li, S. (2013). Inhibition of non-target languages in multilingual word production: Evidence from Uighur–Chinese–English trilinguals. *Acta Psychologica*, 143, 277-283.
- Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, 42, 113-157.
- Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science*, 15, 409-414.
- Hartsuiker, R. (2014). *Monitoring and Control of the Production System*. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.). *The Oxford handbook of language production* (pp. 417-436). Oxford: Oxford University Press.
- Herrmann, M. J., Römmler, J., Ehlis, A. C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, 20, 294-299.
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, 100, 464-482.
- Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 845–865). Cambridge: MIT Press.
- Janssen, N., Carreiras, M., & Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *Neuroimage*, 57, 1243-1250.
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108, 1246-1251.
- Jonides, J., & Smith, E. E. (1997). The architecture of working. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 243-276). Cambridge, MA: MIT Press.
- Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *The Quarterly Journal of Experimental Psychology*, 68, 710-730.

- Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, and Behavioral Neuroscience*, 4, 43-57.
- Kiehl, K. A., Liddle, P. F., & Hopfinger, J. B. (2000). Error processing and the rostral anterior cingulate: An event-related fMRI study. *Psychophysiology*, 37, 216-223.
- Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., & Miyashita, Y. (1998). No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *European Journal of Neuroscience*, 10, 1209-1213.
- Konishi, S., Nakajima, K., Uchida, I., Kikyo, H., Kameyama, M., & Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain*, 122, 981-991.
- Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 3–20). Oxford University Press.
- Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25, 497-514.
- Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9, 119-135.
- Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531-553). Oxford: Oxford University Press.
- Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. *Sentence Processing*, 281-294.
- Laver, J. D. M. (1973). The detection and correction of slips of tongue. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence*. The Hague: Mouton.
- Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand*. New York: Academic Press.
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
- Levelt, W. J. (1989). *Speaking. From intention to articulation*. MA: The MIT Press, Cambridge.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T. & Havinga, J. (1991) The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122–142.
- Liss, J.M. (1998). Error-revision in the spontaneous speech of apraxic speakers. *Brain* and Language*, 62, 342–360.
- MacKay, D. G. (1987). *The organization of perception and action: a theory for language and other cognitive skills*. New York: Springer.
- MacKay, D. G. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition*, 1, 199-225.
- MacKay, D. G. (1992). Errors, ambiguity, and awareness in language perception and production. In B. J. Baars (Ed.) *Experimental slips and human error* (pp. 39-69). New York: Plenum Press.
- Marshall, R. C., Rappaport, B. Z., & Garcia-Bunuel, L. (1985). Self-monitoring behavior

- in a case of severe auditory agnosia with aphasia. *Brain and Language*, 24, 297- 313.
- Marshall, J., Robson J., Pring ,T., & Chiat, S. (1998). Why does monitoring fail in jargon aphasia? Comprehension, judgment, and therapy evidence. *Brain and Language*, 63, 79–107.
- Masaki H, Tanaka H, Takasawa N, Yamazaki K. (2001). Error-related brain potentials elicited by vocal errors. *NeuroReport*, 12, 1851–1855.
- McCarthy, R. A., & Kartsounis, L. D. (2000). Wobbly words: Refractory anomia with preserved semantics. *Neurocase*, 6, 487-497.
- McClain, R., Rossi, E., & Kroll, J.F. (in preparation). *Using the native language to evaluate the consequences of learning to speak a second language.*
- McNamara P, Obler L.K, Au R., Durso R., & Albert, M. L. (1992). Speech monitoring skills in Alzheimer’s disease, Parkinson’s disease, and normal aging. *Brain and Language*, 42, 38–51.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1423-1442.
- Melinger, A., Branigan, H. P., & Pickering, M. J. (2014). Parallel processing in language production. *Language, Cognition and Neuroscience*, 29, 663-683.
- Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, 12, 131-143.
- Miltner, W. H., Lemke, U., Weiss, T., Holroyd, C., Scheffers, M. K., & Coles, M. G. (2003). Implementation of error-processing in the human anterior cingulate cortex: a source analysis of the magnetic equivalent of the error-related negativity. *Biological Psychology*, 64, 157-166.
- Miozzo, M., & Caramazza, A. (2003). When more is less: a counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, 132, 228.
- Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral Cortex*, 25, 3343-3355.
- Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion-symptom mapping. *Nature Communications*, 6, 6762.
- Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, 67, 224-237.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100
- Monsell, S. (1996) Control of mental processes. In V. Bruce (Eds), *Unsolved Mysteries of the Mind: Tutorial Essays in Cognition*, (pp. 93–148). Oxford, Erlbaum.
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, 81, 111-121.

- Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex*, *31*, 209-237.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.) *Consciousness and self-regulation* (vol. 4, pp. 1–18). New York: Plenum.
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*, 1-33.
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227-252.
- Oomen, C. C., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, *30*, 163-184.
- Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 539–557.
- Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica*, *142*, 332-342.
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*, 427–459.
- Pisoni, A., Papagno, C., & Cattaneo, Z. (2012). Neural correlates of the semantic interference effect: New evidence from transcranial direct current stimulation. *Neuroscience*, *223*, 56-67.
- Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proceedings of the National Academy of Sciences*, *91*, 7398-7403.
- Postma, A., & Kolk, H. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech, Language, and Hearing Research*, *35*, 537-544.
- Postma, A., & Kolk, H. (1992). Error monitoring in people who stutter: Evidence against auditory feedback defect theories. *Journal of Speech, Language, and Hearing Research*, *35*, 1024-1032.
- Postma, A., & Kolk, H. (1993). The covert repair hypothesis prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech, Language, and Hearing Research*, *36*, 472-4.
- Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, *39*, 375-392.
- Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, *17*, 682-691.
- Race, D. S., & Hillis, E. (2014). Neural bases of word representations for naming. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.). *The Oxford handbook of language production* (pp. 105 –117). Oxford: Oxford University Press.

- Rapp, B., Benzing, L., & Caramazza, A. (1997). The autonomy of lexical orthography. *Cognitive Neuropsychology*, *14*, 71–104.
- Rapp, B. & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460 – 499.
- Ridderinkhof, R. K. (2002). Micro-and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research*, *66*, 312-323.
- Riès, S. K., Xie, K., Haaland, K. Y., Dronkers, N. F., & Knight, R. T. (2013). Role of the lateral prefrontal cortex in speech monitoring. *Frontiers in Human Neuroscience*, *7*:703.
- Roelofs, A. (1998). Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition*, *1*, 94-95.
- Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, *111*, 561–572.
- Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 353 – 368.
- Rubia, K., Smith, A. B., Brammer, M. J., & Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage*, *20*, 351-358.
- Shao, Z., Roelofs, A., Martin, R., & Meyer, A. S. (2015). Selective inhibition and naming performance in semantic blocking, picture-word interference, and color-word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Schlenk, K., Huber, W., & Wilmes, K. (1987). Prepairs and repairs: different monitoring functions in aphasic language production. *Brain and Language*, *30*, 226-244.
- Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*, 199-227.
- Schotter, E. R., Jia, A., Ferreira, V. S., & Rayner, K. (2014). Preview benefit in speaking occurs regardless of preview timing. *Psychonomic Bulletin & Review*, *21*, 755-762.
- Simonyan, K., & Fuertinger, S. (2015). Speech networks at rest and in action: Interactions between functional brain networks controlling speech production. *Journal of Neurophysiology*, *113*, 2967-2978.
- Slevc, L. R., & Ferreira, V. S. (2006). Halting in single word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language*, *54*, 515-540.
- Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes*, *28*, 597 – 614.
- Strijkers, K., Baus, C., Runnqvist, E., Fitzpatrick, I., & Costa, A. (2013). The temporal dynamics of first versus second language production. *Brain and Language*, *127*, 6-11.
- Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*, 912 – 928.

