16

Defective Neural Motor Speech Mappings As a Source for Apraxia of Speech

Evidence from a Quantitative Neural Model of Speech Processing

BERND J. KRÖGER, NICK MILLER, AND ANJA LOWIT

Introduction

Since apraxia of speech (AOS) emerged as a focus of attention in the 1960s (Darley, Aronson, & Brown, 1975) the design and interpretation of studies has been dogged by variations in definitions used and consequent characteristics of cases investigated. The plethora of psycholinguistic, neurological, and speech motor control models employed to derive or justify diagnostic markers has led to sets of clinical criteria that are not all universally accepted. A major obstacle to investigations is the rarity with which AOS occurs in isolation from other disorders. Nevertheless, it generally is accepted that AOS represents some kind of distinct motor speech disorder (Ogar, Slama, Dronkers, Amici, & Gorno-Tempini, 2005) or a form of a phonetic-motoric disorder (McNeil, 2008), separating it from other neurological disorders of speech and language, most notably aphasia and dysarthria (Jordan & Hillis 2006). From a cognitive or functional point of view, AOS has been labeled or defined as "inefficiencies in the translation of well-formed and -filled phonological frames into previously learned kinematic information used for carrying out intended movements" (McNeil, 2008, p. 264), or

more generally "impairment in the translation of phonological representations into specifications for articulation" (Croot, 2002). On the one hand, these cognitive definitions of AOS separate it from aphasia since linguistic processing (comprising conceptual, lexical, and grammatical processing) is not deemed to be impaired. On the other hand, they separate AOS from the dysarthrias since it is the specification of articulations but not the basic neuromuscular articulatory system *per se*, that is believed to be impaired. Following these definitions, both the linguistic processing system as well as the entire speech production apparatus (lungs, larynx, pharynx, nasal and oral cavity, lower jaw, lips, tongue, and velum) and its muscular system, including (peripheral) neuromuscular activation, are seen as separate and unimpaired. Furthermore, most definitions cite primary peripheral sensory (auditory and somatosensory, i.e., tactile and proprioceptive) processing as unimpaired in AOS (see McNeil, 2008).

Behavioral definitions of AOS rely on descriptions of typical symptoms such as "intra- and inter-articulator temporal and spatial segmental and prosodic distortions, (. . .) distortions of segments and intersegment transitionalization," with errors being "relatively consistent in location within the utterance and invariable in type" (McNeil, Robin, & Schmidt, 1997, p. 329). Earlier summaries pointed to "(1) effortful, trial and error groping articulatory movements and attempts at self-correction, (2) dysprosody unrelieved by extended periods of normal rhythm, stress, and intonation, (3) articulatory inconsistency on repeated productions of the same utterance, (4) obvious difficulty initiating utterances" (Wertz, LaPointe, & Rosenbeck, 1984, p. 81).

A major drawback of behavioral definitions of AOS is that none of these symptoms can be accepted as unambiguous or strong indicators for the disorder (Croot, 2002). Numerous signs, such as inconsistency in production errors, remain highly controversial, and debate continues as to how far many of the perceived segmental errors (insertions, elisions, segmental changes, for example, from voiced to voiceless, etc.) are not actually segmental in nature but are associated with deficits in the overall coordination of articulatory movements.

Furthermore, a lack of a comprehensive definition provides transparency between functional, behavioral, and psycholinguistic conceptualizations, motoric characterizations, and neurophysiological and anatomical specifications of AOS. One barrier to this is that the brain networks associated with translating phonological representations into specifications for articulation are distributed widely over cortical and subcortical regions (Hickok and Poeppel, 2004; Hillis et al. 2004; Miller, 2002; Rieker, Brendel, Ziegler, Erb, & Ackermann, 2008) and distinct brain regions can be active during very different tasks (speech and nonspeech). Nevertheless several writers have called for a definition of AOS based on a detailed and quantitative model of speech processing comprising cognitive as well as sensorimotor aspects of speech production (Croot, 2002; Miller, 2000; Miller, 2002) as the only way to proceed to a full understanding of AOS.

The aim of this chapter is to demonstrate how a quantitative neural model of speech processing, comprising both

cognitive and sensorimotor aspects of speech production, works toward a more comprehensive understanding of AOS. Such a model could lead to insights into possible underlying neural functional processes of speech production and, particularly, the relations between neural dysfunctions in the process of speech production and the resulting articulatory misbehavior. We will proceed by highlighting the main features of the model and then illustrate how so-called "typical symptoms" of apraxia of speech arise from lesions at different points in the model. Through this, we hope to throw more light onto the nature of the mechanisms in speech apraxic break down and the relationship between perceived and underlying disturbance in AOS.

An Action-Based Quantitative Neurocomputational Model of Speech Processing

The neurocomputational action-based model (ACT, Figure 16–1) is described in detail in Kröger, Kannampuzha, and Neuschaefer-Rube (2009) and Kröger, Kannampuzha, Lowit, and Neuschaefer-Rube (2009). As these papers have

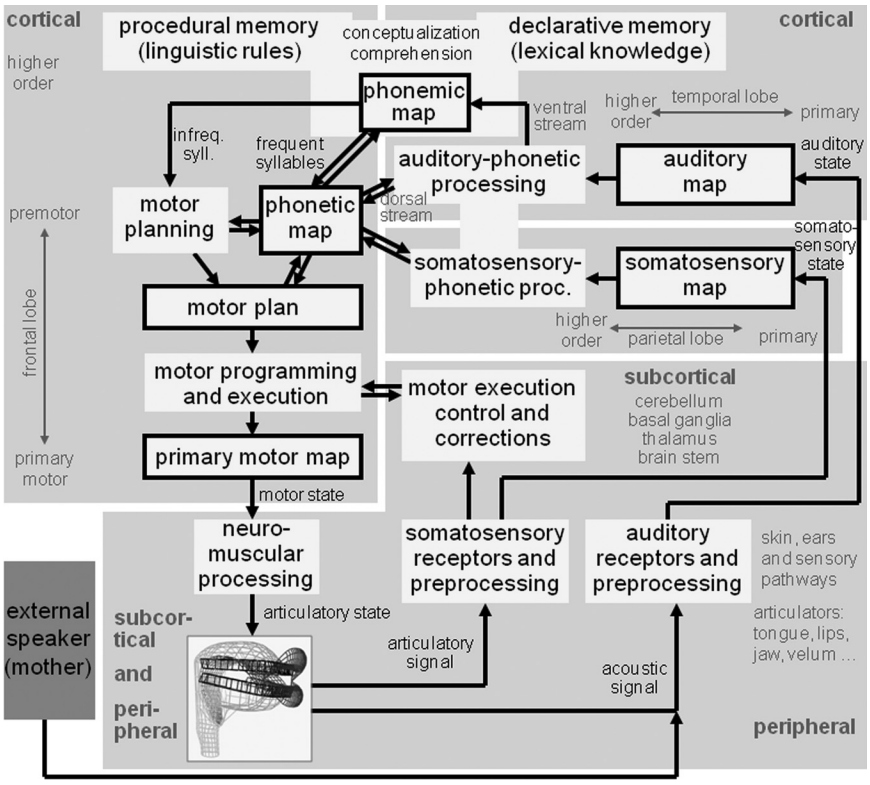

Figure 16–1. Organization of the neural model. Boxes with black outline represent neural maps. Arrows indicate processing paths or neural mappings. Boxes without outline indicate processing modules.

