Speech motor control in relation to phonology
Terband, Hayo Robert

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 18-11-2019
Chapter 3

Testing hypotheses about the neurological mechanisms underlying Childhood Apraxia of Speech

Abstract

**Purpose:** Childhood apraxia of speech (CAS) is a controversial clinical entity with respect to both clinical signs and underlying neuro-motor deficit. In the present study, two hypotheses about the core neurological deficit were investigated using computational modeling with the Directions Into Velocities of Articulators (DIVA) model: reduced/degraded somatosensory information (H1) and increased levels of neural noise (H2).

**Method:** In a series of computer simulations, the involvement of the two hypothesized deficits was systematically varied during the acquisition of systemic and phonemic mappings. The effects of both deficits on speech production if implemented after (undisordered) asymptotic learning was investigated as control conditions.

**Results:** Simulations results showed that both deficits lead to a deterioration of speech production after imitation learning, but for different reasons: H2 predominantly leads to deterioration on the phonetic level (systemic mapping/forward model) and H1 on the phonological level (phonemic mappings).

**Conclusions:** These findings imply a close relation between the quality of the system for self monitoring auditory error and the involvement of phonological vs. motor processes in children with speech sound disorders. It is suggested that H2 might be involved in apraxic speech output disorders and H1 in speech output disorders typically classified as phonological in nature.

### 3.1 Introduction

Childhood apraxia of speech (CAS) is “a neurological childhood (pediatric) speech sound disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits” (ASHA, 2007, pp. 3–4). CAS has been associated with a wide variety of diagnostic descriptions and has been shown to involve different symptoms during successive stages of development (Maassen, 2002; Maassen, Nijland, & Terband, 2010). Perhaps the most prominent of speech characteristics is inconsistent errors on consonants and vowels in repeated productions of syllables or words (ASHA, 2007; Hall, Jordan, & Robin, 2007; Ozanne, 2005). The errors are not typically immature and mainly comprise a large number of consonantal errors in which omissions are more prevalent than substitutions. The vowel errors are mostly distorted productions and reductions. Often, the errors constitute non-phonemic productions that defy accurate transcription, even when using narrow transcription. A second prominent symptom of CAS is deviant or disrupted coarticulatory transitions between sounds and syllables (ASHA, 2007; Hall et al., 2007; Ozanne, 2005). Coarticulation has been found to be both stronger and more extended as well as more segmental (or hyper-articulation) in the speech of children with CAS as compared to normally developing children (Nijland, Maassen, Van der Meulen et al., 2002, 2003). Further important features of the speech of children with CAS include groping or searching-articulatory behavior (both prevocalic and during sound production) and difficulties and low maximum repetition rates in the production of alternate syllables or diadochokinesia (ASHA, 2007; Hall et al., 2007;
Chapter 3

Ozanne, 2005).

The search for the underlying deficit of CAS has not yielded clear-cut results thus far. In psycholinguistic terms, the core impairment can be summarized as an inability to transform an abstract phonological code into motor speech commands (Nijland, Maassen, Van der Meulen et al., 2003). Although the core deficit of CAS is thought to reside at the phonetic programming level (Schmidt & Lee, 1999; Van der Merwe, 1997), CAS has been interpreted at various levels: lexical representation, phonological planning, assembly of a phonetic program, and motor implementation (see Maassen et al., 2010, for an overview). This reflects a fundamental problem in isolating the underlying deficit of CAS: there is a developmental interaction between the different cognitive levels of processing. The progression to the adult system is a gradual and continuous process comprising interactions between emerging modules (Karmiloff-Smith, 2006; Karmiloff-Smith, Scerif, & Ansari, 2003; Karmiloff-Smith & Thomas, 2003; see also Munson, Swenson, & Manthei, 2005). In other words, cognitive modules are the outcome of development rather than the starting point. For example, the developmental trajectory of speech motor control (see e.g., Green & Nip, 2010 for an overview) has general consequences for the sequence of phonological development (Maassen et al., 2010). Another example is that phonological knowledge not only reflects the level of categorical lexical contrasts, but also comprises the incremental acquisition of representations at the acoustic-perceptual and articulatory-motor levels (Edwards, Fourakis, Beckman, & Fox, 1999). As a result, a specific underlying impairment on one cognitive level also affects the development on adjacent levels. For instance, we know that CAS disrupts the development of the lexicon, the phonological system, and auditory processing (Crary, Landess, & Towne, 1984; Marion, Sussman, & Marquardt, 1993). This developmental interaction sets serious limitations to the use of psycholinguistic models in isolating the deficits underlying developmental speech (or language) disorders (Maassen et al., 2010; Terband & Maassen, 2010). To resolve this issue, a different approach is to use neurolinguistic models and to focus on the sensori-motor information processing that is involved in the acquisition of speech motor control (Maassen et al., 2010; Terband & Maassen, 2010). In a previous study in our laboratory (Terband, Maassen, Guenther, & Brumberg, 2009), we utilized this approach and attempted to associate the evolving symptoms of CAS to particular stages of information processing by using computational neural modeling with the DIVA model (Directions Into Velocities of Articulators; Guenther, 1994; Guenther, Ghosh, & Tourville, 2006). A series of computer simulations with the DIVA model tested the hypothesis that the speech production system in CAS suffers from weak feedforward control and consequently an increased reliance on the feedback control subsystem. The simulations accounted for four key characteristics of speech production in CAS: deviant coarticulation, distorted vowel productions, searching articulation, and increased token-to-token variability. Two possible core deficits of degraded feedforward control in CAS were discussed: reduced/degraded somatosensory information (H1) and increased levels of neural noise (H2). The aim of the present study was to test these two specific hypotheses in a series of DIVA simulations. Furthermore, the previous simulation study was limited to the stage of imitation learning. The simulations of the present study start earlier in development and, previous to imitation learning, also include the babbling stage. Before describing the model manipulations, we will first give a short overview of the acquisition of
speech motor control in DIVA.

