A Speech-Based Prognostic Model for Dysarthria Progression in ALS

Gabriela Stegmann[1,2], **Sherman Charles**[2,3], **Julie Liss**[1,2], **Jeremy Shefner**[4], **Seward Rutkove**[5], **Visar Berisha**[1,2]

[1.]Arizona State University, Phoenix, AZ

[2.]Aural Analytics, Scottsdale, AZ

[3.]University of Arizona, Tucson, AZ

[4.]Barrow Neurological Institute, Phoenix, AZ

[5.]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

Abstract

Objective: We demonstrated that it was possible to predict ALS patients' degree of future speech impairment based on past data. We used longitudinal data from two ALS studies where participants recorded their speech on a daily or weekly basis and provided ALSFRS-R speech subscores on a weekly or quarterly basis (quarter-annually).

Methods: Using their speech recordings, we measured articulatory precision (a measure of crispness of pronunciation) using an algorithm that analyzed the acoustic signal of each phoneme in the words produced. First, we established the analytical and clinical validity of the measure of articulatory precision, showing that the measure correlated with perceptual ratings of articulatory precision (r = .9). Second, using articulatory precision from speech samples from each participant collected over a 45-90 day model calibration period, we showed it was possible to predict articulatory precision 30-90 days after the last day of the model calibration period. Finally, we showed that the predicted articulatory precision scores mapped onto ALSFRS-R speech subscores.

Results: the mean absolute error was as low as 4% for articulatory precision and 14% for ALSFRS-R speech subscores relative to the total range of their respective scales.

Conclusion: Our results demonstrated that a subject-specific prognostic model for speech predicts future articulatory precision and ALSFRS-R speech values accurately.

Keywords

Prognosis; speech; articulatory precision; longitudinal; Amyotrophic Lateral Sclerosis (ALS)

Corresponding author: Gabriela Stegmann, Department of Speech and Hearing Science, Lattie F. Coor Hall, 975 S Myrtle Ave, Tempe, AZ 85281, Gabriela.Stegmann@asu.edu.

Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that selectively affects motor neurons in the brain and spinal cord and causes paresis of voluntary muscles [1]. There is great interest in devising clinical tools that accurately predict the rate of decline in ALS. In this study, we developed and evaluated a prognostic model for predicting rate of speech decline in ALS patients.

Loss of speech takes a substantial toll on quality of life and is associated with depression and social isolation [2]. Predicting how quickly a patient's speech will decline is not trivial due in part to heterogeneity of speech symptoms across patients and phenotypes [3;4]. Estimating the amount of time for which spoken communication will be functional is important for patients and their families to know, as well as to provide a timeline for the introduction of assistive and alternative communication technology (AAC). Beyond its clinical utility, a prognostic model that relies on automated speech analysis can improve existing prognosis models of survival. Previous work highlights the ALSFRS-R speech and bulbar subscales as important variables for predicting survival in ALS [5; 6]. However, the ALSFRS-R has been shown to lack sensitivity to small changes in speech decline [2]. An automated and validated objective speech measure that can be collected remotely and at more frequent intervals can augment these predictive models by more precisely characterizing speech decline. Finally, an accurate estimate of the rate of disease decline could be used when designing clinical trials; for example, identifying participants with faster decline.

We developed a subject-specific prognostic model that uses the patient's historical data to predict the patient's speech on a future date. The model is driven by the hypothesis that when a patient's speech is measured frequently, objectively, and reliably over a period of time, it is possible to accurately estimate the rate of decline. We chose *articulatory precision* as the primary speech measure because it has been found to map on to speech severity both cross-sectionally and longitudinally [7; 8]. Articulatory precision was measured algorithmically from recorded speech samples.

Several speech analysis tools and software have been utilized in ALS research, including open-source software such as openSMILE [9], Talk2me [10], and Praat [11]. However, most of these tools were developed for purposes unrelated to disease tracking (e.g., emotion recognition, assessing second language proficiency) [12; 13]. Studies [14] have found most of these speech features to have not undergone rigorous validation in a clinical context and have low reliability, which makes them suboptimal for tracking disease. In our study, before building the prognostic model, we first validated articulatory precision following the V3 validation framework described as best practice by the Digital Medicine Society (DiMe) [15]. Then we demonstrated the feasibility and accuracy of using it in a prognostic model for predicting loss of speech.