shown, this model is capable of producing real articulatory and acoustic signals and simulating the normal speech acquisition process. The model introduced here is computer-implemented and, thus, strongly quantitative. Mappings co-activate neural states in those maps, which are connected by a mapping. Thus, the mappings between phonetic map, motor plan map, and sensory maps (see Figure 16–1) co-activate neural representations of phonemic states, auditory states, somatosensory states, and motor plan states of speech items under production. Thus, we assume that the phonetic map comprises mirror neurons capable of closely linking sensory and motor states. We further assume that the appropriate mappings are distributed widely within frontal, temporal, and parietal cortical regions. It is beyond the scope of this chapter to give a comprehensive discussion of anatomical location of brain functions for speech processing, but it should be noted that many of our neural processing stages are comparable to those defined by Guenther (2006) and Guenther, Ghosh, and Tourville (2006). These authors offer a comprehensive discussion of the anatomical cortical and subcortical location of functional modules, maps, and mappings for speech processing.

ACT is capable of producing articulatory movement patterns and acoustic signals by controlling a 3-D articulatory-acoustic model. In parallel to the quantitative neural model of speech production introduced by Guenther (2006) and Guenther et al. (2006) (DIVA), our model separates feed-forward and feedback control. A major difference of our approach compared to DIVA is the separation of motor planning and motor execution. This separation results from the assumption that speech movements (similar to limb movements) are controlled by action units. Using this concept (Goldstein, Byrd, & Saltzman, 2006; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007), a motor plan (also termed vocal tract action score or gestural score) specified for a speech item under production is the result of action planning. On the motor plan level, all gestures forming an utterance are selected, and their intergestural temporal coordination is specified. Subsequently, these gestures or vocal tract action units are executed; that is, gestures or vocal tract actions are the relevant control units for programming and executing articulatory movements. A differentiation of motor planning, programming, and execution also was introduced in the detailed model of speech production given by van der Merwe (2008). A shortcoming of her approach, however, is that it is not strongly quantitative, and the model cannot be tested by producing or perceiving speech items. Furthermore, it should be noted that van der Merwe's approach is not strongly action based but segment oriented.

The importance of the concept of action (action planning, programming, and execution) in AOS is argued by Miller (2000, 2002). Likewise, the importance of models of speech production embracing cognitive linguistic as well as sensorimotor aspects of speech production has been advocated (Miller, 2002; Croot, 2002). For example, the production of a labial closure or of a glottal opening is a "speech action" or "speech gesture." A complete language-specific system of speech actions is introduced by Kröger and Birkholz (2007) for standard German. Criticism

concerning the action concept, at least in its formulation as articulatory phonology (Browman & Goldstein, 1989, 1992), focused mainly on the lack of integration of the auditory domain into the gestural theory (Kohler, 1992). In our modeling approach, the concept of vocal tract action units or gestures is introduced as a concept for sensorimotor control of speech production. Thus gestures are not interpreted primarily as phonological or linguistic units. Rather, in our modeling approach, no linguistic unit is favored. Different linguistic units (i.e., features, phonological gestures, segments, syllable constituents such as onset, rhyme or coda, syllables, words, and larger prosodic units) are seen as potential and coexisting units of linguistic speech processing. These different units are ordered hierarchically: prosodic units can be subdivided into one or more words, words into one or more syllables and so on. Thus, phonological gestures have an intermediate status between segments and features. A bundle of features determines a gesture (e.g., manner and place features determine the labial closing gesture), and one or more gestures determine a segment (e.g., labial closing and glottal opening gesture determine the phoneme /p/). These linguistic units lead to specifications of sensorimotor action units (i.e., sensorimotor gestures). Moreover, in contrast to Browman and Goldstein (1992), gestural targets or goals are not seen primarily as articulatory targets in our approach but as functional goals that can be specified in a functional manner in the sensory (somatosensory and auditory) domain in our model. It should be emphasized that the phonetic map—that is, the core of the "mental syllabary" in our model (Kröger, ~~B. J.,~~ Kannampuzha, J., & Neuschaefer-Rube, C., 2009)—is organized on a syllabic level.

A further feature of our model is the assumption that a disruption of the phonetic to motor plan network (or mapping) can occur separately for different types of syllables (e.g., V, CV, and CCV) since the neural self-organization of the phonetic maps always leads to topologically connected or continuous subregions for these syllable types (see experiment 1). In addition, we assume that particular neural defects always occur in spatially connected subregions of the phonetic map. This allows the modeling of different degrees of severity when disturbances are introduced into the model, for example, defective mapping of V, CV, and CCV items as opposed to CCV items only (see Experiment 2).

ACT also assumes the existence of a motor planning module (Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C., 2009). The motor planning module forms a neural processing pathway (or route) for infrequent syllables. This arises from the fact that the storage of motor plans and sensory states of syllables as whole patterns within a mental syllabary implies practicing this syllable frequently until the motor pattern is "overlearned" (Levelt, Roelofs, & Meyer, 1999). This cannot be achieved for infrequent syllables. Therefore, a module for the generation of motor plans for infrequent or novel syllables must exist, and by extension, two neural pathways for motor planning are needed (Levelt et al., 1999).

A final feature of our model relevant to the current discussion is the postulation of four processing levels within the motor planning module (Table 16–1). Level 1 of the motor planning module is a generator for proto-gestures, which are

Table 16–1. Knowledge Stored on Different Levels of the Motor Planning Module

Level of Motor Planning Module	Knowledge Stored per Level for the Production of . . .
1	proto-gestures, for example, mouth opening for producing a proto-vowel, tongue elevation for producing a proto-consonant
2	fine-tuned language-specific gestures (see language-specific gesture system in Kröger and Birkholz, 2007)
3	language-specific temporal coordination of gestures for syllables
4	words and sentences, that is, knowledge concerning connecting syllables to words and sentences and modification of gestures and gesture timing with respect to stress and intonation

gestures defined before any language-specific fine-tuning of targets has taken place (proto-gestures are explained in Kröger, Birkholz, Kannampuzha, & Neuschaefer-Rube, 2006). Levels 2 and 3 reflect the organization of (language-specific) motor plans. In order to produce a language-specific speech item, proto-gestures have to be fine-tuned, and intergestural temporal coordination must be fixed. A set of language-specific gestures and its temporal coordination within syllables is established after language-specific imitation training (Kröger et al., 2006, Kröger and Birkholz, 2007). The resulting knowledge concerning the set of language-specific gestures is stored on level 2 of the motor planning module while the knowledge concerning the language specific temporal coordination of gestures within syllables is stored on level 3. Level 4 is involved mainly in modifying gestures and gestural coordination with respect to connecting syllables into words and words into sentences. It, thus, involves specific prosodic categories such as different levels of stress (e.g., unstressed vs. stressed) or types of intonation.

