SIMULATIONS OF FEEDFORWARD AND FEEDBACK CONTROL IN APRAXIA OF SPEECH (AOS): EFFECTS OF NOISE MASKING ON VOWEL PRODUCTION IN THE DIVA MODEL

Hayo Terband[1], Joe Rodd[1] & Edwin Maas[2]

[1]Utrecht Institute of Linguistics - OTS, Utrecht University
[2]Department of Speech, Language, and Hearing Sciences, University of Arizona
h.r.terband@uu.nl; j.j.e.rodd@uu.nl; emaas@email.arizona.edu

ABSTRACT

Apraxia of Speech (AOS) is a motor speech disorder whose precise nature is still poorly understood. A recent behavioural experiment featuring a noise masking paradigm suggests that AOS reflects a disruption of feedforward control, whereas feedback control is spared and plays a more prominent role in achieving and maintaining segmental contrasts [10]. In the present study, we set out to validate the interpretation of AOS as a feedforward impairment by means of a series of computational simulations with the DIVA model [6, 7] mimicking the behavioural experiment.

Simulation results showed a larger reduction in vowel spacing and a smaller vowel dispersion in the masking condition compared to the no-masking condition for the simulated feedforward deficit, whereas the other groups showed an opposite pattern. These results mimic the patterns observed in the human data, corroborating the notion that AOS can be conceptualized as a deficit in feedforward control.

Keywords: apraxia of speech; computational modelling; vowel acoustics; feedback masking.

1. INTRODUCTION

Apraxia of speech (AOS) is a neurogenic motor speech disorder resulting from brain lesions to the left cerebral hemisphere, although more specific lesion locations reported in the literature diverge [4, 8, 18, 19]. The speech of patients with AOS is characterized by slow speech rate, abnormal prosody, abnormal speech sound and syllable segmentation, speech sound distortions, and speech errors that are inconsistently present but relatively consistent in type and location [5, 15]. From a functional point of view, AOS is defined as an impairment in the planning and/or programming of speech movements [3, 5, 11, 13, 23], however, the precise nature of the disorder remains poorly understood.

One of the main difficulties in isolating the underlying deficit(s) is diagnostic circularity. The ability to investigate the characteristics underlying AOS requires 'pure' cases of AOS selected on the basis of unambiguous/clear-cut criteria, which are only available as a result of research. As lesion inducing medical accidents such as strokes, brain injuries, or tumors rarely produce isolated and one-dimensional deficits, pure cases are rare and symptom profiles show considerable variation between individuals and a large overlap in symptomatology with other speech disorders. Additionally, when confronted with a partial breakdown, the speech system itself is likely to adapt to the deviant circumstances and/or compensate for the impediments, and individuals may vary widely in these adaptive and compensatory mechanisms.

This problem of practical-diagnostic circularity results from the behavioural, symptom-oriented approach that is employed [14, 21]. Although the symptomatology might be aspecific, it is possible to describe a specific speech-motor core deficit from the perspective of the underlying cognitive and neurological processes. As such, we argue that to identify underlying deficits, one must begin by deriving detailed, specific hypotheses within the context of a detailed model of the behavioural and cognitive operations involved. These hypotheses should then be tested empirically, and ideally contrasted with alternative hypotheses for underlying deficits (e.g., those presumed to underlie 'neighbouring' impairments such as dysarthria) [21].

One promising, and relatively recent approach to understanding AOS in this respect, relates to the development of the DIVA model, a computationally implemented neural network model of speech acquisition and speech motor control [6, 7]. The main function of computational modelling is to understand the effects of a particular underlying deficit. Currently, clinicians tend to interpret symptoms at face validity (e.g. errors in place of articulation as resulting from motor programming errors). By deductive reasoning, modelling allows us to test such interpretations directly, thereby giving us a powerful tool for validating inductive reasoning (from symptom to deficit) [20, 21]. In the current study, we utilized this modelling approach and investigated the potential role of two deficits that have been hypothesized to underlie AOS [10] in Simulink DIVA [16], a computational implementation of the DIVA model.

2. OVERVIEW OF THE DIVA MODEL

The DIVA model consists of a neural network controller detailing feedforward and feedback control loops that are assumed to be involved in early speech development and mature speech production, focusing

on the sensorimotor transformations underlying the control of articulator movements [6, 7]. The model strives to be neurobiologically plausible and its components have been associated with regions of the cerebral cortex and cerebellum [7]. In order to produce an acoustic signal, DIVA controls the movements of an articulatory synthesizer [12].