3.2 The DIVA model

3.2.1 Speech motor acquisition in DIVA

The DIVA model consists of a neural network controller detailing feedforward and feedback control loops that are involved in early speech development and mature speech production (Figure 3.1). The model strives to be biologically plausible and its components have been associated with regions of the cerebral cortex and cerebellum (Guenther et al., 2006). The final acoustic output of the model is generated through articulatory synthesis (Maeda, 1990) of the DIVA model motor parameters.

![Schematic representation of the DIVA model of speech motor control (Guenther et al., 2006). Sup. = Superior; Inf. = Inferior. Projections to and from the cerebellum are simplified for clarity. Not shown in the figure are the forward and inverse models, which are used by the model to calculate the feedback commands and feedforward commands respectively and in this scheme would be located behind these two.](image)

The acquisition process comprises three stages. In the first stage, semi-random artic-
ulatory movements (resembling the babbling stage) are used to learn the neural relations between motor commands and their auditory and somatosensory consequences (systemic mapping). In the second stage, the model is presented with sample speech sounds and learns an auditory target region for each sound. The targets can consist of a single phoneme, a syllable, a word, or a small phrase. In the final stage, feedforward commands are acquired by the model iteratively attempting to produce sounds that fall into the auditory target region (Figure 3.2). In this way, the model learns a set of phonemic mappings between articulatory movements and language-specific goals. For a detailed description of the model, see Guenther et al. (2006). An overview of the model's parameters and equations can be found in Appendix A.

![Figure 3.2: The DIVA model learning the word baby. The top-panel shows the target regions for F1, F2, and F3 (black) and the model’s auditory realization (red). On the basis of the error information provided by the auditory feedback control subsystem (middle-panel), the feedforward command (bottom-panel) is updated with each attempt, thus becoming more accurate. After approximately five iterations, learning has become asymptotic.](image)

### 3.2.2 Implementation of the hypothesized deficits

The first hypothesis of the present study (H1) seeks the core deficit of CAS in a reduced or degraded oral sensitivity. It has been found that children with CAS have a lowered oral sensitivity of the tongue and palate (e.g., Hall et al., 2007; Ozanne, 2005). In DIVA, such reduction of somatosensory information would have different, but cumulative, effects in successive stages of speech development. First, during the babbling stage, uncertainty in the somatosensory state would cause weak or underspecified somatosensory-to-motor projections (synaptic projections from somatosensory error mappings to the articulator velocity map in motor cortex). Furthermore, a poor estimate of the somatosensory state will cause a poor estimate of the mo-
tor state, which subsequently would slow down the learning of auditory-to-motor projections (synaptic projections from auditory error mappings to the articulator velocity map in motor cortex). This would lead to degraded feedforward control in the imitation and performance stages as the execution of the appropriate feedforward commands depends on knowledge of the current somatosensory state. Furthermore, the unstable somatosensory-to-motor projections would lead to degraded somatosensory feedback control in the performance stage. The deficit of degraded somatosensory information can be simulated in DIVA by adding noise to the somatosensory and motor state representations, which are hypothesized in the DIVA model to be located in the inferior parietal cortex and motor cortex.

The second hypothesis (H2) explains CAS as resulting from an increased level of neural noise throughout the speech motor system. Neural noise has been widely associated with the token-to-token variability that characterizes human motor performance (e.g., Fitts, 1954; Harris & Wolpert, 1998; Perkell & Nelson, 1985). Neural noise is viewed as the primary factor limiting the possibility of simultaneously rapid and accurate movements, forming one of the main arguments for the existence of paired internal models. Through the combination of forward and inverse models the central nervous system can optimally estimate a current state of the system (Wolpert, Ghahramani, & Flanagan, 2001). In the DIVA model, neural noise could affect all of the neural maps. During the babbling stage, uncertainty in the motor, auditory and somatosensory state maps would slow down learning, causing weak or underspecified synaptic projections from both auditory-to-motor and somatosensory-to-motor transformations. In further stages, this would be expected to result not only in incorrect and/or imprecise feedforward commands, but also in poor performance in the auditory and somatosensory feedback control systems. Increased levels of neural noise can be simulated in DIVA by adding noise to all the models state representations. Note the crucial difference with the implementation of the first hypotheses (degraded somatosensory information); to implement increased levels of neural noise, a noise term is added to the auditory state representations in addition to the somatosensory- and motor state representations.

In a series of computer simulations we tested whether these two deficits will lead to CAS-like behavior in the DIVA model. The DIVA model features a noise generator, by which Gaussian random noise can be added to the cell activations of motor-, auditory-, and somatosensory cortices. By specifying the standard deviation of the Gaussian distribution, the level of the noise can be manipulated. In the case of H1, we added noise to the somatosensory-, and motor state representations (since in DIVA a poor estimate of the somatosensory state will cause a poor estimate of the motor state). Regarding H2, we added noise to the motor-, auditory-, and somatosensory state representations. In both cases, the noise was zero-mean noise. The standard deviation was systematically varied from 0% to 25% in 5% steps creating a total of 2x5 experimental conditions (note that 0% noise functioned as a baseline condition). The simulation series comprised two stages, which are elaborated in more detail below. The effect that both deficits have on speech production after (undisordered) asymptotic learning were investigated as control conditions in order to differentiate the effect of noise during acquisition from the effect of noise during production.
Chapter 3

3.3 DIVA simulations: babbling stage

3.3.1 Method and materials

In the first stage of the simulations, we trained a forward model for each of the experimental conditions. DIVA's forward model consists of an adaptive hyperplane radial basis function (AHRBF) network, which is a neural network-like mathematical constructs that is composed of a set of nodes, each corresponding to a different position of the input space. In the training process, the model iteratively learns an optimal mapping between input articulatory movements and auditory outputs, having a minimum total error. The same set of 4000 input patterns was used for all conditions, which in the baseline condition (no deficits) corresponds to asymptotic learning.

The forward models were evaluated by means of the total error on a mutually exclusive test set of input patterns. The total error was calculated by the sum-squared Euclidean error of the trained forward model on the evaluation set. To assess the developmental trajectory, we additionally trained and evaluated forward models in all conditions using 500, 1000, 1500, 2000, and 3000 training items.

