Auditory-Motor Interactions in Pediatric Motor Speech Disorders: Neurocomputational Modeling of Disordered Development

H. Terband[a,b,c], **B. Maassen**[b,c], **F.H. Guenther**[d,e], and **J. Brumberg**[d]

H. Terband: h.r.terband@uu.nl; B. Maassen: b.a.m.maassen@rug.nl; F.H. Guenther: fguenth@gmail.com; J. Brumberg: jbrumberg@gmail.com

[a]Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK, Utrecht, the Netherlands [b]Centre for Language and Cognition, Faculty of Arts, University of Groningen, Oude Kijk in 't Jatstraat 26, 9712 EB, Groningen, The Netherlands [c]Medical Psychology/Pediatric Neurology Centre/ENT, Radboud University Nijmegen Medical Centre, Geert Grooteplein-Zuid 10, 6525 GA, Nijmegen, the Netherlands [d]Department of Speech, Language, & Hearing Sciences and Biomedical Engineering, Boston University, 677 Beacon Street, Boston, MA 02215, USA [e]Division of Health Sciences and Technology, Harvard University/Massachusetts Institute of Technology, 77 Massachusetts Avenue, E25-519, Cambridge, MA 02139, USA

Abstract

Background/Purpose—Differentiating the symptom complex due to phonological-level disorders, speech delay and pediatric motor speech disorders is a controversial issue in the field of pediatric speech and language pathology. The present study investigated the developmental interaction between neurological deficits in auditory and motor processes using computational modeling with the DIVA model.

Method—In a series of computer simulations, we investigated the effect of a motor processing deficit alone (MPD), and the effect of a motor processing deficit in combination with an auditory processing deficit (MPD+APD) on the trajectory and endpoint of speech motor development in the DIVA model.

Results—Simulation results showed that a motor programming deficit predominantly leads to deterioration on the phonological level (phonemic mappings) when auditory self-monitoring is intact, and on the systemic level (systemic mapping) if auditory self-monitoring is impaired.

Conclusions—These findings suggest a close relation between quality of auditory self-monitoring and the involvement of phonological vs. motor processes in children with pediatric motor speech disorders. It is suggested that MPD+APD might be involved in typically apraxic speech output disorders and MPD in pediatric motor speech disorders that also have a phonological component. Possibilities to verify these hypotheses using empirical data collected from human subjects are discussed.

Corresponding author: Hayo Terband, Utrecht Institute of Linguistics OTS, Trans 10, kmr. 0.25, Utrecht, 3512 JK, The Netherlands, Phone: +31(0)633821241, h.r.terband@uu.nl.

1. Introduction

Although infants possess an inborn capacity to acquire verbal communication, not all children go through the learning process successfully. In around 6% of all children, the development of speech production is impaired (Broomfield & Dodd, 2004; Law, Boyle, Harris, Harkness, & Nye, 2000; McKinnon, McLeod, & Reilly, 2007) and a recent account (Priester, Post, & Goorhuis-Brouwer, 2009) indicated that 90% of the children that are in treatment at speech- and language pathologists suffer from such a so-called speech sound disorder (SSD).

The classification and differentiation of speech sound disorders (SSD) in children constitutes one of the main questions in the field of pediatric speech- and language pathology. Especially the differentiation between Phonological Disorder (PD), speech delay and pediatric motor speech disorders (MSD), and among MSD the classification of subtype childhood apraxia of speech (CAS) are controversial (Guyette & Diedrich, 1981; Shriberg et al., 2010). Attempts to find single diagnostic markers for differential diagnosis are hindered by a high overlap in symptoms. Different underlying disorders may show similar symptoms, as many examples have shown (Maassen, Nijland, & Terband, 2010; Weismer & Kim, 2010). Additionally, in pediatric disorders development itself plays a role. Symptoms change along with the different stages of development and, importantly, the different cognitive processes influence each other during development. Due to these developmental interactions, a specific underlying deficit can produce symptoms on (apparently) different levels or domains. For example, phoneme substitutions, which are generally considered phonological symptoms, can follow from acoustic-perceptual and/or articulatory-motor impairment (Edwards, Fourakis, Beckman, & Fox, 1999).

The interaction between perception and production has been a well explored topic of research. Several studies have shown that the articulatory proficiency of speakers in producing a phonological contrast between two speech sounds is related to their perceptual ability to discriminate the same contrast (Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). The 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; Warner-Czyz & Davis, 2008; Warner-Czyz, Davis, & MacNeilage, 2010). A number of findings suggest that there is also a developmental influence of production on perception and that poor articulation affects the perceptual acuity for phonological contrasts. Raaymakers and Crul (1988) investigated the perception of a /-s/-/-ts/ continuum in a group of children who misarticulated the final consonant cluster /-ts/ compared to a group of children who misarticulated other phonemes than the specific /-ts/ cluster, as well as to a group of normally developing children, and adults. The results showed more variability in the perception of the /-s/-/-ts/ contrast specifically in the children with poorer articulation proficiency on the /-ts/ cluster (Raaymakers & Crul, 1988). Furthermore, several studies have found subtle (subclinical) auditory processing deficits in children with CAS. Auditory discrimination of consonants was found to be poorer in children with CAS compared to normally developing controls (Groenen, Maassen, Crul, & Thoonen, 1996), and this finding has been replicated for vowel continua (Maassen, Groenen, & Crul, 2003). Furthermore, the study of Groenen and colleagues showed a specific relation between the perception and production of place-of-articulation in children with CAS. 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 correlated to the frequency of place-of-articulation substitutions in production (Groenen, et al., 1996).

Edwards, Fox, and Rogers (2002) found a similar relationship between perceptual acuity and speech production measures investigating preschool-age children diagnosed with PD, three age groups of typically developing children and adults. In this study, the ability to discriminate CVC words that differed only in the identity of the final consonant was found to be related to 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, the ability to attend to fine phonetic detail, and the acquisition of articulatory-acoustic and acoustic-auditory representations (Edwards, et al., 2002). Nijland (2009) investigated the relation between perception and production on different cognitive levels in a variety of children with SSD as compared to a group of age-matched normally developing controls. The children with speech output disorders were divided into three groups: CAS, PD and a mixed group consisting of children that exhibited characteristics of both disorders and could not be classified. The results provided an interesting pattern: higher-order production disorders (PD) seemed to be linked to higher-order perception problems (e.g., word rhyming), whereas lower-order production disorders (CAS) seemed to be linked to both lower-order perception problems (e.g., non-word discrimination) and higher-order perception problems (Nijland, 2009).

In summary, the studies reported in the literature indicate that there is a close relation between production symptoms and perceptual acuity in children with SSD. The reported literature indicates that lower order problems (such as MSD or hearing impairment) are correlated to both lower- and higher-order difficulties in either direction, from production to perception and from perception to production, but the mechanisms and direction of causality behind this relation are poorly understood. One aspect of perception that might play a role is the auditory acuity in self-monitoring. The work of Perkell (Guenther & Perkell, 2004; Lane & Perkell, 2005; Perkell, Guenther, et al., 2004; Perkell et al., 1997; Perkell, Matthies, et al., 2004) stresses the importance of auditory self-monitoring both in acquiring sensory goals for producing speech sounds and in the coordination of the timing and magnitude of the movements in the different systems involved (i.e., the muscular structures controlling respiration, resonance, phonation, and the articulators). The aim of the present study was to investigate the developmental interaction between auditory self-monitoring or auditory feedback control and a motor processing deficit in a series of computer simulations with a computational neural model of speech motor control called DIVA (Directions Into Velocities of Articulators; Guenther, 1994; Guenther, Ghosh, & Tourville, 2006). Neurocomputational models like DIVA can provide a valuable reference frame for studying the sensorimotor information processing involved in speech production during development. Being computationally implemented, DIVA allows for computer simulations of the speech acquisition and production processes with the possibility of manipulating parameters independently and systematically. Previously, we have employed this to investigate the possible consequences of increased reliance on the feedback control (Terband, Maassen, Guenther, & Brumberg, 2009) and reduction of speech rate (Terband & Maassen, 2010) during imitation learning. The model is linked to an articulatory speech synthesizer, which enables the computer simulations to generate articulatory and acoustic data that can be compared to real speech data.

The present study focused on the influence of the quality of auditory self-monitoring on the trajectory and endpoint of the speech motor acquisition process in the context of a motor processing deficit that (presumably) underlies particular symptoms that are frequently found in pediatric MSD. For this, two stages of speech acquisition are distinguished: a babbling phase of open loop learning and a phase of imitation learning. The babbling phase is a perceptual-motor stage during which the infant acquires control over the articulo-motoric system by establishing an internal model (Jordan, 1990; Wolpert, Ghahramani, & Jordan, 1995), i.e., a systemic mapping between the movements and their auditory and

somatosensory consequences. During the phase of imitation learning, the acquired perceptual-motor skills are exploited to produce the first meaningful utterances. This second phase is a phonological stage, resulting in a set of phonemic mappings between articulatory movements and specific goals. Before describing the model manipulations, we will first give a short overview of the acquisition of speech motor control in DIVA.

2. 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 (Fig. 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 of the DIVA model motor parameters (Maeda, 1990).

The model comprises three stages in development. In the first stage, semi-random[1] articulatory movements (resembling the babbling phase) are used to learn a *systemic* mapping between motor commands and their auditory and somatosensory consequences. In DIVA, this systemic mapping is implemented as an adaptive hyperplane radial basis function (HRBF) network, which is a neural network-like mathematical construct that is composed of a set of nodes, each corresponding to a different position of the input space (Figure 2). In the training process, the model iteratively works towards an optimal mapping between input articulatory movements and their auditory outputs[2].

