



Published in final edited form as:

*J Speech Lang Hear Res.* 2009 December ; 52(6): 1595–1609. doi:10.1044/1092-4388(2009/07-0283).

## Computational Neural Modeling of Speech Motor Control in Childhood Apraxia of Speech (CAS)

H. Terband<sup>1</sup>, B. Maassen<sup>1</sup>, F.H. Guenther<sup>2,3</sup>, and J. Brumberg<sup>2</sup>

<sup>1</sup>Medical Psychology/Pediatric Neurology Centre/ENT, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands <sup>2</sup>Department of Cognitive and Neural Systems, Boston University, Boston MA, USA <sup>3</sup>Division of Health Sciences and Technology, Harvard University/Massachusetts Institute of Technology, Cambridge MA, USA

### Abstract

**Purpose**—Childhood apraxia of speech (CAS) has been associated with a wide variety of diagnostic descriptions and has been shown to involve different symptoms during successive stages of development. In the present study, the authors attempted to associate the symptoms of CAS in a particular developmental stage with particular information-processing deficits by using computational modeling with the Directions Into Velocities of Articulators (DIVA) model. The hypothesis was that the speech production system in CAS suffers from poor feed-forward control and, consequently, an increased reliance on the feedback control subsystem.

**Method**—In a series of computer simulations, the authors systematically varied the ratio between feed-forward and feedback control during production attempts in the acquisition of feed-forward motor commands. The simulations were evaluated acoustically on 4 selected key symptoms of CAS.

**Results**—Results showed that increasing the reliance on feedback control causes increased severity of these 4 symptoms of CAS: deviant coarticulation, speech sound distortion, searching articulation, and increased variability.

**Conclusions**—The findings support the idea that the key symptoms found in CAS could result from an increased reliance on feedback control due to poor feed-forward commands. Two possible root causes of degraded feed-forward control in CAS are discussed: reduced somatosensory information and increased levels of neural noise.

### 1. Introduction

Childhood apraxia of speech (CAS) is considered an impairment of purposeful speech movements (Groenen, Maassen, Crul, & Thoonen, 1996; Hall, Jordan, & Robin, 2007; Maassen, Groenen, & Crul, 2003) and is generally defined as a disorder of motor planning and/or motor programming (e.g., Caruso & Strand, 1999; Hayden, 1994; Maassen, Nijland, & Van der Meulen, 2001; Nijland, Maassen, & Van der Meulen, 2003; Ozanne, 2005; Peter & Stoel-Gammon, 2005; Schmidt & Lee, 1999; Smith, Marquardt, Cannito, & Davis, 1994; Van der Merwe, 1997). More specifically, CAS can be defined as “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). The clinical criteria for diagnosing CAS, however, are controversial. As stated

in ASHA's (2007, p. 4) technical report, "Review of the research literature indicates that, at present, there is no validated list of diagnostic features of CAS that differentiates this symptom complex from other types of childhood speech sound disorders, including those primarily due to phonological-level delay or neuromuscular disorder (dysarthria)." One of the main difficulties is the evolving nature of clinical symptoms (Maassen, 2002; Stackhouse, 1992b). Symptoms vary dramatically at different stages of development, starting with little or no canonical babbling and subsequently less variegated babbling during the first year (for an overview of reports on babbling in CAS, see, e.g., Hall et al., 2007; Ozanne, 2005) and moving to slow expansion of vocabulary during the second year of life (Maassen, 2002). From that age, speech production of children with CAS has been associated with a variety of symptoms (ASHA, 2007; Hall et al., 2007; Shriberg, Aram, & Kwiatkowski, 1997; Stackhouse, 1992a).

Differential diagnostic characteristics comprise inconsistent errors of consonants and vowels in repeated productions of syllables or words and lengthened, disrupted coarticulatory transitions between sounds and syllables, and inappropriate prosody, especially in the realization of lexical or phrasal stress (ASHA, 2007). The speech sound errors mainly comprise a large number of consonantal errors in which omissions are more prevalent than substitutions (Hall et al., 2007; Ozanne, 2005). Further characteristics include groping as well as difficulties and low-maximum repetition rates in the production of alternate syllables or diadochokinesis. The concept of groping is a collective term that is used in the literature to denote two similar but different types of speech motor behavior. First, prevocalic groping or silent posturing can be defined as "a static state of articulatory positioning that occurs without sound production" (Hall et al., 2007, p. 43). The second type of groping constitutes searching articulatory behavior that takes place during sound production "in an attempt to find the desired articulatory position necessary for correct phoneme production" (Hall et al., 2007, p. 43). This type of searching articulatory behavior often involves repetitive attempts to produce the desired target, constituting trial-and-error behavior (e.g., Hall et al., 2007; Ozanne, 2005; Stackhouse, 1992a).

More fine-grained phonetic characteristics include distorted productions of speech sounds (mainly vowels) and deviant coarticulation (ASHA, 2007; Hall et al., 2007; Ozanne, 2005; Stackhouse, 1992a). Speech sound distortions comprise nonphonemic productions of speech sounds that often defy accurate transcription, even when using narrow transcription (Hall et al., 2007). Pollock and Hall (1991) perceptually examined the vowel and diphthong production of five children with CAS and found that all five children made a variety of errors that were categorized as diphthong reduction, tensing, laxing, and derhotacization. Although the perceptual analysis of speech sound errors often is troublesome in pathological speech, in this case the perceptual judgments were acoustically validated by Walton and Pollock (1993). In a study in our laboratory, Nijland et al. (2002) found a reduction in vowel distinction and deviant coarticulation patterns in the speech of children diagnosed with CAS as compared with normally developing children. Although the results showed large individual differences among the children with CAS, the most prevalent finding was a larger within-subject variability of productions for the children with CAS than for the controls. Furthermore, the errors were found not to be typically immature. In a subsequent study that further investigated the coarticulation patterns, coarticulation was found to be stronger and more extended in the speech of children with CAS as compared with normally developing children (Nijland, Maassen, Van der Meulen et al., 2003).

In psycholinguistic models, CAS is considered a multideficit disorder that involves three levels: phonological planning, assembly of a phonetic program, and motor implementation (Ozanne, 2005), with the core deficit at the phonetic programming level (Caruso & Strand, 1999; Ozanne, 2005; Schmidt & Lee, 1999; Van der Merwe, 1997). This can be summarized by stating that CAS is an inability to transform an abstract phonological code into motor speech commands

(Nijland, 2003). To pursue this further, the aim of the present study was to explore which characteristics can be accounted for by computational modeling. For this, we used the computational neural model of speech motor control called the Directions Into Velocities of Articulators (DIVA) model (Guenther, 1994; Guenther, Ghosh, & Tourville, 2006). If the speech characteristics of children with CAS could be simulated with the DIVA model, then this would give valuable insights into the neurological mechanisms and deficits of this disorder. This applies even more when a diverse set of characteristics can be modeled by a single parameter manipulation: If several symptoms that are in a psycholinguistic model hypothesized to emerge from multiple underlying deficits could be explained with a single model manipulation, then this would yield a strong indication for a more general underlying processing principle involved in CAS.

## 2. Overview of the DIVA Model

The DIVA model is a neural network model of speech acquisition and production, the components of which correspond to regions of the cerebral cortex and cerebellum (Guenther et al., 2006). DIVA is a model of speech motor control and focuses on the sensorimotor transformations underlying the control of articulatory movements. According to the model of Levelt (1989), the speech production process comprises the successive stages of conceptualization, formulation (which can be further divided in grammatical and phonological encoding), and articulation. In this framework, the DIVA model comprises the process of articulation.

Figure 1 presents an overview of the model. The DIVA model is computationally implemented and consists of a neural network controller detailing feed-forward and feedback control loops that are involved in early speech development and mature speech production. The boxes in the diagram correspond to maps (sets of neurons), and the arrows correspond to synapses. In order to produce an acoustic signal, DIVA controls the movements of an articulatory synthesizer (Maeda, 1990).