Experiment 1: Learning a Model Language—The Unimpaired Speaker

In order to simulate different types of lesions with our model, it first had to be trained for normal speech production (and perception) as a "model speaker before stroke." The model language consisted of a five vowel system in which Vs were /i/, /e/, /a/, /o/, or /u/ along with nine consonants in which Cs were voiced and voiceless plosives /b/, /d/, /g/, /p/, /t/, and /k/, the nasals /m/ and /n/ as well as the lateral /l/. Consonants could be combined with all vowels and C_1C_2V syllables were trained in which C_1 is /b/, /p/, /g/, or /k/ and in which C_2 was always the lateral /l/, again in combination with all five vowels. The model language thus comprised 5 vow-

els, 15 CV syllables with voiced plosives, 15 CV syllables with voiceless plosives, 10 CV syllables with nasal consonants, and 20 CCV syllables (/plV/, /blV/, /klV/, and /glV/). Furthermore, the model was capable of processing two-syllable words composed from these syllables. All combinations of the 60 syllables occurred as words within the model language and these words were defined as trochee structures (i.e., with stress on the first syllable).

Motor plan states and sensory states of frequent syllables are stored as a whole by the phonetic to motor plan mapping and by the phonetic to sensory mappings (see arrows in Figure 16–1). This results from extensive training of these frequent syllables during speech acquisition (and further during lifetime). Thus, frequent syllables also are called well-practiced, overlearned, or automated syllables in terms of sensorimotor control. Infrequent syllables have to be assembled from subsyllabic parts such as onset, rhyme, or coda, single sound segments, or single vocal tract action units by the motor planning module. The neural pathway, consisting of the motor planning module, may be termed the gestural assembly route, analogous to the idea of a segmental assembly route as introduced by Levelt, Roelofs, and Meyer (1999). In our model language, the production of isolated vowels and most of the CV and CCV syllables was defined as high frequency and these were, therefore, stored in long-term memory—that is in the phonetic to motor plan and phonetic to sensory mappings introduced previously (comparable with the concept of mental syllabary, Levelt & Wheeldon, 1994; Levelt et al., 1999; Indefrey & Levelt, 2004). Only the syllables /lo/ and /ple/ were defined as infrequent syllables in our model language. Accordingly, the motor plan state for these syllables had to be assembled by the motor planning module. A typical ordering of the automated or well-practiced syllables within a 25 × 25 neuron self-organizing phonetic map is shown in Figure 16–2. This results from a babbling and imitation training experiment as described in Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C., 2009. One main result of this experiment was that vowels, CV-, and CCV- items were shown to capture different (cortical) regions within the phonetic map.

Experiment 2: Simulations of Different Types of Breakdown—The Virtual Apraxic Speakers

Different instances of the neural model can be trained by starting from (1) several initial settings of link weight values for the mappings, by using (2) different training items resulting from different randomization procedures, and by (3) varying orderings of training items (Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C., 2009). The resulting "trained models" represent different virtual speech processing units—that is, different virtual listeners and virtual speakers.

In the current study, versions of the model comprising different specific neural disruptions were introduced. In addition to the virtual unimpaired speaker described previously, four "impaired" versions were trained that exhibited a variety of dysfunctions of certain neural maps and mappings and

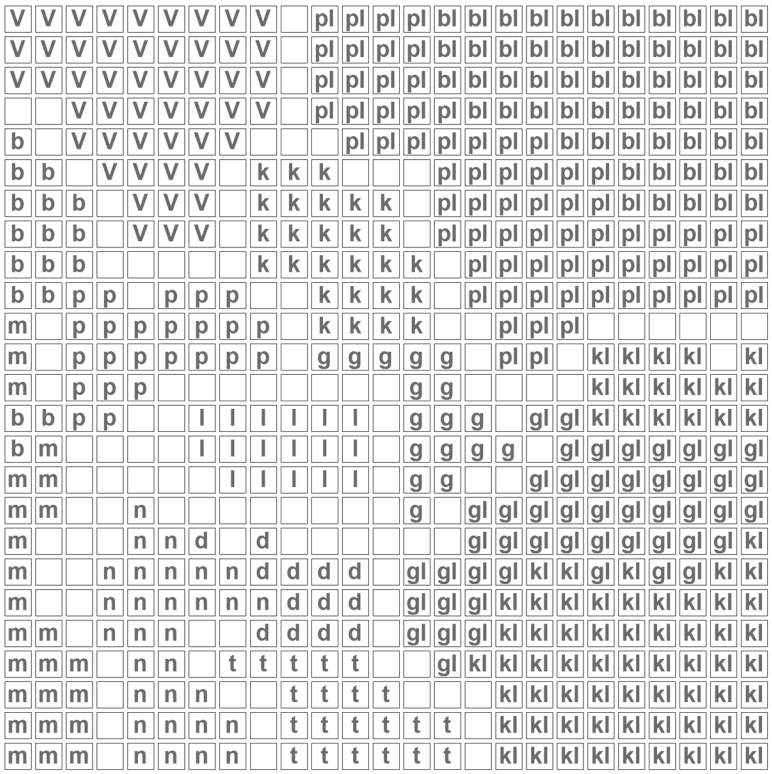

Figure 16–2. 25 × 25 neuron self-organizing phonetic map after babbling and imitation training for the model language (see text). A neuron is marked with a letter V if the neuron represents a vocalic state. Neurons are marked with lower case letters for C or CC if the neuron represents a CV or a CCV state. Clear regions can be found for [V, dV, gV, pV, tV, kV, nV, blV, glV, plV, klV]. A slight mixture of regions occurs for [mV, bV]. Vowels are not broken down to [i, e, a, o, u] in this figure. Unmarked neurons indicate states that cannot be associated clearly with any phonemic state.

varying dysfunctions of particular neural processing modules. The benefit of using a quantitative model of speech processing is that the processing of one speech item (specified on the phonological level) can lead to different anomalous articulation behaviors and different acoustic signals for that speech item due to the specific dysfunction introduced to the model. The resulting phonetic anomalies can be related to the specific neural deficits applied to the model speaker. Table 16–2 subsumes types of combinations of (1) disruptions within the phonetic to motor plan mapping (mental syllabary path for frequent syllables) and (2) disruptions within the motor planning module (gestural assembly path for infrequent syllables). Concerning the phonetic-to-motor plan mapping, we assume that this mapping can be disrupted as a whole or in parts with respect to different groups of speech items such as

Table 16–2. Speech Production Symptoms for Model Instances (Virtual AOS Speakers)

Virtual Speakers	Available (concerning: phonetic-to-motor plan mapping)	Available (concerning: motor planning module)	Symptoms
1 (severe form of AOS)		• Level 1: proto-gestures	• groping for vowels; • no CV, CCV syllables and words; • dysprosody
2	• V	• Level 1: proto-gestures • Level 2: vocalic gestures	• groping for consonants; • gestural timing errors for CV syllables; • no CCV syllables and words; • dysprosody
3a	• V • CV	• Level 1 • Level 2: all language-specific gestures	• gestural timing errors for infrequent CV syllables, • no CCV syllables and words; • dysprosody
3b	• V • CV	• Levels 1 & 2 • Level 3: timing for CV- syllables	• gestural timing errors for CCV- syllables and words; • dysprosody
4a	• V • CV • CCV	• Levels 1 & 2 • Level 3: timing for CV- syllables	• gestural timing errors for infrequent CCV syllables and words; • dysprosody
4b (mild form of AOS)	• V • CV • CCV	• Levels 1, 2 & 3	• gestural timing inaccuracies at syllable boundaries in words (e.g., pauses at syllable boundaries); • dysprosody
normal virtual speaker	• V • CV • CCV	• Level 1, 2, 3 & 4	• none

Note. These symptoms resulted from specific defects of the motor planning module and/or the phonetic to motor plan mapping. It is assumed that all syllables are realized as stressed if the prosodic part of the phonetic-to-motor plan mapping is defective.

