USING COMPUTATIONAL MODELS TO CHARACTERIZE THE ROLE OF MOTOR NOISE IN SPEECH: THE CASE OF AMYOTROPHIC LATERAL SCLEROSIS

Benjamin Parrell[1], Antje Mefferd[2], Sarah Harper[3], Simon Roessig[4], Doris Mücke[5]

[1]Communication Sciences and Disorders, University of Wisconsin-Madison,
[2]Hearing and Speech Sciences, Vanderbilt University Medical Center, [3]Chang Lab, UC San Francisco,
[4]Department of Linguistics, Cornell University, [5]IfL-Phonetics, University of Cologne
bparrell@wisc.edu, doris.muecke@uni-koeln.de

ABSTRACT

Although the speech impairments in neurogenic speech disorders have been characterized using various approaches (kinematic, acoustic, auditory-perceptual), the underlying mechanisms that contribute to the observed symptoms remain poorly understood. A more direct assay of speech motor control is needed to test hypotheses about the underlying cause of the observed changes in speech patterns. Computational models of speech production provide one possibility, as they allow direct manipulation of control parameters.

Here, we present a proof-of-concept of how the Task Dynamic model of speech production can be used to test potential control impairments using an analysis-by-synthesis approach. Specifically, we test how changes in motor noise in speakers with amyotrophic lateral sclerosis may account for altered patterns of spatial variability in this population. This work establishes the potential of exploring how noise and bias in motor control may help uncover the mechanisms underlying speech impairments in this and other disorders.

Keywords: variability, ALS, motor noise, motor control, kinematics

1. INTRODUCTION

One of the major challenges in understanding neurogenic speech disorders has been establishing the precise nature of the underlying motor impairments. Typically, these disorders are characterized by auditory-perceptual speech characteristics, such as the Mayo clinic dysarthria classification system [1]. However, many perceptual speech features overlap across speakers with different underlying etiologies and dysarthria types (e.g., “consonant imprecision”). To better understand these disorders, a growing body of research has focused on the characterization of motor impairment patterns by direct examination of speech acoustics or kinematics rather than listener impression. While this work has promising potential to provide a more precise description of the motor speech impairments in these speakers, acoustic and kinematic analyses can ultimately only address the observable symptoms stemming from the neurologic damage to the speech production system. Alternative approaches are necessary to reveal the pathomechanisms that underlie those symptoms.

One potential way to move beyond description of symptoms to a full characterization of the control impairments in neurogenic disorders is through the use of computational models of speech production. For example, selective impairments in either the feedforward or feedback pathways of the DIVA model have been used to replicate speech impairments seen in stuttering [2] and apraxia of speech [3]. Generally, however, existing simulations replicate gross structural changes in speech (e.g. phoneme repetition or substitution). Here, we present a proof-of-concept for a new approach. Using the Task Dynamics model of speech production [4]–[6] , we modify parameters regulating speech motor control and execution to assess their ability to replicate established gradient patterns of kinematic impairment in a specific speech motor disorder—amyotrophic lateral sclerosis (ALS). Critically, this analysis-by-synthesis approach allows us to test specific hypotheses about control deficits in particular speech disorders rather than relying only on inferences drawn from kinematic patterns.

2. SPEECH IMPAIRMENTS IN AMYOTROPHIC LATERAL SCLEROSIS

ALS is a neurodegenerative disease that results in damage to motor neurons, the nerve cells that convey information from the brain to the muscles. ALS can affect both upper and lower motor neurons, with both spinal and bulbar pathways being affected eventually. This damage results in muscle weakness and loss of motor control broadly, typically resulting in spastic-flaccid dysarthria [1], [7]. Because the impairment in motor control in ALS – impairments in descending pathways from cortex to the orofacial musculature — is relatively simple compared to disorders that involve subcortical structures or cortical damage outside motor neuron projections to the periphery, it

provides a good test case for assessing neurogenic control impairments in speech through the use of computational models.

Although the general neurologic impairment in ALS is well-established, precisely how degeneration of motor neurons affects speech motor control is still unclear. Looking at the broader limb control literature, one intriguing possibility is that damage to motor neurons may results in increased motor variability even when there is no major deficit in muscular strength. Motor noise—variability in muscular contraction—is typically thought to be signal-dependent, such that generating larger forces also results in more variable forces. While this relationship generally holds, there is evidence from limb control that noise depends not only on total force but also on muscular strength [8]. For a given force output, stronger muscles (which recruit more motor neurons firing at lower rates) produce *less* variability than weaker muscles (where fewer motor neurons must fire faster to produce equivalent force output). It follows, then, that loss of motor neurons in ALS would result not only in weakness but also increased variability when producing equivalent forces to healthy controls. Indeed, this seems to be the case when generating isometric force with the fingers, at least when lower motor neurons are affected [9]. Although such direct evidence for this hypothesized relationship is not available in speech, consistent with this idea, the severity of speech symptoms in individuals with ALS is predicted by the variability in kinematic acceleration of the speech articulators [10]. Together, these results suggest that increased motor noise (as measured through the acceleration signal in speech kinematics) may be the driving force behind the high degree of variability in this population.