In the first stage of learning in the model, semi random articulatory movements (meant to roughly correspond to early stages of infant babbling) are used to learn relations between motor commands and their auditory and somatosensory consequences. The information is stored in the synapses between motor, auditory, and somatosensory cortical areas, represented by arrows in Figure 1. In particular, the model learns mappings between auditory and somatosensory information and the corresponding motor commands. For example, the model learns how to transform detected auditory errors, which represent desired changes to the formant frequencies, into motor commands that will reduce these errors. That is, it learns a mapping from changes in formant frequencies into motor commands that affect these changes. Later during development, this mapping allows the DIVA model to compute the motor commands that are necessary to achieve the desired auditory consequences for new speech sounds (i.e., to reach the auditory target region for the sound currently being produced).

In the second stage of learning (meant to correspond to the imitation stage), the model is presented with sample speech sounds and learns an auditory target for each sound, which can consist of a single phoneme, a syllable, a word, or a small phrase. The auditory targets are effectively stored in the synaptic projections from speech sound map cells in left ventral premotor cortex to cells in the higher order auditory cortical areas.

After learning the auditory target for a new speech sound, feed-forward commands for the sound are learned during repeated attempts to produce the sound by the model. These commands are encoded by synaptic weights projecting from speech soundmap cells in the left ventral premotor cortex to articulator velocity and position cells in the primary motor cortex bilaterally (labeled Feedforward commands in Figure 1). On the basis of the information

provided by the auditory feedback control subsystem, which compares the actual auditory signal with the desired auditory target, the feed-forward command is updated with each attempt, thus becoming more accurate. Eventually, the feed-forward command is accurate enough for the model to produce the sound without generating auditory errors and thus without invoking the auditory feedback control subsystem (resembling normal/adult production). During acquisition of the feedforward command, the model also learns a somatosensory target region for the sound encoding the tactile and proprioceptive information that accompanies correct production of the sound. This somatosensory target is used in the somatosensory feedback control subsystem during normal production. For a detailed description of the DIVA model, its neural correlates, and its mathematical and computational implementation, see Guenther et al. (2006). An overview of the model's parameters and equations is presented in Appendices A and B.

### 3. CAS in DIVA

According to the DIVA model, the acoustic signal plays a central role in speech acquisition and production, as embodied by the model's auditory targets and auditory feedback control subsystem. Regions in auditory perceptual space form the targets that the model strives to achieve in order to produce a particular speech sound. If the actual auditory signal does not correspond to the desired auditory target, then an online corrective motor command will be calculated. In addition, the feed-forward command is updated for the next attempt (in order to become more accurate). Because the auditory signal not only serves as a “teaching” signal but also is the input for the calculation of feed-forward commands, auditory targets in the DIVA model correspond to what is traditionally seen as motor plans or phonetic plans. In the DIVA model, the motor commands for different speech sounds are stored in feed-forward projections that specify articulatory trajectories that will produce the desired auditory target. These feed-forward projections correspond to motor programs in traditional psycholinguistic models (e.g., in the model of Van der Merwe, 1997), although it should be noted that in DIVA motor programming and execution are not strictly separated.

A problem in the transformation of an abstract phonological code into motor speech commands, hypothesized to underlie CAS (Caruso & Strand, 1999; Nijland, 2003; Ozanne, 2005; Schmidt & Lee, 1999; Van der Merwe, 1997), can thus be simulated by poor feed-forward control in the DIVA model. In DIVA, the feed-forward control subsystem establishes the coupling between the auditory target (phonetic plan) and the motor command (motor program) that will produce it. If feed-forward commands are incorrect and/or imprecise, then the articulatory movements are off target. If the articulatory movements are off target, then this leads to error cell activation in the auditory and somatosensory error maps, causing the auditory and somatosensory feedback systems to generate corrective commands (labeled Feedback commands in Figure 1). The introduction of errors due to poor (incorrect and/or imprecise) feed-forward control thus causes the system to rely more heavily on the sensory feedback control subsystems. Therefore, the model predicts that the speech motor control of children with CAS is biased toward feedback control.

To test whether an overreliance on feedback control due to impaired feed-forward control can account for the speech characteristics in CAS, we conducted a series of computer simulations with the DIVA model in which we systematically varied the ratio between feed-forward and feedback control. The computationally implemented model is equipped with a standard mapping of the relations between motor commands and their auditory and somatosensory consequences, which was acquired during a training stage of semirandom articulatory movements (resembling infant babbling). The model observations were done during production attempts in the acquisition of feed-forward commands to produce prespecified auditory targets. In children, this corresponds to imitation learning, from early attempts to

produce a new meaningful utterance (e.g., a word) until skillful production without online feedback control.

Four key characteristics of CAS were identified for evaluation: deviant coarticulation, speech sound distortion, searching articulation, and increased variability. There are several reasons for focusing on this specific set of speech motor characteristics of CAS. First, because sound targets are predefined in the present computational implementation of the model, deviant suprasegmental characteristics (e.g., inappropriate prosody) or phonological characteristics (e.g., speech sound omissions, substitutions and transpositions) could not be investigated. For the same reason, a number of speech motor characteristics also could not be investigated, such as prevocalic groping and speech sound prolongations. Therefore, in these simulations, we focused on the more fine-grained phonetic characteristics of CAS speech, which are largely independent of phonological development and can be simulated with DIVA. These characteristics are indicative of CAS during an extended period of speech development, comprising the whole stage of imitation learning.

The effects of impaired feed-forward control and an overreliance on feedback in the DIVA model has been previously studied by Max, Guenther, Gracco, Ghosh, and Wallace (2004) in the context of stuttering. In their simulations, Max et al. implemented a bias toward feedback control as the basic deficit. This overreliance on feedback could explain the observation that auditory masking, frequency-altered, and delayed auditory feedback sometimes have a fluency-enhancing effect (e.g., Ingham, Moglia, Frank, Ingham, & Cordes, 1997; Postma & Kolk, 1992). Presumably, in these cases the modification or absence of auditory feedback prevents the use of auditory feedback control. To explain the typical stuttering phenomena (blocking, sound repetitions), Max et al. additionally implemented a reset function. Triggered by sensory errors exceeding a certain threshold, the reset signal orders the system to abort the current production and start over. Although no detailed results were presented, the authors reported that the combination of increased sensory error and a reset function lead to stuttering behavior, in particular sound repetitions. The present study differs in two important aspects from this previous work. Max et al. did not investigate the effects of impaired feed-forward control processes on speech acquisition (investigating only the mature system). Furthermore, their simulations involved a reset function; the typical stuttering phenomena found in their study are the results of restart based on reset. In the present study, we focused on the effects of the amount of feedback on a selection of typical CAS symptoms. However, the underlying parallel is that similar underlying mechanisms (i.e., overreliance on feedback control) might play a role in different speech production disorders.

## 4. DIVA Simulations

### 4.1 Method

The simulation series was performed on a personal computer running the 2007 distribution of the computational implementation of the DIVA model, written in the MATLAB programming environment.<sup>1</sup> In the simulation series, the ratio between feed-forward and feedback control was varied during production attempts in the acquisition of feed-forward commands. The acquisition process comprised 10 iterations, which corresponds to approximately asymptotic learning. The feed-forward/feedback ratio was manipulated ranging from the standard level of 90/10 to 55/45 in 5% steps.<sup>2</sup> An overview of the model's parameters and equations can be found in Appendices A and B. VICV2 utterances were used as speech sound targets, comprising all combinations of /a/, /i/, and /u/ for the vowels (V1 and V2) and /b/, /d/, and /g/ for the consonants (C), forming a total number of 27 sound targets. The targets had a fixed

<sup>1</sup>The source code is available from <http://speechlab.bu.edu/DIVAcodes.php>.

<sup>2</sup>Driven by the experience that lower feed-forward/feedback ratios do not yield intelligible productions, 55/45 was chosen as a limit.



duration of 560 ms, consisting of 200 ms for the vowels, 100 ms for the consonant,<sup>3</sup> and 30 ms for each of the transitions. The spectral characteristics of the auditory targets were specified by trajectories of the first three formant frequencies, that is, the target for a sound consisted of timevarying regions in F1; F2; F3-space. The simulations produced an acoustic realization of each target VCV utterance.