Once the systemic mapping is acquired during the babbling phase (the system 'knows' what movements to make in order to produce an auditory target, and can predict what will be the auditory outcome of a certain motor configuration)[3], it is used in the next developmental phase for the purpose of imitation. In this second phase, the model learns a set of *phonemic* mappings between articulatory movements and language-specific goals, which have the form of auditory target regions (Guenther, Hampson, & Johnson, 1998). The targets can consist of a single phoneme, a syllable, a word, or a small phrase. The feedforward commands are acquired by the model iteratively attempting to produce sounds that fall into the auditory target region. The auditory feedback control subsystem compares the actual auditory signal to the desired auditory target and the feedforward command is updated with each attempt, until it is accurate enough for the model to produce the sound without generating auditory errors (for an example of the iterative process see Figure 3). When the feedforward command is accurate enough for the model to produce the sound without generating auditory errors, the model then enters the third phase, i.e. the stage of normal (mature) production.

Recent behavioral experiments have shown that somatosensory targets constitute a goal for speech movements that is separate from the auditory targets (Tremblay, Shiller, & Ostry, 2003). Correspondingly DIVA also features a somatosensory reference frame, which has a similar function, except for the auditory reference frame being the teaching signal (Guenther, et al., 1998). The somatosensory targets encode the tactile (place and pressure of tongue-palate and lip contact) and proprioceptive (muscle length) information that

---

[1]Limited by biomechanical constraints.

[2] In the training process, the network is given a set of input articulatory movements together with their true/ideal auditory outputs. Training involves iteratively adjusting the weights of the network nodes as to minimize the error between the true/ideal auditory output for the given input articulatory movement and model's predicted acoustic output based on the current mapping iteration.

[3] The technical terms of these two faces of the systemic mapping are inverse mapping and forward mapping respectively. Note that in the computational implementation of DIVA, the forward mapping is trained, from which the inverse mapping is derived by calculating the Jacobian pseudo-inverse.

accompanies the correct production of the speech sound and is used in the somatosensory feedback control subsystem during normal production. Thus, in DIVA the reference frames for speech production are multimodal with an underlying hierarchy among modalities. Similar views have also been disseminated elsewhere (Perrier, 2005). The somatosensory target for each speech sound is learned during acquisition of the feedforward command. This specificity of somatosensory target information has recently been confirmed in behavioral experiments (Tremblay, Houle, & Ostry, 2008).

3. Method and materials

3.1 Implementation

In a series of computer simulations, we investigated the effect of a motor processing deficit alone (MPD), and the effect of a motor processing deficit in combination with an auditory processing deficit (MPD+APD) on the trajectory and endpoint of speech motor development 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. The motor processing deficit was implemented in the model by adding noise to the motor-, and somatosensory state representations. (In DIVA, the motor state and the somatosensory state are closely coupled and in the feedback control subsystem, the motor command derives from somatosensory state information.) The auditory processing deficit was implemented by adding noise to auditory state representations. For both MPD and MPD+APD, the noise was uncorrelated zero-mean noise while the standard deviation was systematically varied from 5% to 25% (of the range of activations per model cell) in 5% steps, thus creating a total of 2×5 experimental conditions. 0% noise, corresponding to no deficits implemented, functioned as a baseline condition. The simulation series comprised two stages, which are elaborated in more detail below. In order to differentiate the effect of noise on the acquisition process from the effect of noise during production attempts, the effects that the deficits have on speech production when implemented after asymptotic learning (i.e. the developmental trajectory was fully completed without any deficits implemented) were investigated as control conditions.

3.2 Babbling stage

In the first stage of the simulations (babbling stage), we trained a HRBF network (systemic mapping) for each condition (2×5 experimental conditions plus the baseline condition). The same set of 4000 input patterns was used in all cases, which in the baseline condition (no deficits implemented) corresponds to asymptotic learning. To assess the developmental trajectory and the effect of underlearning, we separately trained and evaluated HRBF networks in all conditions using 500, 1000, 1500, 2000, and 3000 training items. The quality of the systemic mappings was evaluated by means of the sum-squared Euclidean error of the trained network on a new set of 4000 input patterns (that was similar to, but different from, the set of input patterns used for training).

3.3 Imitation learning stage

In the second stage (imitation learning), a series of simulations we systematically varied system noise levels during production attempts during the acquisition of feedforward commands (phonemic mappings). For each of the conditions, we used the systemic mapping generated in the corresponding babbling simulations (i.e. the babbling simulations with the same deficit and noise level). The methodology of the imitation learning simulations was largely the same as in Terband et al. (2009). [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. The sound targets consisted of time

varying regions in F1; F2; F3-space specifying the trajectories of the first three formant frequencies with a fixed duration of 560 ms (consisting of 200 ms for the vowels, 100 ms for the consonant, and 30 ms for each of the transitions). The same auditory targets were used in both conditions. The acquisition of the motor commands for each sound target was repeated 6 times. The simulations produced acoustic realizations of each target VCV utterance which 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 during production alone, the effects that both deficits have on speech production after asymptotic learning (i.e., learning was completed) were investigated as control conditions. In this case, the deficits were introduced only after training the model (i.e., after the developmental trajectory was fully completed at the baseline condition and the feedforward commands had been acquired without any deficits implemented; all noise levels at 0%) during production attempts in the performance stage. The same set of 27 sound targets was used, and each sound target was repeated 6 times.

3.3.1 Perceptual evaluation—Poor intelligibility is a core characteristic of any speech disorder and in pediatric MSD, speech intelligibility is thought to be closely related to the severity of the disorder (e.g., Hall, Jordan & Robin, 2007). A perceptual scaling experiment was used to assess the intelligibility and the speech quality of the model's speech output across the varied noise levels for both MPD and MPD+APD. 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 intelligibility and quality of the synthesizer output across the varied levels of severity.

A total of 14 subjects (age 18–21 years) 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 (MPD and MPD+APD) × 5 (severity levels) × 6 (repetitions of each target word) 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. intelligibility (How clearly is the second utterance recognizable 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 smoothness and naturalness?). 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 intelligibility and quality judgments, the mean scores, per subject, of the 6 repetitions of each target word were calculated. Subsequently,

statistical testing was accomplished using Linear Mixed Model (a form of general linear model that does not assume homogeneity of variance, sphericity, or compound symmetry, and, moreover, allows for missing data, Max & Onghena, 1999; Quené & van den Bergh, 2004) analyses using a 0.05 level of significance. First, main effects of severity level were investigated for MPD and MPD+APD in separate analyses featuring subject as correlated term, and target utterance and severity level as correlated residuals, Additionally, MPD and MPD+APD were compared in a Linear Mixed Model analysis featuring subject as correlated term, and target utterance, severity level and condition as correlated residuals.

3.3.2 Acoustical evaluation—The acoustical evaluation conformed to previous DIVA simulation studies (Terband & Maassen, 2010; Terband, et al., 2009) and focused on four common phonetic characteristics of MSD: deviant coarticulation, speech sound distortion, searching articulation, and increased token-to-token variability (ASHA, 2007; Crary, 1993; Dodd, 2005; Hall, Jordan, & Robin, 2007; Maassen, et al., 2010). Because in the present computational implementation of the model the speech sound targets are predefined, deviant suprasegmental characteristics (e.g., inappropriate or lack of prosody) or phonological characteristics (e.g., speech sound omissions, substitutions and transpositions) could not be investigated. Additionally, slow speech rate and speech motor characteristics such as prevocalic groping and speech sound prolongations could not be investigated for the same reason. Therefore, we focused on these four more fine-grained phonetic characteristics.

For each speech sound of each VCV utterance, the formant frequencies (F1, F2, F3) were measured at three points in time: at the beginning, in the middle, and at the end of the steady-state area (see Figure 4). The formant values measured in Hz were normalized using a log10(x) transformation in order to focus on relative rather than absolute frequency differences. Deviant coarticulation was measured by the absolute differences in mean formant frequencies of a particular speech sound across all possible vowel contexts, speech sound distortion was measured by the absolute differences in mean formant frequencies of each produced speech sound relative to the frequencies of the target speech sound, online searching articulatory behavior was measured by the differences in formant frequencies at the beginning, middle, and end of each speech sound, and token-to-token variability was measured by the standard deviations in the mean formant frequencies of repeated productions of the same speech sound. A detailed description of the measures, accompanied by formulas, can be found in Terband et al. (2009).

4. Results and discussion

4.1 Babbling stage

The results of the evaluation of the babbling stage show a decrease in the quality of the systemic mappings (i.e., an increase in sum-squared Euclidean error) with increased severity for both MPD and MPD+APD (Figure 5, left panel), but the effect is much larger for MPD +APD. The quality of the systemic mapping at different stages in the training process shows the effects to be stable throughout (Table 1). Compared to the effect of underlearning in the baseline condition (Figure 5, right panel; 500–1000 training items), the results show a very drastic and consistent decrease in mapping quality with increased severity for MPD+APD. In the case of a MPD only, 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 error only starts to increase at noise levels of 15% and higher. At these higher noise levels, the error-increase appears to remain small, though it does correspond to substantial underlearning when compared to the baseline condition.