In the DIVA model, production of a speech sound begins with activation of a speech sound map (SSM) cell in left inferior frontal cortex. SSM cells represent speech sounds (which may range in size from phonemes, to syllables, to frequent words and phrases) and are presumed to be activated by higher-level input from the phonological encoding stage [2, 7]. The activated SSM cell then activates a feedforward control system and a feedback control system, whose motor commands are combined in primary motor cortex. Feedback control involves comparing actual auditory and somatosensory feedback signals to expected auditory and somatosensory consequences, and generating corrective motor commands to motor cortex when a mismatch (error) is detected. Expected sensory consequences are encoded as regions in auditory space (superior temporal gyrus) and somatosensory space (postcentral and supramarginal gyri). Feedforward control involves predictive motor commands from the SSM to motor cortex. Feedforward commands are learned by incorporating the feedback system's corrective commands from previous productions. With sufficient practice, the feedforward commands generate little to no errors, so that contributions of the feedback control system are minimal during normal speech, although feedback may be continuously monitored for deviations from expectations, even in adult speakers [22].

3. THE PRESENT STUDY

As noted above, the current consensus is that AOS is defined as a speech motor planning and/or programming disorder, or, more specifically, an inability to transform an abstract linguistic code involving intact phonological representations into spatially and temporally coordinated patterns of muscle contractions that produce speech movements [9, 11, 15]. Within this accepted consensus, Maas et al. [10] proposed two alternative hypotheses with respect to the underlying mechanisms. One suggestion was that the underlying (core) deficit in AOS may be viewed as one of impaired feedforward control (*Feedforward System Deficit Hypothesis*; FF hypothesis). The disruption in feedforward processing would cause the motor commands to be inappropriate or underspecified, thereby introducing errors, which in turn would increase the contribution of feedback-based corrective commands to the overall motor command. In other words, according to the FF hypothesis the speech system would show an increased reliance on feedback control and the role of feedback control would be facilitatory in achieving and maintaining segmental contrast in AOS.

The other suggestion was that AOS may involve impaired feedback control (*Feedback System Deficit Hypothesis*; FB hypothesis). In this case, the disruption of feedback processing would be in deriving error information from mismatching feedback and/or generating corrective commands on the basis of such errors, for example because incorrect target regions are activated, because the internal model that governs corrections is damaged, or because feedback commands cannot be integrated with feedforward commands.

Maas and colleagues [10] investigated these two hypotheses in a behavioural experiment featuring an auditory feedback masking paradigm. The rationale is that noise masking effectively prevents auditory feedback control, forcing reliance on feedforward control (and somatosensory feedback control). If the feedforward system is impaired, and people with AOS rely primarily on auditory feedback control to maintain segmental contrast, then such removal of auditory feedback would reveal the – impaired – feedforward system. On the other hand, if symptoms of AOS reflect interference from the auditory feedback signal (e.g., due to generating unnecessary or inadequate corrective commands), then removing the auditory feedback should improve speech performance in terms of segmental contrast and stability.

Findings from vowels produced by six speakers with AOS revealed that vowel spacing (acoustic contrast) was more reduced under noise conditions than in control speakers, consistent with the hypothesis of a feedforward deficit. In addition, a marginal interaction between group and condition emerged for vowel dispersion (the token-to-token variability of a vowel around its mean location in F1 x F2 space [17]), hinting at greater dispersion for the AOS group than the controls in the clear (no-masking) condition but comparable dispersion in the noise condition. This pattern would be expected if speakers with AOS rely to a greater extent than controls on auditory feedback control, which tends to be more variable due to on-line corrections to the motor commands. In conclusion, these findings provide support for the notion of an impaired feedforward control system in AOS [10].

In the present study, we set out to investigate the interpretation of AOS as a feedforward impairment further by means of a series of computational simulations mimicking the experiment of auditory feedback masking in speakers with AOS by Maas and colleagues [10].

4. DIVA SIMULATIONS

4.1. Experimental paradigm

Modified versions of the Simulink DIVA model [16] were derived from a pre-trained model that in its

'healthy' state produces stable, mature output for the targets that it is tested against. These targets were the same as those used in the Maas et al experiment [10]; namely /bVt/ tokens, which are reliably synthesised with the Maeda [12] articulatory synthesiser.