CCV syllables, CV syllables, or V-items (vowels) since these groups of speech items form spatially connected regions at the level of the self-organizing phonetic map (see Figure 16–2).

As regards the motor planning module, we assume it may be disrupted in a top-down direction—that is, the motor planning module can be disrupted at higher levels of motor plan specification while lower levels of motor plan specification remain available. The aim of these simulation experiments with contrasting types of "virtual AOS speakers" was to describe the varying speech outcomes from the different lesions and to compare these to features claimed as characteristic of AOS in real speakers. Thereby, we aimed to contribute to the debate on possible origins of speech apraxic disruptions within models of speech output in relation to varying types of neural disruption. An overview of virtual speakers exhibiting specific neural defects and of the resulting AOS signs or symptoms is provided in Table 16–2.

Virtual Speaker 1: Lesion at the Level of Phonetic-to-Motor Plan Mapping for Vowels and Consonants

In this version of the model, the phonetic-to-motor plan mapping was cut off entirely, rendering the speaker incapable of executing any language-specific motor plans for vowels or syllables (Experiment 1). However, it is assumed that the phonetic-to-auditory and the phonetic-to-somatosensory mappings remain unaffected. Thus, despite the fact that the virtual speaker is unable to produce language-specific speech items, he still knows what a vowel or a frequent syllable should sound like and how a vowel or frequent syllable feels during production when he activates this speech item at the level of the phonemic map. It is assumed that for this type of speaker, the motor planning module is disrupted starting at level 2, rendering him incapable of producing language-specific speech items, even via the motor planning module (vocal tract action assembly path). Only proto-actions can be activated from the motor planning module (see Table 16–2).

In order to investigate the effects of this disruption, the production of a simple task was simulated—that is, to produce a realization of the vowel /u/. This leads to the activation of the neuronal representation of the phonemic /u/ state within the phonemic map. Subsequently, this brings about coactivation of the auditory and somatosensory state of the stored /u/ realization via the phonetic map as was trained during speech acquisition (Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C., 2009). Thus, the virtual speaker activates the auditory, tactile, and proprioceptive neural state for /u/ while the motor plan state for a /u/-realization is inaccessible. Despite the fact that speaker 1 is not capable of producing the appropriate motor plan state, he still "feels" (1) the vocal tract state from the prestored proprioceptive activation pattern and (2) the tactile contact pattern from the prestored tactile activation pattern. Since the motor plan state of an /u/ realization is not available to speaker 1, he activates what he is able to—that is, several proto-vocalic actions—and compares the resulting somatosensory (i.e., tactile and proprioceptive) states of his current production trials with the somatosensory target state for a /u/ action. Figure 16–3 gives an

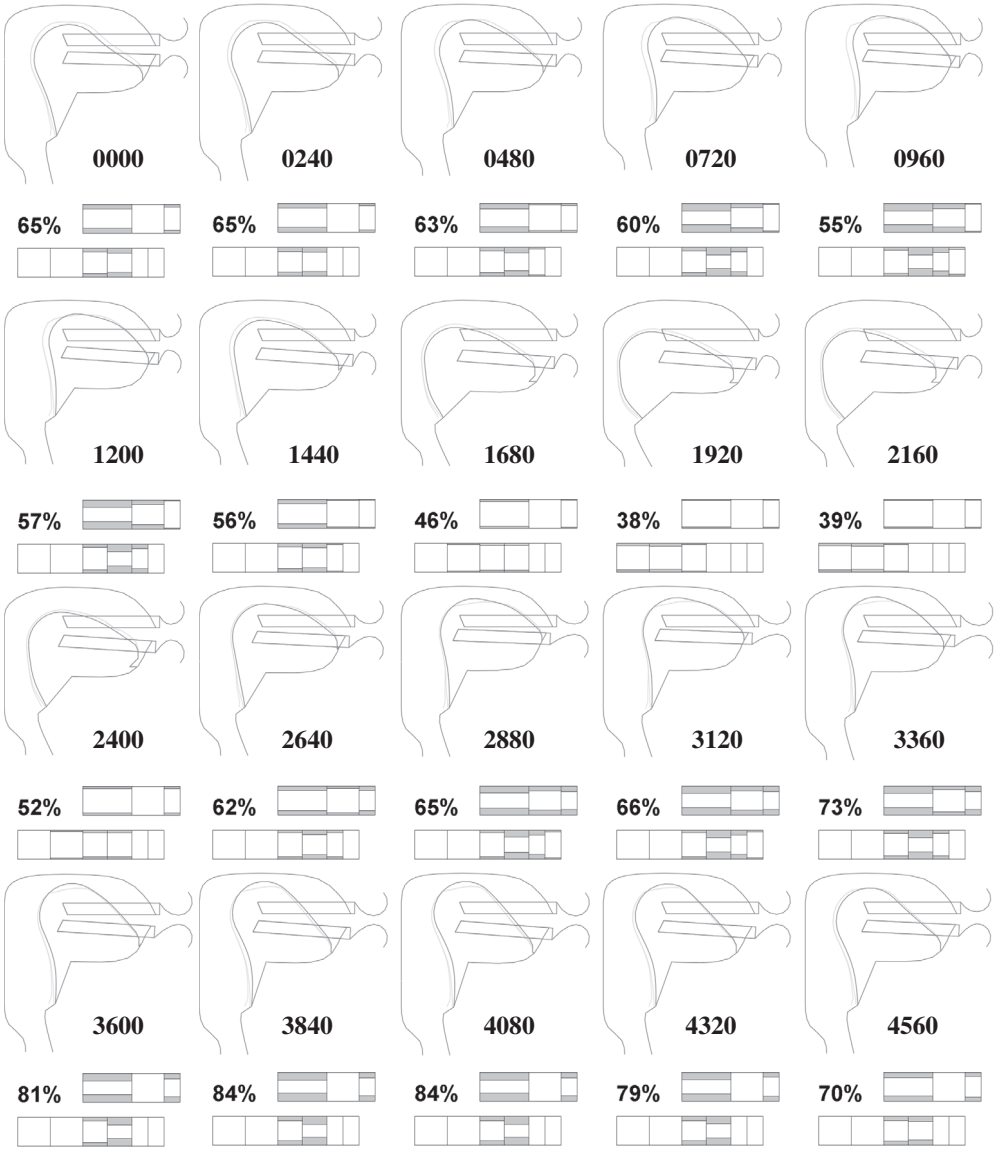

Figure 16–3. Trial and error groping of speaker 1. The speaker tries to articulate a realization of /u/. Time is indicated in ms. Degree of similarity of current somatosensory state with respect to somatosensory state of the /u/ realization is given in percent. The shaded areas in the tactile contact pattern indicate contact of vocal tract organs or articulators (adjacent to the percentage; from left to right: contact area of tongue body, tongue tip, lips) with regions of the vocal tract wall (below; from left to right: lower pharyngeal, upper pharyngeal, velar, palatal, post-alveolar, and alveolar region).

example for the resulting vocal tract movements. It can be seen that groping behavior occurs for this speaker. He successively produces a number of proto-vocalic actions, first a front-high-unrounded action (from 0 ms to about

960 ms, Figure 16–5), followed by a low-unrounded (from 960 ms to about 1920 ms), a front-high-rounded (from 1920 ms to about 3120 ms) and a back-high-rounded action (from 3120 ms to about 4080 ms).

