

# 14

## **Relations between Speech Production and Speech Perception: Some Behavioral and Neurological Observations**

Willem J. M. Levelt

### **One Agent, Two Modalities**

There is a famous book that never appeared: Bever and Weksel (shelved). It contained chapters by several young Turks in the budding new psycholinguistics community of the mid-1960s. Jacques Mehler's chapter (coauthored with Harris Savin) was entitled "Language Users." A normal language user "is capable of producing and understanding an infinite number of sentences that he has never heard before. The central problem for the psychologist studying language is to explain this fact—to describe the abilities that underlie this infinity of possible performances and state precisely how these abilities, together with the various details . . . of a given situation, determine any particular performance." There is no hesitation here about the psycholinguist's core business: it is to explain our abilities to produce and to understand language. Indeed, the chapter's purpose was to review the available research findings on these abilities and it contains, correspondingly, a section on the listener and another section on the speaker.

This balance was quickly lost in the further history of psycholinguistics. With the happy and important exceptions of speech error and speech pausing research, the study of language use was factually reduced to studying language understanding. For example, Philip Johnson-Laird opened his review of experimental psycholinguistics in the 1974 *Annual Review of Psychology* with the statement: "The fundamental problem of psycholinguistics is simple to formulate: what happens if we understand sentences?" And he added, "Most of the other problems would be half-way solved if only we had the answer to this question." One major other

problem is, of course, How do we produce sentences? It is, however, by no means obvious how solving the issue of sentence or utterance understanding would halfway solve the problem of sentence or utterance production.

The optimism here is probably based on the belief that utterance production is roughly utterance understanding in reverse. In fact, that has long been a tacit belief among psycholinguists in spite of serious arguments to the contrary, such as these: “An ideal delivery in production requires completeness and well-formedness at all linguistic levels involved. The pragmatics should be precisely tuned to the discourse situation. The words, phrases, sentences should be accurate and precise renditions of the information to be expressed. Syntax and morphology have to be complete and well-formed and the same holds for the segmental and suprasegmental phonology of the utterance. Finally, the phonetic realization has to conform to the standards of intelligibility, rate, formality of the speech environment” (Levelt, 1996, p. x). These representations are generated completely and on the fly, that is, incrementally. This multi-level linguistic completeness and well-formedness is in no way required for successful utterance understanding. “Almost every utterance that we encounter is multiply ambiguous, phonetically (*I scream*), lexically (*the organ was removed*), syntactically (*I enjoy visiting colleagues*), semantically (*there are two tables with four chairs here*) or otherwise. As listeners we hardly notice this. We typically do not compute all well-formed parses of an utterance, even though ambiguities can produce momentary ripples of comprehension. Parsing is hardly ever complete. Rather, we go straight to the one most likely interpretation, given the discourse situation” (Levelt, 1996, p. x). In other words, the aims of the two systems are deeply different: attaining completeness is a core target of production; ambiguity is hardly ever a problem. Attaining uniqueness in the face of massive ambiguity is a core target of speech perception; completeness of parsing should definitely be avoided—it would make the system explode.

Meanwhile, the unbalance of comprehension vs. production perspectives in psycholinguistics has been somewhat redressed. After three decades, the title (and the content) of Herbert Clark’s 1996 treatise *Using Language* returns us to the language user who is as much a speaker as a

listener. It is not in detail Mehler and Savin's language user; that one was mainly concerned with relating surface phonetic and underlying semantic representations, but these are still core ingredients of a language user in Clark's sense, a participant in intentional, joint action.

It is, in my opinion, a major theoretical and empirical challenge to reconcile the unicity of the language user as an agent with the fundamental duality of linguistic processing, speaking, and understanding. In the following, I first consider some perception-production relations from the perspective of our speech production model, which has been taking shape over the years. I then change the perspective to cognitive neuroscience, not only because the recent neuroimaging literature often provides additional support for the existing theoretical notions but also because it provides new, additional challenges for how we conceive of the production-perception relations.

### **Speech Perception in a Model of Production**

The theory of production proposed in Levelt (1989) and the further modeling of its lexical access component in Levelt, Roelofs, and Meyer (1999) postulate three types of relation between the productive and receptive mechanisms of the language user. They reside in the "perceptual loop," in the concept-lemma system, and in connections at the form level. I consider them in turn.