4.2 Imitation learning stage

4.2.1 Perceptual evaluation—The results of the perceptual evaluation of the model's speech output after imitation learning are presented in Figure 6. In comparison, the perceptual evaluation yields a more consistent pattern for MPD than for MPD+APD. In the case of MPD, results show a steady decrease of both intelligibility and speech quality with increased severity. For MPD+APD, 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 intelligibility. Statistical analyses revealed highly significant main effects of noise level/severity on the intelligibility and quality judgments for both deficits [all p values < .001].

Closer perceptual inspection of the individual utterances reveals that from 15% noise and higher, the model's speech output is highly deteriorated. This finding applies to both deficits, but manifests itself differently, which seems to be related to the quality of the systemic mapping. In the case of MPD, the quality of the speech output starts to decrease at noise levels of 15% and higher. In the case of MPD+APD, the quality of the systemic mapping shows a very strong and consistent decrease with increased noise level, and at a noise level of 15% the sum-squared Euclidean error is already twice as large as in the baseline condition. Under these circumstances, the systemic mapping becomes so unspecified that every utterance is strongly neutralized and reduces to schwa. Hence the inverse relation between speech quality and intelligible for noise levels of 15% and higher: the model's productions reduce to schwa, but these schwa's are produced well without much distortion or irregularities. In the case of MPD, results do not show this strong effect of underspecification of the systemic mapping and correspondingly no strong neutralization effect in the speech production after imitation learning. Under circumstances of a motor processing deficit only (MPD), systemic mapping quality decreases from noise levels of 15% and up, but only relatively mildly. Thus, the effect of increased severity is different from the effect regarding MPD+APD. For MPD, the increase in severity causes primarily a strong decrease in speech quality. When noise levels exceed 10%, the speech production features so much distortion and irregularities that it does not resemble speech anymore. For these reasons, we limited the further analysis of both implemented deficits to the first two levels of severity (noise levels of 5% and 10%).

With respect to the quantitative comparison between conditions, the statistical analyses revealed the intelligibility and quality judgments to be significantly higher for MPD as compared to MPD+APD [p > .001 and p < .05 respectively]. Statistical analyses also revealed a highly significant condition x noise level/severity interaction [both p values < .001], indicating a stronger decline in intelligibility and quality for MPD than for MPD +APD.

4.2.2 Acoustical evaluation—Figure 7 presents deviant 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 speech symptoms for both deficits, but the increase is much larger for MPD than for MPD+APD.

A series of Kruskal-Wallis tests for independent samples was conducted to test whether any of the effects was significant. Test statistics, presented in Table 2, show almost all main 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 online searching articulatory behavior in V1 (Table 2). Second, for MPD+APD speech sound distortion tends to be *smaller* in the noise conditions than in the baseline condition (noise level of 0), especially in C and V2. A third striking feature is that in the

case of MPD+APD the results show no systematic difference in the acoustical measures between consonant and vowels whereas in the case of MPD, all CAS symptoms are stronger in the consonant than in the vowels (Figure 7). As this is seen especially at 5% noise, we suspect that this could be related to the quality of the systemic mapping which was found for MPD to be higher (lower SSE) at 5% noise as compared to not only 10% but also 0% noise (babbling results; Figure 5). As a result, even the smallest differences in auditory space may evoke an articulatory compensation. The fact that this is seen only in the consonant could be due to the fact that zero-mean noise can lead to a non-zero systematic error due to cell activations in DIVA being limited between 0 and 1 and protractions in the space the cells encode. For example, if an activation of 1 represents a full bilabial closure, adding 5% has no effect as the activation cannot exceed 1. On the other hand, subtracting 5% can have a huge effect as an activation of 0.95 could mean the lips are not closed at all.

To differentiate the effects of the deficits during production attempts from the effects on the acquisition process, we compared the effect that both deficits have on speech production when present during the whole learning process versus when implemented after the developmental trajectory was fully completed without any deficits implemented (Figure 8). A series of Mann-Whitney U tests for 2 independent samples was conducted to test whether any of the during- versus after-learning differences was significant. Test statistics are presented in Table 3. The comparisons show deviant coarticulation to be much larger if the deficit is present during the whole learning process than if implemented after learning for MPD. For MPD+APD results show no differences in the amount of deviant coarticulation if the deficit is implemented during as compared to after learning. Interestingly, with respect to speech sound distortion, the results show an opposite pattern with few significant differences for MPD and many for MPD+APD (Table 3). Although for the consonant in MPD, the amount of speech sound distortion appears to be much larger if the deficit is present during learning as compared to when implemented after learning, these differences fail to reach significance. In MPD+APD, the amount of speech sound distortion tends to be larger if the deficit is implemented after learning was completed than when present during the whole acquisition process.

The results concerning online searching articulatory behavior show larger effects if the deficit was present during learning for MPD in the consonants, but results show no differences in the vowels. For MPD+APD, the during- versus after-learning comparison shows no differences in online searching articulatory behavior in general. Furthermore, where significant differences are found (namely for MPD+APD 5% in V1 and C; Figure 8B), the differences are 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 learning process. This effect is larger for MPD (Figure 8A).

Together, these results suggest that when the auditory feedback loop is intact, noise in the motor state representation affects the process of imitation learning to a larger extent, causing the learned motor commands to be inherently unstable or deviant. When both feedforward control and auditory self-monitoring are impaired on the other hand, the increase in anticipatory coarticulation, speech sound distortion and token-to-token variability could be attributed mainly to production/execution processes.

To explore this further, we examined the acquired feedforward commands by letting the DIVA model with no deficits implemented run the feedforward commands that were acquired with deficits implemented without online feedback control. 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 9. In the case of MPD, the auditory feedback loop is intact and, as was shown in the babbling simulations, the

systemic mapping (that is used in the auditory feedback loop to calculate online corrections) remains relatively intact. Subsequently during imitation learning, the model is sensitive to the articulatory errors introduced by the noise in the motor state representation and induces a corrective response. 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 imitation learning trajectory and its endpoint.

In the case of MPD+APD, on the other hand, auditory state representations are additionally affected by noise. However, in this case, the systemic mapping is of poor quality (see babbling imitations) which works as a facilitating mechanism, dampening online feedback corrections. Due to the underspecification of the systemic mapping, small differences in auditory space do not translate to differences in motor space; therefore, do not provoke an articulatory compensation. This mechanism expresses itself in lower values for the online searching articulatory behavior measure (see Figure 7). As a result, the influence of errors introduced by noise in motor state representations during the update of the motor command is decreased. Since the noise is zero-mean, in the long run errors introduced by noise in motor state representations tend to average out. Consequently, although these errors still express themselves in the model's synthesized speech output, the endpoint of imitation learning (the acquired phonemic mappings) is not inherently deviant from the baseline condition.

Finally, we compared the two control conditions in which the deficits were implemented after the developmental trajectory was fully completed without any deficits implemented. The results showed no differences between MPD and MPD+APD in their effect on the four acoustic measures when implemented after learning was fully completed (Figure 8; test statistics are presented in Table 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.

5. General discussion

5.1 Summary of findings

The developmental interaction between the quality of auditory self-monitoring or auditory feedback control and motor processing deficits was investigated in a series of computer simulations with the DIVA model. The involvement of two deficits (a motor processing deficit alone [MPD], and a motor processing deficit in combination with an auditory processing deficit [MPD+APD]) was systematically simulated during the model's acquisition processes of babbling and imitation learning. The effects of both deficits on speech production after normal asymptotic learning were investigated as control conditions. With respect to simulated speech output after the fully completed developmental trajectory, results did not differentiate qualitatively between the two deficits; both led to degraded speech output with similar characteristics on both perceptual and acoustical evaluation, i.e. a decrease in intelligibility and quality (perceptual measures) and an increase in deviant coarticulation, online searching articulatory behavior, and token-to-token variability (acoustic measures) with increased severity.

Nevertheless, quantitative comparison between conditions indicated higher scores combined with a stronger decline in intelligibility and quality judgments for MPD than for MPD +APD. Furthermore, the results provided clear differences between MPD and MPD+APD in the 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 systemic mapping quality between the two: MPD+APD caused a strong deterioration of the systemic mapping whereas for MPD, this representation developed normally (Figure 5).

Second, results after imitation learning showed large differences in the phonemic mappings between the two deficits: MPD caused the acquired motor commands to be inherently deviant or unstable, whereas for MPD+APD, the acquired motor commands were not inherently deviant from the baseline condition (Figures 8–9).

In summary, within the present simulations, MPD led predominantly to deterioration of what are called phonemic mappings in DIVA, whereas MPD+APD led predominantly to deterioration of the systemic mapping. These results can be interpreted as MPD mainly affecting the process of imitation learning while for MPD+APD the majority of symptoms should be attributed mainly to production/execution processes. In DIVA the acquired phonemic mappings are stored in synaptic projections emanating from neurons in the speech sound map. They 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 corresponds to the phonetic level, comprising language independent relations between motor commands and their auditory and somatosensory consequences in which articulatory parameters are automatically organized in task-specific groupings.

5.2 Speech sound disorders

The present results indicate that DIVA model's simulations of early speech development predict a close relation between perceptual acuity and production symptoms. Specifically, these findings would suggest that in children with speech sound disorders, the quality of the system for auditory self-monitoring is one of the determinants of the involvement of phonological vs. speech-motor processes. If the system for self-monitoring auditory error is intact (as in MPD), the impairment of the feedforward control subsystem leads to additional problems at the phonological level, while it leads to problems at the systemic level if the auditory feedback control subsystem is also impaired (as in MPD+APD).