While the speaker is executing these proto-actions, the somatosensory state is monitored online and compared to the prelearned somatosensory state of a typical /u/ action. This prelearned state is activated in parallel throughout the duration of the whole trial-and-error process since the speaker permanently activates the /u/ neuron within the phonemic map, leading to prelearned sensory co-activations. The comparison of current and prelearned somatosensory states is performed within the somatosensory phonetic processing module. If the current and the prestored somatosensory states are comparable (e.g., degree of similarity of neural states higher than 80% in the case of our simulation) the current proto-action and its intragestural parameter setting are retained in short-term memory, and the speaker now endeavors to co-activate pulmonary initiation and glottal phonation in order to make vowel production audible. A further refinement of the vocalic action realization toward the prestored /u/ can be attempted through comparisons of auditory states. It should be noted that proto-gestures are permitted to vary with respect to all intragestural parameters. Thus, the motor plan state of the vowel /u/ is re-attained by trial-and-error groping. Due to the severe neural defects of this virtual speaker, the production of motor plan items more complex than isolated vowels is assumed to involve too excessive demands and no co-articulation with other phonemes is deemed possible.

Virtual Speaker 2: Lesion at the Level of Phonetic–to-Motor Plan Mapping for Consonants Only

The neural defects modeled for speaker 2 are identical to those of speaker 1 with the exception that the vocalic part of the phonetic-to-motor plan mapping is unaffected (see Table 16–2). This virtual speaker is, therefore, capable of activating prestored motor plans for language-specific vowels from the phonetic-to-motor plan mapping, but unable to activate such plans for CV or CCV syllables, neither via the phonetic map nor the motor planning module. Similar to speaker 1, speaker 2 is able to activate the sensory states of all frequent syllables but in addition, also all vowels. In relation to consonants, only proto-actions can be produced since only level 1 of the motor planning module remains unaffected. As a result, the speaker will start to grope for the correct gesture during consonant production. For example, for the realization of the syllable /pa/, the speaker will start with successive productions of randomly chosen proto-consonantal labial, apical, and dorsal closing actions. Through being able to monitor productions in the somatosensory domain and possessing the facility to compare current productions and a prestored /pa/ realization, the speaker will maintain the proto-labial closing actions in short-term memory. However, to produce an acceptable realization of /pa/, speaker 2 now has to find the correct interaction temporal coordination for all actions of /pa/ (i.e., labial closing action, glottal opening action, and vocalic action) through trial-and-error productions. Typical segmental effects resulting

from these trial–and-error productions as generated by the model are provided in Figure 16–4.

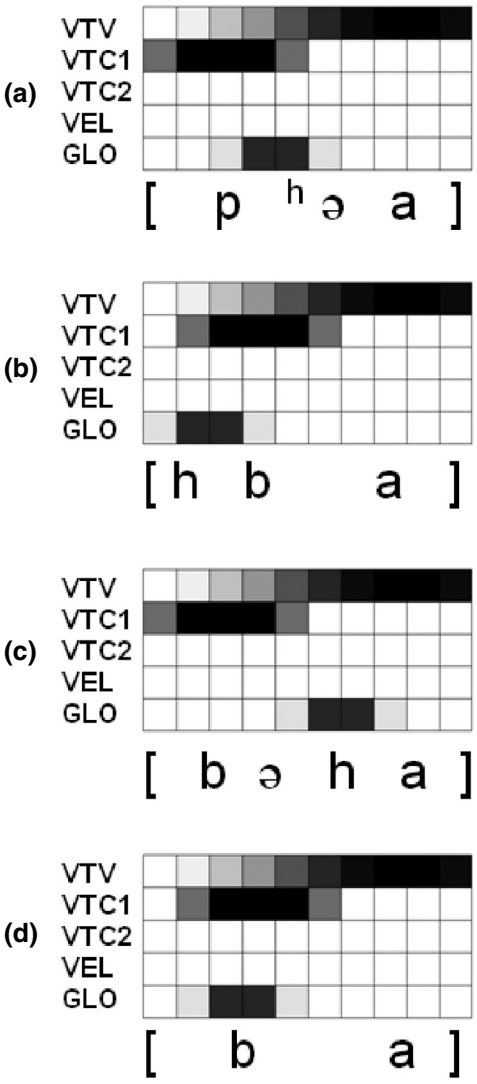

Figure 16–4. CV motor plan temporal activation patterns and appropriate phonetic transcriptions of the resulting acoustic speech signal produced by model speaker 2. This model speaker tries to produce an acceptable /pa/ realization. Trial **(a)**–**(d)**: severe mistiming of gestures resulting in deviating phonetic transcription for /pa/. *continues*

The neural activation pattern of the motor plans for /pa/ and the resulting acoustic signal are listed in Figure 16–4: (a) If the timing of the glottal opening action with respect to the consonantal vocal tract closing action is correct, but if the vocalic tract forming action starts too late with respect to the consonantal closing action, the perceptual impression of an inserted schwa-sound arises. (b) If the timing of the consonantal closing action and of the vocalic action is correct, but the glottal opening action starts too early with respect to the consonantal closing action, pre-aspiration can occur, and the consonant may be perceived as voiceless. (c) If the glottal opening action together with the vocalic action starts too late with respect to the consonantal closing action, schwa-insertion and [h]-insertion are perceived. (d) If the timing of the consonantal closing action and the vocalic action is

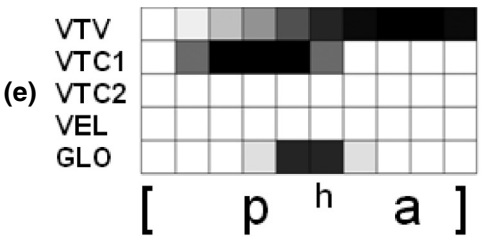

Figure 16–4. *continued* Trial **(e)**: proper gestural timing for a /pa/ realization. The gestural timing for **(a)**–**(e)** is illustrated by the motor plan neural activation patterns. These patterns are described in the text. Each box represents a motor plan neuron (white to black: no activation to full activation). The rows indicate gestural activation for actions; from top: vocalic tract forming action as part of V (VTV), consonantal vocal tract closing action as part of C1 and as part of C2 (VTC1, VTC2), velum action (VEL), and glottal action (GLO).

correct, but the glottal opening action is produced temporally synchronous with the consonantal closing action, the perceptual impression of a voiced plosive occurs. (e) The production of an acceptable /pa/ realization is achieved only if the temporal ending of the glottal opening gesture (i.e., time instant of maximum glottal opening) coincides in time with the termination of the consonantal closing action (i.e., time instant of release of consonantal closure). It is assumed that speaker 2 is capable of producing CV syllables after several trials at finding the correct temporal coordination.