In the masking condition, no auditory feedback was provided to the model from the articulatory synthesiser. Each impairment × masking condition pair was captured in a separate, appropriately modified implementation of the model. Additionally, a third condition, severity, parametrically varied the degree of signal degradation caused by the impairments, at three levels (5%, 10%, 15%).

4.2 Feedforward impairment condition

The hypothesised feedforward deficit was modelled by creating two separate speech sound map components, one connected to the feedback systems and one connected to the feedforward system. The output to the feedforward system (cerebellum and articulatory and position velocity maps) was distorted by (1) cancelling activation of a random subset of cells and (2) adding random, signal-independent noise. The activation in these cells represents on a binary level the activation state of the neurons that link to synaptic networks encoding the weightings of individual productions. Higher levels of the severity condition resulted in a greater percentage of activation-cancellation, and a higher level of signal-independent noise.

4.3 Feedback impairment condition

Two different feedback deficits were modelled, one as a disruption prior to, and one as a disruption after, the integration of feedback and feedforward commands in motor cortex; in both cases, the feedback deficit was simulated by the same mechanism: adding random signal-independent noise to the feedback-based error signals (one per auditory dimension), which are derived from discrepancies between intended and actual speech targets. The noise was the form of a random multiplier between 0 and 2 applied arbitrarily to a number of the error feedback signals, in proportion to the severity level.

4.4 Target items

The target items that the model produces are specified by various time-aligned minimum and maximum limits; in Hz for F_0 and the vowel formants; and on an arbitrary -1 to 1 scale for the articulatory dimensions of pressure, voicing, and closure at pharyngeal, uvular, palatal, alveolar-dental and labial places of articulation. Vowel formant targets were derived from 95^{th} and 5^{th} percentile LPC from twenty natural productions of each vowel. The articulatory targets for the consonants were derived from sample target items distributed with the computational DIVA model. All the items were time-normalised; meaning vowel onset and offset times were the same in all items.

4.5 Simulation procedure

Forty-one repetitions were performed of each the possible combinations of conditions: item × group (impairment) × masking × severity. Ten simulation cycles of each condition combination were performed, to allow the performance of the model to stabilise in its new, impaired condition. The productions of the tenth cycle were stored as WAV files for analysis.

4.6 Acoustic analysis

Acoustic analysis was conducted in Praat [1] using a customised script. First, the vowel component of each output item was identified based on the intensity contour. Items where the longest of these detected potential vowels was shorter than 5 analysis frames (0.125s), or whose mean intensity was below the 10th percentile of the whole dataset were excluded. A sample (n = 40) of the remaining tokens were manually inspected; no errors were found.

Formants were then measured in each item using LPC analysis and converted to Mel-space. LPC settings were tuned manually. A median value for each formant was calculated, and items where the variability in pitch measurements within any one formant was above the 95^{th} percentile of the whole dataset were additionally excluded.

As in Maas et al. [10], vowel spacing was calculated as the mean Euclidean distance between the means of each possible pair of vowels. Vowel dispersion was calculated as the average of the Euclidian distances between each vowel token and that vowel's mean.

5. RESULTS & DISCUSSION

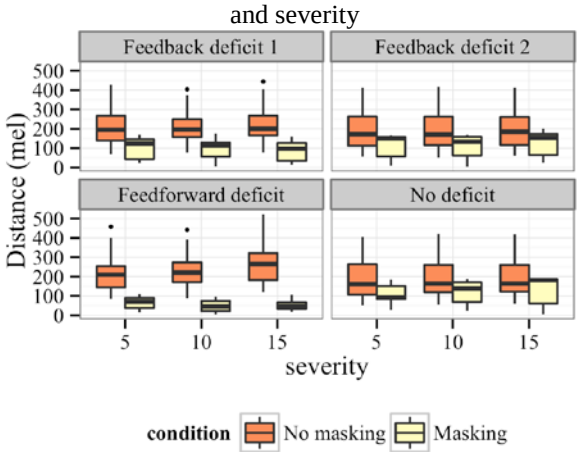
5.1 Average vowel spacing

Figure 1 presents the results on average vowel spacing, demonstrating a reduced vowel space in the masking condition compared to the no-masking condition in all the group (impairment) × severity conditions. The effect is most pronounced for the feedforward deficit simulations.