The formant frequencies (F1, F2, F3) of each speech sound of each VCV utterance<sup>4</sup> were measured at three points in time: at the beginning, in the middle, and at the end of the steady-state area (see Figure 2). In order to focus on relative rather than absolute frequency differences, the formant values measured in Hz were normalized using a  $\log_{10}(x)$  transformation.

Coarticulation refers to the phenomenon that the specific properties of articulator movements are context dependent. Acoustically, this manifests itself as the realizations of consecutive speech segments affecting each other mutually. This is exemplified in Figure 3. Coarticulation was measured in both vowels and consonants. Coarticulation effects usually change the characteristics of a speech sound in the direction of the neighboring speech sound, but the deviant coarticulation patterns of children with CAS found by Nijland and colleagues (Nijland et al., 2002; Nijland, Maassen, Van der Meulen et al., 2003) also contained hyperarticulation (change in the opposite direction: enhanced contrasts). Therefore, the amount of coarticulation was estimated by the absolute differences in mean formant frequencies of a particular speech sound across all possible vowel contexts. Mean absolute differences across the three vowel contexts were calculated for each formant. Subsequently, a coarticulation index was calculated by averaging over the three formants. In this way, both anticipatory coarticulation (i.e., differences in the realization of V1 [intersyllabic] and C [intrasyllabic] in the three vowel contexts V1C/i/ vs. V1C/u/ vs. V1C/a/) and carry-over coarticulation (i.e., differences in the realization of V2 and C [both intersyllabic] in the three vowel contexts /i/CV2 vs. /u/CV2 vs. /a/CV2) were measured. Formulas of the calculations can be found in Appendices A and B.

The distortion of speech sounds was calculated by averaging the absolute differences in mean formant frequencies (F1, F2, and F3) of each produced speech sound relative to the frequencies of the target speech sound (a formula is included in Appendices A and B). In this way, an index is created of how much the realized speech sound deviates from the intended target sound.

It is not possible to simulate prevocalic groping or silent posturing in the present computational implementation of the DIVA model. Therefore, the focus was on searching articulatory behavior that takes place during sound production in an attempt to achieve the desired target. As was mentioned in the introduction, this type of searching articulatory behavior often involves repetitive attempts to produce the desired target. However, in the present computational implementation of the DIVA model, this trial-and-error aspect could not be investigated. Therefore, the amount of groping or online searching articulatory behavior was quantified by the comparison of the formant frequencies at the beginning, middle, and end of each speech sound. In this way, an indication of the change in formant frequencies during the course of each speech sound is obtained, which corresponds to the online articulatory adjustments that constitute this type of groping or searching articulatory behavior. First, for each formant of each speech sound, the standard deviation of its values at the beginning, middle,

<sup>3</sup>A 100-ms consonant duration is decomposed into a 70-ms closure period followed by a 5-ms smooth transition to a 25-ms pre-release voicing period that simulates the effect of air pressure in the vocal tract below the closure point. This procedure is suggested by Maeda (1996) for phoneme concatenation of vowels and unvoiced stop consonants and was adapted to the voiced stop consonant VCV utterances used in these simulations.

<sup>4</sup>A large number of studies focus on F2 trajectories. Although this is often not substantiated, the reason for this is likely to be the study of Öhman (1966), in which he showed that for VCV utterances, F2 is the most indicative of the first three formants. However, because all formants do contain spectral information, and because DIVA does not discriminate between formants (i.e., does not weigh the auditory error signal of different formant frequencies differently), the acoustic measures used in this study take all three formants into account.

and end of the production was calculated. These three values were then averaged to create an index of the general amount of groping or online searching articulatory behavior (a formula can be found in Appendices A and B).

Finally, the variability between productions was measured by the standard deviations in the mean formant frequencies of repeated productions of the same speech sound (see Appendices A and B for a formula of the calculations). Note that this is a very rough index of token-to-token variability that also encompasses variability due to coarticulation and speech sound distortions. In addition, the standard deviations of the coarticulation, speech sound distortion, and groping measures were examined as variability indices.

To test whether any of the effects were significant, analyses of variance were conducted first, with feed-forward/feedback ratio as a factor and vowel and consonant identity as covariates. However, the assumption of homogeneity of variance failed for all variables, as the Levene's tests turned out to be significant. Because parametric testing was not justified, nonparametric tests (Kruskal–Wallis) were conducted.

## 4.2 Results

Figure 4 presents the mean coarticulation in V1, C, and V2 as a function of feed-forward/feedback ratio. The results show an increase in coarticulation as the feed-forward/feedback ratio decreases. Among the vowels, anticipatory coarticulation is stronger than carryover coarticulation. In the consonant, the opposite effect was found: Carryover coarticulation is stronger than anticipatory coarticulation, but this difference only manifests itself from a feed-forward/feedback ratio of 80/20 (see Figure 4).

The course of coarticulation along the utterance is depicted in Figure 5. The data indicate that anticipatory coarticulation is larger in V1 than in C, whereas carryover coarticulation is larger in C than in V2. Note that the results show especially high values at the beginning of V1 (see Figure 5, left panel), which corresponds to utterance onset. Utterance onset appears to constitute an exceptional case, which is discussed in detail below.

Nonparametric testing (a series of Kruskal–Wallis tests for  $x$  independent samples) showed that for V1, C, and V2, all effects of feed-forward/feedback ratio were significant at a  $p < .001$  level. For anticipatory and carryover coarticulation, we also tested the effect separately for the beginning, middle, and end of speech sounds. With the exception of anticipatory coarticulation at the beginning of V1 ( $p = .067$ ), all effects of feed-forward/feedback ratio were significant (all  $ps < .001$ ).

For speech sound distortion, the results show an increase for V1, C, and V2 (see Figure 6). Apart from the feed-forward/feedback ratio of 60/40, the amount of distortion is larger in V1 than in C and V2. Between C and V2, the results do not show large differences. A series of Kruskal–Wallis tests for  $x$  independent samples showed the effects of feed-forward/feedback ratio for V1, C, and V2 to be significant at  $p < .001$ .

The amount of groping or searching articulatory behavior was captured by the standard deviations of the formant values at the beginning, middle, and end of each speech sound. The results show an increase in searching articulatory behavior as the reliance on feedback control increases (see Figure 7). Overall, the mean variability in the course of the productions was largest in V1, but the difference between V1 and C decreased as the reliance on feedback increased. From feed-forward/feedback ratios of 60/40 to 90/10, the results show no difference between V1 and C. V2 shows the smallest amount of variability in the course of its production. Statistical analysis (a series of Kruskal–Wallis tests for  $x$  independent samples) showed the effect of feed-forward/feedback ratio to be significant for V1, C, and V2 (all  $ps < .001$ ).

Figure 8 presents the variability between productions as a function of feed-forward/feedback ratio. The results show an increase in variability as the reliance on feedback control increases, both in terms of the standard deviation of mean formant frequencies (see Figure 8, left panel) and in terms of the standard deviations of the coarticulation, speech sound distortion, and searching indices (see Figure 8, right panel). A series of Kruskal–Wallis tests for  $x$  independent samples showed the effects of the feed-forward/feedback ratio on the standard deviation of mean formant frequencies to be significant for C ( $p < .05$ ) and V2 ( $p < .01$ ) but not for V1 ( $p = .087$ ).

On the basis of the differences in standard deviation shown in Figure 8 (right panel), the aforementioned lack of homogeneity of variance is to be expected, but interacting factors might be present underneath. Therefore, we investigated the effects of vowel and consonant identity. We conducted a series of homogeneity tests for each vowel and each consonant, separately for V1, C, and V2. All proved to be significant; thus, the differences in variance cannot be accounted for by interactions with vowel or consonant identity. Furthermore, the results showed no large differences in variance between V1, C, and V2. We therefore conclude that the differences in variance (see Figure 8, right panel) are significant differences that reflect a main effect of feed-forward/feedback ratio.