Virtual Speakers 3a and 3b: Modeling of Gestural Timing Errors for CCV Syllables

The next progression in the location of neural defects is to have levels 1 and 2 intact, but a breakdown at level 3. That means that phonetic-to-motor plan mapping is functioning for V-items and CV syllables (see Table 16–2). The speaker is, therefore, capable of producing vowels and consonants and combining these to form syllables. The latter ability might be restricted to frequent syllables only (speaker 3a) or can include the production of infrequent syllables as well (speaker 3b). In the case of speaker 3a, infrequent CV syllables are produced by taking the information concerning action timing from phonetically similar frequent CV syllables—that is, by taking the knowledge from the motor part of the intact part of the phonetic map. Irrespective of the ability to produce such CV syllables, neither of the speakers is able to activate motor plans for CCV syllable via the phonetic map, but speaker 3b tries to produce CCV syllables since he is capable of producing all CV syllables. Typical errors in temporal coordination of actions for the CCV syllable for speaker 3b are illustrated in Figure 16–5 in which the model is instructed to produce /gla/. It is capable of producing

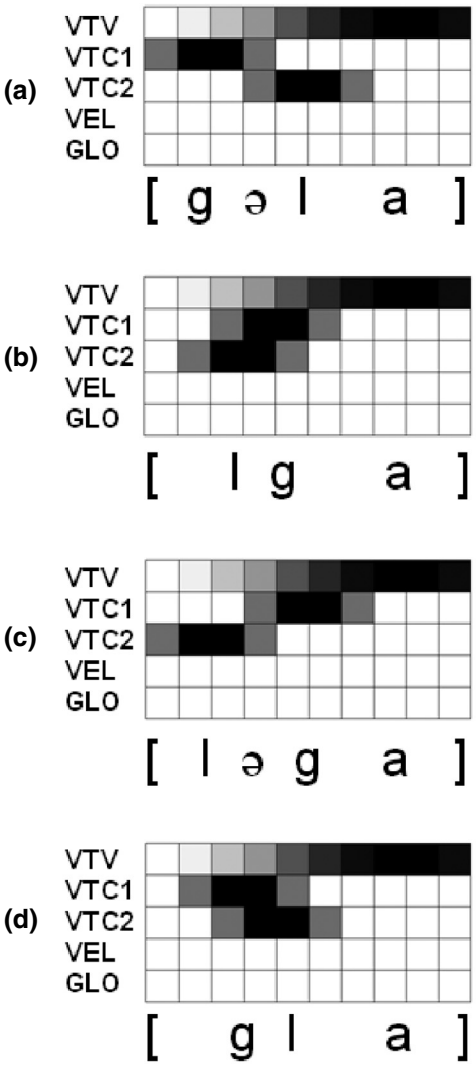

Figure 16–5. (a)–(c): CCV motor plan temporal activation patterns for typical timing errors occurring in realization of /gla/ for model speaker 3b (see text). **(d)**: correct neural activation pattern for the motor plan of /gla/.

the syllables /ga/ or /la/, but as can be seen, the timing errors of actions for /CCV/ syllables lead to segmental effects such as schwa-insertion in the consonant cluster or to metathesis of the two initial consonants with or without schwa-insertion.

Virtual Speakers 4a and 4b: Modeling of Gestural Timing Errors for Infrequent CCV Syllables and Dysprosody

Speakers 4a and 4b present a further decrease in severity of speech problem. Compared to speaker 3b, speaker 4a now has intact motor plans for all (frequent) V, CV, and CCV syllables. However, his knowledge of the correct temporal coordination for infrequent CCV syllables is still unavailable (see Table 16–2). The speech deviations for such syllables thus are comparable to those illustrated for speaker 3b (Figure 16–5). In the case of speaker 4a infrequent CCV syllables are produced by taking the information concerning action timing from phonetically similar frequent CCV syllables—that is, by taking the knowledge from the motor part of the phonetic map.

As a result of disruption of level 4 of the motor planning module, the main defect for speaker 4b, which also occurs as a side defect for all other speakers (1, 2, 3a, 3b, and 4a) in as far as these speakers are capable of producing at least two consecutive syllables, affects the prosodic make up of words. Model speaker 4b is, thus, characteristic of someone with a mild form of AOS. The difference between this speaker and the unimpaired model is that he is not capable of accessing information on how to produce an unstressed version of a syllable (all syllables within the phonetic map are assumed to be stressed versions of syllables) and how to connect two syllables correctly in order to form a word within the model language. Thereby, speaker 4b is capable of producing the syllables of the model language but realizes a bisyllabic word as two equally stressed syllables with an unnatural pause in between.

Discussion

The simulation experiments described in this paper have demonstrated that signs of AOS such as trial-and-error groping or different segmental errors (or effects) resulting from errors in temporal coordination can result from particular neural defects in mappings (here the phonetic-to-motor plan mapping) or processing modules (here, the motor planning module) within our neural model. These findings lend weight to speculations aired by Miller (2000) that within a perspective on speech motor control that integrates cognitive (e.g., attentional, short-term/ working memory, processing capacity), motor as well as neural strands (e.g., interconnectivity), speech derailments would be emergent properties of (disordered) interaction within and across tiers in the speech production system. This is even before one starts to ponder the possibilities for break down that could occur when considering possible interactions between speech processing and broader language processing (e.g., word retrieval; sentence stress assignment). One would not need to posit separate phonological or phonetic ordering, deletion, or insertion processes, and the like to explain perceived speech errors. This

is amply illustrated in virtual speaker 2 where schwa-insertion, apparent substitutions, additions, and omissions emerge from disruptions to the timing between gestures. Such a perspective also affords a transparency between perceived speech derailments and understanding them in terms of underlying alterations in neural functioning, such as levels of activation and inhibition, integration of feed-forward and feedback processes and perceptual and output processes. However, the virtual speakers illustrated in Table 16–2 give only a first very broad indication of possible subtypes of AOS speech disruptions. Basic disruption to the phonetic map (V, CV, CCV, prosody) could lead to further subtypes.

It is beyond the scope of this chapter to discuss the different levels of knowledge in specifying the temporal coordination (or intergestural timing) of all gestures forming a specific syllable or speech item. One would need to attend (1) to the detailed knowledge for each syllable attained by looking for the phonetically most similar syllable occurring within the mental syllabary (phonetic-to-motor plan mapping) and by adapting the intra- and intergestural parameters of this specific syllable for the target under production, or (2) to the broader knowledge differentiating types of syllable (e.g., CV with C = nasals; or C = voiced plosives; or C = voiceless plosives; or CCV with C1 = voiced plosives and C2 = /l/; or CCV with C1 = voiceless plosives and C2 = /l/, etc.), which can be generalized from the phonetic-to-motor plan mapping by processing gestural parameters over all CV syllables belonging to a specific syllable type. Although on the one hand the knowledge of how to specify the gestural parameters of a specific syllable in detail needs the online availability of the phonetic-to-motor plan mapping (i.e., needs a nondisrupted phonetic to motor plan mapping) in the current version of ACT; on the other hand, the knowledge for syllable types can be seen as a generalization of motor planning knowledge, and it can be assumed to be available even if phonetic to motor planning mapping is disrupted. This generalization of knowledge for the temporal coordination of gestures is not implemented in our model currently. It would lead to a further separation of the four severity stages suggested in Table 16–2 and is an area for future investigation.