Materials and Methods

Participants

Two observational, longitudinal studies conducted remotely of ALS participants were used. Study 1 sample consisted of participants with a diagnosis of ALS within five years and no additional known neurological disorder. Participants were requested to provide daily speech samples and weekly ALSFRS-R scores (self-assessed) for the first three months, and for the next 6 months the speech samples were requested twice weekly. More details in this study can be found in [16]. Study 2 sample consisted of participants with definite, probable, or possible ALS according to the El Escorial Criteria, including 6 participants with symptoms of frontotemporal cognitive dysfunction noted either by themselves or a caregiver. The cognitively unimpaired and impaired participants did not differ significantly in the ALSFRS-R speech subscores. Participants were requested to provide weekly speech samples (collected at home) and quarterly ALSFRS-R scores (collected in clinic) for up to 12 months. For both studies, approval was granted by the Institutional Review Board of St. Joseph's Hospital and Medical Center, and all subjects signed an informed consent document. Participants were allowed to receive assistance from their caregivers if needed. In our current study, each analysis used all the available data.

Measurements

Articulatory Precision—Articulatory precision is a measure of how accurately the participant pronounced each phoneme in the sentences. This tracks with previous accounts of the effects of ALS on speech motor control, such as reduction in consonant precision [17], centralization of vowels [18], speed of articulatory movements [19], speaking rate [20], and impaired vocal quality [21]. We measured it on recorded speech samples collected via a smartphone app [16; 7]. The algorithm is an extension of the work in [22] and [23], which compares the acoustics of the speech produced in the sample to the expected acoustics derived from a large normative database. A detailed description of the algorithm is provided in [7]. The values were scaled to be between 0 and 10, where high values indicated high articulatory precision and low values indicated low articulatory precision.

ALSFRS-R Speech—Participants were instructed to provide weekly self-assessed (Study 1) or quarterly in-clinic (Study 2) ALSFRS-R speech subscores. The ALSFRS-R speech subscore was used to quantify ALS speech severity.

Clinical ratings—Clinical ratings were provided by a trained speech scientist after the data collection. The articulatory precision for each token was rated on a 0-4 scale with 0 as typical and 4 as severe impairment [24]. The ratings focused on the production of consonants. The rater was blinded to the algorithmic measurements of articulatory precision. The articulatory precision clinical ratings were used to conduct the analytical validation of the algorithmically measured articulatory precision.

Statistical Analyses

Validation of articulatory precision—We used the V3 framework described by DiME [15] to validate articulatory precision. For the analytical validation, we showed that the

algorithmic measurement of articulatory precision corresponded with the human perception of articulatory precision. A sample of 628 sets of sentences from 22 participants were selected to cover a wide range of speech severity. The speech samples were rated, and the correlation between the algorithmic and clinically rated articulatory precision values was computed. For the clinical validation, we demonstrated its clinical usefulness by showing that it had high test-retest reliability, it correlated with the ALSFRS-R speech subscores, and it declined over time. We demonstrated that it declined longitudinally using a mixed-effects growth curve model [25]. Any correlations computed were based on [26], given that participants had repeated measurements.

Development and Evaluation of the Prognostic Model—The proposed model captured the speech measure values from a given individual patient over a period of time which we call the *model calibration period*. The speech samples collected during the model calibration period were used to predict speech after a *waiting period* after the last day in the model calibration period, as illustrated in Figure 1. This was a subject-specific prognostic model to predict the degree of speech impairment in the future. To facilitate clinical interpretation, a second model was built to estimate the ALSFRS-R speech subscore using the predicted speech value as the predictor. By combining the two models, we first used the prognostic model to predict the speech measure at future time points, and then we used these predicted scores as the input to obtain the corresponding estimated ALSFRS-R speech subscores. The Supplementary Materials provide a detailed description of the model-building process and the cross-validation used to avoid overfitting.

We evaluated model calibration periods of 45, 60, 75, and 90 days and waiting periods of 30, 45, 60, 75, and 90 days after the last day in the model calibration period. A minimum of 10 observations were required during the model calibration period to create a prediction. Longer waiting periods were not possible due to lack of sufficient data resulting from patient attrition and irregular compliance throughout the study.

To evaluate the prediction accuracy, we compared the observed articulatory precision values and the ALSFRS-R speech subscores with the predicted values estimated based on the prognosis model by computing the mean absolute error (MAE), MAE% (MAE divided by the range of the corresponding scale), and the correlation [26] between the two.