Given this hypothesis, it would be expected that speech kinematic variability would be higher in ALS patients than in neurotypical controls. There is limited evidence that this may indeed be the case, at least for some speakers [11], [12]. Although speakers with mild ALS produced slightly *less* variable movement patterns than neurotypical controls when speaking at a comfortable rate, their habitual speech rates were slower than those of control speakers. When speakers with ALS were prompted to speak as fast as possible, their speech rates were comparable to the control speakers' habitual rate. However, their speech movement patterns during the fast rate were significantly more variable than those produced by controls during their habitual rate. This finding suggests that speakers with ALS may produce speech at an abnormally slow, habitual rate as a strategy to maximize articulatory control. Indeed, the slowing-down of the speech rate has been frequently discussed as a compensation mechanism to decrease spatial variability across various motor speech impairments as well as aging [13].

However, previous findings on movement pattern variability must be taken with some caution as 1) they used a combined metric of both spatial and temporal variability, the *Spatiotemporal Index* (STI), rather than focusing on the purely spatial effects which could be directly attributed to changes in motor unit recruitment without other changes in control, 2) the sample sizes and number of repetitions from each speaker were relatively small given the heterogenous and idiosyncratic progression of ALS, 3) the increase in STI with speech rate in ALS speakers has not been consistently found [12].

3. MODELLING CONTROL IMPAIRMENTS IN ALS

In the Task Dynamics model [4], the functional unit of control is the speech gesture—movements to create linguistically relevant constrictions of specific sizes at specific locations in the vocal tract. Each gesture is modelled as a point attractor with 2nd order dynamics (essentially, a mass-spring system) that controls the movements of the speech articulators. This model relies on a hierarchical feedback control architecture, where, at each time step, the current positions and velocities of the speech articulators (mobility state) are first used to estimate the size of constrictions (task state). The current task state is then compared against the desired task state (gestural goal) to generate a task-state error, which is fed into the mass-spring dynamic equations to generate a desired change in task state (represented as a velocity given the 2nd order dynamics). This desired task-state velocity is then converted to a "motor command" (mobility-space acceleration) to move the speech articulators in the system.

Signal-dependent motor noise is not currently implemented in the computational implementation of Task Dynamics, Task Dynamics Application (TADA, [5]). However, it is a relatively straightforward modification to add noise to the mobility-space acceleration signal—the output of the control algorithm that is the final movement command to the plant. Essentially, the acceleration signal at each time step is multiplied by a weighting term drawn from a normal distribution, $N(0,1)$:

$$(1) \quad \ddot{u} = J^{-1}(u)\left[\ddot{x} - \dot{J}(u,\dot{u})(u)\right]\left[1 + \alpha N(0,1)\right]$$

where α is the weighting term that determines the amount of signal dependent noise, $\ddot{u}$ is the mobility-space acceleration or motor command, u is the current mobility-space position, $\dot{u}$ the current mobility space velocity, J is the Jacobean mapping

changes in mobility space to changes in task space, J^{-1} is the pseudo-inverse of this Jacobean, and $\dot{J}$ is its derivative. When α is set to 0, this equation yields the same output as in the standard model.

Our hypothesis is that increased spatial kinematic variability in ALS compared to neurologically healthy control when speaking at the same absolute rate is due to increased signal-dependent noise subsequent to motor neuron generation (due to compensation mechanisms in impaired speech, the fast speech rate of ALS speakers is comparable to the habitual rate of controls). We tested this by evaluating how changes in the weighting term α in Eq. 1 affect the variability of the kinematic output of the TADA model. We varied α from 0 to 0.4 in steps of 0.05. In order to account for speech rate changes, we multiplied the frequency of the planning oscillators which determine the time periods of gestural activation by r = [0.7 0.85 1.15 1.3], where setting r to 1 is equal to the default speech rate of the model, values of r greater than 1 speed up the speech rate, and values of r smaller than 1 slow the speech rate. Our preliminary modelling suggested that modifying this parameter alone does not substantially affect the movement kinematics, as the dynamics of the gestures themselves are still stable across speech rates. It has been observed that the stiffness of speech movements (estimated via the ratio of peak velocity relative to displacement) consistently increases as movement durations become shorter, despite variability across individuals in whether increased speech rate is accomplished though faster movements, smaller movements, or both [14]. To replicate this in TADA, the stiffness parameter of each gesture was multiplied by r, such that fast speech rates resulted in both faster and shorter movements. This change was sufficient to replicate the basic pattern observed in human data, where shorter movements with equivalent amplitudes as longer movements have higher peak velocities.

