INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Relation of Automatically Extracted Formant Trajectories with Intelligibility Loss and Speaking Rate Decline in Amyotrophic Lateral Sclerosis

Rachelle L. Horwitz-Martin[1,2], Thomas F. Quatieri[1,2], Adam C. Lammert[1], James R. Williamson[1], Yana Yunusova[3], Elizabeth Godoy[1], Daryush D. Mehta[1,4,5], Jordan R. Green[2,4]

[1] MIT Lincoln Laboratory, Lexington, MA, USA
[2] Speech and Hearing Bioscience and Technology Program, Harvard-MIT, Cambridge, MA, USA
[3] Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada
[4] Massachusetts General Hospital Institute of Health Professions, Boston, MA, USA
[5] MGH Voice Center, Massachusetts General Hospital, Boston, MA, USA

[rachelle.horwitz, quatieri, adam.lammert, jrw, elizabeth.godoy]@mit.ll.edu, yana.yunusova@utoronto.ca, mehta.daryush@mgh.harvard.edu, jgreen2@mghihp.edu

Abstract

Effective monitoring of bulbar disease progression in persons with amyotrophic lateral sclerosis (ALS) requires rapid, objective, automatic assessment of speech loss. The purpose of this work was to identify acoustic features that aid in predicting intelligibility loss and speaking rate decline in individuals with ALS. Features were derived from statistics of the first (F_1) and second (F_2) formant frequency trajectories and their first and second derivatives. Motivated by a possible link between components of formant dynamics and specific articulator movements, these features were also computed for low-pass and high-pass filtered formant trajectories. When compared to clinician-rated intelligibility and speaking rate assessments, F_2 features, particularly mean F_2 speed and a novel feature, mean F_2 acceleration, were most strongly correlated with intelligibility and speaking rate, respectively (Spearman correlations > 0.70, $p < 0.0001$). These features also yielded the best predictions in regression experiments ($r > 0.60$, $p < 0.0001$). Comparable results were achieved using low-pass filtered F_2 trajectory features, with higher correlations and lower prediction errors achieved for speaking rate over intelligibility. These findings suggest information can be exploited in specific frequency components of formant trajectories, with implications for automatic monitoring of ALS.

Index Terms: speech analysis, formant frequencies, disordered speech, amyotrophic lateral sclerosis

1. Introduction

A common symptom of amyotrophic lateral sclerosis (ALS) is reduced intelligibility and speaking rate [1–2]. Previous studies have identified acoustic differences that contribute to speech decline in ALS. Many of these differences result from variations in formant frequencies. A reduced second formant (F_2) trajectory, for example, has been noted in several studies of ALS speech [3–10]. When correlated against intelligibility scores, F_2 slope has Spearman correlations of 0.82 and 0.85 for females and males, respectively, with ALS [5]. Although previous studies have demonstrated potential diagnostic value of formants, to our knowledge, they have not used automatically extracted formant features. Automated analyses are needed by clinicians who are pressed for time during evaluations and who often lack the expertise required to extract formants from prerecorded speech samples. In addition, although the slope of F_2 appears to be a good predictor of intelligibility [3–10], prior work has not examined the predictive value of features extracted from the derivatives of these trajectories. For example, acceleration features may be particularly useful for characterizing impaired articulatory control and dyscoordination due to ALS, because these features may encode important information about the initiation and termination of articulator motion.

Based on observations that speed (magnitude of velocity), duration, and extent of tongue, lip, and jaw movements change as the disease progresses [10–11], the finding that jaw and lip strength is less affected by ALS than that of the tongue [12–13], and the assumption that different speech articulators are assumed to have different velocity profiles, we hypothesize that bulbar motor deterioration will differentially affect frequency content of formant trajectories. In this study, we examine different frequency bands of the formant trajectories using low- and high-pass filters, and also their velocities and accelerations. After extracting statistical features from the trajectories, we evaluate correlations between each of the features and both intelligibility and speaking rate, and use the features to predict the assessment metrics.