3.3.2 Results and discussion

Results show decreased forward model performance (i.e., an increase in the sum-squared Euclidean error on the test set of input patterns) with increased severity for both deficits (Figure 3.3, left panel), but the effect is much larger for H2 (increased levels of neural noise). The results thus provide a clear difference between the two hypotheses: H2 (increased levels of neural noise) leads to a poor representation at the phonetic level (systemic mapping or forward model), whereas, for H1 (reduced/degraded somatosensory information) the representation at the phonetic level remains relatively intact. The performance of the forward model at different stages in the training process shows the effects to be stable throughout (Table 3.1). Compared to the effect of underlearning in the baseline condition (Figure 3.3, right panel), the results show a very drastic and consistent decrease in forward model performance with increased noise level for H2 (increased levels of neural noise). In the case of H1 (reduced/degraded somatosensory information), the results are less consistent. The sum-squared Euclidean error is slightly smaller for noise levels of 5% and 10%, respectively, as compared to the baseline condition and forward model performance only starts to decrease at noise levels of 15% and higher. At these higher noise levels, the decrease in performance appears to remain small, though it does represent substantial underlearning when compared to the baseline condition (Figure 3.3, right panel).
Neurological mechanisms underlying CAS

![Graph showing forward model performance (SSE; sum-squared Euclidean error) after the full training process (i.e., 4000 training items) in relation to the severity of deficits. Right: forward model performance in relation to number of training items in the baseline condition (no deficits implemented).]

3.4 DIVA simulations: imitation learning

3.4.1 Method and materials

The methodology of the imitation learning simulations was largely the same as in Terband et al. (2009). In a series of simulations we systematically varied system noise levels during production attempts during the acquisition of feedforward commands. For each of the conditions, we used the corresponding forward model generated in babbling simulations. V1CV2 utterances were used as speech sound targets, comprising all combinations of /a/, /i/, and /u/ for the vowels and /b/, /d/, and /g/ for the consonants, forming a total of 27 sound targets. Each sound target was repeated 6 times. The simulations were evaluated both perceptually and acoustically; the procedures are described in detail below.

In order to differentiate the effect of noise during acquisition from the effect of system noise alone, the effect that both deficits have on speech production after (undisordered) asymptotic learning were investigated as control conditions. In this case, the deficits were introduced only during production attempts in the performance stage, after the feedforward commands had been acquired without any deficits implemented (i.e., the baseline condition; all noise levels at 0%). The same set of 27 sound targets was used, and each sound target was repeated 6 times.

Perceptual evaluation

A perceptual scaling experiment was used to assess the recognizability and the speech quality of the speech output across the varied levels of severity. It should be noted that although DIVA is able to simulate particular aspects of speech, i.e., formant frequency trajectories of single
words or simple phrases, at the present stage the resulting synthesized speech does not reach the naturalness needed for a detailed perceptual evaluation of pathological characteristics. Therefore, we limited ourselves to a perceptual evaluation of the general recognizability and quality of the synthesizer output across the varied levels of severity.

A total of 14 subjects (age 18-21) with no reported hearing problems participated in the listening experiment. Testing took place in a sound-treated booth. The signal was delivered binaurally by headphones (Beyerdynamic DT250/80) through a RME digi96/8 PAD sound module and 2 Marantz MA6100 mono amplifiers. The sound level was fixed at a comfortable level (approximately 70 dBA).

The perceptual evaluation was done according to a fixed anchor 2AX-paradigm. Stimuli were presented in pairs consisting of two versions of the same word. The second of each pair formed the target utterance, whereas the first of each pair formed the ideal and consisted of the production of the standard model (noise level 0), resulting in 2 (conditions) x 5 (severity levels) x 6 (versions of the same target words) comparisons. Six target words were selected for perceptual evaluation (/abi/, /agi/, /iba/, /ida/, /ubi/, and /ugi/) and the total number of 360 stimuli was presented in a pseudo-random order. Twelve stimuli comprising words that were not used in the experiment were presented as practice materials.

The participants received written instructions and were asked to judge the second utterance of each pair on two aspects: 1. recognizability (How clear is the second utterance as a version of the first utterance?); 2. quality (How is the speech quality of the second utterance in comparison with the first in terms of distortion and dysfluencies?). The judgments were administered on two 7-point scales on which 1 denoted “not good at all” and 7 denoted “very
good”. The subjects were instructed to try to use the whole scale.

To investigate the effect of deficit severity on recognizability and quality judgments, the mean scores, per subject, of the 6 versions of the same target words were calculated. Subsequently, statistical testing was accomplished using a Linear Mixed Model analysis featuring subject as correlated term, and target utterance and distance as correlated residuals, using a 0.05 level of significance. The two conditions were investigated in separate analyses.

Acoustical evaluation

The simulations were evaluated acoustically on four selected key symptoms of CAS. Since DIVA is a model of speech motor control, we focused on the more fine-grained phonetic characteristics of CAS-speech, deviant coarticulation, speech sound distortion, searching articulation, and increased variability. These four key characteristics are indicative of CAS during the stage of imitation learning and are largely independent of phonological development. Coarticulation was measured by the absolute differences in mean formant frequencies across contexts. The amount of vowel distortion was calculated by averaging the absolute differences in mean formant frequencies of each produced vowel relative to the frequencies of the target vowel. Finally, variability was measured by the error-variances in mean formant frequencies of repeated productions. A more detailed description of the measures, accompanied by formulas, can be found in Appendix B.

To test whether any of the effects were significant, analyses of variance were conducted first. However, Levene’s tests turned out to be significant, meaning the assumption of homogeneity of variance failed. Since parametric testing was not justified, nonparametric tests (Kruskal-Wallis and Mann-Whitney U) were conducted.

3.4.2 Results and discussion

Perceptual evaluation

The results of the perceptual evaluation are presented in Figure 3.4. In comparison, the perceptual evaluation yields higher recognizability and quality judgments and a more consistent pattern for H1 (degraded somatosensory information) than for H2 (increased levels of neural noise). In the case of H1, results show a steady decrease of both recognizability and speech quality with increased severity. For H2, the results follow the same trend for noise levels up to 10%, but then the pattern becomes more complicated. For noise levels of 15% and higher, results show an inverse relation between speech quality and recognizability. Statistical analyses revealed the effects of noise level/severity on the recognizability and quality judgments to be highly significant for both deficits [all p values <.001].