All analyses were conducted using R Programming. The mixed-effects models were performed using the lme4 [27] and nlme [28] packages, which are two widely used R packages.

Results

Sample Description

The sample consisted of 110 ALS participants (84 from Study 1; 26 from Study 2) and a total of 7,637 speech sessions (6750 from Study 1; 887 from Study 2). Of those, 83 participants provided a total of 537 ALSFRS-R speech subscores (482 from Study 1; 55 from Study 2). Table 1 shows the description of the sample.

To help the reader visualize the data, examples of speech trajectories of five participants are shown in Figure 2. In each plot, the *x*-axis is the number of days since the beginning of the data collection; the *y*-axis of the top plots shows the algorithmic measure of articulatory precision; the *y*-axis of the bottom plots show the ALSFRS-R speech subscores; each column of plots shows each participant. The participants in the sample were chosen to cover a wide range of speech impairments according to the ALSFRS-R speech subscores.

Validation of articulatory precision

In the analytical and clinical validation study, we found that articulatory precision correlated with clinical ratings (r = .90), was reliable (ICC = .97), correlated with ALSFRS-R speech (r = .82), it declined over time with disease progression, and the rate of change correlated moderately [29] with the rate of decline of ALSFRS-R speech (r = .37) and ALSFRS-R bulbar (r = .41) subscales. For the comparison of the rates of decline, only participants who had at least 8 observations and over 60 days of data were considered (n = 28 participants). The Supplementary Materials provide the growth curve parameters from the longitudinal model and plots for each analysis.

Evaluation of prognostic model

Next, we evaluated the performance of the prognostic model. The MAE was between .44 (95% CI .38-.50) and 1.3 (95% CI .7-1.8), the MAE% was between 4% (95% CI 3-5%) and 13% (95% CI 7-18%), and the correlation between the predicted and observed values was between r = .7 and r = .96. Shorter waiting periods corresponded with higher prediction accuracy. The supplementary materials provide figures with the results.

When evaluating the prognostic model based on the predicted values of the ALSFRS-R speech subscores, the MAE was between .55 (95% CI .5-.6) and .81 (95% CI .66-.96), the MAE% was between 14% (95% CI 13-15%) and 19% (95% CI 17-24%), and the correlation between the predicted and observed values was between r = .53 and r = .78.

Discussion

We first conducted a validation study following best practices [15]. Articulatory precision was shown to map onto the perceptual ratings of articulatory precision, was reliable, correlated with ALSFRS-R speech impairment, and declined in ALS patients, thus indicating that it was both accurate and useful as a clinical measure for tracking speech impairment and decline in ALS. Given the strong correlation between articulatory precision and the ALSFRS-R speech subscore, we believe the low correlations between rates of decline are at least in part due to correlation attenuation resulting from measurement error from both measures [30]. Next, we showed that using a model calibration period where participants provided regular speech samples, it was possible to predict the degree of speech impairment both in terms of the articulatory precision values and according to the ALSFRS-R speech up to 90 days after the last model calibration period measurement. This indicates that this tool could be useful for accurately tracking and predicting speech decline in ALS.

We anchored the analyses to the ALSFRS-R speech subscore given its clinical acceptance as a reliable measure of speech function in ALS. However, it is clear from the 5 examples

in Fig. 2 that articulatory precision declines more reliably at the individual participant level than ALSFRS-R. The models predicting future speech impairment using ALSFRS-R speech as the target metric had two sources of accuracy loss: (a) the accuracy loss resulting from using a prior time window of articulatory precision scores to predict future articulatory precision scores, and (b) mapping the articulatory precision scores to ALSFRS-R speech subscores. For (a), the future predicted articulatory precision scores were highly accurate (MAE% from 4% to 13%). This suggests that the majority of the accuracy loss came from mapping the future articulatory precision to the observed ALSFRS-R speech subscores (MAE% from 14% to 20%). A likely reason for this is that the ALSFRS-R speech subscores are not as granular as articulatory precision and they are subjective ratings rather than an objective measure. This suggests that, rather than relying on the ALSFRS-R speech subscore, the articulatory precision score itself is a more reliable measure of speech decline.