2. Data collection and pre-processing

2.1. Data collection

The University of Nebraska-Lincoln, the Massachusetts General Hospital Institute of Health Professions, and the University of Toronto have collected longitudinal data from 123 subjects with ALS. Details of the collection protocols are described in [1] and [14]. We select subjects from this database who met the clinical, linguistic, and literacy criteria described in [15]. A particular session for a given subject is

* This work is sponsored by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract FA8721-05-C-0002, and NIH Grants R01 DC009890 and R01 DC0135470. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

http://dx.doi.org/10.21437/Interspeech.2016-403

included in this study if, during the session, the subject was evaluated for speaking rate and intelligibility, and uttered at least three repetitions of the standard phrase “Buy Bobby a puppy” (BBP) in succession. Additional criteria are an absence of severe clipping, background noise, or speech signal distortion in the recordings. This results in 136 sessions from 34 subjects (16 males, 18 females). The mean and standard deviation of number of sessions attended per subject are 4.00 and 2.96, respectively, and the range spans 1 to 13 sessions.

In the longitudinal database, intelligibility is quantified by the percentage of words that can be accurately identified by a listener, and speaking rate is quantified in terms of words spoken per minute (wpm). Both measures were obtained using the Sentence Intelligibility Test [16], where normal intelligibility is defined to be >97% and severe intelligibility loss is <87% [17]. Figure 1 displays the intelligibility and speaking rate scores of all subjects across the 136 sessions. As can be observed in Figure 1, after speaking rate declines to below 100-120 wpm, intelligibility declines rapidly [1,17,18].

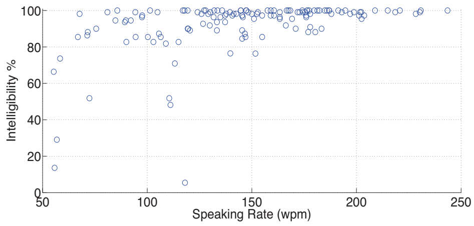

Figure 1: *Intelligibility vs. speaking rate for each session.*

We analyze BBP because it is the most frequently spoken utterance across all sessions that contains dynamic formant motion. Critically, this utterance contains rapid, continuous formant movement in the diphthong [ay] in “buy,” and in the coarticulation of [iy-ax] or [iy-ey] in “Bobby a,” where the lack of a pause leads to a pronounced diphthong.

2.2. Pre-processing

The pre-processing steps consist of converting the stereo files to mono, if necessary; manually identifying and removing the clinicians’ speech from the recordings; downsampling from various initial sampling rates to 16 kHz, if necessary; and applying adaptive wideband noise reduction [15–16].

3. Feature selection and extraction

3.1. Formant estimation

We use a Kalman-based autoregressive modeling and inference algorithm (KARMA) [22] to track F_1 and F_2. The analysis window length is set to 20 ms, with 50% overlap.

To exclude unvoiced regions from feature computation, we utilize the energy-based voicing activity detector (VAD) within KARMA, smooth the output from the VAD with a median filter, and interpolate the formants over the unvoiced regions using the shape-preserving piecewise cubic Hermitian Interpolation function (pchip) in MATLAB. The interpolation yields a continuous trajectory with a continuous derivative.

3.2. Feature extraction

After obtaining the estimated formant tracks for the first two formants, we apply a Chebyshev Type II low-pass filter (LPF) in both forward and reverse directions to the entire utterance of 3 to 11 successive BBPs, resulting in zero-phase LPFed formant trajectories. We obtain the complementary component from the high-pass filter (HPF) by subtracting the LPFed component from the unfiltered formant trajectory.

As depicted in Figure 2, in addition to the unfiltered, LPFed, and HPFed trajectories, we compute the derivative (velocity) and second derivative (acceleration) of each of the three trajectories, leading to 9 representative trajectories [23].

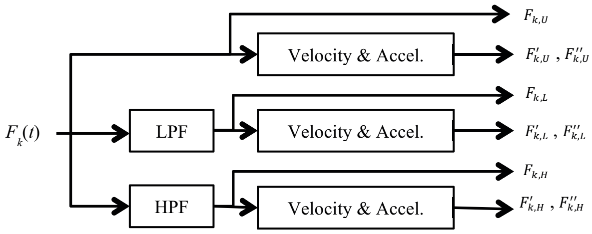

Figure 2: *Derivation of the 9 representative trajectories per formant per frame.* k *is the index of the formant* (k=1 *or* k=2). *After LPFing and HPFing the formant trajectory, velocities* (′) *and accelerations* (″) *are also computed.*

We then compute the mean absolute value (MAV) and the variance of each of the 9 trajectories, only on voiced frames, leading to 18 features per formant. Since we evaluate both F_1 and F_2, a total of 36 features is evaluated at a particular filter cutoff frequency.