Coarticulation usually changes the characteristics of a speech sound in the direction of the neighboring speech sound, but the deviant coarticulation patterns of children with CAS also contained hyper-articulation, i.e., change in the opposite direction or enhanced contrasts (ASHA, 2007; Nijland, Maassen, Van der Meulen et al., 2002, 2003).
Figure 3.4: Perceptual evaluation: recognizability and quality judgments of the models speech output in relation to the severity of deficits.

Closer perceptual inspection of the individual utterances reveals that from 15% noise and higher, the quality of the models speech is very bad and, in isolation, the utterances are not recognizable anymore. This finding applies to both deficits, but manifests itself in different ways, which seems to be related to forward model performance. In the case of degraded somatosensory information (H1), forward model performance starts to decrease at noise levels of 15% and higher. In the case of increased levels of neural noise (H2), forward model performance shows very strong decrease right from the start and for noise levels of 15% and higher, the sum-squared Euclidean error is already more than 50% larger. Under these circumstances, the forward model becomes so unspecified that every utterance is strongly neutralized and reduces to schwa. Hence the inverse relation between speech quality and recognizability for noise levels of 15% and higher: the model's productions reduce to schwa, but these schwa's are produced fluently without much distortion or interruptions. In the case of degraded somatosensory information (H1), results do not show this strong effect of underspecification of the forward model and correspondingly no strong neutralization effect in the speech production after imitation learning. Under circumstances of degraded somatosensory informa-
Neurological mechanisms underlying CAS

forward model performance decreases from noise levels of 15% and up, but only relatively mildly. The effect of increased severity is different from the effect regarding H2 (increased levels of neural noise). For H1 (reduced/degraded somatosensory information) the increase in noise in the somatosensory and motor representations causes primarily a strong decrease in speech quality. When noise levels exceed 10%, the speech production features so much distortion and dysfluencies that, although relatively recognizable, does not resemble speech anymore. For these reasons, we limited the further acoustical analysis of both hypothesized deficits to the first two levels of severity (noise levels of 5% and 10%).

Acoustical evaluation

Figure 3.5 presents coarticulation, speech sound distortion, searching articulation, and token-to-token variability of the model's productions after asymptotic imitation learning as a function of the severity of the two deficits. In general, the results show increased severity of the selected key-symptoms of CAS for both deficits, but on different scales. In the case of reduced/degraded somatosensory information (H1), the results show a large increase in the symptoms, whereas for increased levels of neural noise (H2), the increase is more moderate.

A series of Kruskal-Wallis tests for x independent samples was conducted to test whether any of the effects were significant. Test statistics, presented in Table 3.2, show almost all effects of noise level to be significant at a p<.001 level. Three aspects of the results are especially striking. First, for both deficits, the results show no significant differences in the amount of searching articulatory behavior in V1 (Table 3.2). Second, for H2 speech sound distortion in all positions tends to be smaller in the noise conditions than in the control condition (no involvement of any deficit). A third striking feature is that in the case of reduced/degraded somatosensory information (H1), all CAS symptoms are stronger in the consonant than in the vowels. In the case of increased levels of neural noise (H2), however, the results show no structural difference in the acoustical measures between consonant and vowels (Figure 3.5).

To differentiate the effect of noise during acquisition from the effect of system noise alone, we compared the effect that both deficits have on speech production present during the whole learning process versus implemented after (undisordered) asymptotic learning (Figure 3.6). A series of Mann-Whitney U tests for 2 independent samples was conducted to test whether any of the differences in the effects of the deficits during and after learning were significant. Test statistics are presented in Table 3.3.

The comparisons show coarticulation to be much larger if the deficit is present during the whole learning process for reduced/degraded somatosensory information (H1). For increased levels of neural noise (H2) results show no differences in the amount of coarticulation if the deficit is implemented during or after learning.

Interestingly, with respect to speech sound distortion, the results show an opposite pattern with few significant differences for H1 (reduced/degraded somatosensory information) and many for H2 (increased levels of neural noise). Although the amount of speech sound distortion appears to be much larger for C in reduced/degraded somatosensory information (H1) if the deficit is implemented during compared to after learning, these differences fail to reach significance. In increased levels of neural noise (H2), the amount of speech sound dis-
Figure 3.5: Acoustical evaluation of normal productions with online feedback after asymptotic imitation learning for degraded somatosensory information (H1; continuous lines) and increased levels of neural noise (H2; dashed lines). Top: coarticulation (left) and searching articulatory behavior (right). Bottom: speech sound distortion (left) and token-to-token variability of mean formant frequencies (right).

tortion tends to be larger if the deficit is implemented after (undisordered) asymptotic learning than when present during the acquisition process.

The results concerning searching articulatory behavior show larger effects if the deficit was implemented during learning for H1 (reduced/degraded somatosensory information) in the consonants, but results show no differences in the vowels. For increased levels of neural noise (H2), the comparison between the deficit implemented during or after learning shows no differences in searching articulatory behavior in general. Furthermore, where significant (for H2; 5% in V1 and C) the difference is very small and the results yield no clear pattern. Finally, results show token-to-token variability predominantly to be larger for both deficits if present during the whole learning process. This effect is larger for reduced/degraded somatosensory information (H1; Figure 3.6).

Together, these results suggest that for increased levels of neural noise (H2), the increase in anticipatory coarticulation should be attributed mainly to production/execution processes, and the same applies to speech sound distortion and token-to-token variability. Reduced/degraded somatosensory information (H1) appears to affect the acquisition processes to a larger extent, causing the learned motor commands to be inherently unstable or deviant.

To explore this further, we let the DIVA model (the standard model; no deficits implemented) run stored feedforward command after asymptotic learning without online feedback
### Table 3.2: Results of Kruskal-Wallis tests for the acoustic measures for the two implemented deficits.