Although the results from this study are promising, the current data availability only allowed us to predict only a short period of time into the future (90 days), which limits clinical utility). In addition, it would also be of interest to compare these findings to using the ALSFRS-R bulbar subscale for predicting future impairment; in this study, the ALSFRS-R measures were not collected at the required frequency to conduct this comparison. Having to observe a patient for 3 months is limiting in clinical trials. A future study should use a larger sample size, both in terms of the number of participants and the length of time that the participants are tracked, in order to be able to answer these questions. Finally, a future study could explore similar types of predictions for other outcomes, such as predicting future forced vital capacity using maximum phonation time, since other studies have found these to be related [31].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

[1]. Rowland LP, Shneider NA (2001) Amyotrophic lateral sclerosis. New England Journal of Medicine, 34(22), 1688–1700.

[2]. Yunusova Y, Plowman EK, Green JR, Barnett C, & Bede P (2019). Clinical measures of bulbar dysfunction in ALS. Frontiers in Neurology, 10, 106–106. 10.3389/fneur.2019.00106 [PubMed: 30837936]

[3]. Chiò A, Calvo A, Moglia C, Mazzini L, Mora G, & PARALS study group. (2011). Phenotypic heterogeneity of amyotrophic lateral sclerosis: A population based study. Journal of Neurology, Neurosurgery & Psychiatry, 82(7), 740–746. 10.1136/jnnp.2010.235952 [PubMed: 21402743]

[4]. Stipancic KL, Yunusova Y, Campbell TF, Wang J, Berry JD, & Green JR (2021). Two distinct clinical phenotypes of bulbar motor impairment in Amyotrophic Lateral Sclerosis. Frontiers in Neurology, 12, 664713. 10.3389/fneur.2021.664713 [PubMed: 34220673]

[5]. Rascovsky K, Xie S, Boller A, Han X, McCluskey L, Elman L, & Grossman M (2014). Subscales of the ALS Functional Rating Scale (ALSFRS-R) as determinants of survival in amyotrophic lateral sclerosis (ALS). Neurology, 82(10 supplement)

[6]. Westeneng H-J, Debray T, Visser A, van Eijk R, Rooney J, Calvo A, Martin S McDermott CJ, Thompson AG, Pinto S, Kobeleva X, Rosenbohm A, Stubendorff B, Sommer H, Middelkoop BM, Dekker AM, van Vugt J, van Rheenen W, Vajda A, Heverin M, …, van den Berg L (2018). Prognosis for patients with amyotrophic lateral sclerosis: Development and validation of a personalised prediction model. Lancet Neurology, 17(5).

[7]. Stegmann G, Hahn S, Liss J, Shefner J, Rutkove S, Shelton K, Duncan CJ, Berisha V (2020). Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis. Nature Digital Medicine, 3(1), 1–5.

[8]. Borrie S, Wynn C, Berisha V, & Barrett T (2022). From speech acoustics to communicative participation in dysarthria: Toward a causal framework. Journal of Speech, Language, and Hearing Research, 65, 405–418.

[9]. Eyben F, Wöllmer M, Schuller B (2010). OpenSMILE: the Munich versatile and fast open-source audio feature extractor. Proceedings of the International Conference on Multimedia. Florence, Italy. New York: ACM Press.

[10]. Komeili M, Pou-Prom C, Liaqat D, Fraser KC, Yancheva M, Rudzicz F. Talk2Me: Automated linguistic data collection for personal assessment. Greatorex Riches N, ed. PLoS ONE. 2019;14(3):e0212342. doi:10.1371/journal.pone.0212342 [PubMed: 30917120]

[11]. Boersma P (2001). Praat, a system for doing phonetics by computer. Glot Int.,5, 341–5

[12]. Eyben F, Huber B, Marchi E, Schuller D, Schuller B (2015). Real-time robust recognition of speakers' emotions and characteristics on mobile platforms. International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 778–80.

[13]. Lu X (2012). The Relationship of lexical richness to the quality of ESL learners' oral narratives. Mod Lang J, 96(2), 190–208.