Moreover, it should be noted that due to the speaker's ability to activate the somatosensory state of a syllable during production, groping too could be advantageous in order to prepare the production of a syllable silently. We assume that the somatosensory pattern can be used for selecting proto-gestures as was the case in our silent groping simulations, but we assume that somatosensation does not provide a sufficiently detailed or strong signal to set the correct timing of all gestures of a syllable in advance during silent groping, especially in the case of glottal and velopharyngeal gestures. This offers one explanation of why audible trial and error productions often follow after silent groping and why certain segments may appear more problematic than others.

The Dual Route Paradigm

In agreement with Levelt, Roelofs, and Meyer (1999), we assume a "dual route model" (or as it would be labeled in our

approach: dual neural pathway model) for translating phonological specifications of syllables into articulatory specifications. In agreement with these authors we assume one neural pathway for frequent and another for infrequent syllables. The frequent pathway comprises phonemic-to-motor mapping, which is comparable to the mental syllabary. This neural pathway (also called syllabic encoding route or in terms of our model: syllabic motor plan storage) is implemented in our model by a self organizing (phonetic) map, and its bilateral mappings, which associate phonemic, motor plan, and sensory states. All motor plans and sensory states for frequent syllables are stored within the phonetic to motor plan and phonetic to sensory mappings. For the second neural pathway (also termed subsyllabic encoding route, or in terms of our model motor planning module, generator of motor plans, or generator of gestural scores) we assume that motor plans here are assembled from smaller subsyllabic units (Levelt et al., 1999). These smaller units, though, are basically vocal tract action units (or gestures) in our approach and not necessarily segments as is claimed by other authors. In some cases, one gesture represents one segment (e.g., some vowels such as /i/, /a/; voiced plosives) and so in these cases one might interpret the gesture-by-gesture assembly of motor plans within the motor planning module as segment-by-segment assembly. However, in other cases segments are ensembles or groups of gestures (vowels such as /o/, /u/ comprise lip and tongue gestures; voiceless consonants comprise vocal tract constricting or closing and glottal opening gestures; nasals comprise vocal tract constricting or closing and velopharyngeal opening gestures). Furthermore, it is possible that the assembly of syllabic motor plans is based on larger subsyllabic units, for example, syllable onset consonantal clusters, since intragestural timing first of all relates all gestures constituting this cluster and then relates the gestures for this cluster with other constituents forming a syllable.

A major difference between our approach and the dual route concepts discussed previously is that a strong neural association between mental syllabary and the gestural assembling module (motor planning module) is assumed in our model (see Figure 16–1). In our approach, it is assumed that the knowledge for assembling infrequent syllables always stems from knowledge stored within the mental syllabary. For example, in the case of the realization of the infrequent syllable /lo/ (designated as low frequency in our model language), the phonemic activation of /lo/ leads to an activation of all frequent /lV/ syllables (i.e., /lu/ as well as /la/, /le/ and /li/). All these syllables are phonetically similar and thus one can assume that the temporal gestural frame of these frequent syllables /lV/, i.e. the specification of all temporal inter- and intra-gestural parameters for the /lV/ syllable type is copied from the syllabary in order to have a prototypical gestural motor plan for /lo/. Thus, only the spatial target information of /l/ and /o/ need be added as "spatial content" to this "temporal gestural frame" (cf., MacNeilage, 1970, for segments) in order to specify completely the motor plan for this infrequent syllable.

Hence, a model of AOS that assumes total impairment of the mentally syllabary and a need to compensate by

using an "indirect route" is not compatible with our model. First, knowledge from the mental syllabary is needed to assemble gestural motor plans of infrequent syllables in our approach. Second, the gestural assembly route (i.e., the route using the motor planning module in our approach) cannot be interpreted as an indirect or "second choice" route if the main route, that is, the mental syllabary, is defective. The gestural assembly pathway in our approach (i.e., the motor planning module) is as important as the mental syllabary itself, being, for example, the neural pathway for modifying temporal and spatial gestural parameters in relation to prosody (e.g., speaking rate, emphatic stress).

Moreover, in our approach it is not assumed that one of the motor planning routes can be completely defective. For instance, it is assumed that parts of the mental syllabary (for example, CV syllables or CCV syllables) and in parallel parts (or levels) of the motor planning module are defective. One of the main results of this study, therefore, is that the functional picture of neural defects in terms of our model is not as simplistic as in other approaches (e.g., Varley & Whiteside 2001). Experimental data from speakers with AOS compatible with the assumptions and results given by our model appear for example in Aichert and Ziegler (2004). They conclude that not only the mental syllabary but also the "indirect route must be disturbed as well in AOS patients" (p. 154) and that "a disturbance of the indirect route should not be explained by side-effects . . . of speech motor programming" (p. 154). Moreover, Aichert and Ziegler claim that "patients with AOS do access the mental syllabary . . ." (p. 156). This assumption is again in accordance with the assumptions of our model. Furthermore, Aichert and Ziegler (2004) postulate that AOS speakers retrieve corrupted entries from the mental syllabary. This has not been modeled in our experiments to date, but could lead to more types of segmental errors than were generated in our model so far, for example, by varying the temporal coordination of gestures.

Motor Planning versus Motor Programming

Although motor planning is a central concept in our neural model, motor programming is not addressed here explicitly. It is striking in the literature that the terms motor planning and motor programming currently refer to different control concepts or control models. Thus, Darley, Aronson, and Brown (1975) claimed a three-step model separating language processing, motor programming, and execution. Van der Merwe (2008) introduced a four-stage sensorimotor model of speech production and separated linguistic processing, motor planning, motor programming, and motor execution. She claimed that motor planning involves (1) activating and organizing the temporal and spatial specifications or the production of sequences of phonemes and (2) adaptation of core motor plans for particular phonemes to specific speech contexts in which they will appear by entering into subroutines that enable movement of the articulatory structures. Further, motor programming means that the motor plan subroutines are fed-forward to the motor programming system in which muscle-specific motor programs

for articulatory movements are selected. Following this approach, AOS would be a motor planning disorder while the dysarthrias would be motor programming disorders (cf., Peach, 2004). Kent (2000) separates "the planning and preparation of movements (sometimes called motor programming) and the execution of movement plans to result in muscle contractions and structural displacements" (p. 391). He maintains that "acquired AOS . . . impairs especially the process of planning or programming speech movements . . . " and states that "the dysarthrias affect the execution of movements" (p. 403). Although the model introduced in this paper differentiates only planning and execution, by introducing the motor plan level, it is one of our future tasks to specify the execution module in more detail, for example, in relation to a differentiation of this module with respect to action unit programming and execution (Maas, Robin, Wrigth, & Ballard, 2008; Wright et al., 2009). Currently, we allocate programming between planning and execution, which is in agreement with the model of van der Merve (2008).