### 4.3 Discussion

In the simulation series, we systematically varied the ratio between feed-forward and feedback control during production attempts in the acquisition of feed-forward motor commands to test the prediction that the speech motor control of children with CAS is biased toward feedback control. High feed-forward/feedback ratios in the simulation series correspond to the speech acquisition of normally developing children. The models standard setting of feed-forward/feedback is 90/10, but slightly lower ratios are presumed to still correspond to normal development. Lower feed-forward/feedback ratios (in these simulations from 70/30 downward), however, are presumed to correspond to speech acquisition in CAS, with a decrease of the feed-forward/feedback ratio reflecting increased severity of the disorder.

The simulation results showed an increase in coarticulation, speech sound distortion, groping, or searching articulatory behavior as the reliance on feedback control increased. Furthermore, the results showed an effect of feed-forward/feedback ratio for variability in terms of the standard deviations of the coarticulation, speech sound distortion, and groping indices, and for the variability of mean formant frequencies. Overall, the simulation results indicated that an increased reliance on the feedback control subsystem can account for at least four key characteristics of CAS, thus indicating that the symptoms of CAS could result from impaired (incorrect and/or imprecise) feed-forward commands.

We found a large difference between the magnitude of the effects on V1 as compared with the speech sounds in the middle and final positions of the utterance. The amounts of coarticulation, speech sound distortion, searching articulatory behavior, and variability all were larger in V1 than in C and V2. Furthermore, with respect to the standard deviation of mean formant frequencies, the effect of feed-forward/feedback ratio proved not to be significant for V1. As a whole, these findings raise suspicion that a different mechanism could be at work regarding V1. A closer look at the results on the course of coarticulation may provide an explanation. The results showed a significant effect of feed-forward/feedback ratio on anticipatory and carryover coarticulation over the whole course of the utterance (beginning, middle, and end of each speech sound), with the exception of anticipatory coarticulation at the beginning of V1, in which case the effect of feed-forward/feedback ratio proved not to be significant. An interpretation of these findings could be that the present state of the model is poorly defined at the onset of the utterance. At speech onset, the model lacks afferent information about the present state of the vocal tract, which causes the model's information of vocal tract positioning



to be less precise. In turn, this causes an imprecision in the feed-forward commands, whose specifics are dependent on vocal tract state. This baseline error has a magnifying effect on the production errors and thus predicts speech sound distortion to be larger in initial position. Furthermore, as a larger error in turn invokes a larger correction, it also predicts a larger variability in initial position. This corresponds to findings that in CAS, different error patterns are found for word-initial as compared with word-final position. Thus, higher error rates and more diversity of error types have been reported for speech sounds and speech sound clusters in initial than in final position (e.g., Shriberg et al., 1997; Thoonen, Maassen, Gabreels, & Schreuder, 1994). In this respect, a parallel with stuttering becomes apparent. The high uncertainty at utterance initial position might provide an explanation for the specific utterance onset problems found in stuttering.

With respect to anticipatory and carryover coarticulation, the simulations yield interesting results. Given the models' particular behavior at utterance onset (as discussed above), the vowel data do not provide a balanced comparison, but the data on the consonants do. Down to a feed-forward/feedback ratio of 80/20, the results showed no clear difference between anticipatory and carryover coarticulation in the consonant. However, carryover coarticulation got stronger as the feed-forward/feedback ratio went from 80/20 to a feed-forward/feedback ratio of 55/45. In this respect, the simulation results lead to the straightforward testable prediction that both types of coarticulation are increased in the speech of children with CAS in comparison to normally developing speakers but that the increase in carryover coarticulation is larger than the increase in anticipatory coarticulation.

Although coarticulation in the speech of children with CAS has been investigated in some studies (e.g., Nijland, Maassen, & Van der Meulen, 2003; Nijland et al., 2002; Nijland, Maassen, Van der Meulen et al., 2003; Sussman, Marquardt, & Doyle, 2000), anticipatory and carryover coarticulation have not yet been compared in a single experimental design for this pathological group. Also, more generally in pathological speech or in speech in abnormal circumstances (e.g., fast or slow speech), the number of studies in which anticipatory and carryover coarticulation have been directly compared is limited. Hertrich and Ackermann (1995) found in slow as compared with normal speech rate of normally speaking adults a decrease in carryover coarticulation while anticipatory coarticulation remained the same. According to the authors, this indicates that these two types of coarticulation are controlled by different mechanisms. Carryover coarticulation is thought to reflect biomechanical or motor constraints (e.g., inertia, transmission delays in neuromotor control processes), whereas anticipatory coarticulation seems to represent higher level phonetic processing. Supplementary evidence for this view was found in the coarticulation patterns of speakers with ataxic dysarthria (Hertrich & Ackermann, 1999). Because the DIVA model does not contain inertia-type biomechanics, we hypothesize from these data that in our simulations, neuromotor constraints (i.e., transmission delays that are incorporated in the control loops) are the main determinants of the increase in carryover coarticulation in lower feed-forward/feedback ratios. This leads to the testable hypothesis that the difference between carryover and anticipatory coarticulation in the speech of children with CAS in comparison to normally developing controls disappears when speech rate is slowed down.

#### 4.4 What causes degraded feed-forward commands?

We present two hypotheses regarding the underlying neurological mechanisms that cause the degraded feed-forward commands. The first hypothesis is derived from observations of children with CAS having a lowered oral sensitivity of the tongue and palate (Hall et al., 2007; Ozanne, 2005; Stackhouse, 1992a). The importance of somatosensory information in speech motor control has been known for some time (e.g., Lindblom, Lubker, & Gay, 1979) and has recently been emphasized in work by Ostry and colleagues (Nasir & Ostry, 2006;

Ostry & Clark, 2005; Tremblay, Shiller, & Ostry, 2003).<sup>5</sup> In this view, the core deficit of CAS lies in a reduced or degraded oral sensitivity. In DIVA, a lack of somatosensory information would have different effects in successive stages of speech development that have a cumulative effect. In the babbling stage, it causes weak or underspecified synaptic projections from somatosensory error mappings to the articulator velocity map in motor cortex, subsequently leading to degraded somatosensory feedback control in the imitation stage. Furthermore, readout of the appropriate feed-forward commands depends on knowing the current somatosensory state. A poor estimate of somatosensory state will lead to impaired readout of feed-forward commands, resulting in degraded feed-forward commands and thus an increased reliance on the feedback control subsystem.

This view converges with observations of deaf and/or hearing-impaired (HI) speech. Through extensive training, children who are born deaf or (severely) hearing impaired can learn how to speak intelligibly. The characteristics of CAS and deaf or HI speech show specific contrasts and correspondences. For example, regarding plosives, the forceful articulation of HI speech contrasts with the imprecise articulation found in CAS, but both populations have trouble in controlling voiceonset time (VOT). It is clear that in HI, speech motor control, the learning of feed-forward commands is based mainly or even solely on somatosensory feedback. In CAS—as in normal development—the learning of feedforward commands is based mainly on auditory feedback, and due to the lack of somatosensory information, the appropriate feed-forward commands cannot be read out properly. In this view, mastering VOT control requires both sources of information to be intact. It is well known that the acquisition of the voicing contrast requires precise timing control between laryngeal and oral structures (Gracco, 1994; Grigos, Saxman, & Gordon, 2005; Munhall, Löfqvist, & Kelso, 1994). In Lane and Perkell's (2005) review article on VOT control in the absence of hearing, they pointed out that “achieving an appropriate voicing contrast involves an intricate coordination of the timing and magnitude of movements in the respiratory, laryngeal, and supraglottal systems” (p. 1339). Although Lane and Perkell mainly stressed the importance of self-hearing, it is clear that such a coordination also requires adequate somatosensory and proprioceptive information of the current state of the systems that are involved. It appears that if the system fails on one of the two requirements, then a proper coordination cannot be established, and VOT control cannot be mastered fully.