#### **The Perceptual Loop**

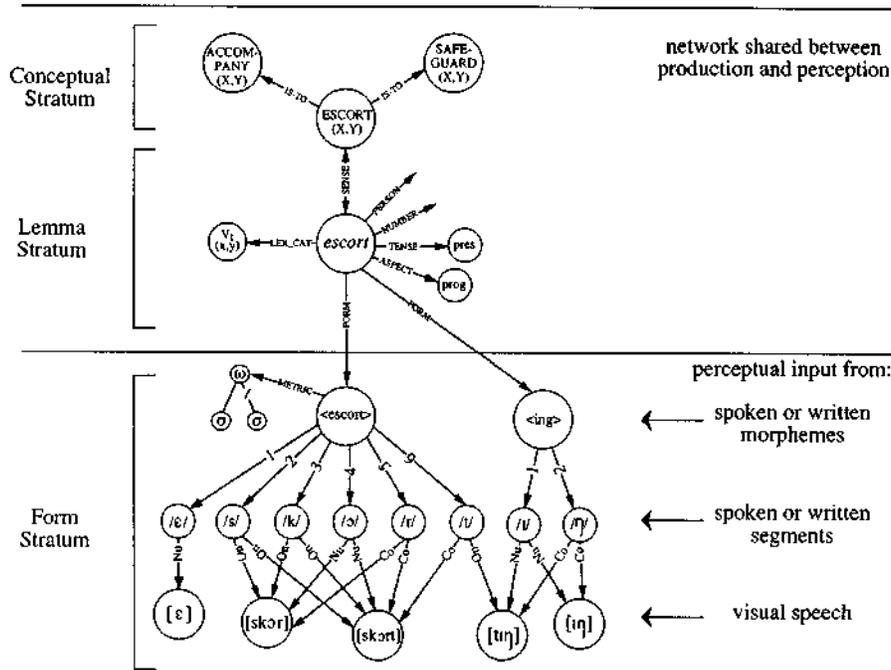
The global architecture of the speaker as proposed in Levelt (1989) is essentially a feedforward system. Utterances are incrementally produced going through stages of conceptual preparation, grammatical encoding, phonological encoding, and articulation. The perceptual component in this architecture consists of a dual-feedback loop. As a speaker you are normally hearing your own overt speech and that will be parsed just as any other-produced speech you hear or overhear. The one crucial difference is that the attribution of the speech is to self. That being the case, the speaker may use it for self-monitoring. As in any complex motor action, speaking involves some degree of output control. If the self-perceived speech is disruptively deviant from the intended delivery, you

may self-interrupt and make a repair (see Levelt, 1983, for details of this self-monitoring mechanism). As a speaker you can, in addition, monitor some internal, covert representation of the utterance being prepared. The 1989 model identified this “internal speech” as the phonetic code that serves as input to the articulatory mechanism, that is, the stuff that you can temporarily store in your “articulatory buffer” (Morton, 1970). Experimental evidence obtained by Wheeldon and Levelt (1994), however, supports the notion that the code is more abstract. Effective self-monitoring is not wiped out when the articulatory buffer is filled with nuisance materials. The more likely object of internal self-monitoring is the self-generated phonological representation, a string of syllabified and prosodified phonological words. Both the external and the internal feedback loops feed into the language user’s normal speech-understanding system. The model is maximally parsimonious in that it does not require any reduplication of mechanisms (as is often the case with alternative models of self-monitoring).

### **Shared Lemmas and Lexical Concepts**

The 1989 model follows Kempen and Hoenkamp (1987) in defining *lemmas* as the smallest units of grammatical encoding. These lemmas were semantic and syntactic entities. Retrieving lemmas for content words from the mental lexicon involved matching a conceptual structure in the input message to the semantic structure of the lemma. A major observation at that stage of theory formation was that all existing theories of lexical selection in speech production were deeply flawed. They all ran into the “hyperonym problem”: the speaker would select an item’s superordinates instead of, or in addition to, the target itself. This was, essentially, due to the fact that if a word’s critical semantic features are all activated or selected, then the critical features of any of its superordinates (or “hyperonyms”) are necessarily also activated or selected; they form, after all, a subset of the target word’s critical features. Roelofs (1992) solved this problem by splitting up the lemma into a “lexical concept” and a “syntactic lemma” (or “lemma” for short). If the terminal elements of the speaker’s message are “whole” lexical concepts, lexical selection is, essentially, a one-to-one mapping of lexical concepts to syntactic lemmas. Lexical concepts are no longer sets or bundles of features. Their semantics