An ANOVA revealed no main effect for group, indicating that the different impaired models did not differ from each other or the unimpaired model overall. Comparisons within each impairment revealed that the difference between the masking and non-masking conditions was highly significant for the feedforward deficit group; $F(1,93) = 102.12$ $p = 0$ ***. In contrast to Maas et al. [10], there was still an effect for condition within the healthy control group, but the effect was smaller than for the feedforward deficit group; $F(1,82) = 10.08$ $p = 0.0021$ **. The interaction

of model and condition was very highly significant; $F(1,177) = 11.76$ p = 8e-04 ***. For the feedforward deficit model, there is a statistically significant interaction of severity of the deficit and condition; $F(3,91) = 3.67$ $p = 0.0153$ *. This was not the case for either of the feedback deficit models, nor was there a model x condition interaction for the feedback conditions.

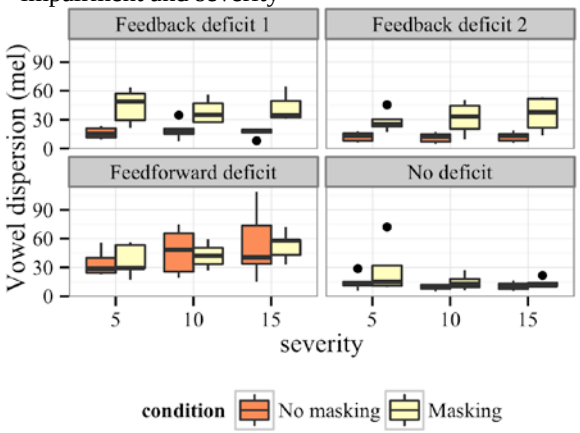
Figure 1: Average vowel spacing by impairment and severity



5.2 Vowel dispersion

Results on average vowel dispersion are presented in Figure 2, demonstrating a smaller vowel dispersion in the masking condition compared to the no-masking condition for the simulated feedforward deficit, whereas the other groups showed an opposite pattern (greater vowel dispersion in the masking condition compared to the no-masking condition).

Figure 2: Average vowel dispersion by impairment and severity



Maas and colleagues [10] found a marginal effect for impairment group on vowel dispersion, and a marginal group (impairment) × condition interaction. In this investigation, we find a highly significant effect for impairment when comparing the feedforward impaired model and the healthy model; $F(1,81) = 61.5$ p = 0 ***. Interpretation of this main effect is superseded by a small impairment × condition interaction effect; $F(1,81) = 4.75$ $p = 0.0323$ *. Examining only the feedforward model finds no effect for condition, however; $F(1,41) = 1.97$ $p = 0.1685$, meaning that the interaction effect is driven by the difference in the healthy model rather than that in the feedforward model.

For the feedback models and the control model, there is a significant interaction effect of model × condition on vowel dispersion $F(2,122) = 4.68$ $p = 0.0111$ *, indicating a disproportionate increase in dispersion with masking for the feedback deficit models.

When the no-masking outputs of the models are compared, there is a highly significant effect for impairment type (across all models, $F(3,91) = 36.67$ p = 0 ***). Looking only at the feedforward deficit and control models, there is again a very highly significant effect; $F(1,45) = 43.58$ p = 0 ***. The interaction with severity is also significant; $F(3,43) = 3.39$ $p = 0.0273$ *. There is also a very highly significant effect looking at the control and feedback deficit 1 models; $F(1,46) = 18.52$ p = 1e-04 ***, but for the control and feedback deficit 2 models, there is no effect; $F(1,46) = 0.05$ $p = 0.8328$.

6. CONCLUSIONS

Results from the simulations support the interpretation of the human observations, in that the model with the simulated feedforward deficit demonstrated a disproportionate reduction in acoustic vowel contrast compared to the intact model. Although the simulated feedback deficits also resulted in a reduction of acoustic vowel contrast with masking, this magnitude of this reduction did not differ from that in the intact model.

In addition, with respect to vowel dispersion the simulations revealed a small group x condition interaction such that vowel dispersion was greater in the clear (no-masking) condition than in the masking condition for the simulated feedforward deficit, whereas the intact model (and the feedback deficit models) showed, numerically, an opposite pattern (greater dispersion in the masking than in the clear condition). Again, this pattern mimics the patterns observed in the human data, corroborating the notion that AOS can be conceptualized as a deficit in feedforward control.

7. ACKNOWLEDGEMENTS

This work was supported by the Netherlands Organization for Scientific Research (NWO-VENI grant 275-89-016 awarded to the first author', and the National Institutes of Health (National Institute of Deafness and Other Communication Disorders grant K01-DC010216 awarded to the third author).