The second hypothesis is that CAS can be explained as resulting from an increased level of neural noise. Neural noise has been suggested to be the primary factor limiting the possibility of simultaneously rapid and accurate movements (Fitts, 1954) and has been widely associated with the token-to-token variability that characterizes human motor performance (Fitts, 1954; Harris & Wolpert, 1998; Meyer, Abrams, Kornblum, Wright, & Smith, 1988; Perkell & Nelson, 1985). Wolpert, Ghahramani, and Flanagan (2001) presented neural noise as the main argument for the existence of paired (forward and inverse) internal models, by whose combination the central nervous system can optimally estimate a current state of the system. In the DIVA model, neural noise could affect all of the neural maps in Figure 1. This would be expected to result not only in incorrect and/or imprecise feed-forward commands but also in poor performance in the auditory and somatosensory feedback control systems.

## 5. Conclusions

In the present study, we attempted to simulate the imitation learning stage of speech acquisition in CAS using the neurocomputational DIVA model. In a series of computer simulations, the

---

<sup>5</sup>Somatosensory information plays two different roles in the DIVA model. Within the somatosensory feedback control system, somatosensory feedback is compared with expected somatic sensations (somatosensory targets) to detect somatosensory errors that are corrected by the somatosensory feedback system. In feed-forward control, somatosensory feedback is needed to indicate the current state of the vocal tract in order to choose the appropriate feed-forward commands. Here, we refer to the latter of these two functions.

hypothesis was tested that the speech production system in CAS suffers from impaired feed-forward commands, and consequently an increased reliance on the auditory feedback control subsystem. The simulation results showed an increase in coarticulation, speech sound distortion, searching articulatory behavior, and variability as the reliance on feedback control increased. These results support the idea that the key symptoms found in CAS could result from an increased reliance on auditory feedback control due to incorrect and/or imprecise feed-forward commands.

For high feed-forward/feedback ratios, anticipatory coarticulation in the consonants was found to be equal to carryover coarticulation. With feed-forward/feedback ratios lower than 80/20, however, carryover coarticulation was found to be stronger than anticipatory coarticulation. In this respect, the simulations predict that in the speech of children with CAS, the difference in carryover coarticulation in comparison to normally developing speakers is relatively large compared with the difference in anticipatory coarticulation. Furthermore, it is predicted that this interaction disappears when speech rate is slowed down. Both the relation between anticipatory and carryover coarticulation and the effect of speech rate on speech motor control have not yet been studied in this pathological group.

Two hypotheses were presented for the underlying mechanisms causing the weak feed-forward control: reduced somatosensory information and an increased level of neural noise. We are presently working on the implementation of these mechanisms in DIVA.

Although the work we presented here is based on experimental findings that have been reported in the literature, it should be noted that it is of a theoretical nature. Further investigation is needed, and the observations need to be verified in children with CAS. Nonetheless, the present work confirms that simulation studies can provide valuable insights into the neurological mechanisms and deficits that underlie speech disorders.

## Acknowledgments

This study was funded by the Netherlands Organization for Scientific Research (NWO) and in part by National Institutes of Health Grant R01 DC02852 and Center of Excellence in Learning, Science, and Technology (CELEST), a National Science Foundation Science of Learning Center (Grant NSF SBE-0354378).

## Appendix

### DIVA model

DIVA model parameters		
Name	Default Value	Description
$\alpha_{ff}$	0.9	Contribution of feedforward command to total command; feedforward gain
$\alpha_{fb}$	0.1	Contribution of the feedback command; feedback gain
$\tau_{MAr}$	5ms	Transmission delay from motor cortex cell activity to physical movement of articulators
$\tau_{ArS}$	10ms	Transmission delay from movement of articulators to feedback signals in somatosensory cortex
$\tau_{ArAu}$	25ms	Transmission delay from movement of articulators to feedback signals in auditory cortex
$\tau_{PM}$	5ms	Transmission delay from premotor (speech sound map) to motor cortex
$\tau_{PS}$	20ms	Transmission delay from premotor to somatosensory cortex
$\tau_{PAu}$	35ms	Transmission delay from premotor to auditory cortex
$\tau_{SM}$	5ms	Transmission delay from somatosensory to motor cortex

DIVA model parameters		
Name	Default Value	Description
$\tau_{AuM}$	5ms	Transmission delay from auditory to motor cortex

DIVA model equations	
Equation	Description
$M(t) = M(0) + \alpha_{ff} \int_0^t \dot{M}_{ff}(t)g(t)dt + \alpha_{fb} \int_0^t \dot{M}_{fb}(t)g(t)dt$	Motor cortex position map
$\dot{M}_{fb}(t) = \Delta Au(t - \tau_{AuM})z_{AuM} + \Delta S(t - \tau_{SM})z_{SM}$	Feedback motor command
$\dot{M}_{ff}(t) = P(t)z_{PM}(t) - M(t)$	Feedforward motor command
$\Delta Au(t) = Au(t) - P(t - \tau_{PAu})z_{PAu}(t)$	Auditory error map activity
$\Delta S(t) = S(t) - P(t - \tau_{PS})z_{PS}(t)$	Somatosensory error map activity
$Au(t), S(t)$	Auditory and somatosensory state activity
$z_{AuM}, z_{SM}$	Synaptic weights that transform auditory and somatosensory error into corrective motor velocities for a speech sound
$z_{PM}$	Synaptic weights encoding feedforward commands for a speech sound
$z_{PAu}, z_{PS}$	Synaptic weights encoding auditory and somatosensory expectation for a speech sound
$P(t) = \begin{cases} 1 & \text{If sound is being produced or perceived} \\ 0 & \text{otherwise} \end{cases}$	Speech sound map activity
$g(t)$	Go signal

$\alpha_{ff}$  and  $\alpha_{fb}$  were varied systematically in the simulation series. Further details concerning the parameters and equations can be found in Guenther et al. (2006).

### Computation of coarticulation, speech sound distortion, searching articulatory behavior, and variability indices

Legend	
Parameter/function	Description
$CA$	Coarticulation index
$SSD$	Speech sound distortion index
$SAB$	Searching articulatory behavior index
$VAR$	Variability index
$V_{1j}$	Vowel 1
$j = \{/a/, /i/, /u/\}$	
$C_k$	Consonant
$k = \{/b/, /d/, /g/\}$	
$V_{2l}$	Vowel 2
$l = \{/a/, /i/, /u/\}$	
$S = \{V_{1j}, C_k, V_{2l}\}$	$S$ = list of all speech sounds
$W_{i,k,l} = V_{1j}C_kV_{2l}$	$W$ = list of all possible words (e.g. $W_{1,1,1}$ is /aba/; $W_{2,k,l}$ are all words beginning with /i/.)

Legend	
Parameter/function	Description
$F_{i,m}(S \langle W_{j,k,l} \rangle)$ $i = \{F1, F2, F3\}$ $m = \{beginning, middle, end\}$ $T = Target(F)$	Formant $i$ at measurement point $m$ of speech sound $S$ in the context of word $W$ . (Note that the formant values measured in Hz were normalized using a $\log_{10}(x)$ -transformation.)  Target value of formant $F$
$StDev = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$	Standard deviation of set $x$ with number of elements $n$