- Strijkers, K., Yum, Y. N., Grainger, J., & Holcomb, P. J. (2011). Early goal-directed top-down influences in the production of speech. *Frontiers in Psychology, 2*, 371.
- Ullsperger, M., & von Cramon, D. Y. (2001). Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *Neuroimage, 14*, 1387-1401.
- Van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 14*, 593-602.
- Vigliocco, G., & Hartsuiker, R.J. (2002). The interplay of meaning, sound and syntax in sentence production. *Psychological Bulletin, 128*, 442-472.
- Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*, 101-175.
- Weissberger, G. H., Wierenga, C. E., Bondi, M. W., & Gollan, T. H. (2012). Partially overlapping mechanisms of language and task control in young and older bilinguals. *Psychology and Aging, 27*, 959.
- Woollams, A. M. (2015). Lexical is as lexical does: Computational approaches to lexical representation. *Language, Cognition and Neuroscience, 30*, 395-408.

Figure Captions

Figure 1. An illustration of components within the monitoring system.

Figure 2. Graphic illustration of questions for future research. Three panels show pairs of topics that interface with one another—from left to right. The topmost portion of each panel illustrates a topic that is well-understood in the field, while the bottom illustrates a question that is less-understood.

Figure 1

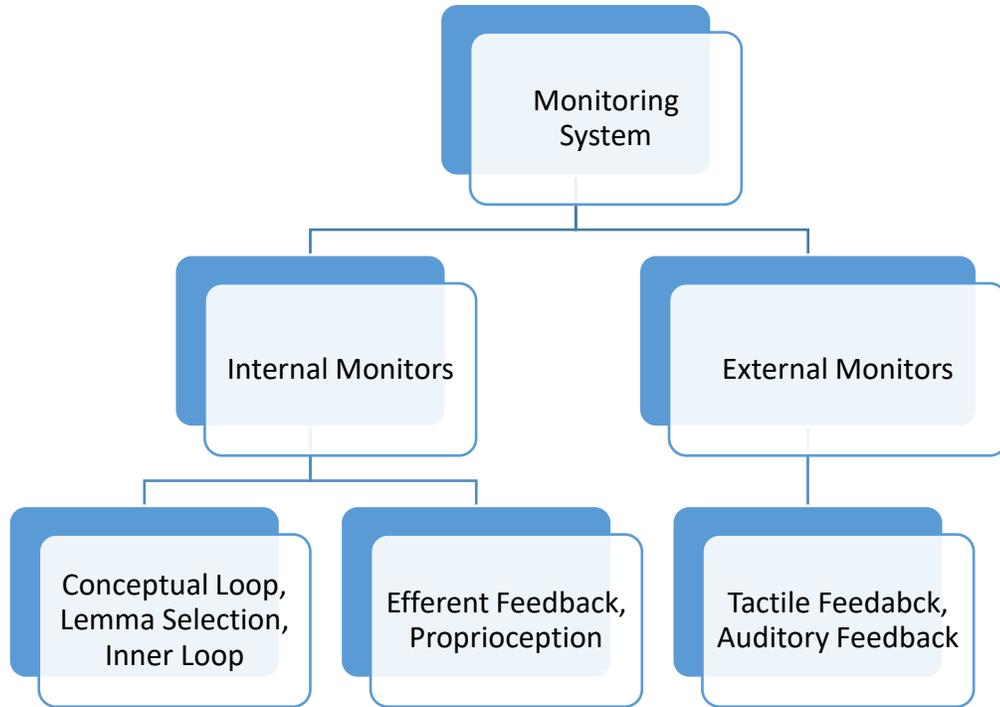


Figure 2

<i>Current focus of field</i>	The dynamics of conceptual, lexical, phonological activation during speech production	The integration of monolingual and bilingual perspectives	The division of labor between monitoring components
<i>Future directions</i>	Extent to which intention enhances word planning	Influence of attention and executive control on lexical selection	Influence of task on monitoring