8. REFERENCES

[1] Boersma, P., Weenink, D., 2013. Praat: doing phonetics by computer: Retrieved March 10, 2013, from http://www.praat.org/.

[2] Bohland, J.W., Bullock, D., Guenther, F.H. 2010. Neural representations and mechanisms for the performance of simple speech sequences, *Journal of Cognitive Neuroscience*, 22, 1504-29.

[3] Deger, K., Ziegler, W. 2002. Speech motor programming in apraxia of speech, *Journal of Phonetics*, 30, 321-335.

[4] Dronkers, N.F. 1996. A new brain region for coordinating speech articulation., *Nature*, 384, 159-161.

[5] Duffy, J.R. 2005. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St Louis, MO: Mosby.

[6] Guenther, F.H. 1994. A neural network model of speech acquisition and motor equivalent speech production, *Biological Cybernetics*, 72, 43-53.

[7] Guenther, F.H., Ghosh, S.S., Tourville, J.A. 2006. Neural modeling and imaging of the cortical interactions underlying syllable production, *Brain and Language*, 96, 280-301.

[8] Hillis, A.E., et al. 2004. Re-examining the brain regions crucial for orchestrating speech articulation, *Brain*, 127, 1479-1487.

[9] Jacks, A., Mathes, K.A., Marquardt, T.P. 2010. Vowel Acoustics in Adults With Apraxia of Speech, *Journal of Speech, Language, and Hearing Research*, 53, 61-74.

[10] Maas, E., Mailend, M.L., Guenther, F. in press. Feedforward and feedback control in apraxia of speech (AOS): Effects of noise masking on vowel production, *Journal of Speech, Language and Hearing Research*,

[11] Maas, E., et al. 2008. Motor programming in apraxia of speech, *Brain Lang*, 106, 107-118.

[12] Maeda, S., 1990. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In: *Speech production and speech modeling*, W. Hardcastle, A. Marchal, Eds. Boston: Kluwer Academic Publishers.

[13] McNeil, M.R., et al., in press. Apraxia of Speech Theory, Assessment, Differential Diagnosis and Treatment: Past, Present and Future In: *Speech Motor Control in normal and disordered speech: Future developments in theory and methodology*, P. Van Lieshout, B. Maassen, H. Terband, Eds. Rockville, MD: American Speech-Language-Hearing Association.

[14] McNeil, M.R., Pratt, S.R., Fosset, T.R.D., 2004. The differential diagnosis of apraxia of speech In: *Speech Motor Control in Normal and Disordered Speech*, B. Maassen, et al., Eds. Oxford: Oxford University Press, 389-414.

[15] McNeil, M.R., Robin, D.A., Schmidt, R.A., 2009. Apraxia of speech: definition and differential diagnosis. In: *Clinical Management of Sensorimotor Speech Disorders*, McNeil, Ed. New York, NY: Thieme, 249-268.

[16] Nieto-Castanon, A., 2011. Simulink DIVA model: Speech Lab, Department of Speech, Language & Hearing Sciences, Sargent College, Boston University.

[17] Perkell, J., et al. 2007. Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users, *The Journal of the Acoustical Society of America*, 121, 505-18.

[18] Schiff, H.B., et al. 1983. Aphemia: clinical-anatomic correlations, *Archives of neurology*, 40, 720.

[19] Taubner, R., Raymer, A.M., Heilman, K.M. 1999. Frontal-opercular aphasia, *Brain Lang*, 70, 240-261.

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

[21] Terband, H., Maassen, B., Maas, E., in press. Towards a model of pediatric speech sound disorders (SSD) for differential diagnosis and therapy planning In: *Speech Motor Control in normal and disordered speech: Future developments in theory and methodology*, P. Van Lieshout, B. Maassen, H. Terband, Eds. Rockville, MD: American Speech-Language-Hearing Association.

[22] Tourville, J.A., Reilly, K.J., Guenther, F.H. 2008. Neural mechanisms underlying auditory feedback control of speech, *NeuroImage*, 39, 1429-43.

[23] Van der Merwe, A., 1997. A theoretical framework for the characterization of pathological speech sensorimotor control In: *Clinical Management Of Sensorimotor Speech Disorders*, M. R. McNeil, Ed. New York: Thieme Medical Publishers Inc, 1-25.