Calculations
<p>Mean formant frequency</p> $V_1:F_i(V_{1j} \langle W_{j,k,l} \rangle) = 1 / 3 \sum_{m=1}^3 F_{i,m}(V_{1j} \langle W_{j,k,l} \rangle)$ $C:F_i(C_k \langle W_{j,k,l} \rangle) = 1 / 3 \sum_{m=1}^3 F_{i,m}(C_k \langle W_{j,k,l} \rangle)$ $V_2:F_i(V_{2l} \langle W_{j,k,l} \rangle) = 1 / 3 \sum_{m=1}^3 F_{i,m}(V_{2l} \langle W_{j,k,l} \rangle)$
<p>Anticipatory coarticulation</p> $V_1:CA(V_{1j} \langle W_{j,k} \rangle) =$ $1 / 3 \sum_{i=1}^3 1 / 3 \sum_{l,l'=1}^3 (F_i(V_{1j} \langle W_{j,k,l} \rangle) - F_i(V_{1j} \langle W_{j,k,l'} \rangle)) \text{ with } l \neq l'$ $C:CA(C_k \langle W_{j,k} \rangle) =$ $1 / 3 \sum_{i=1}^3 1 / 3 \sum_{l,l'=1}^3 (F_i(C_k \langle W_{j,k,l} \rangle) - F_i(C_k \langle W_{j,k,l'} \rangle)) \text{ with } l \neq l'$
<p>Carry-over coarticulation</p> $C:CA(C_k \langle W_{k,l} \rangle) =$ $1 / 3 \sum_{i=1}^3 1 / 3 \sum_{j,j'=1}^3 (F_i(C_k \langle W_{j,k,l} \rangle) - F_i(C_k \langle W_{j',k,l} \rangle)) \text{ with } j \neq j'$ $V_2:CA(V_{2l} \langle W_{k,l} \rangle) =$ $1 / 3 \sum_{i=1}^3 1 / 3 \sum_{j,j'=1}^3 (F_i(V_{2l} \langle W_{j,k,l} \rangle) - F_i(V_{2l} \langle W_{j',k,l} \rangle)) \text{ with } j \neq j'$



**Calculations**

Speech sound distortion

$$V_1:SSD(V_{1j} \langle W_{j,k,l} \rangle) = \frac{1}{3} \sum_{i=1}^3 |F_i(V_{1j} \langle W_{j,k,l} \rangle) - T(F_i(V_{1j} \langle W_{j,k,l} \rangle))|$$

$$C:SSD(C_k \langle W_{j,k,l} \rangle) = \frac{1}{3} \sum_{i=1}^3 |F_i(C_k \langle W_{j,k,l} \rangle) - T(F_i(C_k \langle W_{j,k,l} \rangle))|$$

$$V_2:SSD(V_{2l} \langle W_{j,k,l} \rangle) = \frac{1}{3} \sum_{i=1}^3 |F_i(V_{2l} \langle W_{j,k,l} \rangle) - T(F_i(V_{2l} \langle W_{j,k,l} \rangle))|$$

Searching articulatory behavior

$$V_1:SAB(V_{1j} \langle W_{j,k,l} \rangle) = \frac{1}{3} \sum_{i=1}^3 StDev(\{F_{i,1}(V_{1j} \langle W_{j,k,l} \rangle), F_{i,2}(V_{1j} \langle W_{j,k,l} \rangle), F_{i,3}(V_{1j} \langle W_{j,k,l} \rangle)\})$$

$$C:SAB(C_k \langle W_{j,k,l} \rangle) = \frac{1}{3} \sum_{i=1}^3 StDev(\{F_{i,1}(C_k \langle W_{j,k,l} \rangle), F_{i,2}(C_k \langle W_{j,k,l} \rangle), F_{i,3}(C_k \langle W_{j,k,l} \rangle)\})$$

$$V_2:SAB(V_{2l} \langle W_{j,k,l} \rangle) = \frac{1}{3} \sum_{i=1}^3 StDev(\{F_{i,1}(V_{2l} \langle W_{j,k,l} \rangle), F_{i,2}(V_{2l} \langle W_{j,k,l} \rangle), F_{i,3}(V_{2l} \langle W_{j,k,l} \rangle)\})$$

Variability

$$V_1:VAR(V_{1j}) = \frac{1}{3} \sum_{i=1}^3 StDev(\{F_i(V_{1j} \langle W_{j,k,l} \rangle)\}) \text{ with } k, l=1, \dots, 3$$

$$C:VAR(C_k) = \frac{1}{3} \sum_{i=1}^3 StDev(\{F_i(C_k \langle W_{j,k,l} \rangle)\}) \text{ with } j, l=1, \dots, 3$$

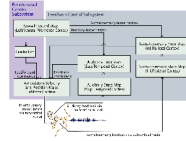
$$V_2:VAR(V_{2l}) = \frac{1}{3} \sum_{i=1}^3 StDev(\{F_i(V_{2l} \langle W_{j,k,l} \rangle)\}) \text{ with } j, k=1, \dots, 3$$

## References

- ASHA. Childhood Apraxia of Speech [technical report]. American Speech-Language-Hearing Association; 2007. Available from [www.asha.org/policy](http://www.asha.org/policy)
- Caruso, A.; Strand, E. Clinical management of motor speech disorders in children. New York: Thieme; 1999.
- Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol* 1954;47(6):381–391. [PubMed: 13174710]
- Gracco VL. Some Organizational Characteristics of Speech Movement Control. *Journal of Speech and Hearing Research* 1994;37(1):4. [PubMed: 8170129]
- Grigos MI, Saxman JH, Gordon AM. Speech motor development during acquisition of the voicing contrast. *J Speech Lang Hear Res* 2005;48(4):739–752. [PubMed: 16378470]
- Groenen P, Maassen B, Crul T, Thoonen G. The specific relation between perception and production errors for place of articulation in developmental apraxia of speech. *J Speech Hear Res* 1996;39(3):468–482. [PubMed: 8783127]
- Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. *Biol Cybern* 1994;72(1):43–53. [PubMed: 7880914]
- Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 2006;96(3):280–301. [PubMed: 16040108]
- Hall, P.; Jordan, L.; Robin, D., editors. *Developmental Apraxia of Speech: Theory and Clinical Practice*. 2nd. Austin, TX: Pro-ed; 2007.
- Harris CM, Wolpert DM. Signal-dependent noise determines motor planning. *Nature* 1998;394(6695):780–784. [PubMed: 9723616]
- Hayden D. Differential diagnosis of motor speech dysfunction in children. *Clinics in Communication Disorders* 1994;4:119–141. [PubMed: 7920576]
- Hertrich I, Ackermann H. Coarticulation in slow speech: durational and spectral analysis. *Lang Speech* 1995;38(Pt 2):159–187. [PubMed: 8867759]
- Hertrich I, Ackermann H. Temporal and spectral aspects of coarticulation in ataxic dysarthria: an acoustic analysis. *J Speech Lang Hear Res* 1999;42(2):367–381. [PubMed: 10229453]
- Ingham R, Moglia R, Frank P, Ingham J, Cordes A. Experimental investigation of the effects of frequency altered auditory feedback on the speech of adults who stutter. *Journal of Speech, Language, and Hearing Research* 1997;40:361–372.
- Lane H, Perkell JS. Control of voice-onset time in the absence of hearing: a review. *J Speech Lang Hear Res* 2005;48(6):1334–1343. [PubMed: 16478375]
- Levelt, WJM. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press; 1989.
- Lindblom B, Lubker J, Gay T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* 1979;7:147–161.
- Maassen B. Issues contrasting adult acquired versus developmental apraxia of speech. *Semin Speech Lang* 2002;23(4):257–266. [PubMed: 12461725]
- Maassen B, Groenen P, Crul T. Auditory and phonetic perception of vowels in children with apraxic speech disorders. *Clinical Linguistics & Phonetics* 2003;17(6):447–467. [PubMed: 14564831]
- Maassen B, Nijland L, Van der Meulen S. Coarticulation within and between syllables by children with developmental apraxia of speech. *Clinical Linguistics & Phonetics* 2001;15(1-2):145–150.
- Maeda, S. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In: Hardcastle, W.; Marchal, A., editors. *Speech production and speech modeling*. Boston: Kluwer Academic Publishers; 1990.
- Max L, Guenther F, Gracco V, Ghosh S, Wallace M. Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders* 2004;31:105–122.
- Meyer DE, Abrams RA, Kornblum S, Wright CE, Smith JE. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychol Rev* 1988;95(3):340–370. [PubMed: 3406245]