For each pair of parameters r and α, we generated 20 simulations of the phrase "Buy Bobby a puppy", which has been commonly used to measure STI in both healthy and disordered speech [11], [15], [16]. To measure spatial variability, we analysed the resulting trajectories of the lower lip vertical position parameter in the CASY plant model used in TADA [17], as it gives a good comparison to human data, where the vertical position of the lower lip is often used. Other parameters of the CASY model (jaw angle, tongue body angle and length from the temporomandibular joint) are less directly comparable to human kinematics. No prosodic modulation was applied, resulting in output where each syllable in the phrase as equal duration.

Because we are particularly interested in the effects of signal-dependent motor noise on spatial variability, the frequently-used STI metric is inappropriate given its confounding of both spatial and temporal variability. Instead, we used a metric of the residual spatial variability after temporal variability has been accounted for through non-linear time warping using functional data analysis [18]–[20]. Briefly, this method first linearly normalizes articulatory trajectories to the span [0,1] after each trajectory has been converted from raw positions to z-scores. Trajectories are then transformed into a set of basis functions and are non-linearly aligned to a common reference trajectory, here the barycenter average of all trajectories [21]. The average variability of the aligned trajectories can then be taken as a measure of purely spatial variability, the *Index of Amplitude Variability* (IAV, [18]). Note that the time-warping functions can be used separately to generate a measure of temporal variability, the *Index of Phase Variability* (IPV).

4. RESULTS

To first validate the IAV metric, tongue body trajectories from one control speaker and one speaker with ALS with moderate dysarthria severity were used to calculate both STI and IAV (Figure 1). Each speaker produced 7 (control) or 9 (ALS) repetitions of the phrase "Buy Kaia a kite". The vertical position of the tongue body was tracked using electromagnetic articulography (EMA), trimmed from the time of maximum vertical position of the lower lip during the initial [b] in "buy" to time of the maximum vertical position of the tongue tip during the [t] in "kite". The tongue body was chosen because the signal from the jaw sensor had relatively little movement in the control participant. Results show that the speaker with ALS has a higher STI value than the control speaker, indicating a higher degree of variability. FDA analyses suggest that IAV (and IPV) is higher as well.

Simulated lower lip trajectories for one level of motor noise (α = 0.2) at the default speech rate (r = 1) are plotted in Figure 2 (top panel). Figure 2 additionally shows results from all simulations for STI (middle panel) and IAV (bottom panel). To replicate the relatively low number of kinematic trajectories typically used in studies with patients (typically 5-10), for each pair of α and r values, 8 out of the 20 trajectories were randomly drawn without replacement and used to calculate STI and IAV. This procedure was repeated 20 times for each pair of values to generate a mean and standard deviation.

For both STI and IAV, variability increases with both motor noise and speech rate. These simulations demonstrate that there is a trade-off between speech

rate and the amount of signal dependent noise on the resulting spatial variability. Critically, the increase in spatial variability that would be expected due to higher levels of signal-dependent noise can be countered by reducing the speech rate, which results in slower movements and so, less motor noise. For example, the IAV at a speech rate of 1.15 is 0.095 when α = 0.15 and 0.125 when α = 0.3. However, slowing the rate down to 0.85 results in an IAV of 0.066 when α = 0.3, less than the IAV for α = 0.15 at the original rate. This pattern qualitatively replicates the data in [11], where the ALS patients had lower variability and slower rates than controls, but higher variability than controls when pushed to speak at similar rates.

5. CONCLUSIONS

Our simulations have shown that increased motor noise due to muscular weakness in ALS may provide an explanation for what has been a curious pattern: at least in one study, speakers with ALS show reduced variability compared to controls at their (abnormally slow) habitual rate, but increased variability when speaking at the same rate as controls. However, this proof-of-concept relies on several as yet untested assumptions. First, existing studies report only STI, which conflates spatial and temporal variability. More detailed analyses using metrics such as IAV and IPV are needed to separate spatial and temporal changes in motor speech disorders. Second, increased variability at a fast speech rate in ALS has not been consistently found. Studies with larger samples, and more trials per participant are needed (especially given the high variability found in our simulations using 8 trajectories). Finally, our simulations of rate assume that individuals with ALS show the same increase in stiffness as duration decreases as controls—this assumption should be examined using speech kinematic data from this population. In sum, our results here suggest that analysis-by-synthesis is a promising method for testing hypotheses about the control deficits underlying observable impairments in motor speech disorders.

6. ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation (DFG) in the collaborative research center SFB1252 Prominence in Language (Project-ID 281511265) and project RO 6767/1-1 (Walter Benjamin program) as well as NIH NIDCD grants R01 DC017091 and R01DC019648.

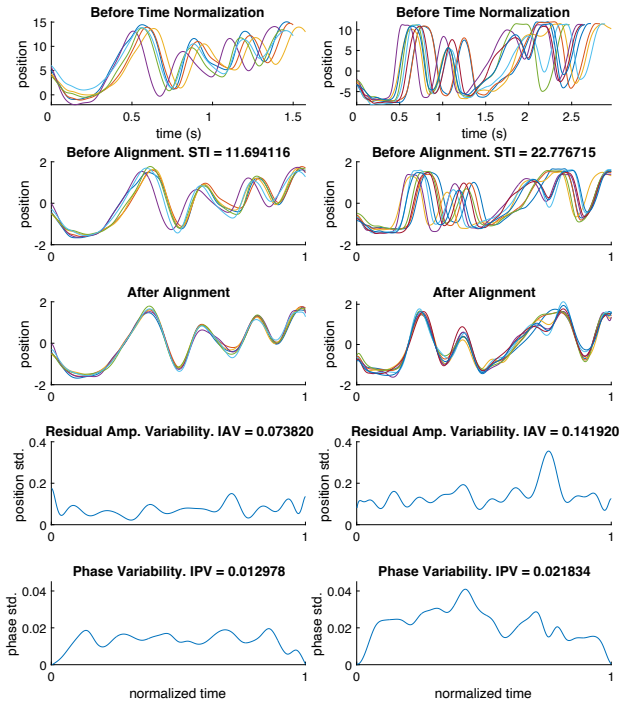

Figure 1: Variability in tongue body movement from a control speaker (left) and a speaker with ALS (right). Top row: original tongue body trajectories, measured with EMA. Second row: z-score and linearly time-alignment trajectories used both as first step of FDA process and to calculate the STI. Third row: non-linearly aligned trajectories. Fourth row: residual spatial variability after time-alignment, used to calculate the IAV. Bottom row: variability of time-warping functions generated in non-linear time alignment process, used to calculate the IPV.

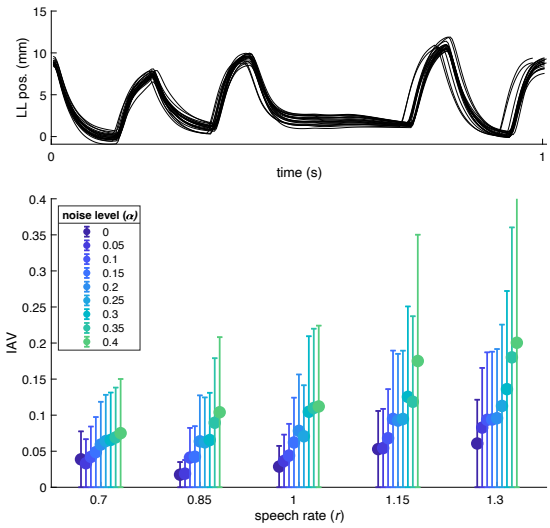

Figure 2: Top panel: Example simulated kinematic trajectories showing the vertical position of the lower lip during 20 repetitions of the phrase "Buy bobby a puppy", trimmed from the time of maximum vertical position of the lower lip during the [b] in "Buy" to the time of maximum lower lip position during the second [p] in "puppy" (α = 0.1, r = 1). Bottom panel: Mean and standard deviation of IAV across noise levels (colors) and speech rates (on x axis). IAV generally increases with both motor noise and speech rate. However, note the large variability in IAV measures with only 8 repetitions.

7. REFERENCES

[1] F. L. Darley, A. E. Aronson, and J. R. Brown, *Motor speech disorders*. Philadelphia: Saunders, 1975.

[2] O. Civier, D. Bullock, L. Max, and F. H. Guenther, "Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation," *Brain Lang.*, vol. 126, no. 3, pp. 263–278, Sep. 2013, doi: 10.1016/j.bandl.2013.05.016.