3.3. Selection of cutoff frequency

To address our hypothesis that there is different information in different frequency components of the formant trajectories, we perform low- and high-pass filtering in this preliminary analysis. This requires that we identify a potential cutoff frequency to use for both low- and high-pass filters. To accomplish this, we compute Spearman correlations between each of the features and both intelligibility and speaking rate, sweeping the cutoff frequency from 1 Hz to 20 Hz. Results for the mean acceleration features are displayed in Figure 3.

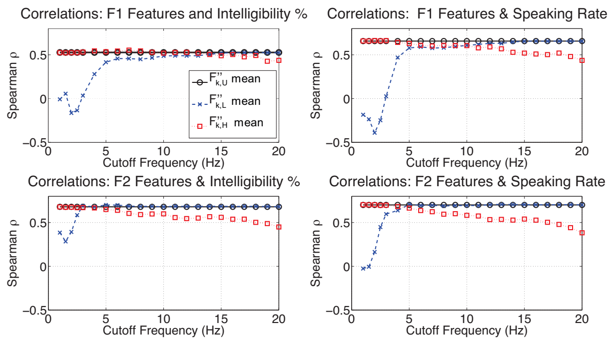

Figure 3: *Spearman correlation between features* ($F''_{k,U}$, $F''_{k,L}$, *and* $F''_{k,H}$; *top:* F_1, *bottom:* F_2) *and assessment metrics (left: intelligibility, right: speaking rate), vs. cutoff frequency.*

There is clearly significant variation in the performance of the components below 5 Hz. Since the correlations tend to asymptote at a cutoff frequency near 5 Hz for both formants and assessment metrics, we select 5 Hz as the cutoff frequency that divides that spectrum between low-pass and high-pass components. A further justification for using a 5 Hz cutoff is that the syllabic rate in English is approximately 6 syllables/s

[24], and the average syllabic rate is expected to be reduced in speakers with ALS.

4. Results

We first compute Spearman correlations between each feature and both intelligibility and speaking rate. We then perform regression experiments to evaluate the accuracy of predicting the assessment metrics, given a feature or set of features.

To predict intelligibility and speaking rate, we perform linear regression through leave-one-subject-out cross validation, training on the data collected from all subjects except the subject under test, and testing on a single session. Metrics include mean average error (MAE), root-mean-squared error (RMSE), Pearson correlation (r), and Spearman correlation (ρ), each computed between the actual assessment metric and predicted value. If the predicted intelligibility is >100%, the prediction is set to 100%.

We perform univariate linear regression on each of the 36 features, and multivariate linear regression on sets of features, such as velocity, acceleration, formant index, and frequency component. Sets of features are analyzed to obtain the best possible prediction and to obtain a generalizable, more interpretable representation of the features. When comparing features or groups of features to determine whether the difference in predictions is statistically significant, we perform paired sample t-tests between the squares of the residuals.

4.1. Correlations between formant trajectory features and both intelligibility and speaking rate

Spearman correlations between each of the 36 features and both intelligibility and speaking rate, are shown in Figure 4.

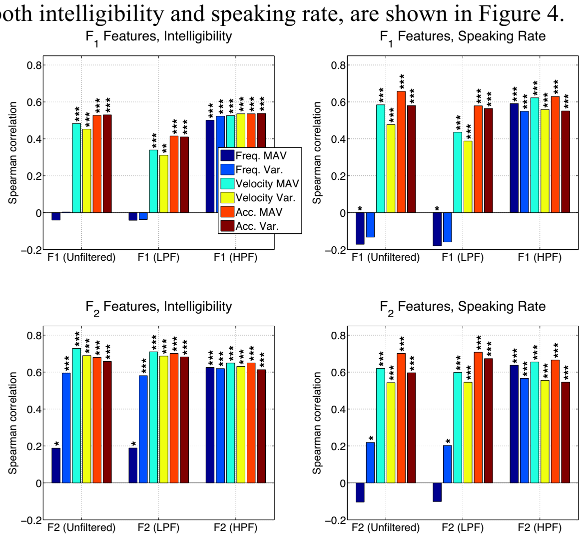

Figure 4: *Spearman correlations between each of the features (F_1: top, F_2: bottom), and assessment metrics (left: intelligibility, right: speaking rate). *: $0.01 \leq p < 0.05$; **: $10^{-4} \leq p < 0.01$; ***: $p < 10^{-4}$*