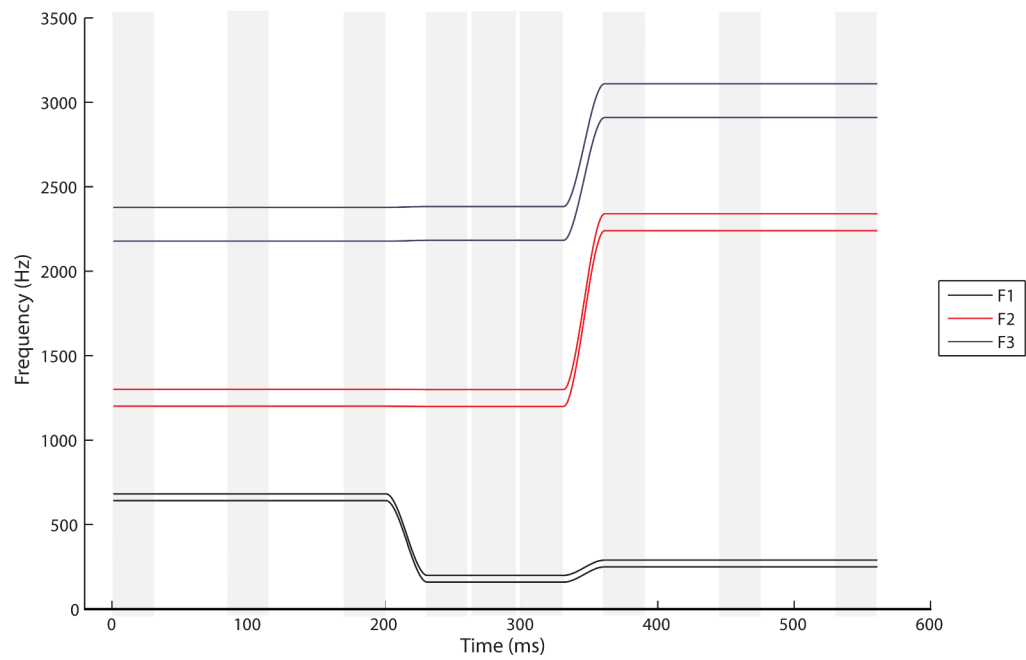
- Munhall K, Löfqvist A, Kelso J. Lip-larynx coordination in speech: Effects of mechanical perturbations of the lower lip. *The Journal of the Acoustical Society of America* 1994;95:3605–3616. [PubMed: 8046150]
- Nasir SM, Ostry DJ. Somatosensory precision in speech production. *Curr Biol* 2006;16(19):1918–1923. [PubMed: 17027488]
- Nijland, L. Developmental apraxia of speech: deficits in phonetic planning and motor programming. Radboud University Nijmegen Medical Centre; Nijmegen, the Netherlands: 2003. Unpublished doctoral dissertation
- Nijland L, Maassen B, Van der Meulen S. Evidence of motor programming deficits in children diagnosed with DAS. *J Speech Lang Hear Res* 2003;46(2):437–450. [PubMed: 14700384]
- Nijland L, Maassen B, Van der Meulen S, Gabreels F, Kraaimaat FW, Schreuder R. Coarticulation patterns in children with developmental apraxia of speech. *Clin Linguist Phon* 2002;16(6):461–483. [PubMed: 12469451]
- Nijland L, Maassen B, Van der Meulen S, Gabreels F, Kraaimaat FW, Schreuder R. Planning of syllables in children with developmental apraxia of speech. *Clin Linguist Phon* 2003;17(1):1–24. [PubMed: 12737052]
- Ostry, D.; Clark, H. Somatosensory contributions to speech motor control: Science and clinical application. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association; 2005.
- Ozanne, A. Childhood apraxia of speech. In: Dodd, B., editor. *Differential Diagnosis and Treatment of Children with Speech Disorder*. Second. London: Whurr; 2005.
- Perkell J, Nelson W. Variability in production of the vowels /i/ and /a/. *J Acoust Soc Am* 1985;77(5):1889–1895. [PubMed: 3998298]
- Peter B, Stoel-Gammon C. Timing errors in two children with suspected childhood apraxia of speech (scas) during speech and music-related tasks. *Clinical Linguistics & Phonetics* 2005;19:67–87. [PubMed: 15704499]
- Pollock K, Hall P. An analysis of the vowel misarticulations of five children with developmental apraxia of speech. *Clinical Linguistics & Phonetics* 1991;5:207–224.
- Postma A, Kolk H. Error monitoring in people who stutter: Evidence against auditory feedback defect theories. *Journal of Speech and Hearing Research* 1992;35:1024–1032. [PubMed: 1447913]
- Schmidt, R.; Lee, T. *Motor Control and Learning: A Behavioral Emphasis*. Champaign, IL: Human Kinetics; 1999.
- Shriberg LD, Aram DM, Kwiatkowski J. Developmental apraxia of speech: I. Descriptive and theoretical perspectives. *J Speech Lang Hear Res* 1997;40(2):273–285. [PubMed: 9130199]
- Smith, B.; Marquardt, T.; Cannito, M.; Davis, B. Vowel variability in developmental apraxia of speech. In: Till, JA.; Yorkston, KM.; Beukelman, DR., editors. *Motor speech disorders: Advances in assessment and treatment*. Baltimore: Paul H. Brookes; 1994. p. 81-89.
- Stackhouse J. Developmental verbal dyspraxia. I: A review and critique. *Eur J Disord Commun* 1992a;27(1):19–34. [PubMed: 1446093]
- Stackhouse, J. Developmental verbal dyspraxia: A longitudinal case study. In: Campbell, R., editor. *Mental lives: Case studies in cognition*. Oxford, United Kingdom: Blackwell; 1992b. p. 84-98.
- Sussman H, Marquardt T, Doyle J. An acoustic analysis of phonemic integrity and contrastiveness in developmental apraxia of speech. *Journal of Medical Speech-Language Pathology* 2000;8:301–313.
- Thoonen G, Maassen B, Gabreels F, Schreuder R. Feature analysis of singleton consonant errors in developmental verbal dyspraxia (DVD). *J Speech Hear Res* 1994;37(6):1424–1440. [PubMed: 7533219]
- Tremblay S, Shiller DM, Ostry DJ. Somatosensory basis of speech production. *Nature* 2003;423(6942):866–869. [PubMed: 12815431]
- Van der Merwe, A. A theoretical framework for the characterization of pathological speech sensorimotor control. In: McNeil, MR., editor. *Clinical management of sensorimotor speech disorders*. New York: Thieme Medical Publishers Inc; 1997. p. 1-25.
- Walton J, Pollock K. Acoustic validation of vowel error patterns in developmental apraxia of speech. *Clinical Linguistics & Phonetics* 1993;7:95–111.

Wolpert DM, Ghahramani Z, Flanagan JR. Perspectives and problems in motor learning. *Trends Cogn Sci* 2001;5(11):487–494. [PubMed: 11684481]

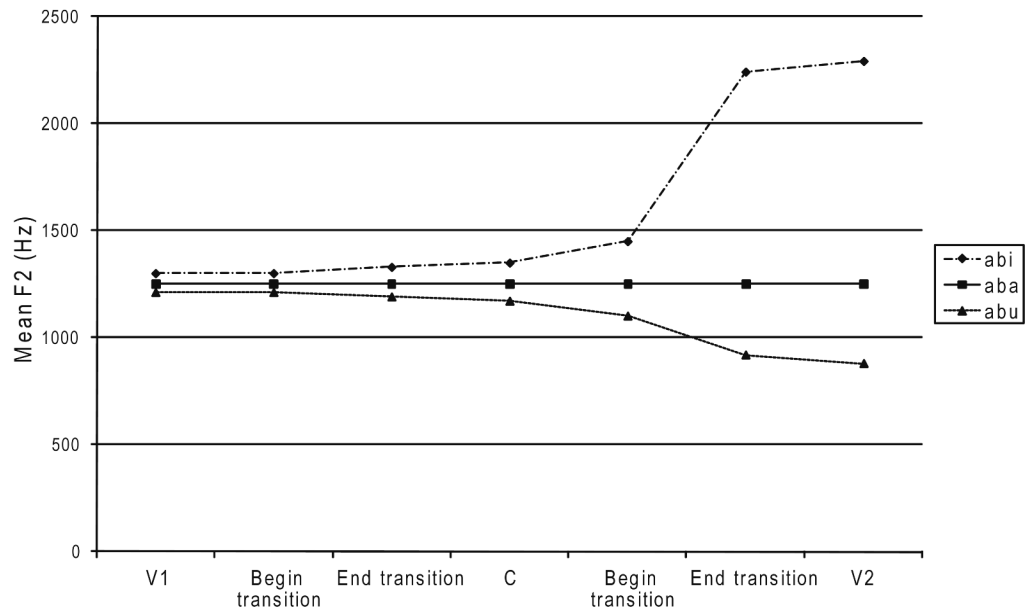


**Fig. 1.** Schematic representation of the Directions Into Velocities of Articulators (DIVA) model of speech motor control. Projections to and from the cerebellum are simplified for clarity (Guenther et al., 2006). Sup. = Superior; Inf. = Inferior.