<table>
<thead>
<tr>
<th>Implemented deficit</th>
<th>H1: reduced/degraded somatosensory information</th>
<th>H2: increased levels of neural noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chi-square</td>
<td>p</td>
</tr>
<tr>
<td>Coarticulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ant V1</td>
<td>23.676</td>
<td>0.000</td>
</tr>
<tr>
<td>Ant C</td>
<td>63.806</td>
<td>0.000</td>
</tr>
<tr>
<td>Co C</td>
<td>45.994</td>
<td>0.000</td>
</tr>
<tr>
<td>Co V2</td>
<td>50.173</td>
<td>0.000</td>
</tr>
<tr>
<td>Searching articulatory behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAB V1</td>
<td>4.896</td>
<td>0.086</td>
</tr>
<tr>
<td>SAB C</td>
<td>231.579</td>
<td>0.000</td>
</tr>
<tr>
<td>SAB V2</td>
<td>254.669</td>
<td>0.000</td>
</tr>
<tr>
<td>Speech sound distortion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD V1</td>
<td>1.486</td>
<td>0.476</td>
</tr>
<tr>
<td>SSD C</td>
<td>0.134</td>
<td>0.935</td>
</tr>
<tr>
<td>SSD V2</td>
<td>57.919</td>
<td>0.000</td>
</tr>
<tr>
<td>Token-to-token variability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTV V1</td>
<td>54.776</td>
<td>0.000</td>
</tr>
<tr>
<td>TTV C</td>
<td>53.649</td>
<td>0.000</td>
</tr>
<tr>
<td>TTV V2</td>
<td>63.957</td>
<td>0.000</td>
</tr>
</tbody>
</table>

In this way, the model executes the motor command without any alterations and we were able to extract the acquired feedforward commands. An example is presented in Figure 3.7. In the case of reduced/degraded somatosensory information (H1), the auditory feedback loop is intact and, as was shown in the babbling simulations, the forward model (that is used in the auditory feedback loop to calculate online corrections) remains relatively intact. As a result, the model is much more sensitive to the articulatory errors introduced by the noise in the motor state representation and induces a corrective response. This also explains why coarticulation, speech sound distortion, and token-to-token variability was found to be larger for the consonants compared to the vowels; consonant errors are more difficult to correct with auditory feedback than vowel errors. Furthermore, as the update of the motor command (for the next attempt to produce the target) is based on auditory error (difference between the realized formant trajectories and the target formant trajectories), these articulatory errors affect the update of the motor command and thus affect the learning trajectory and its endpoint.

In the case of increased levels of neural noise (H2), on the other hand, auditory state representations are additionally affected by noise. Since the noise is zero-mean, in the long run errors introduced by noise in motor state representations tend to average out. Consequently, the endpoint of learning trajectories is not inherently deviant from the control condition. The poor quality of the forward model (see babbling imitations) works as a facilitating mechanism, as it dampens online feedback corrections. Due to the underspecification of the forward model, small differences in auditory space do not translate to differences in motor space; therefore, do not provoke an articulatory compensation. This mechanism expresses
Table 3.3: Results of Mann-Whitney U tests for the acoustic measures comparing the involvement of the two deficits if implemented during the whole learning process and if implemented after (undisordered) asymptotic learning.

<table>
<thead>
<tr>
<th>Implemented deficit</th>
<th>5% Noiselevel</th>
<th>10% Noiselevel</th>
<th>5% Noiselevel</th>
<th>10% Noiselevel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarticulation</td>
<td>H1: reduced/degraded somatosensory information</td>
<td>H2: increased levels of neural noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ant V1</td>
<td>-2.476 0.013  -3.300 0.001</td>
<td>-0.823 0.410  -1.069 0.285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ant C</td>
<td>-4.731 0.000  -4.854 0.000</td>
<td>-0.074 0.941  -0.117 0.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co C</td>
<td>-3.865 0.000  -4.295 0.000</td>
<td>-0.664 0.507  -0.817 0.414</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co V2</td>
<td>-0.842 0.400  -1.966 0.049</td>
<td>-2.175 0.030  -3.005 0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Searching articulatory behavior</td>
<td>SAB V1</td>
<td>-1.298 0.194  -0.901 0.367</td>
<td>-2.065 0.039  -1.465 0.143</td>
<td></td>
</tr>
<tr>
<td>SSD V1</td>
<td>-0.983 0.325  -1.051 0.293</td>
<td>-6.142 0.000  -5.321 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD C</td>
<td>-0.619 0.536  -1.355 0.176</td>
<td>-5.623 0.000  -0.701 0.483</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD V2</td>
<td>-5.737 0.000  -4.055 0.000</td>
<td>-4.041 0.000  -5.593 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Token-to-token variability</td>
<td>TTV V1</td>
<td>-6.237 0.000  -5.925 0.000</td>
<td>-3.763 0.000  -2.379 0.017</td>
<td></td>
</tr>
<tr>
<td>SSD C</td>
<td>-6.081 0.000  -5.026 0.000</td>
<td>-5.268 0.000  -1.289 0.197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD V2</td>
<td>-6.081 0.000  -5.026 0.000</td>
<td>-5.268 0.000  -1.289 0.197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTV V1</td>
<td>-2.033 0.042  -2.517 0.012</td>
<td>-0.874 0.382  -2.880 0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTV C</td>
<td>-2.033 0.042  -2.517 0.012</td>
<td>-0.874 0.382  -2.880 0.004</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results showed no differences between the two deficits in their effect on the four acoustic measures when implemented after (undisordered) asymptotic learning (Figure 3.6; test statistics are presented in Table 3.4). This is an intuitive finding since the motor command is in principle on target, and thus the error in the motor command is limited in advance to 5 or 10%. Given the feedforward/feedback ratio of 85/15 (which was kept to the standard value in the current simulations), this reduces the influence of the auditory feedback control to a large extent.

Itself in lower values for the searching articulatory behavior measure (see Figure 3.4). As a result, the influence of errors introduced by noise in motor state representations during the update of the motor command is decreased. (Note that nevertheless, these errors still express themselves in the model’s synthesized speech output.)
Table 3.4: Results of Mann-Whitney U tests for the acoustic measures comparing the two deficits if implemented after (undisordered) asymptotic learning.