The picture that thus emerges is one in which MPD+APD might be responsible for pediatric MSD that are classified as typically apraxic and MPD only for speech output problems that also have a phonological component. There is some behavioral evidence that corroborates this idea. First, an important feature of the model under MPD+APD is that the simulation results indicated the majority of symptoms could be mainly attributed to production processes (planning and execution) rather than stored motor-patterns. This corresponds to the interpretation of MPD+APD resulting in apraxia of speech, and might explain why the difference in number of errors between words and non-words is smaller in CAS compared to normally developing children, as if all utterances are new (Thoonen, Maassen, Gabreels, Schreuder, & de Swart, 1997).

Second, Munson, Edwards, and Beckman (2005) found that the speech-production problems in children with phonological symptoms are associated with difficulties forming robust representations of the acoustic/auditory and articulatory characteristics of speech sounds. This appears to be exactly what happened in DIVA in the current simulations under the MPD deficit, as this deficit caused the phonemic mappings between language-specific goals and articulatory movements to be inherently deviant.

In conclusion, the results of our current computer simulations suggest that the auditory perception impairments that have been found in children with CAS (Groenen, et al., 1996;

Maassen, et al., 2003; Marion, Sussman, & Marquardt, 1993; Nijland, 2009) are not derived or consequential, but that the deficit in auditory processing might play a fundamental role in the emergence of the disorder.

5.3 Predictions for auditory and articulatory perturbation experiments

It should be noted that these results are hypotheses obtained from a computer simulation and further verification is required using empirical data collected from human subjects. In this respect, the present findings lead to directly testable predictions for auditory and articulatory perturbation experiments. One 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 and presented to the speaker through headphones. This creates an apparent mismatch between the speech sound the speaker intended to produce and what he/ she hears back. If the system for auditory self-monitoring is intact, the speaker is able to generate a compensatory response in order to match the intended target. Healthy adults show compensation in their speech in the direction opposite to the induced shift within approximately 140 ms of onset (Tourville, Reilly, & Guenther, 2008). If the systemic mapping and the auditory feedback control subsystem are intact, this would allow a relatively normal compensatory response to the shifts in F1. On the other hand, if the systemic mapping and/or the auditory feedback control subsystem are impaired, this would prevent a consistent compensatory response.

Another possibility is to compare speech production in the condition of normal sensory feedback with a condition in which proprioceptive feedback is blocked by masseter tendon vibration, which is applied through a tool altering the activity of muscle spindles (comp. Loucks & De Nil, 2001; Namasivayam, van Lieshout, McIlroy, & De Nil, 2009). Since the estimation of the motor state that is necessary for the readout of the appropriate feedforward commands largely depends on proprioceptive information, perturbing this information through tendon vibration would impair the readout of feedforward commands, thus causing articulatory errors. If the systemic mapping is intact, the speech system can use auditory feedback to try to compensate for the introduction of errors. Experimental data of jaw-opening movements of normal speakers showed that tendon vibration typically causes a reduction in the amplitude and velocity, or movement undershoot (Loucks & De Nil, 2001). So although complete compensation is not to be expected in any speech motor system, in such an experiment, blocking proprioceptive information is predicted to affect speech output more negatively if the core impairment is in the systemic mapping, with the stored motor commands (relatively) intact, than if the stored motor commands are impaired, irrespective of the systemic mapping.

5.4 Limitations of the DIVA model

With respect to speech output after asymptotic learning, both deficits led to degraded speech output with similar characteristics. In this study, we focused on four more fine-grained phonetic characteristics mainly motivated by the fact that these symptoms can be simulated with DIVA. The question that arises consequently is what about other symptoms? A number of speech motor characteristics of MSD (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 at the segmental level; higher order suprasegmental characteristics (e.g., inappropriate prosody) and errors in phoneme selection and sequencing (e.g., speech sound omissions, substitutions and transpositions) are beyond its scope and could not be investigated. Therefore, it is necessary to expand our modeling studies to these higher order aspects of the speech production process. For example, phoneme selection and sequencing could be investigated with the gradient order DIVA model (GODIVA, Bohland, Bullock, &

Guenther, 2010), a neurocomputational model extending the existing DIVA model with the assembly 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).

In addition to speech production, also speech perception processes need to be further investigated. The present study focused on speech motor acquisition and the creation of auditory targets or the acquisition of phonemic contrasts was not investigated. It is not fully clear how an auditory processing deficit (as investigated in the current study to modulate the quality of auditory self-monitoring) would affect these acquisition processes. Based on the work of Perkell, one would predict speakers who have a less acute perception of fine acoustic detail to learn larger goal regions (Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). However, it is also known that input variability is required to form stable representations (e.g., Wolpert, Ghahramani, & Flanagan, 2001) and increased variability (through the introduction of errors caused by the auditory processing deficit) could also facilitate the acquisition process, as seen in the current study in the systemic mapping acquired under MPD. Expanding the scope of our modeling studies to other processing levels of speech production and perception is a straightforward direction for further research.

5.5 Conclusions

In a series of computer simulations with the DIVA model both a motor processing deficit alone (MPD) and a motor processing deficit in combination with an auditory processing deficit (MPD+APD) led to an increase in deviant coarticulation, speech sound distortion, online searching articulatory behavior, and token-to-token variability after imitation learning, but for different reasons: MPD+APD predominantly led to deterioration at the phonetic level (systemic mapping) whereas MPD predominantly led to deterioration at the phonological level (phonemic mappings). These findings suggest that the quality of the system for auditory self-monitoring is an important determinant for the involvement of phonological vs. motor processes in children with speech sound disorders. It is suggested that MPD+APD might be involved in typically apraxic speech output disorders and MPD in pediatric motor speech disorders that also have a phonological component.

The current study exemplifies how characteristics that seem to apparently reside at the level of stored representations could actually have a motoric origin. Different underlying disorders may show similar symptoms (Maassen, et al., 2010; Weismer & Kim, 2010) and in our DIVA simulations this is expressed by the fact that after completed learning, both MPD and MPD+APD lead to degraded speech output with similar characteristics. The high overlap in symptomatology necessitates an approach that aims at describing the *development of processing deficits* to characterize disorders rather than focusing on single diagnostic markers for differential diagnosis. It has been argued that such a process-oriented approach holds important advantages for the diagnosis and treatment of pediatric speech disorders (Terband & Maassen, 2012; Terband, 2013; Maassen & Terband, in press). Offering direct leads for treatment aimed at the specific underlying impairment, a process-oriented approach allows a treatment planning tailored to the specific needs of the individual and provides a better starting point to evaluate and adjust the therapeutic approach in the course of the speech disorder. A more purposeful treatment would render therapy more effective and efficient, yielding better outcome and increased quality of life.

Acknowledgments

Abbreviations

MPD   motor processing deficit

APD   auditory processing deficit

References

ASHA. Childhood Apraxia of Speech [technical report]. American Speech-Language-Hearing Association; 2007. Available from www.asha.org/policy

Bohland JW, Bullock D, Guenther FH. Neural representations and mechanisms for the performance of simple speech sequences. Journal of Cognitive Neuroscience. 2010; 22(7):1504–1529. [PubMed: 19583476]

Broomfield J, Dodd B. Children with speech and language disability: caseload characteristics. International Journal of Language and Communication Disorders. 2004; 39:1–22. [PubMed: 14660185]

Browman CP, Goldstein LM. Articulatory phonology: an overview. Phonetica. 1992; 49(3–4):155–180. [PubMed: 1488456]

Browman, CP.; Goldstein, LM. The gestural phonology model. In: Hulstijn, W.; Peters, HFM.; van Lieshout, PHHM., editors. Speech production: Motor control, brain research, and fluency disorders. Amsterdam: Elsevier Sciences Publishers; 1997. p. 57-71.

Crary, MA. Developmental Motor Speech Disorders. San Diego, CA: Singular Publishing Group; 1993.

Dodd, B., editor. Differential Diagnosis and Treatment of Children with Speech Disorder. 2. London: Whurr; 2005.

Edwards J, Fourakis M, Beckman ME, Fox RA. Characterizing knowledge deficits in phonological disorders. Journal of Speech, Language, and Hearing Research. 1999; 42(1):169–186.

Edwards J, Fox RA, Rogers CL. Final consonant discrimination in children: effects of phonological disorder, vocabulary size, and articulatory accuracy. Journal of Speech, Language, and Hearing Research. 2002; 45(2):231–242.

Ertmer DJ, Young NM, Nathani S. Profiles of Vocal Development in Young Cochlear Implant Recipients. Journal of Speech, Language, and Hearing Research. 2007; 50(2):393–407.

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. Journal of Speech and Hearing Research. 1996; 39(3):468–482. [PubMed: 8783127]

Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics. 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 and Language. 2006; 96(3):280–301. [PubMed: 16040108]

Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review. 1998; 105(4):611–633. [PubMed: 9830375]

Guenther, FH.; Perkell, JS. A neural model of speech production and its application to studies of the role of auditory feedback in speech. In: Maassen, B.; Kent, R.; Peters, HFM.; van Lieshout, PHHM.; Hulstijn, W., editors. Speech Motor Control in Normal and Disordered Speech. Oxford, UK: Oxford University Press; 2004. p. 29-50.

Guyette, T.; Diedrich, WM. A critical review of developmental apraxia of speech. In: Lass, NJ., editor. Speech and Language. Advances in basic research and practice. New York: Academic Press Inc; 1981. p. 1-49.

Hall, P.; Jordan, L.; Robin, D., editors. Developmental Apraxia of Speech: Theory and Clinical Practice. 2. Austin, TX: Pro-ed; 2007.