is handled by the relational, labeled network that connects them. The hyperonym problem vanishes, and a more realistic issue arises: What happens if more than a single lexical concept, in particular semantically related concepts, are (co-)activated during conceptual preparation of the message? Multiple activation of lemmas will be the result and there is a selection problem that should be solved “in real time.” The computational model, now called WEAVER (Roelofs, 1992, 1997; see figure 14.1 for a fragment), handles this issue in detail and is meanwhile supported by a plethora of chronometric experimental data. The (syntactic) lemma is the unit of grammatical encoding. Lemmas are, essentially, lexical



**Figure 14.1** Fragment of the WEAVER lexical network (for lexical item *escort*), displaying the input connections from the perceptual network. The top half of the network is shared between perception and production. The bottom half of the network is specific to production, but is three-way-sensitive to input from the perceptual system. The perceptual form network is not shown. It produces perceptual input to the lemma stratum. Adapted from Levelt et al., 1999.

syntactic trees, called “lexical frames” by Kempen (2000; see also Vosse and Kempen, 2000). Kempen models grammatical encoding as a unification of incrementally selected lexical frames. The output is a surface syntactic tree for the utterance as a whole.

A standard method in word production research is the picture-word interference paradigm. The subject is presented with a picture to be named. At the same time, a distractor word is presented, either auditorily or printed in the picture. A semantically related distractor can affect the response latency. In the WEAVER model this is due to activation of the corresponding lemma. Levelt et al. (1999) proposed to account for this effect of perception on production by assuming that lemmas are shared between speech perception and production (leaving undecided how in detail orthographic input affects the corresponding lemma in the speech network). The obvious further step was to claim that the lexical networks for spoken word perception and spoken word production are shared from the lemma level upward. This rather drastic theoretical merger has an inevitable consequence. The feedforward from lexical concepts to lemmas, required for production, is now complemented by feedback from lemmas to lexical concepts, required for spoken word understanding, that is, concept-to-lemma connections are bilaterally activating. We are, as yet, not aware of empirical counterevidence to this proposition.

It would contribute to the aesthetics of this grand unification to make the further claim that grammatical encoding and grammatical decoding are one. This is exactly the step taken by Kempen (2000; see also Vosse and Kempen, 2000). In both encoding and decoding, lexical frames are incrementally unified. In the production case, the frames (lemmas) are conceptually selected (see above). In the perceptual case, the lexical frames are selected on the basis of a recognized phonological code. But then, in both modalities, the selected lexical frames unify incrementally, growing a syntactic tree, “from left to right.” Although this is a well-argued and attractive proposal, its empirical consequences need further scrutiny. For instance, it should be impossible for the language user to simultaneously encode and decode an utterance; that would mix up the lemmas in the unification space. Can you parse your interlocutor’s utterance while you are speaking yourself?

### Connections at the Form Level

Levelt et al. (1991) proposed that lemma-to-lexeme (= word form) connections are unilateral. Upon selection of the lemma, activation spreads to the form node, but there is no feedback. Both the selection condition and the nonfeedback claim have become controversial issues, leading to ever-more sophisticated experiments (see Levelt et al., 1999, and Levelt, 1999, for reviews). If the nonfeedback claim is correct, then the two networks cannot share the input and output form level nodes. In other words, phonological codes for perception and for production are not identical. Dell (1986) proposed a feedback mechanism to explain unmistakable facts of speech error distributions. There is usually a lexical bias in speech errors (they result in real words a bit more often than random segment changes would predict). And there is a statistical preponderance of mixed errors (such as *cat* for *rat*), where the error is both semantically and phonologically related to the target. But these properties can also be handled in terms of the internal loop monitoring mechanism discussed above. A real word and a word in the correct semantic field have a better chance of slipping through the monitor than a nonword or an odd word (Levelt, 1989). Another reason for keeping the perception and production form nodes apart are reports in the aphasiological literature of selective losses of word perception vs. word production (cf. Caplan, 1992).