Moreover, it should be noted that the model introduced in this paper is a model of speech processing that includes production and perception components. The DIVA model also includes feedback loops, and through this, introduces self-perception. Despite the fact that perception is seldom mentioned in descriptions of AOS, it is assumed in our perspective that peripheral and central sensory speech processing are not affected in AOS. This is important for example in our modeling of groping, since the notion of the auditory and somatosensory target of the speech item is the driving force for groping in our approach.

We set out to introduce the main features of ACT and to examine within this action, gestural- as opposed to segmental-based model, whether lesioning would produce sound derailments compatible with perceptual and instrumental descriptions of apraxic speech derailments that were emergent properties of the system without having to posit a linear, segmental organization of speech output. In as far as the speech "errors" produced in model speakers 1–4 reflected the kinds of disruption reported for people with AOS, we believe we have achieved this. Speech derailments could be seen, for instance, to arise from overall problems in inter- and intragestural timing, in degraded access to computational elements, and so forth. This opens up an avenue of enquiry that can supplement and complement studies of AOS. ACT offers the possibility on the one hand of testing out theoretical predictions concerning the nature of AOS and on the other hand informing the interpretation of apraxic speech output, informing construction of tasks to test out with real AOS speakers and against real lesions. It was seen that predictions and results from ACT were compatible with contemporary models of apraxia of speech in terms of nonlinear dynamics in output and notions of gestural internal versus external breakdown of organization (Maas et al., 2008; Wright et al., 2009; Ziegler, 2005, 2009).

However, it should be noted that the neurocomputational model of speech processing introduced here as well as our modeling of AOS should be interpreted only as an initial endeavor. While the model is among the most detailed quantitative neural models of speech processing currently in existence and is

in accordance with very recent results of functional imaging and behavioralistic studies (Eickhoff, Heim, Zilles, & Amunts, 2009; Moser et al., 2009; Martins & Ortiz, 2009), it nevertheless still delivers only a very broad picture of the true complexity of human neural processes in speech processing. Further, the neural deficits introduced here are related strongly to the organization of the model and are highly schematic. In reality, it can be assumed that true neural deficits are far more complex. Thus, this model and the resulting modeling of signs of AOS should be interpreted as only a starting point in understanding AOS. More detailed modeling as well as true clinical studies must focus on this topic in order to gain a better and more detailed understanding of AOS.

Acknowledgment. This work was supported in part by the German Research Council Project Nr. Kr 1439/13-1 and project Nr. Kr 1439/15-1.

References

Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain & Language, 88*, 148–159.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.

Croot, K. (2002). Diagnosis of AOS: Definition and criteria. *Seminars in Speech and Language*, *23*, 267–279.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). Motor speech disorders. Philadelphia, PA: Saunders.

Eickhoff, S. G., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical transactions of the Royal Society B: Biological Science*, *367*, 2399–2421

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge, UK: Cambridge University Press.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, L., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*, 386–412.

Guenther, F.H. (2006). Cortical interaction underlying the production of speech sounds. *Journal of Communication Disorders, 39*, 350–365.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain & Language, 96*, 280–301.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*, 67–99.

Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., & Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain, 127*, 1479–1487.

Indefrey, P., Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.

Jordan, L. C , & Hillis, A. E. (2006). Disorders of speech and language: Aphasia, apraxia and dysarthria. *Current Opinion in Neurology, 19*, 580–585.

Kent, R. (2000). Research on speech motor control and its disorders: a review and prospective. *Journal of Communication Disorders*, *33*, 391–428.

Kohler, K. J. (1992). Gestural reorganization in connected speech: A functional view-

point on "articulatory phonology." *Phonetica, 49*, 205–211.

Kröger, B. J., Birkholz, P., Kannampuzha, J., & Neuschaefer-Rube, C. (2006). Learning to associate speech-like sensory and motor states during babbling. *Proceedings of the 7th International Seminar on Speech Production* (pp. 67–74). Brazil: Belo Horizonte.

Kröger, B. J., & Birkholz, P. (2007). A gesture-based concept for speech movement control in articulatory speech synthesis. In A. Esposito, M. Faundez-Zanuy, E. Keller, & M. Marinaro (Eds.), *Verbal and nonverbal communication behaviours LNAI 4775* (pp. 174–189) Berlin, Heidelberg: Springer.

Kröger, B. J., Kannampuzha, J., Lowit, A., & Neuschaefer-Rube, C. (2009). Phonetotopy within a neurocomputational model of speech production and speech acquisition. In S. Fuchs, H. Loevenbruck, D. Pape, & P. Perrier (Eds.), *Some aspects of speech and the brain* (pp. 59–90). Berlin: Lang.

Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C. (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication 51*, 793–809.

Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*, 239–269.

Maas, E., Robin, D. A., Wright, D. L., & Ballard, K. J. (2008). Motor programming in apraxia of speech. *Brain & Language*, *106*, 107–118.

Martins, F. C., & Ortiz, K. Z. (2009). The relationship between working memory and apraxia of speech. *Arq Neuropsiquiatr, 67*, 843–848.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review, 77*, 182–196.

McNeil, M. R. (Ed.) (2008). *Clinical management of sensorimotor speech disorders*. New York, NY: Thieme.

McNeil, M. R., Robin, D. A., & Schmidt, R. A. (1997). Apraxia of speech: Definition, differentiation, and treatment. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 311–344). New York, NY: Thieme.

Miller, N. (2000). Changing ideas in apraxia of speech. In I. Papathanasiou (Ed.), *Acquired neurogenic communication disorders* (pp. 173–202). London, UK: Whurr.

Miller, N. (2002). The neurological bases of apraxia of speech. *Seminars in Speech and Language*, *23*, 223–230.

Moser, D., Fridriksson, J., Bonila, L., Healy, E. W., Baylis, G., Baker, J. M., & Rorden, C. (2009). Neural recruitment for the production of native and novel speech sounds. *Neuroimage, 46*, 549–557.

Ogar, J., Slama, H., Dronkers, N., Amici, S., & Gorno-Tempini, M. L. (2005). Apraxia of speech: An overview. *Neurocase, 11*, 427–432.

Peach, R. K. (2004). Acquired apraxia of speech: Features, accounts, and treatment. *Topics in Stroke Rehabilitation, 11*, 49–58.

Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain & Language, 107*, 102–113.

van der Merwe, A. (2008). Theoretical framework for the characterization of pathological speech sensorimotor control. In M. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (2nd ed., pp. 3–18). New York, NY: Thieme.

Varley, R., & Whiteside, S. (2001). What is the underlying impairment in acquired apraxia of speech? *Aphasiology, 15*, 39–49.

Wertz, R. T, LaPointe, L. L., & Rosenbeck, J. C. (1984). *Apraxia of speech in adults: The disorder and its management*. Orlando, FL: Grune and Stratton.

Wright, D., Robin, D., Rhee, J., Vaculin, A., Jacks, A., & Guenther, F. (2009). Using

the self-select paradigm to delineate the nature of speech motor programming. *Journal of Speech, Language, and Hearing Research, 52*, 755–765.

Ziegler, W, (2005), A nonlinear model of word length effects in apraxia of speech. *Cognitive Neuropsychology, 22*, 603–623.

Ziegler, W, (2009), Modelling the architecture of phonetic plans: Evidence from apraxia of speech. *Language and Cognitive Processes*, *24*, 631–661.