Kröger, BJ.; Birkholz, P.; Lowit, A.; Neuschaefer-Rube, C. Phonemic, sensory, and motor representations in an action-based neurocomputational model of speech production (ACT). In: Maassen, B.; Van Lieshout, P., editors. Speech motor control: New developments in basic and applied research. Oxford: Oxford University Press; 2010. p. 23-36.

Kröger, BJ.; Kannampuzha, J.; Lowit, A.; Neuschaefer-Rube, C. Phonetotopy within a neurocomputational model of speech production and speech acquisition. In: Fuchs, S.; Loevenbruck, H.; Pape, D.; Perrier, P., editors. Some aspects of speech and the brain. Frankfurt: Peter Lang; 2009. p. 59-90.

Kröger BJ, Kannampuzha J, Neuschaefer-Rube C. Towards a neurocomputational model of speech production and perception. Speech Communication. 2009; 51:793–809.

Lane H, Perkell JS. Control of voice-onset time in the absence of hearing: a review. Journal of Speech, Language, and Hearing Research. 2005; 48(6):1334–1343.

Law J, Boyle J, Harris F, Harkness A, Nye C. Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. International Journal of Language and Communication Disorders. 2000; 35(2):165–188. [PubMed: 10912250]

Loucks TMJ, De Nil LF. The effects of masseter tendon vibration on non-speech oral movements and vowel gestures. Journal of Speech, Language and Hearing Research. 2001; 44(2):306–316.

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.; Terband, H. Developmental models of Childhood Apraxia of Speech. In: Maassen, B.; Van Lieshout, P., editors. Speech motor control: New developments in basic and applied research. Oxford: Oxford University Press; 2010. p. 243-258.

Maassen, B.; Terband, H. Process-oriented diagnosis of Childhood and adult Apraxia of Speech (CAS & AOS). In: Redford, MA., editor. Handbook of Speech Production. Wiley-Blackwell; (in press)

MacNeilage, PF.; Davis, BL. Acquisition of speech production: Frames, then content. In: Jeannerod, M., editor. Attention and performance XIII: Motor representation and control. Hillsdale, NJ: Lawrence Erlbaum; 1990. p. 453-476.

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.

Marion MJ, Sussman HM, Marquardt TP. The perception and production of rhyme in normal and developmentally apraxic children. Journal of Communication Disorders. 1993; 26(3):129–160. [PubMed: 8227501]

Max L, Onghena P. Some Issues in the Statistical Analysis of Completely Randomized and Repeated Measures Designs for Speech, Language, and Hearing Research. Journal of Speech, Language, and Hearing Research. 1999; 42(2):261–270.

McKinnon DH, McLeod S, Reilly S. The prevalence of stuttering, voice, and speech-sound disorders in primary school students in Australia. Language, Speech, and Hearing Services in Schools. 2007; 38(1):5–15.

Munson B, Edwards J, Beckman ME. Relationships Between Nonword Repetition Accuracy and Other Measures of Linguistic Development in Children With Phonological Disorders. Journal of Speech, Language and Hearing Research. 2005; 48(1):61–78.

Nam, H.; Goldstein, L.; Browman, C.; Rubin, P.; Proctor, M.; Saltzman, E. TADA (TAsk Dynamics Application) Manual. Haskins Laboratories, Inc; 2007.

Namasivayam AK, van Lieshout P, McIlroy WE, De Nil L. Sensory feedback dependence hypothesis in persons who stutter. Human Movement Science. 2009; 28(6):688–707. [PubMed: 19692132]

Nijland L. Speech perception in children with speech output disorders. Clinical Linguistics & Phonetics. 2009; 23(3):222–239. [PubMed: 19283579]

Perkell J, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, et al. The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. The Journal of the Acoustical Society of America. 2004; 116(4 Pt 1):2338–2344. [PubMed: 15532664]

Perkell J, Matthies M, Lane H, Guenther F, Wilhelms-Tricarico R, Wozniak J, et al. Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. Speech Communication. 1997; 22(2–3):227–250.

Perkell J, Matthies ML, Tiede M, Lane H, Zandipour M, Marrone N, et al. The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. Journal of Speech, Language, and Hearing Research. 2004; 47(6):1259–1269.

Perrier P. Control and representations in speech production. ZAS Papers in Linguistics. 2005; 40:109–132.

Priester GH, Post WJ, Goorhuis-Brouwer SM. Problems in speech sound production in young children. An inventory study of the opinions of speech therapists. International Journal of Pediatric Otorhinolaryngology. 2009; 73(8):1100–1104. [PubMed: 19439370]

Quené H, van den Bergh H. On multi-level modeling of data from repeated measures designs: a tutorial. Speech Communication. 2004; 43(1–2):103–121.

Raaymakers EM, Crul TA. Perception and production of the final /s-ts/ contrast in Dutch by misarticulating children. Journal of Speech, Language and Hearing Research. 1988; 53(3):262–270.

Saltzman EL, Munhall KG. A dynamical approach to gestural patterning in speech production. Ecological Psychology. 1989; 1(4):333–382.

Schorr EA, Roth FP, Fox NA. A Comparison of the Speech and Language Skills of Children With Cochlear Implants and Children With Normal Hearing. Communication Disorders Quarterly. 2008; 29(4):195–210.

Shriberg LD, Fourakis M, Hall S, Karlsson HK, Lohmeier HL, McSweeny J, et al. Extensions to the Speech Disorders Classification System (SDCS). Clinical Linguistics & Phonetics. 2010; 24:795–824. [PubMed: 20831378]

Terband, H. Neurocognitive-Behavioral Research in Childhood Apraxia of Speech: where do we go from here?. CASANA Childhood Apraxia of Speech Research Symposium; Atlanta GA. February 21–22; 2013.

Terband H, Maassen B. Speech motor development in Childhood Apraxia of Speech (CAS): generating testable hypotheses by neurocomputational modeling. Folia Phoniatrica et Logopaedica. 2010; 62:134–142. [PubMed: 20424469]

Terband H, Maassen B. Spraakontwikkelingsstoornissen: Van symptoom- naar procesdiagnostiek. Logopedie en Phoniatrie. 2012; 7–8:229–234.

Terband H, Maassen B, Guenther FH, Brumberg J. Computational neural modeling of Childhood Apraxia of Speech (CAS). Journal of Speech, Language and Hearing Research. 2009; 52(6):1595–1609.

Thoonen G, Maassen B, Gabreels F, Schreuder R, de Swart B. Towards a standardised assessment procedure for developmental apraxia of speech. European Journal of Disorders of Communication. 1997; 32(1):37–60. [PubMed: 9135712]

Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. NeuroImage. 2008; 39(3):1429–1443. [PubMed: 18035557]

Tremblay S, Houle G, Ostry DJ. Specificity of speech motor learning. Journal of Neuroscience. 2008; 28(10):2426–2434. [PubMed: 18322088]

Tremblay S, Shiller DM, Ostry DJ. Somatosensory basis of speech production. Nature. 2003; 423(6942):866–869. [PubMed: 12815431]

von Hapsburg D, Davis BL, MacNeilage PF. Frame Dominance in Infants With Hearing Loss. Journal of Speech, Language, and Hearing Research. 2008; 51(2):306–320.

Warner-Czyz AD, Davis B. The emergence of segmental accuracy in young cochlear implant recipients. Cochlear Implants International. 2008; 9(3):143–166. [PubMed: 18925564]

Warner-Czyz AD, Davis BL, MacNeilage PF. Accuracy of Consonant-Vowel Syllables in Young Cochlear Implant Recipients and Hearing Children in the Single-Word Period. Journal of Speech, Language, and Hearing Research. 2010; 53(1):2–17.

Weismer, G.; Kim, Y. Classification and Taxonomy of Motor Speech Disorders: What are the Issues?. In: Maassen, B.; Van Lieshout, P., editors. Speech motor control: New developments in basic and applied research. Oxford: Oxford University Press; 2010. p. 229-242.

Wolpert DM, Ghahramani Z, Flanagan JR. Perspectives and problems in motor learning. Trends in Cognitve Sciences. 2001; 5(11):487–494.

Highlights

Developmental interaction between neurological deficits in auditory and motor processes is addressed.

Neurological deficits are simulated with a neurocomputational model of speech motor development called DIVA.

Model simulations are evaluated perceptually and acoustically.