<table>
<thead>
<tr>
<th></th>
<th>Noiselevel</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5%</td>
<td>10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z</td>
<td>p</td>
<td>Z</td>
<td>p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coarticulation</td>
<td>Ant V1</td>
<td>-0.160</td>
<td>0.873</td>
<td>-0.326</td>
<td>0.745</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ant C</td>
<td>-1.057</td>
<td>0.291</td>
<td>-1.020</td>
<td>0.308</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Co C</td>
<td>-0.442</td>
<td>0.658</td>
<td>-0.571</td>
<td>0.568</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Co V2</td>
<td>-0.602</td>
<td>0.547</td>
<td>-1.352</td>
<td>0.176</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Searching articulatory behavior</td>
<td>SAB V1</td>
<td>-0.294</td>
<td>0.769</td>
<td>-0.202</td>
<td>0.840</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SAB C</td>
<td>-0.231</td>
<td>0.817</td>
<td>-0.045</td>
<td>0.964</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SAB V2</td>
<td>-0.189</td>
<td>0.850</td>
<td>-0.463</td>
<td>0.644</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech sound distortion</td>
<td>SSD V1</td>
<td>-0.221</td>
<td>0.825</td>
<td>-0.720</td>
<td>0.472</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSD C</td>
<td>-0.814</td>
<td>0.416</td>
<td>-0.466</td>
<td>0.641</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSD V2</td>
<td>-0.122</td>
<td>0.903</td>
<td>-0.077</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Token-to-token variability</td>
<td>TTV V1</td>
<td>-0.221</td>
<td>0.825</td>
<td>-0.720</td>
<td>0.472</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TTV C</td>
<td>-0.814</td>
<td>0.416</td>
<td>-0.466</td>
<td>0.641</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TTV V2</td>
<td>-0.122</td>
<td>0.903</td>
<td>-0.077</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.5 General discussion

3.5.1 Summary of findings

Two hypotheses about the core neurological deficit in CAS, reduced/degraded somatosensory information (H1) and increased levels of neural noise (H2), were investigated in a series of computer simulations with the DIVA model. The involvement of these two deficits was systematically varied during the model acquisition processes of babbling and imitation learning. The effects of both deficits on speech production after normal (undisordered) asymptotic learning were investigated as control conditions. With respect to simulated speech output after asymptotic learning, results did not differentiate directly between the hypotheses. Both deficits led to degraded speech output with similar characteristics on both perceptual and acoustical evaluation, i.e., a decrease in recognizability and quality (perceptual measures) and an increase in coarticulation, searching articulatory behavior, and token-to-token variability (acoustic measures) with increased severity.

Nevertheless, the results did provide clear differences between the two hypotheses in the neural mappings that result from speech acquisition and underlie production. First, focusing on the first two levels of severity (5 and 10%)—results after babbling showed large differences in forward model performance between the two deficits. Compared to the baseline condition, the sum-squared Euclidean error was only slightly smaller for H1 (reduced/degraded somatosensory information), whereas for H2 (increased levels of neural noise) results showed
a dramatic increase in the sum-squared Euclidean error with increased involvement of the deficit. These findings indicate that increased neural noise has a larger effect than reduced/degraded somatosensory information on the forward model learned during babbling.

Second, the comparison between effects of both deficits on speech production when present during versus after (undisordered) asymptotic learning showed a large difference for reduced/degraded somatosensory information (H1), but not for increased levels of neural noise (H2). In the case of H1 (reduced/degraded somatosensory information), results showed an increased severity of the four selected key-symptoms of CAS in C and V1 if the deficit was present during the whole learning process. These results indicate that reduced/degraded somatosensory information (H1) caused the acquired motor commands to be inherently deviant or unstable, whereas for increased levels of neural noise (H2) the majority of symptoms should be attributed mainly to production/execution processes.

In summary, within the present simulations, reduced/degraded somatosensory information (H1) led predominantly to deterioration of the phonemic mappings (acquired feedforward commands; Figures 3.6 and 3.7), whereas increased levels of neural noise (H2) led predominantly to deterioration of the systemic mapping (forward model; Figure 3.3). In DIVA the acquired phonemic mappings are stored in synaptic projections emanating from neurons in the speech sound map and consist of language specific articulatory and auditory goals as well as feedforward commands specific to a particular phoneme or syllable and are thus phonological in nature. The systemic mapping on the other hand comprises language independent neural relations between motor commands and their auditory and somatosensory consequences in which articulatory parameters are automatically organized in task-specific groupings and is located at a phonetic level.

The picture that emerges is one in which increased levels of neural noise (H2) might be responsible for apraxic speech output problems and reduced/degraded somatosensory information (H1) might be at work in speech output disorders typically classified as phonological in nature. In this respect, an important feature of the model under increased levels of neural noise (H2) is that the simulation results indicated the majority of symptoms should be mainly attributed to production/execution processes. This might explain why the difference in number of errors between words and nonwords is smaller in CAS compared to normally developing children, as if all utterances are new (Thoonen, Maassen, Gabreëls, Schreuder, & de Swart, 1997) and can be interpreted as supportive evidence for H2 (increased levels of neural noise).

However, with respect to speech output after (disordered) asymptotic learning, both deficits led to degraded speech output with similar characteristics. The question that arises consequently is how the impaired speech systems that result from the two hypothesized deficits relate to CAS characteristics other than the four investigated key symptoms. The reason we focused on these four more fine-grained phonetic characteristics of CAS speech is because that they are largely independent of phonological development and are indicative of CAS during an extended period of speech development, comprising the whole stage of imitation learning (Terband et al., 2009). Equally important, these symptoms can be simulated with DIVA. A number of speech motor characteristics (e.g., prevocalic groping and speech sound prolongations) could not be investigated because the present computational implementation
of the model uses sound targets that are predefined. Furthermore, DIVA is only a model of speech motor control; higher order suprasegmental and phonological characteristics (e.g., inappropriate prosody, speech sound omissions, substitutions and transpositions) are beyond its scope and could not be investigated. Therefore, it is necessary to expand modeling studies of CAS to the phonological and suprasegmental levels of the speech production process. For example, phonological planning and sequencing could be investigated with the gradient order DIVA model (GODIVA, Bohland, Bullock, & Guenther, 2010), a neural model extending the existing DIVA model of speech production with the assembly and execution of multisyllabic speech plans. Other possibilities are the action based ACT model (Kröger, Birkholz, Lowit, & Neuschaefer-Rube, 2010; Kröger, Kannampuzha, & Neuschaefer-Rube, 2009), which has been used to model the acquisition of phonological maps comprising syllabic and phonemic categories and contrasts (Kröger, Kannampuzha, Lowit, & Neuschaefer-Rube, 2009), and the TADA model (TAsk Dynamics Application, Nam et al., 2007), the computational implementation of the Linguistic Gestural Model (LGM, Browman & Goldstein, 1992, 1997; Saltzman & Munhall, 1989). Expanding the scope of the simulation studies to higher processing levels of speech production is a straightforward direction for further research.