[14]. Stegmann GM, Hahn S, Liss J, Shefner J, Rutkove SB, Kawabata K, Bhandari S, Shelton K, Duncan CJ, & Berisha V (2020). Repeatability of commonly used speech and language features for clinical applications. Digital Biomarkers, 4(3), 109–122. 10.1159/000511671 [PubMed: 33442573]

[15]. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, Godfrey A., Godino JG, Gujar N, Izmailova E, Manta C, Peterson B, Vandendriessche B, Wood WA, Wang KW, & Dunn J (2020) Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). NPJ Digital Medicine, 3(1):55. 10.1038/s41746-020-0260-4. [PubMed: 32337371]

[16]. Rutkove SB, Qi K, Shelton K, Liss J, Berisha V, & Shefner J. (2019) ALS longitudinal studies with frequent data collection at home: study design and baseline data. Amyotroph. Lateral Scler. Frontotemporal Degener. 20,61–67. [PubMed: 30486680]

[17]. Kent RD, Sufit RL, Rosenbek JC, Kent JF, Weismer G, Martin RE, Brooks BR (1991) Speech deterioration in amyotrophic lateral sclerosis: A case study. Journal of Speech and Hearing Research, 34:1269–1275. [PubMed: 1787708]

[18]. Sapir S, Ramig LO, Spielman JL, & Fox C (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. Journal of Speech, Language, and Hearing Research, 53(1), 114–125. 10.1044/1092-4388(2009/08-0184)

[19]. Green JR, Yunusova Y, Kuruvilla MS, Wang J, Pattee GL, Synhorst L, Zinman L, & Berry JD (2013). Bulbar and speech motor assessment in ALS: Challenges and future directions. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 14(7–8), 494–500. 10.3109/21678421.2013.817585 [PubMed: 23898888]

[20]. Ball LJ, Beukelman DR, & Pattee GL (2002). Timing of speech deterioration in people with amyotrophic lateral sclerosis. Journal of Medical Speech-Language Pathology, 10(4), 231–235.

[21]. Ramig LO, Scherer RC, Klasner ER, Titze IR, & Horii Y (1990). Acoustic analysis of voice in Amyotrophic Lateral Sclerosis: A longitudinal case study. Journal of Speech and Hearing Disorders, 55(1), 2–14. 10.1044/jshd.5501.02 [PubMed: 2299836]

[22]. Sohn J, Kim N & Sung W (1999) A statistical model-based voice activity detection. IEEE Signal Process. Lett, 6,1–3.

[23]. Tu M, Grabek A, Liss J, & Berisha V (2018). Investigating the role of L1 automatic pronunciation evaluation of L2 speech. Interspeech.

[24]. Duffy JR (2020). Examination of motor speech disorders. Motor speech disorders: Substrates, differential diagnosis, and management, 57–89.

[25]. Grimm KJ, Ram N & Estabrook R (2017) Growth Modeling: Structural Equation and Multilevel Modeling Approaches. New York: Guilford.

[26]. Lorah J (2018) Effect size measures for multilevel models: definition, interpretation, and TIMSS example. Large Scale Assess Educ., 6, 1–11.

[27]. Bates D, Maechler M, Bolker B & Walker S (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw, 67,1–48.

[28]. Pinheiro J, Bates D, DebRoy S, Sarkar D & R Core Team. (2021). nlme: Linear and Nonlinear Mixed Effects Models. https://CRAN.R-project.org/package=nlme.

[29]. Cohen J (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed). L. Erlbaum Associates.

[30]. Crocker L, & Algina J (1986). Introduction to Classical and Modern Test Theory. Fort Worth, TX: Holt, Rinehart and Winston.