Results suggest a close relation between quality of auditory self-monitoring and the involvement of phonological vs. motor processes in developmental motor speech disorders.
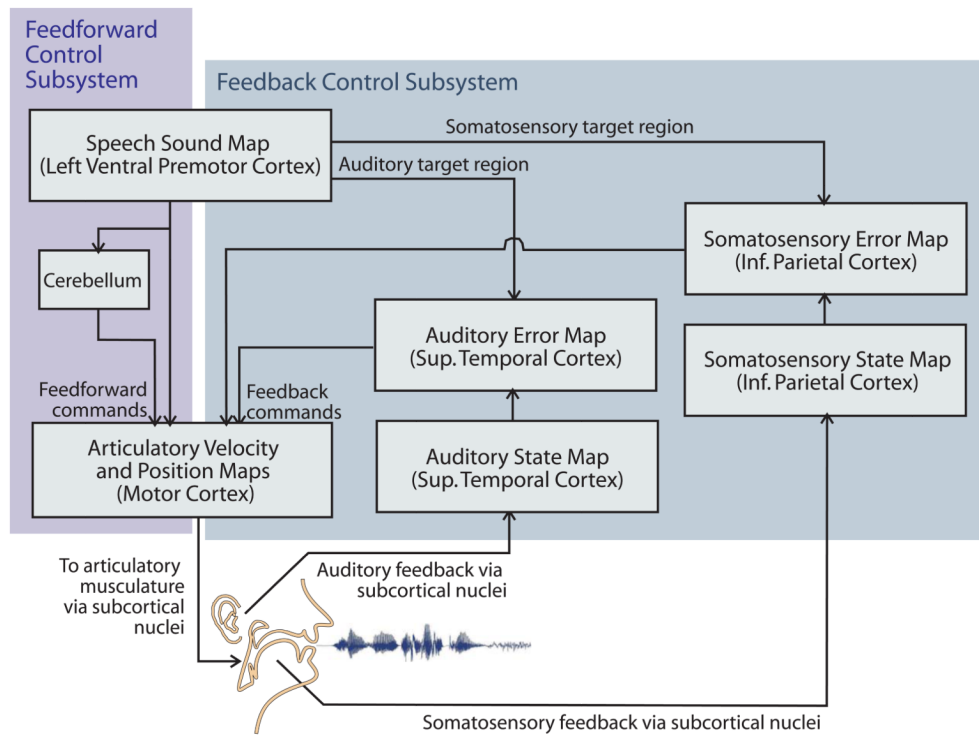
Figure 1.
Schematic representation of the DIVA model of speech motor control (Guenther et al., 2006). Projections to and from the cerebellum are simplified for clarity. Not shown in the figure is the systemic mapping (forward and inverse models), which is used by the model to calculate the *feedback commands* and *feedforward commands* and in this scheme would be located behind these two.
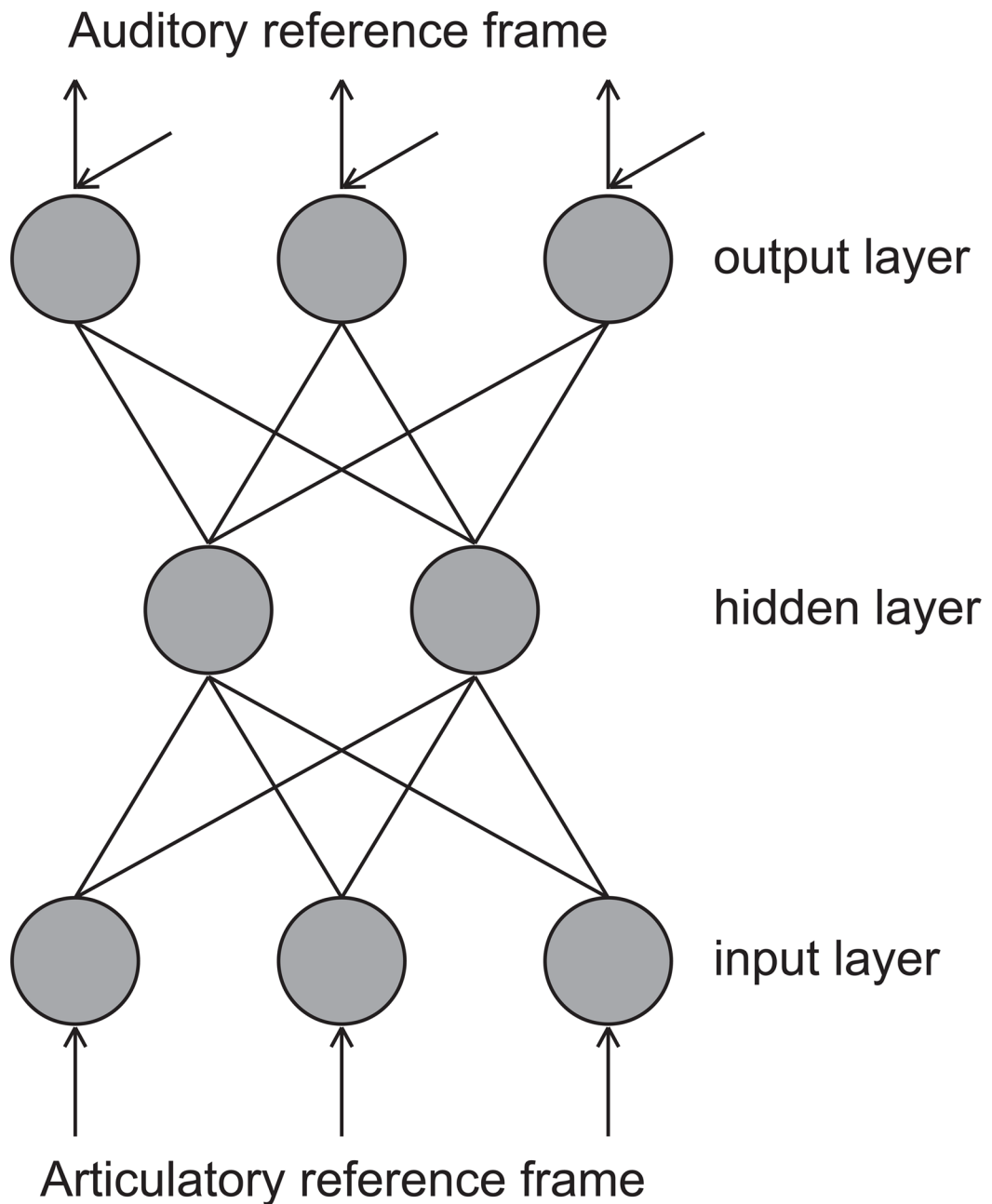
Auditory reference frame