Still, the form networks for perception and production must be connected in some way. Given the evidence from picture-word interference experiments, Levelt et al. (1999) made the following two assumptions: First, a distractor word, whether spoken or written, affects the corresponding morpheme node (which represents the phonological code) in the production network; the details of the perceptual mechanism were left unspecified (see figure 14.1). Second, active phonological segments in the perceptual network can also affect the corresponding segment nodes in the production lexicon. Again, the precise perceptual mechanism was left unspecified (see figure 14.1). A possible further specification may be achieved in terms of the merge model of Norris, McQueen, and Cutler (Meyer and Levelt, 2000). In short, we assume the existence of close connections at the form level, though without sharing of segmental or morphological nodes. What this means in neuroarchitectonic terms is a fascinating issue, to which I return below.

Still entirely open are potential relations at the phonetic level. Gestural scores in the production model, whether for segments, syllables, or whole words, are abstract motor representations. Levelt and Wheeldon (1994) proposed the existence of a “syllabary,” a repository of gestural scores for high-frequency syllables. English (just as Dutch or German) speakers do about 80 percent of their talking with no more than 500 different syllables. One would expect such highly overlearned motor actions to be appropriately stored in the frontal region of the brain. In our model (Levelt et al., 1999), these syllabic gestures are accessed on the fly, as phonological encoding proceeds. As soon as a selected lemma has activated its phonological code, the activated phonological segments are incrementally packaged into phonological syllables, often ignoring lexical boundaries (as in *I'll-sen-dit*). Each such composed syllable accesses “its” gestural score, which can be almost immediately executed by the articulatory system. We have, so far, not made any assumptions about potential relations between retrieving gestural scores and perceptual representations of syllables or other sublexical units. I presently return to that issue.

### **Some Neurophysiological Observations**

Mehler, Morton, and Jusczyk concluded their extensive paper “On Reducing Language to Biology” (1984) as follows: “We have argued that if a mapping between psychological processing and neurophysiological structures is possible, it will only come about after the key theoretical constructs are established for each level of explanation. In the interim, there are restricted circumstances in which neurophysiological observations can play a role in the development of psychological models. But these circumstances require the consideration of such evidence, not only in terms of the physiological organization of the brain, but also in terms of its functional organization” (p. 111). This statement has gained new force in the present era of cognitive neuroimaging. What emerges as relevant and lasting contributions after a decade of exploring the new tools are those studies that were theoretically driven. They are the studies where the experimental and control tasks are derived from an explicit theory of the underlying cognitive process. Where this was not the case, as in most neuroimaging studies of word production, a post hoc theoretic-

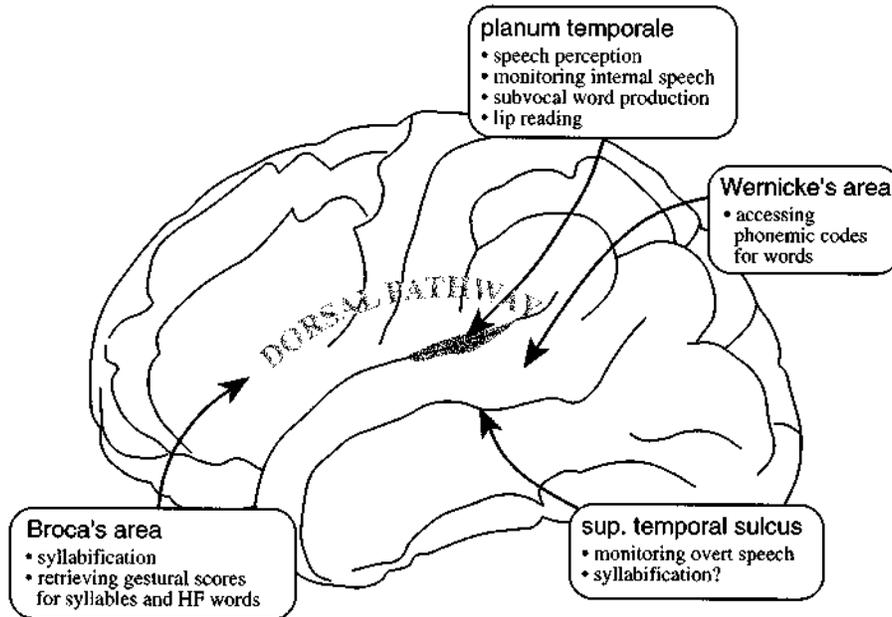
cally driven meta-analysis can still reveal patterns in the data that were not apparent in any of the individual studies themselves (Indefrey and Levelt, 2000; Levelt and Indefrey, 2000).