[31]. Stegmann G, Hahn S, Duncan CJ, Rutkove S, Liss J, Shefner J, Berisha V (2021). Estimation of forced vital capacity using speech acoustics in patients with ALS. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 22(sup1), 14–21. [PubMed: 34348537]
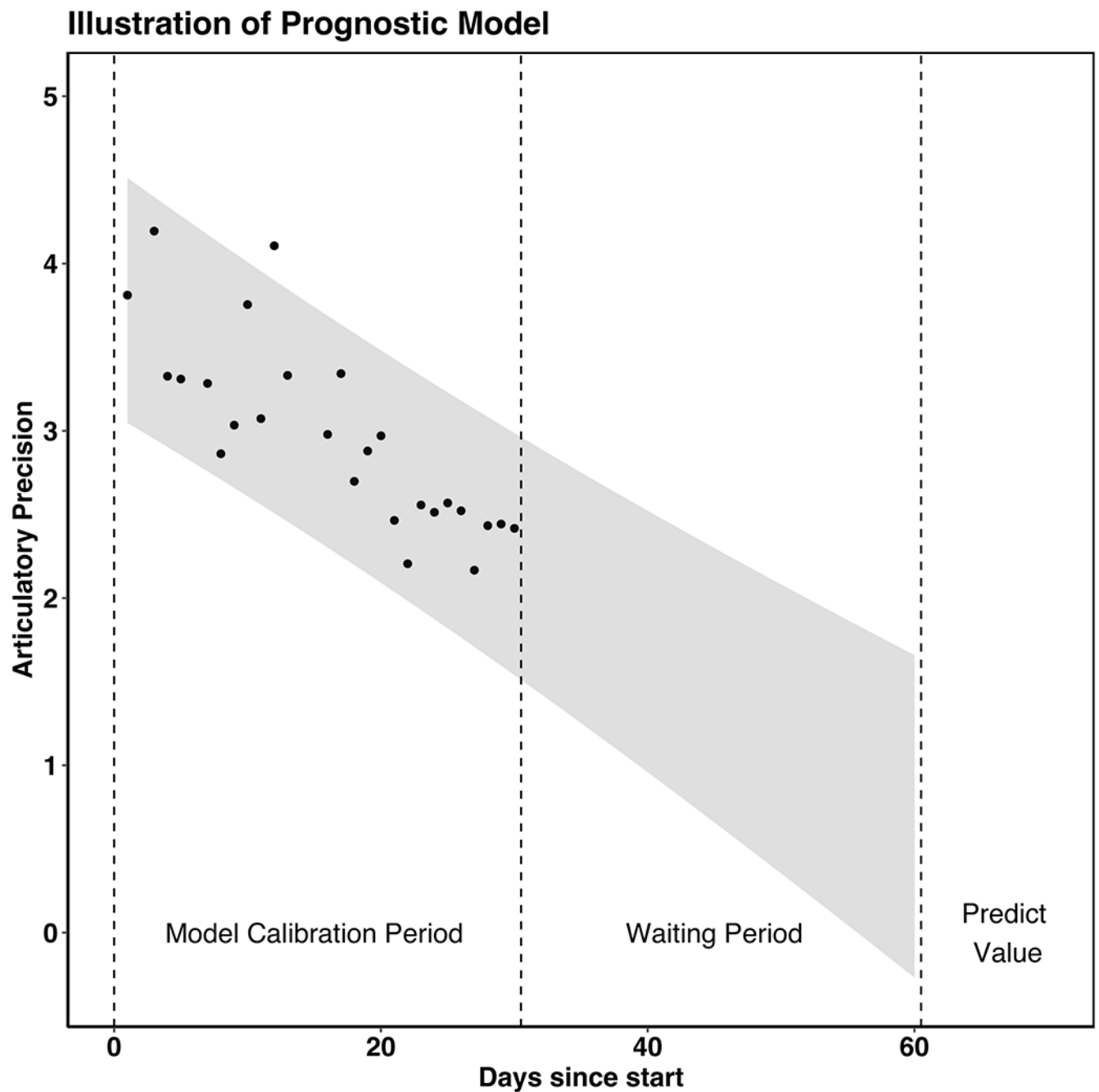
Illustration of Prognostic Model