Figure 2.
Architecture of a radial basis function network for learning a systemic mapping. The network comprises an input stage and an output stage that are connected by a hidden layer with a non-linear activation function. Training is done in an iterative manner by presenting the network with a set of input patterns consisting of articulatory movements together with its auditory effect (generated by the articulatory synthesizer). Each input pattern activates an articulatory positioning and the model adjusts the adaptive synaptic connections between the hidden layer and the output stage as to minimize the difference between the network's associated auditory output (consisting of a position in F1; F2; F3-space) and the auditory output given by the input pattern. During the learning process the mapping works towards the minimal total difference on the whole training set of input patterns.
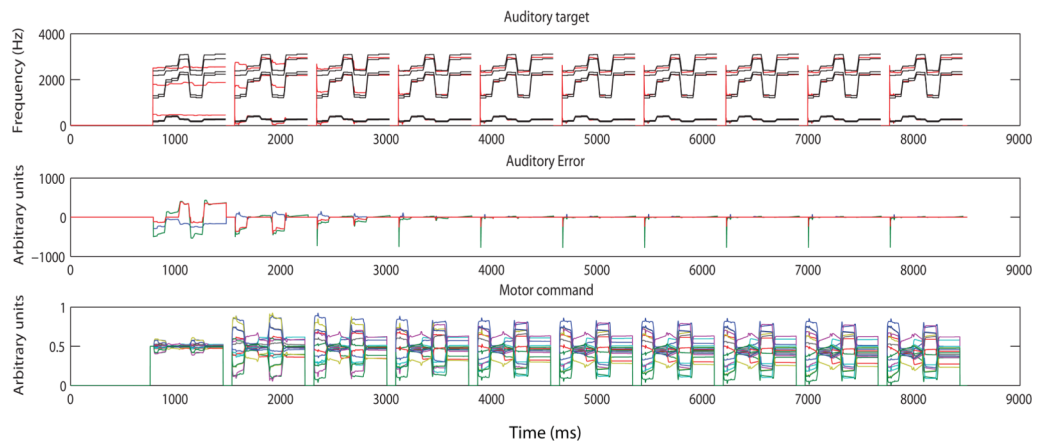
Figure 3.
The DIVA model learning the word 'baby'. The top-panel shows the target regions for F1, F2, and F3 (black) and the model's auditory realization (red). On the basis of the error information provided by the auditory feedback control subsystem (middle-panel), the feedforward command (bottom-panel) is updated with each attempt, thus becoming more accurate. After approximately five iterations, learning has become asymptotic.
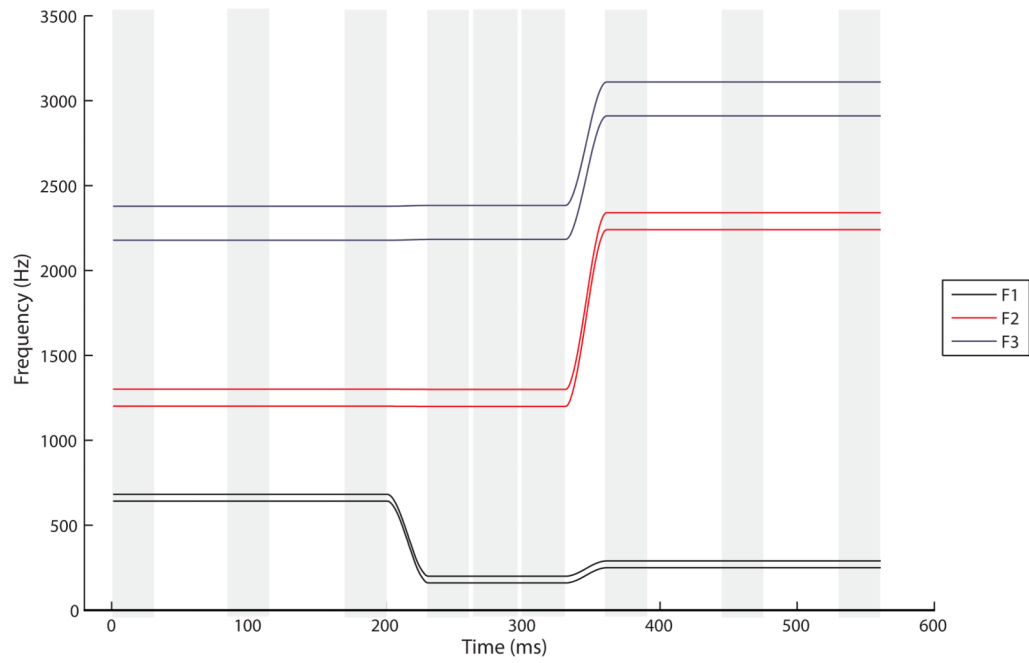
Figure 4.
Schematic representation of the speech sound target for /abi/ (Terband et al., 2009). The light gray columns indicate the measurement points. At each of these points, the mean formant value was calculated over a 30-ms time window (three measurements with 10-ms time intervals).
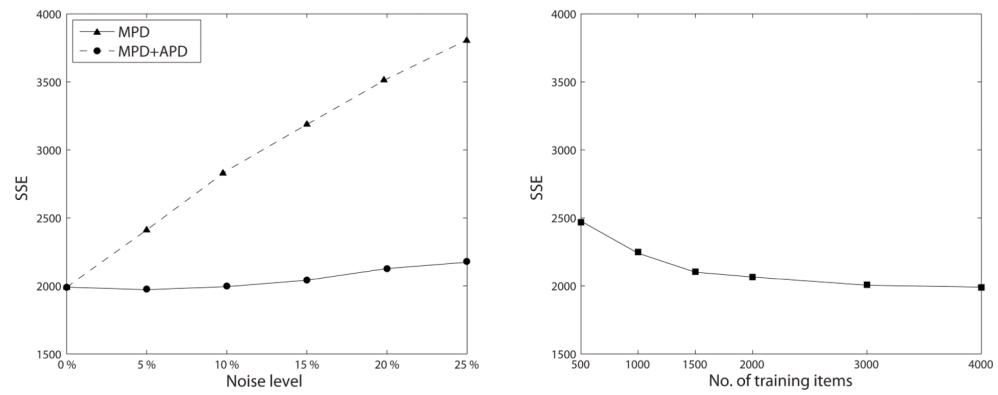
Figure 5.
Left: systemic mapping quality after the full training process (i.e., 4000 training items) in relation to severity of deficits. Right: systemic mapping quality in relation to number of training items in the baseline condition (no deficits implemented). The comparison with the effect of underlearning in the baseline condition (right panel) indicates that the decrease in mapping quality with increased severity for MPD+APD (left panel) is very strong.
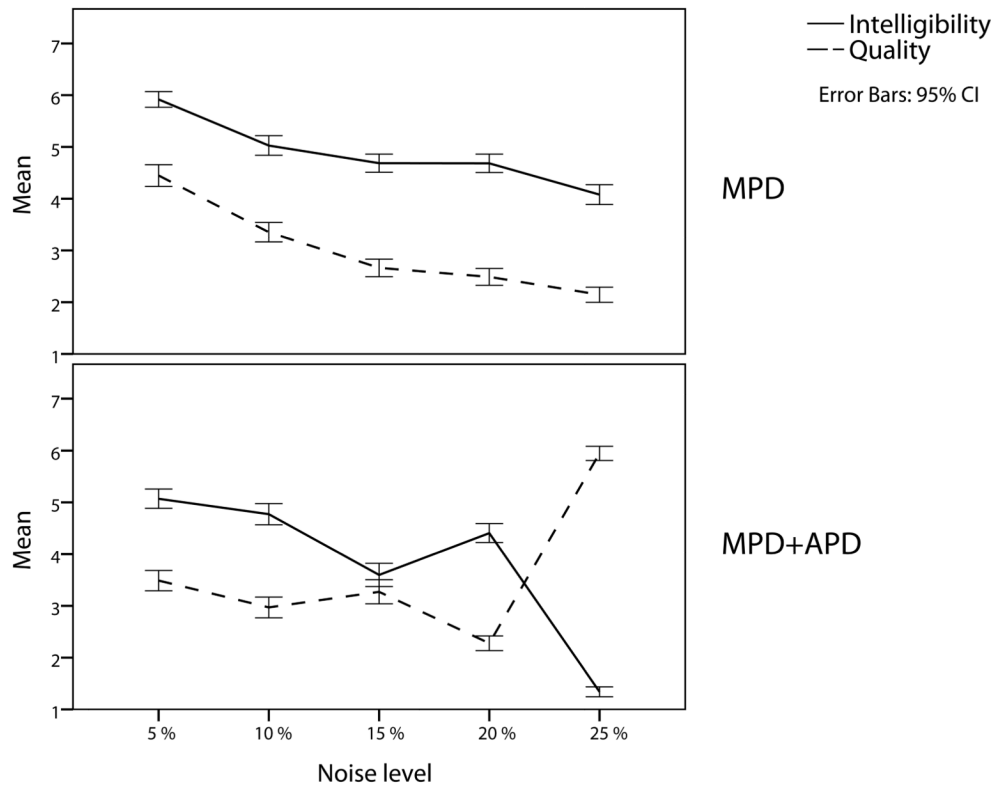
Figure 6.
Perceptual evaluation: intelligibility and quality of the models speech output in relation to severity of deficits. The perceptual judgments indicate a decrease of both intelligibility and speech quality with increased severity for both deficits and for noise levels of 15% and higher, the quality of the models speech is very bad and the utterances are barely recognizable anymore.
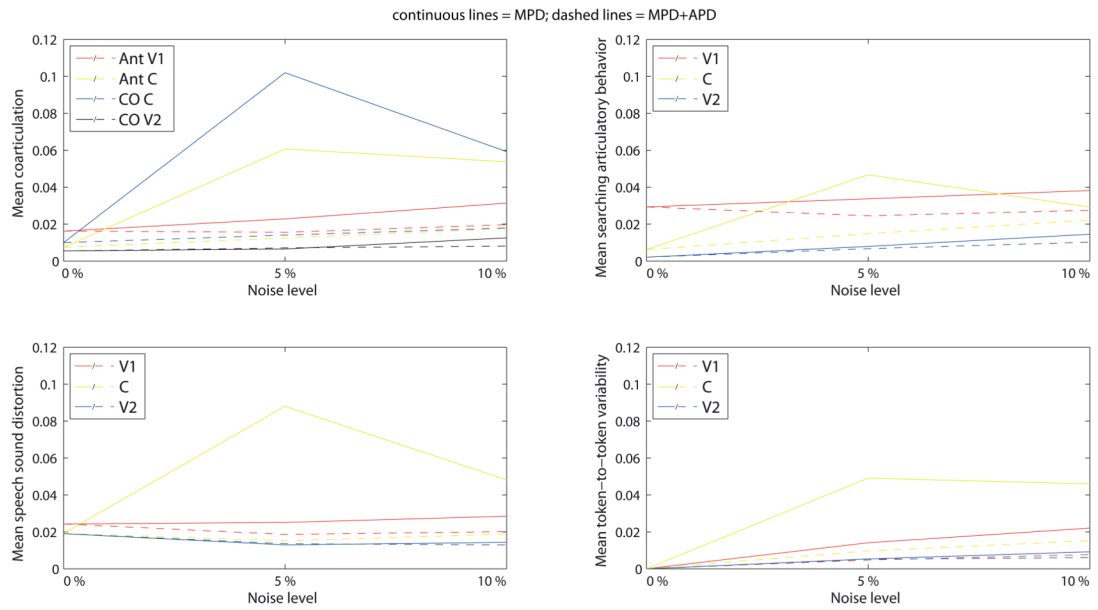
Figure 7.
Acoustical evaluation of normal productions with online feedback after asymptotic imitation learning for the motor processing deficit (MPD; continuous lines) and the motor + auditory processing deficits (MPD+APD; dashed lines). Top: coarticulation (left) and searching articulatory behavior (right). Bottom: speech sound distortion (left) and token-to-token variability of mean formant frequencies (right). Both deficits lead to an increase in the four selected key phonetic characteristics of MSD after imitation learning with increased severity.