Since 1984 major strides have been made in unraveling the functional organization of both speech production and perception, in particular as far as the production and perception of spoken words are concerned. In addition, as discussed in the previous section, the issues about the relations between these two processing modalities can now be stated in explicit theoretical terms. In other words, the “interim” condition in the conclusion of Mehler et al. is sufficiently satisfied to scan the recent neuroimaging evidence for suggestions that can further the modeling of production-perception relations. Here are a few such “neurophysiological observations”:

### **Self-Monitoring**

According to the perceptual loop hypothesis, reviewed above, both the external and the internal feedback are processed by the language user’s normal speech-understanding system. This hypothesis is controversial in the literature (MacKay, 1992; Levelt 1992; Hartsuiker and Kolk, 2001), not only for its own sake but also because of its role in accounting for lexical bias and mixed errors in the speech error literature. McGuire, Silbersweig, and Frith (1996) have provided PET data that strongly support the notion of the speech perceptual system being involved in self-monitoring. If self-generated speech is fed back in a distorted fashion (pitch-transformed), there is increased bilateral activation of lateral temporal cortex (BA 21/22), and in particular of the left superior temporal sulcus (see figure 14.2). A highly similar activation of temporal areas is obtained when not the own voice but an alien voice is fed back to the speaker. These and other findings lead to the conclusion “that i) self- and externally-generated speech are processed in similar regions of temporal cortex, and ii) the monitoring of self-generated speech involves the temporal cortex bilaterally, and engages areas concerned with the processing of speech which has been generated externally” (McGuire et al., 1996, p. 101).

However, these data only concern the external loop. Is the auditory perceptual system also involved when the subject speaks “silently,” not



**Figure 14.2**

Schematic lateral view of the left hemisphere with pointers to regions discussed in the text.

producing an overt auditory signal? Paus et al. (1996) have provided relevant evidence on this point. In their PET experiment the subject whispered a string of syllables, with rate of production as the independent variable (30 through 150 syllables per minute). The authors observed a concomitant increase of activation in left auditory cortex, affecting two regions in particular: (1) an auditory region on the planum temporale, just posterior to Heschl's gyrus, and (2) an auditory region in the caudal portion of the Sylvian fissure. Hence, these regions might qualify as candidates for the reception of the speaker's internal feedback. Finally, Levelt and Indefrey (2000) discuss the possibility that the midsuperior temporal gyrus activation that is obtained even in nonword production tasks (as in nonword reading) may reflect overt or covert self-monitoring.

The Indefrey and Levelt meta-analysis has not provided the final answer with respect to the localization of rapid syllabification. As discussed

below, the imaging evidence points to both Broca's region and the mid-superior temporal gyrus (see figure 14.2). We do know, however, that internal self-monitoring concerns a syllabified phonological representation (Wheeldon and Levelt, 1995). If that representation is created in Broca's area, it must be fed back to the relevant areas in the superior temporal gyrus where the monitoring takes place. Which anatomical pathway could be involved here? There is less of an anatomical problem if phonological syllabification itself involves the mid or posterior regions of the superior temporal gyrus (or as suggested by Hickok and Poeppel, 2000, some inferior parietal region).

### **Phonological Codes**

The meta-analysis by Indefrey and Levelt (2000) strongly supports Wernicke's original notion that retrieving a word's phonological code for production involves what has since been called Wernicke's area. The neuroimaging data show, in particular, that if word production tasks (which all involve retrieving the words' phonological codes) are compared to nonword reading tasks (which do not require access to a phonological code), the critical difference in activation concerns Wernicke's area. In an MEG study of picture naming, Levelt et al. (1998) also showed dipoles in Wernicke's area, more precisely in the supratemporal plane vicinity, becoming active during a time interval in which phonological access is achieved. As discussed above, phonological codes in the production system can be perceptually primed, both by auditory and visual word or segment/letter stimuli. This may well involve part of Wernicke's area. Which part? Zattore et al. (1992) and Calvert et al. (1997) have shown the involvement of the left posterior supratemporal plane (STP), or *planum temporale*, in the perception of speech, and even in active lip reading. Recently, Hickok et al. (2000), in an fMRI study of picture naming, showed this area also to be involved in the subvocal production of words. If this is the region where the supposed linkage between the perceptual and production systems is established, then it is important to note that the relevant perceptual output is quite abstract. It may as well result from speech input as from reading the lips. Can it also be the output of seeing a printed word? If so, we are probably dealing with a phonemic level of representation.