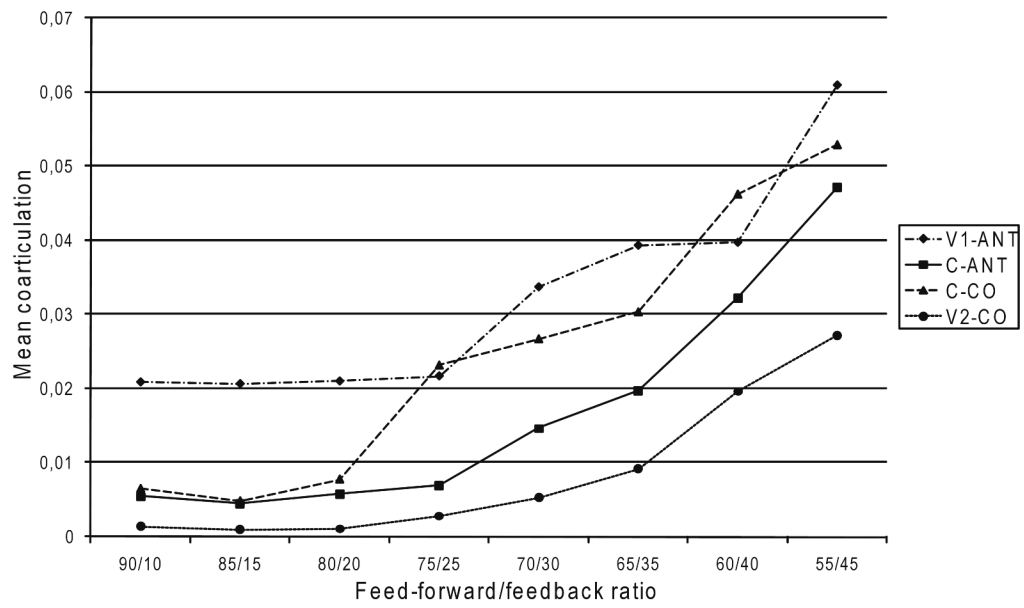




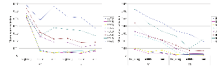
**Fig. 2.** Schematic representation of the speech sound target for /abi/. The light gray columns indicate the measurement points. At each of these points, the mean formant value was calculated over a 30-ms time window (three measurements with 10-ms time intervals).



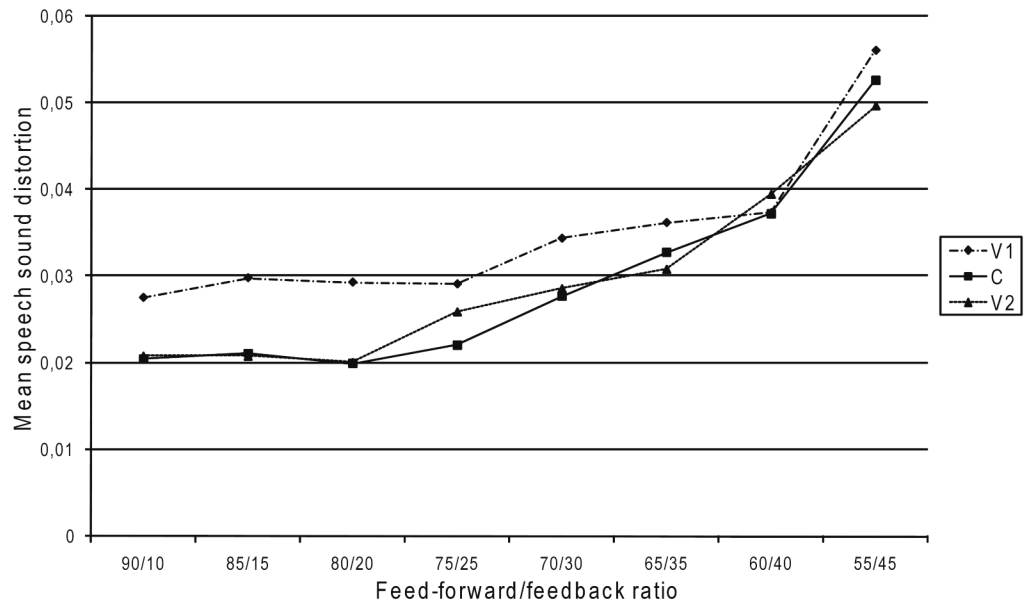
**Fig. 3.** Example of coarticulation: The F2 values of V1 and C differ depending on V2. V = vowel; C = consonant.



**Fig. 4.** Anticipatory (ANT) and carryover (CO) coarticulation for V1, C, and V2 in relation to feed-forward/feedback ratio.

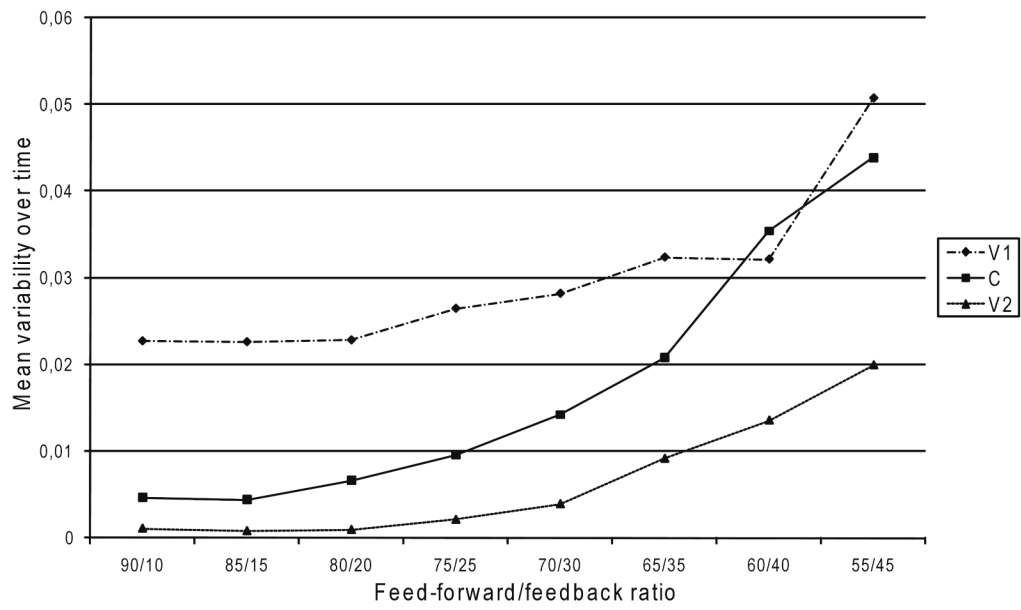


**Fig. 5.** Anticipatory (left panel) and carryover (right panel) coarticulation for different feed-forward/feedback ratios.

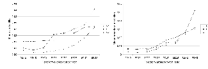


**Fig. 6.** Speech sound distortion in relation to feed-forward/feedback ratio.





**Fig. 7.** Groping: variability over the course of the production of speech sounds in relation to feed-forward/feedback ratio.



**Fig. 8.** Token-to-token variability: mean variability of mean formant frequencies (left panel) and mean standard deviation of the coarticulation (COART), speech sound distortion (SSD), and groping or searching articulatory behavior (G/SAB) indices (right panel) between productions of speech sounds.