A. MPD — continuous lines = deficit implemented during learning; dashed lines = deficit implemented after learning

B. MPD+APD — continuous lines = deficit implemented during learning; dashed lines = deficit implemented after learning
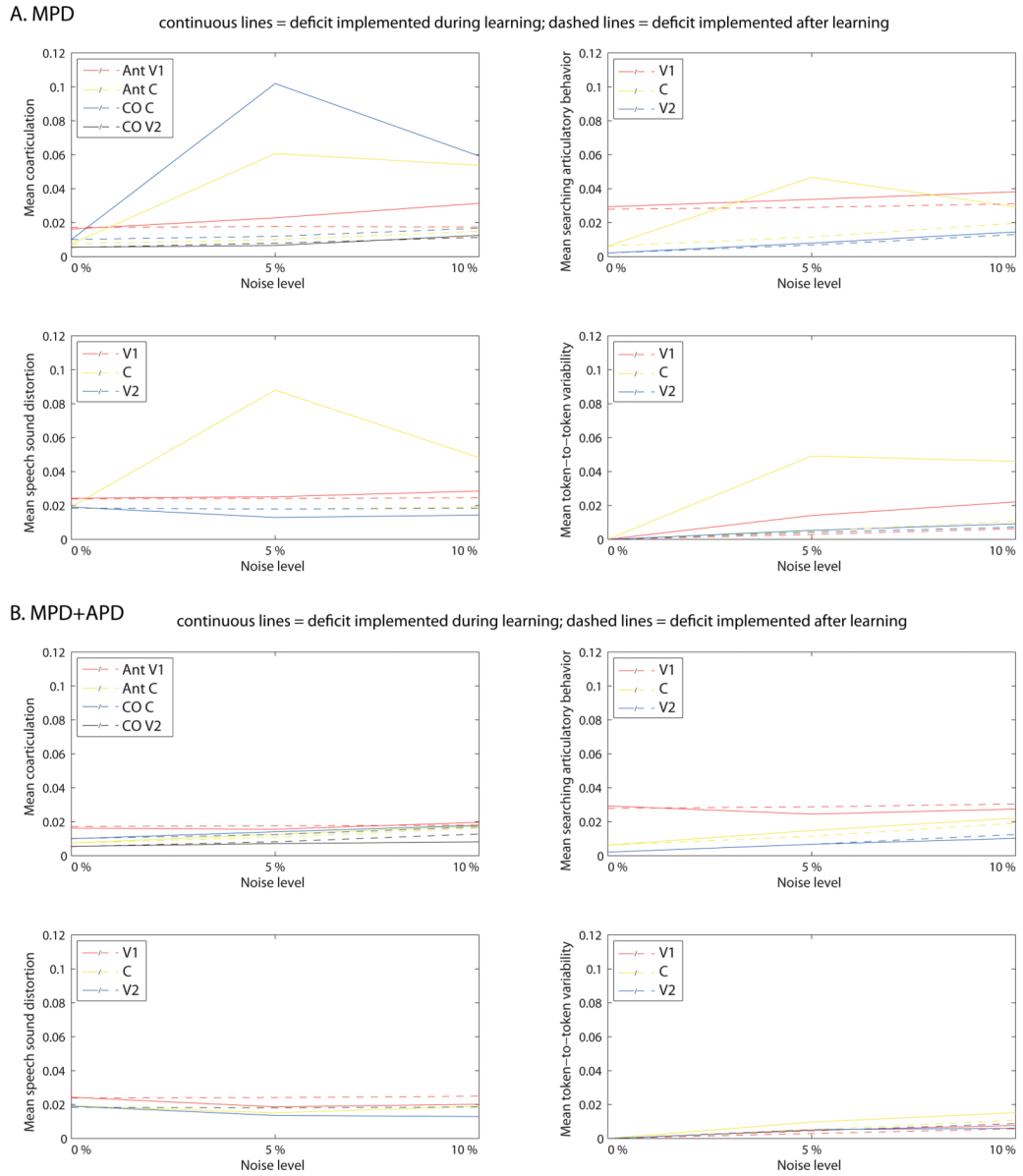
Figure 8.
Comparison between the effect of the deficits if present during the whole learning process (continuous lines) and if implemented after asymptotic learning under zero-noise conditions (dashed lines), for the motor processing deficit (MPD; top four panels) and the motor + auditory processing deficits (MPD+APD; bottom four panels). Clockwise, starting at upper left panel, the panels show coarticulation, searching articulatory behavior, token-to-token variability of mean formant frequencies, and speech sound distortion in relation to severity of deficits. The results show large differences in speech output between the deficit present during and after learning for MPD, but not for MPD+APD.
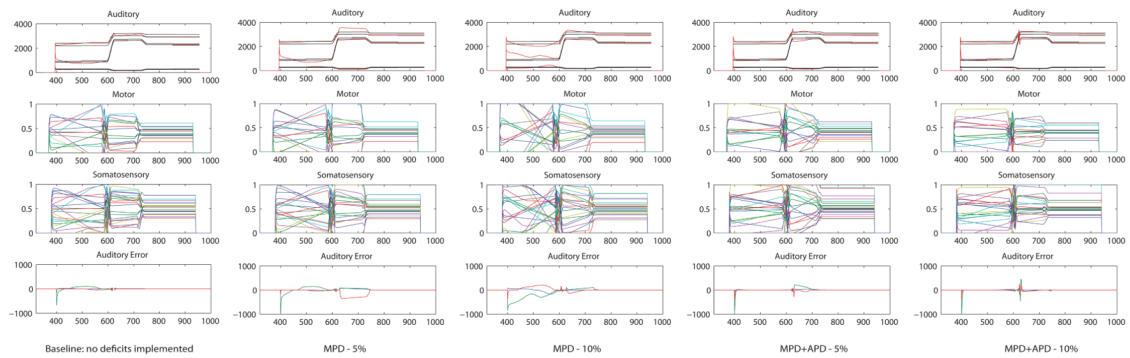
Figure 9.
Acquired forward commands for the word /ugi/ for the two implemented deficits in comparison with the standard (baseline) model. The acquired forward commands are extracted by production attempts in the performance stage with the feedback control subsystem turned off, i.e. letting the model 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. The large auditory error for MPD indicates that the acquired forward command does not correspond to the auditory target.

Table 1

Sum-squared errors of the HRBF networks for the implemented deficits at different stages in the training process. A noise level of 0 denotes the baseline condition (no deficits implemented).

Implemented deficit	No. of training items	Noise level						
		0	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**
MPD	500	2477	2463	2480	2541	2656	2709
	1000	2241	2223	2258	2338	2390	2457
	1500	2101	2103	2138	2194	2292	2370
	2000	2066	2049	2084	2145	2234	2290
	3000	2006	2018	2032	2089	2136	2220
	4000	1991	1973	1990	2048	2110	2184
MPD+APD	500	2477	3218	3842	4171	4360	4711
	1000	2241	2806	3388	3793	4074	4253
	1500	2101	2669	3194	3630	3862	4191
	2000	2066	2530	3069	3449	3767	4045
	3000	2006	2437	2940	3290	3683	3934
	4000	1991	2382	2849	3338	3603	3770

Table 2

Results of Kruskal-Wallis tests investigating possible main effects of the implemented deficits (noise levels of 5% and 10%) on the acoustic measures.

| | | Implemented deficit | | | |
| | | MPD | | MPD+APD | |
		Chi-square	p	Chi-square	p
Coarticulation	V1 Anticipatory	23.676	0.000	12.632	0.002
	C Anticipatory	63.806	0.000	36.682	0.000
	C Carry-over	45.994	0.000	16.363	0.000
	V2 Carry-over	50.173	0.000	21.457	0.000
Searching articulatory behavior	V1	4.896	0.086	4.756	0.093
	C	231.579	0.000	207.640	0.000
	V2	254.669	0.000	232.375	0.000
Speech sound distortion	V1	1.486	0.476	40.345	0.000
	C	0.134	0.935	47.684	0.000
	V2	57.919	0.000	54.807	0.000
Token-to-token variability	V1	54.776	0.000	61.740	0.000
	C	53.649	0.000	57.376	0.000
	V2	63.957	0.000	57.486	0.000

Table 3

Results of Mann-Whitney U tests for the acoustic measures comparing the involvement of the deficits if implemented during the whole learning process and if implemented after asymptotic learning.

		Implemented deficit							
		MPD				MPD+APD			
		Noiselevel				Noiselevel			
		5%		10%		5%		10%	
		Z	p	Z	p	Z	p	Z	p
Coarticulation	V1 Anticipatory	−2.476	0.013	−3.300	0.001	−0.823	0.410	−1.069	0.285
	C Anticipatory	−4.731	0.000	−4.854	0.000	−0.074	0.941	−0.117	0.907
	C Carry-over	−3.865	0.000	−4.295	0.000	−0.664	0.507	−0.817	0.414
	V2 Carry-over	−0.842	0.400	−1.966	0.049	−2.175	0.030	−3.005	0.003
Searching articulatory behavior	V1	−1.298	0.194	−0.901	0.367	−2.065	0.039	−1.465	0.143
	C	−6.647	0.000	−4.337	0.000	−3.930	0.000	−0.655	0.513
	V2	−1.847	0.065	−1.276	0.202	−0.915	0.360	−1.884	0.060
Speech sound distortion	V1	−0.983	0.325	−1.051	0.293	−6.142	0.000	−5.321	0.000
	C	−0.619	0.536	−1.355	0.176	−5.623	0.000	−0.701	0.483
	V2	−5.737	0.000	−4.055	0.000	−4.041	0.000	−5.593	0.000
Token-to-token variability	V1	−6.237	0.000	−5.925	0.000	−3.763	0.000	−2.379	0.017
	C	−6.081	0.000	−5.026	0.000	−5.268	0.000	−1.289	0.197
	V2	−2.033	0.042	−2.517	0.012	−0.874	0.382	−2.880	0.004

Table 4

Results of Mann–Whitney U tests for the acoustic measures comparing the motor processing deficit (MPD) and the motor + auditory processing deficit (MPD+APD) if implemented after asymptotic learning.

| | | Noiselevel | | | |
| | | 5 % | | 10 % | |
		Z	p	Z	p
Coarticulation	V1 Anticipatory	−0.160	0.873	−0.326	0.745
	C Anticipatory	−1.057	0.291	−1.020	0.308
	C Carry-over	−0.442	0.658	−0.571	0.568
	V2 Carry-over	−0.602	0.547	−1.352	0.176
Searching articulatory behavior	V1	−0.294	0.769	−0.202	0.840
	C	−0.231	0.817	−0.045	0.964
	V2	−0.189	0.850	−0.463	0.644
Speech sound distortion	V1	−0.221	0.825	−0.720	0.472
	C	−0.814	0.416	−0.466	0.641
	V2	−0.122	0.903	−0.077	0.939
Token-to-token variability	V1	−0.221	0.825	−0.720	0.472
	C	−0.814	0.416	−0.466	0.641
	V2	−0.122	0.903	−0.077	0.939