3.5.2 Perception-production

Since the main difference between the two underlying deficits is the quality of the auditory feedback control subsystem, the present findings imply that in children with speech sound disorders this quality is one of the determinants of the involvement of phonological vs. motor processes. If this system is intact (as in H1; reduced/degraded somatosensory information), the impairment of the feedforward control subsystem leads to problems at the phonological level, while it leads to problems at the phonetic level if the auditory feedback control subsystem is also impaired (as in H2; increased levels of neural noise). Several studies have shown that the articulatory proficiency of speakers producing a contrast is related to their perceptual ability to discriminate the contrast, both in normal (Perkell, Guenther et al., 2004; Perkell, Matthies et al., 2004) and in disordered speech (Raaymakers & Crul, 1988). However, it should be noted that the current results do not make claims about auditory discrimination; although the quality of the auditory feedback control subsystem is likely to be related to perceptual acuity, auditory discrimination and identification. The DIVA model focuses on speech production and in the computational version of the model, the auditory target formant trajectories are pre-defined by the user. DIVA can model the acquisition of phonological representations in the sense of feedforward motor commands of the phonological building blocks underlying speech production (whether these are phoneme-, syllable-, or word sized chunks), but does not encompass the build-up of phonological contrasts.

With respect to building up phonological representations, a number of findings indicate that there is interdependence between perception and production. Specifically, poor articulation affects the perceptual acuity for phonological contrasts, and vice versa2. For example,

---
2The evidence for the developmental influence of perception on production is abundant (e.g., Ertmer, Young, & Nathani, 2007; MacNeilage & Davis, 1990; Schorr, Roth, & Fox, 2008; Von Hapsburg, Davis, & MacNeilage, 2008;
Groenen, Maassen, Crul, and Thoonen (1996) found subtle (subclinical) auditory processing deficits in children with CAS; auditory discrimination was found to be poorer compared to normally developing controls. Furthermore, the results showed a specific relation between the perception and production of place-of-articulation. In particular, the degree to which the auditory processing of place-of-articulation cues was affected (assessed by the mean discriminability of a /bak/-/dak/ stimulus continuum) was found to be related to the frequency of place-of-articulation substitutions in production. No such relation was found for a voicing contrast (Groenen et al., 1996). In a follow-up study, Maassen, Groenen, and Crul (2003) found poorer identification and discrimination functions for the perception of two vowel continua for children with CAS as compared to normally developing controls. These results indicate a developmental interdependence between production and perception. Therefore the relation between the quality of the auditory feedback control subsystem and the involvement of phonological vs. motor processes as implied by the present findings should be visible on the symptom level. A number of behavioral studies have found evidence for such a relation between perceptual acuity and production symptoms in children with developmental speech disorders, on several cognitive levels.

Nijland (2009) investigated the relation between perception and production on different cognitive levels in a variety of children with speech output disorders (speech sound disorder; SSD) as compared to a group of age-matched normally developing controls. The children with speech output disorders were divided into three groups: childhood apraxia of speech, phonological disorder (PD) and a mixed group consisting of children that exhibited characteristics of both and could not be classified. The results provided an interesting pattern. Higher-order production disorders seemed to be linked to higher-order perception problems, whereas lower-order production problems did not seem to be linked to lower-order perception problems only (Nijland, 2009).

In essence, these results are the perceptual parallel of the results we found for speech production in the present simulation study. If, in addition to impairment of the feedforward control subsystem, the auditory feedback subsystem is impaired—as in the increased levels of neural noise (H2) condition—a poor systemic mapping results (and consequently poor speech output after imitation learning). In this condition, there is a direct auditory impairment on a low cognitive level, which affects low level speech perception and might spread up and cause perception problems on higher levels. In the reduced/degraded somatosensory information condition, however, the impairment of the feedforward control subsystem in combination with the auditory feedback control subsystem being intact leads to deviant phonemic mappings (higher cognitive level). In this latter case, the acquired phonemic representations are deviant, which will affect speech perception on the phonemic level.

Investigating preschool-age children diagnosed with PD, three age groups of typically developing children and adults, Edwards, Fox, and Rogers (2002) found a relationship between the ability to discriminate CVC words that differed only in the identity of the final consonant, and receptive vocabulary size as well as articulatory accuracy across all subjects. These results were interpreted to suggest that there is a complex relationship among word learning skills,

Warner-Czyz & Davis, 2008; Warner-Czyz, Davis, & MacNeillage, 2010).
the ability to attend to fine phonetic detail, and the acquisition of articulatory-acoustic and acoustic-auditory representations (Edwards et al., 2002). Munson, Edwards, and Beckman (2005) investigated the relationship between the frequency effect in a nonword repetition task and other measures of linguistic ability in 40 young children diagnosed with PD compared to 40 age-matched typically developing controls. Results showed that children in both groups repeated low-frequency sequences less accurately than high-frequency sequences. The children with PD were less accurate overall but did not show a larger frequency effect. These results suggest that the speech-production problems in children with PD are not associated with deficits in the formation of autonomous categorical phonemic representations, but rather with difficulties forming robust representations of the acoustic/auditory and articulatory characteristics of speech in the primary sensory and motor domains (Munson et al., 2005). Note that this appears to be exactly what happened in DIVA in the current simulations under the reduced/degraded somatosensory information (H1) deficit, which caused learned feedforward motor commands (phonemic mappings between articulatory movements and language-specific goals) to be inherently deviant.