A review by Hickok and Poeppel (2000), combining several sources of patient and imaging data, suggests that the auditory-to-motor interface involves a “dorsal pathway” from the just-mentioned region in the left posterior superior temporal gyrus through the inferior parietal lobe toward the frontal motor speech areas (see figure 14.2). They make the explicit suggestion that the auditory-motor interface operations have their site in the left inferior parietal lobe. Although this issue is far from solved, one must be careful not to confuse auditory phonemic codes and phonemic codes for production. They are clearly linked, behaviorally and in terms of functional anatomy, but they are not identical.

### **Syllabification and Phonetic Encoding**

The above-mentioned meta-analysis of word production studies (Indefrey and Levelt, 2000) pointed to the left posterior inferior frontal lobe and the left midsuperior temporal gyrus as being involved in phonological encoding, which is largely rapid syllabification in word production tasks. Not surprisingly, the same study showed bilateral, mostly ventral sensorimotor area involvement in actual articulation. Some speculation should be allowed in a Festschrift. The mentioned inferior frontal lobe involvement may extend beyond strict phonological encoding and also include accessing the gestural scores for successive syllables. In other words, the region would somehow store, retrieve, and concatenate our overlearned articulatory patterns for syllables and other high-frequency articulatory units, such as whole bi- or trisyllabic high-frequency words. More specifically, one would conjecture the involvement of Broca’s area, which is, in a way, premotor area. It is, after all, the human homologue of premotor area F5 in the macaque (Broca’s area is, however, different from F5 in terms of its cytoarchitecture; it contains a layer IV, which is absent in our and the macaque’s premotor cortex; K. Zilles, personal communication). Having mentioned F5, a discussion of mirror neurons is unavoidable (see Gallese and Goldman, 1999, for a review). If (the larger) Broca’s region is involved in rapidly accessing syllabic gestures, and if it shares with F5 its mirror-neuron character, one would expect the area to represent *perceived* articulatory gestures as well, that is, what one sees when looking at a speaking face. As early as 1984, Mehler et al. discussed such a possibility in connection with Ojemann’s (1983) observation that the

(larger) cortical area “has common properties of speech perception and generation of motor output.” These common properties, according to Ojemann, may serve functions “described by the motor theory of speech perception” (cf. Liberman, 1996).

If this speculation is of any value, one should be able to prime the production of a syllable by having a speaker look at a speaking face (on the monitor) that produces the same target syllable. That experiment, with all its controls, was recently run by Kerzel and Bekkering (2000), with the predicted result. Still to be shown is that the effect involves Broca’s area or immediately neighboring regions. If so, we are back to Wernicke in a new guise. Wernicke located the auditory word images in the posterior superior temporal area and the motor word images in Broca’s area. He supposed the existence of a connection between the two areas, now known to be the arcuate fascicle. That connection is, according to Wernicke, involved in spoken word repetition: the auditory word image activates the corresponding motor word image. Our chronometric work on phonological encoding suggests that an “auditory image” can affect the activation of individual segments in the same or another word’s phonological code, that is, the code on which word production is based. As mentioned above, retrieving this detailed, segmented code seems to involve Wernicke’s area.

In the WEAVER model, these phonological segments directly affect the activation state of all gestural scores in which they participate. These whole “motor images” for high-frequency syllables are now suggested to be located in the Broca or premotor region. If they are indeed “mirror images,” they represent both our own overlearned articulatory gestures and the ones we perceive on the face of our interlocutors. However, they are not auditory images. The auditory-phonemic link between spoken word perception and word production can stay restricted to Wernicke’s area, the midsuperior temporal gyrus, the planum temporale, and maybe the inferior parietal region, as long as no evidence to the contrary appears.