[3] H. E. Miller and F. H. Guenther, "Modelling speech motor programming and apraxia of speech in the DIVA/GODIVA neurocomputational framework," *Aphasiology*, vol. 0, no. 0, pp. 1–18, May 2020, doi: 10.1080/02687038.2020.1765307.

[4] E. Saltzman and K. G. Munhall, "A Dynamical Approach to Gestural Patterning in Speech Production," *Ecol. Psychol.*, vol. 1, no. 4, pp. 333–382, 1989.

[5] H. Nam, L. Goldstein, E. Saltzman, and D. Byrd, "TADA: An enhanced, portable Task Dynamics model in MATLAB," *J. Acoust. Soc. Am.*, vol. 115, p. 2430, 2004.

[6] H. Nam, L. Goldstein, and E. Saltzman, "Self-organization of syllable structure: a coupled oscillator model," in *Approaches to phonological complexity*, F. Pellegrino, E. Marisco, I. Chitoran, and C. Coupé, Eds. Berlin/New York: Mouton de Gruyter, 2009, pp. 299–328.

[7] J. R. Duffy, *Motor speech disorders: substrates, differential diagnosis, and management*, Third edition. St. Louis, Missouri: Elsevier, 2013.

[8] A. F. de C. Hamilton, K. E. Jones, and D. M. Wolpert, "The scaling of motor noise with muscle strength and motor unit number in humans," *Exp. Brain Res.*, vol. 157, no. 4, pp. 417–430, Aug. 2004, doi: 10.1007/s00221-004-1856-7.

[9] P. K. Kasi *et al.*, "Motor unit firing characteristics in patients with amyotrophic lateral sclerosis," in *2009 IEEE 35th Annual Northeast Bioengineering Conference*, Apr. 2009, pp. 1–2. doi: 10.1109/NEBC.2009.4967680.

[10] K. J. Teplansky *et al.*, "Tongue and Lip Acceleration as a Measure of Speech Decline in Amyotrophic Lateral Sclerosis," *Folia Phoniatr. Logop.*, Jun. 2022, doi: 10.1159/000525514.

[11] A. S. Mefferd, G. L. Pattee, and J. R. Green, "Speaking rate effects on articulatory pattern consistency in talkers with mild ALS," *Clin. Linguist. Phon.*, vol. 28, no. 11, pp. 799–811, Nov. 2014, doi: 10.3109/02699206.2014.908239.

[12] M. Kuruvilla-Dugdale and A. Mefferd, "Spatiotemporal movement variability in ALS: Speaking rate effects on tongue, lower lip, and jaw motor control," *J. Commun. Disord.*, vol. 67, pp. 22–34, May 2017, doi: 10.1016/j.jcomdis.2017.05.002.

[13] T. Thies, A. Hermes, and D. Mücke, "Compensation in Time and Space: Prominence Marking in Aging and Disease," *Languages*, vol. 7, no. 1, p. 21, Jan. 2022, doi: 10.3390/languages7010021.

[14] D. Ostry and K. Munhall, "Control of rate and duration of speech movements," *J. Acoust. Soc. Am.*, 1985.

[15] A. Smith and L. Goffman, "Stability and patterning of speech movement sequences in children and adults," *J. Speech Lang. Hear. Res.*, vol. 41, pp. 18–30, 1998.

[16] A. Smith, L. Goffman, HowardN. Zelaznik, G. Ying, and C. McGillem, "Spatiotemporal stability and patterning of speech movement sequences," *Exp. Brain Res.*, vol. 104, no. 3, Jun. 1995, doi: 10.1007/BF00231983.

[17] P. Rubin, E. Saltzman, L. Goldstein, R. McGowan, M. T. Tiede, and C. P. Browman, "CASY and extensions to the task-dynamic model," in *Proceedings of the 4th Speech Production Seminar*, 1996.

[18] J. C. Lucero, "Comparison of measures of variability of speech movement trajectories using synthetic records.," *J Speech Lang Hear Res*, vol. 48, no. 2, pp. 336–44, 2005, doi: 10.1044/1092-4388(2005/023).

[19] J. Lucero, G. K. Munhall, V. L. Gracco, and O. J. Ramsay, "On the registration of time and the patterning of speech movements," *J. Speech Lang. Hear. Res.*, vol. 40, pp. 1111–1117, 1997.

[20] O. J. Ramsay, G. K. Munhall, V. L. Gracco, and D. Ostry, "Functional data analyses of lip motion," *J. Acoust. Soc. Am.*, vol. 99, pp. 3718–3727, 1996.

[21] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, Mar. 2011, doi: 10.1016/j.patcog.2010.09.013.