Figure 1.
Illustration of the prognostic model. This graph illustrates how the prognostic model works on a given patient. The x-axis shows the number of days since the beginning of data collection. The y-axis shows the articulatory precision values. Each point shows the articulatory precision value obtained each day that the participant provided a speech sample. Using the articulatory precision values during the model calibration period, a model is built to predict a future articulatory precision value.
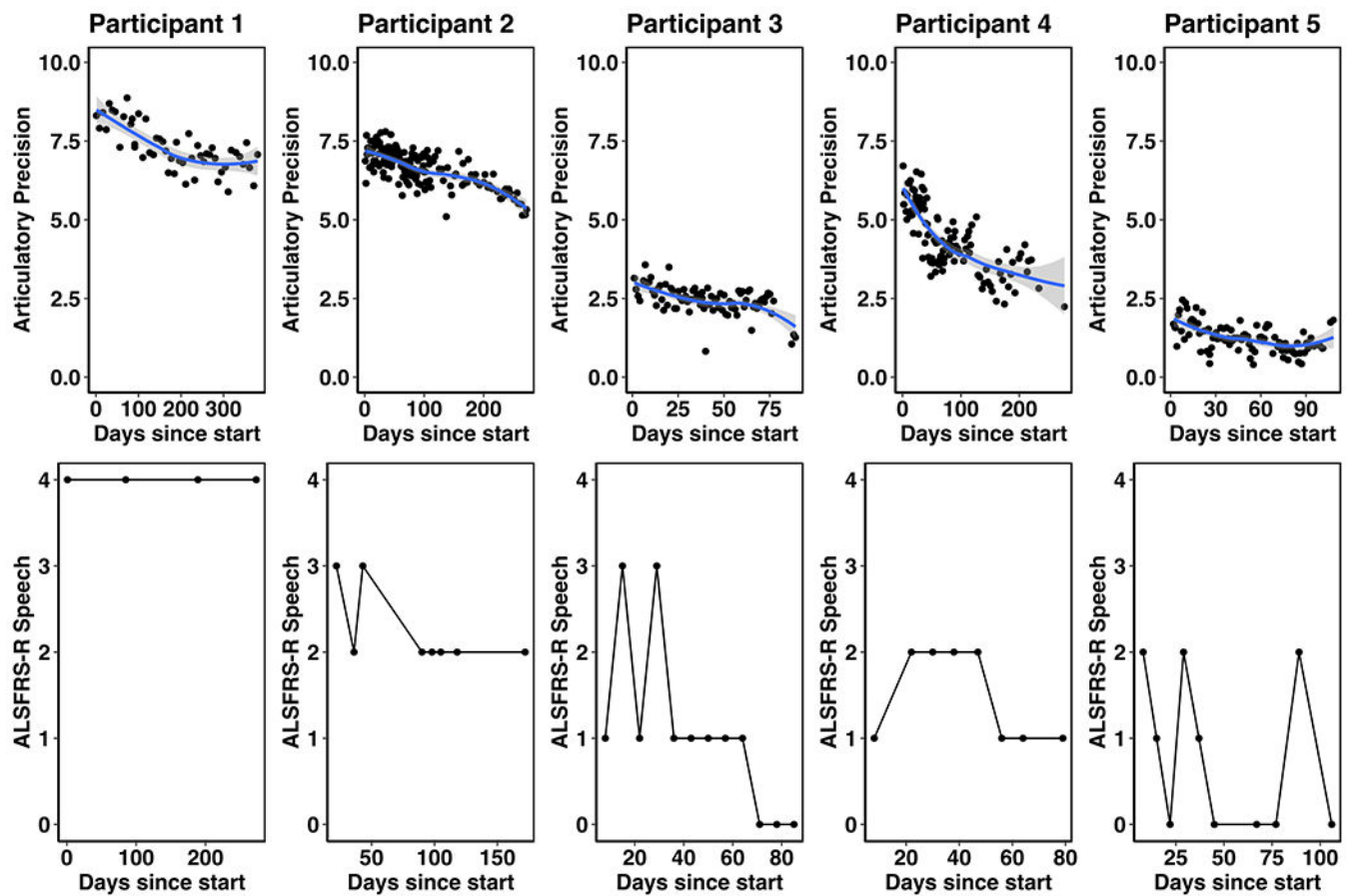
Figure 2.
Individual participants' articulatory precision and ALSFRS-R speech subscores as a function of time. Each plot illustrates speech data for a given participant. Each column represents each participant. For each plot, the x-axis represents the number of days since the study enrollment. For the top plots, the y-axis represents the articulatory precision values; for the bottom plots, the y-axis represents the ALSFRS-R speech subscores. Each point represents the value from each speech outcome at each given time point.

Table 1.

Sample description

Variable	Study 1	Study 2
Age	Mean = 59.2; SD = 10.7	Mean = 65.8; SD = 11.7
Gender	32 F; 50 M; 2 Unknown	8 F; 18 M
ALSFRS-R speech subscores	Mean = 2.9; SD = 1.2	Mean = 3.4; SD = .7
Site of Onset	21% Bulbar; 79% Axial/Other	Unknown
Days since symptom onset	Mean = 1,025; SD = 689	Mean = 924; SD = 544
Number of observations per participant	Mean = 80.4; SD = 55.1	Mean = 34.1; SD = 20.3
Length of time in study (in days)	Mean = 165.6; SD = 110.2	Mean = 286.6; SD = 139.3