Finally, it should be noted that among these syllabic scores are all the articulatory scores for high-frequency monosyllabic *words*. It is a minor step to suppose that there are also whole stored scores for high-frequency combinations of syllables, in particular of disyllabic or even trisyllabic

high-frequency words. If, as Wernicke supposed, these motor images can also be directly activated by “object images,” that is, conceptually, spoken word production is to some extent possible without input from Wernicke’s area. However, the fine-tuning of phonological word encoding requires access to detailed phonological codes. Precise syllabification without paraphasias, resyllabification in context (such as *liaison* in French), and correct stress assignment will always depend on accurate phonemic input from our repository of phonological codes, which, I suggest, involves Wernicke’s area. Indefrey et al. (1998) provide fMRI evidence that their assembly involves left lateralized premotor cortex. Phonotactic assembly and the ultimate access to stored articulatory patterns are somehow handled by posterior inferior frontal areas.

### Acknowledgments

I am grateful to Gregory Hickok for an e-mail exchange on the functions of Wernicke’s area, and to Peter Indefrey for many helpful comments on a draft version of this paper.

### References

- Bever, T., and Weksel, W. (Eds.) (shelved). *The structure and psychology of language*.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., and David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276, 593–596.
- Caplan, D. (1992). *Language: Structure, processing, and disorders*. Cambridge, MA: MIT Press.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493–501.
- Hartsuiker, R., and Kolk, H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, 42, 113–157.
- Hickok, G., and Poeppel, D. (2000). Towards a functional anatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.

- Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A. K., Naeve-Velguth, S., Strupp, J. P., Strick, P. L., and Ugurbil, K. (2000). An fMRI study of the role of left posterior superior temporal gyrus in speech production: Implications for the explanation of conduction aphasia. *Neuroscience Letters*, 287, 156–160.
- Indefrey, P., and Levelt, W. J. M. (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences*. Cambridge, MA: MIT Press.
- Indefrey, P., Gruber, O., Brown, C., Hagoort, P., Posse, S., and Kleinschmidt, A. (1998). Lexicality and not syllable frequency determines lateralized pre-motor activation during the pronunciation of word-like stimuli—An fMRI study. *NeuroImage*, 7, 54.
- Johnson-Laird, P. N. (1974). Experimental psycholinguistics. *Annual Review of Psychology*, 25, 135–160.
- Kempen, G. (2000). Human grammatical coding. Unpublished manuscript, Leiden University, Leiden, Netherlands.
- Kempen, G., and Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201–258.
- Kerzel, D., and Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus-response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 26-2, 634–647.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). The perceptual loop theory not disconfirmed: A reply to MacKay. *Consciousness and Cognition*, 1, 226–230.
- Levelt, W. J. M. (1996). Foreword. In T. Dijkstra, and K. de Smedt (Eds.), *Computational psycholinguistics*. London: Taylor & Francis.
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3, 223–232.
- Levelt, W. J. M., and Indefrey, P. (2000). The speaking mind/brain: Where do spoken words come from? In A. Marantz, Y. Miyashita, and W. O’Neil (Eds.), *Image, language, brain*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Praamstra, P., Meyer, A. S., Salmelin, R., and Kiesela, P. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, 10, 553–567.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., and Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122–142.

- Levelt, W. J. M., and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.
- Liberman, A. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- McGuire, P. K., Silbersweig, D. A., and Frith, C. D. (1996). Functional anatomy of verbal self-monitoring. *Brain*, 119, 101–111.
- MacKay, D. G. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition*, 1, 199–225.
- Mehler, J., and Savin, H. (shelved). In T. Bever, and W. Weksel (Eds.), *The structure and psychology of language*.
- Mehler, J., Morton, J., and Jusczyk, P. W. (1984). On reducing language to biology. *Cognitive Neuropsychology*, 1, 83–116.
- Meyer, A. S., and Levelt, W. J. M. (2000). Merging speech perception and production. Commentary on target paper by Norris, D., McQueen, J. M., and Cutler, A. *Behavioral and Brain Sciences*, 23, 339–340.
- Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press.
- Ojemann, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Sciences*, 6, 189–206.
- Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., and Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8, 2236–2246.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284.
- Vosse, T., and Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–143.
- Wheeldon, L. R., and Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311–334.
- Zattore, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846–849.