In summary, these results indicate that there is a close relation between perceptual acuity and production symptoms in children with developmental speech disorders. Lower order problems (such as CAS) can lead to lower and higher order difficulties in either direction, from production to perception and from perception to production. Differential diagnosis of developmental speech disorders thus is a precarious exercise. A recent study investigating the stability and composition of functional synergies for speech movements in children with developmental speech disorders found the differences in speech motor characteristics between SSD and subtype CAS to be qualitative rather than quantitative and suggested that both groups increase movement amplitude as an adaptive strategy to increase articulatory stability (Terband, Maassen, Van Lieshout, & Nijland, 2011). A high overlap in symptomatology appears to be the rule rather than the exception, thus necessitating an approach that aims at describing profiles of symptoms to characterize disorders rather than focusing on single diagnostic markers for differential diagnosis (Maassen et al., 2010). Weismer and Kim (2010) have made a similar argument with respect to acquired motor speech disorders. Based on the clinical observation that the variety of neurological diseases resulting in dysarthria and apraxia of speech produce a core of similar speech symptoms, they argue for a taxonomical approach to the study of motor speech disorders, as an alternative for the dominant classification approach.

3.5.3 Predictions for auditory and articulatory perturbation experiments

Back to the two hypotheses about the core neurological deficit in CAS. The simulation results in this study provided clear differences between the two deficits in the effects on the speech system: reduced/degraded somatosensory information (H1) led predominantly to deterioration at the phonological level, whereas increased levels of neural noise (H2) led predominantly to deterioration at the phonetic level. However, the simulation results did not differentiate directly between the hypotheses as both deficits led to degraded speech output with similar characteristics after the completed acquisition process. To decide between the two
hypotheses, measurements of speech production in children with CAS and SSD under specifically manipulated circumstances are needed.

In this respect, the present findings lead to directly testable predictions for auditory and articulatory perturbation experiments. One possibility is to compare speech production in the condition of normal auditory feedback with a condition in which the auditory feedback control loop is blocked by masking noise (applied through headphones). In such an experiment, the lack of auditory feedback control is predicted to make a difference for increased levels of neural noise (H2), but not for degraded somatosensory information (H1). If children with CAS store inherently deviant motor commands compared to normally developing children, as was found in the simulations for degraded somatosensory information (H1), the speech system cannot compensate using auditory feedback (as normally developing speakers do) when it is masked. Consequently, auditory deprivation will affect the speech output. In contrast, for increased levels of neural noise (H2), the stored motor commands were found to be relatively intact with the speech problems stemming mainly from motor production/execution processes. If this were the primary neurological deficit in children with CAS, auditory feedback masking would not cause further deterioration of speech production.

Another possibility is an auditory feedback perturbation experiment in which an acoustic cue (e.g., the first formant frequency) is unexpectedly shifted in real time during speech production. The speech output can then be compared to the speech output in a condition of normal auditory feedback. Healthy adults show compensation to the shift by altering the F1 of their speech in the direction opposite the induced shift within approximately 136 ms of onset (Tourville, Reilly, & Guenther, 2008). If the forward model and the auditory feedback control subsystem are intact in children with CAS, as proposed in condition H1 (reduced/degraded somatosensory information), this would allow a relatively normal compensatory response to the shifts in F1. On the other hand, if the forward model and the auditory feedback control subsystem are impaired, as in H2 (increased levels of neural noise), this would prevent a consistent compensatory response.

3.5.4 Conclusions

In conclusion, the results of a series of computer simulations with the DIVA model showed that both deficits that were investigated lead to an increase in CAS characteristics after imitation learning, but for different reasons: increased levels of neural noise (H2) predominantly leads to deterioration on the phonetic level (systemic mapping or forward model) whereas reduced/degraded somatosensory information (H1) predominantly leads to deterioration on the phonological level (phonemic mappings). These findings lead to interpretations on the specificity of CAS and yield directly testable predictions for auditory and articulatory perturbation experiments. Furthermore, the results imply a close relation between the quality of the system for self monitoring auditory error and the involvement of phonological vs. motor processes in children with speech sound disorders.

Although limited to the more fine-grained phonetic speech characteristics, the current work confirms that simulation studies can give valuable insights into the neurological mechanisms and deficits that underlie speech disorders. We are currently working on expanding the
Neurological mechanisms underlying CAS

scope of our modeling studies to the phonological and suprasegmental levels of the speech production process.

Acknowledgements

This study was funded by the Netherlands Organization for Scientific Research (NWO), and in part by NIH grant R01 DC07683, the Fulbright Association, and CELEST, a National Science Foundation Science of Learning Center (SMA-0835976). The authors thank Jason Tourville and Satrajit Ghosh for helpful discussions about the DIVA model. Furthermore, the authors gratefully thank the UIL-OTS and Iris Mulders (Utrecht University) for their facilitation of the perception experiment.

References


Chapter 3


Neurological mechanisms underlying CAS


Figure 3.6: Comparison between the effect of the deficits if present during the whole learning process (continuous lines) and if implemented after (undisordered) asymptotic learning (zero-noise; dashed lines), for degraded somatosensory information (H1; top four panels) and increased levels of neural noise (H2; bottom four panels). Clockwise, starting at upper left panel, the panels show coarticulation (ANT = anticipatory; CO = carry-over), searching articulatory behavior, token-to-token variability of mean formant frequencies, and speech sound distortion in relation to severity of deficits.
Figure 3.7: Acquired forward commands for the word /ugi/ for the two implemented deficits in comparison with the standard (baseline) model. The acquired forward commands are extracted by production attempts in the performance stage with the feedback control sub-system turned off, i.e., letting the model run the stored feedforward command without online feedback control. The two middle panels show the forward (motor) command that the model acquired for this target and its corresponding somatosensory state representation. The top-panel shows the target regions for F1, F2, and F3 (in black) and the auditory consequences that correspond to the model's motor command (in red). The auditory error that is inherent to the acquired forward command is depicted in the bottom panel.