In general, for the unfiltered and LPFed features, velocity and acceleration features are more strongly correlated with both intelligibility and speaking rate than are the formant frequency displacement features. Also, although F_1 correlations are generally lower than the F_2 correlations, the F_1 correlations often reach statistical significance, even with a Bonferroni-corrected α=6.9e-4. The strongest correlation is between mean unfiltered F_2 speed and intelligibility (ρ=0.73, $p < 10^{-23}$).

4.2. Predicting intelligibility and speaking rate

4.2.1. *Intelligibility*

Table 1 displays a subset of the results from the intelligibility regression experiments. The individual features from the univariate regressions that yield the best prediction of intelligibility are mean F_2 speed from both the unfiltered and LPFed trajectories. The difference in their predictions is statistically insignificant (p=0.38). Incorporating all six F_2 velocity features into the multivariate regression does not lead to a statistically significant prediction improvement (p=0.80).

The intelligibility prediction from the multivariate regression using all six LPFed F_2 features is similar to the result obtained using all six unfiltered F_2 features, indicating that the LPFed component of F_2 seems to capture all relevant intelligibility information. To evaluate the performance of the HPFed features, we compare mean unfiltered F_2 speed to the single HPFed feature that leads to the most accurate prediction, MAV acceleration of HPFed F_2. The statistically significant difference (p<0.03) implies that HPFed features generally contribute less than LPFed or unfiltered features.

There is also a statistically significant difference between the F_2 features listed in the table and their corresponding F_1 features (p<0.02). Thus, F_2 features contribute more to intelligibility prediction than F_1 features, which is consistent with previous studies [4,6].

Table 1. *Subset of intelligibility prediction results. All Pearson and Spearman p-values are $< 10^{-4}$. Above dashed line: individual features. Below: sets of features*

Feature Name	MAE (wpm)	RMSE (wpm)	r	ρ
Mean $\lvert F'_{2,U}\rvert$	6.23	11.93	0.60	0.73
Mean $\lvert F'_{2,L}\rvert$	6.28	12.07	0.59	0.71
Mean $\lvert F''_{2,H}\rvert$	7.06	12.81	0.52	0.63
All 6 F_2 vel. features	6.63	11.81	0.61	0.63
All 6 unfilt. F_2 features	6.39	11.74	0.61	0.69
All 6 LPFed F_2 features	6.01	11.13	0.66	0.71
All F_1 features	8.38	14.27	0.31	0.36
All F_2 features	6.26	11.47	0.64	0.65

Table 2. *Subset of speaking rate prediction results. All Pearson and Spearman p-values are $< 10^{-4}$.*

Feature Name	MAE (wpm)	RMSE (wpm)	r	ρ
Mean $\lvert F''_{2,U}\rvert$	24.48	30.16	0.69	0.67
Mean $\lvert F''_{2,L}\rvert$	24.40	30.09	0.69	0.67
Mean $\lvert F'_{2,U}\rvert$	27.21	33.34	0.60	0.58
Mean $\lvert F'_{2,L}\rvert$	28.13	34.01	0.58	0.55
All 6 unfilt. F_2 features	24.67	32.51	0.64	0.63
All 6 LPFed F_2 features	21.39	26.91	0.76	0.73
All 12 unfilt. features	21.45	27.34	0.75	0.72
All 12 LPFed features	21.64	27.47	0.76	0.72

4.2.2. *Speaking rate*

Table 2 displays a subset of the results from the speaking rate experiments. The individual features that most accurately predict speaking rate when used in univariate linear regression are mean F_2 acceleration of both the unfiltered and LPFed trajectories. Of all feature groups attempted, each of three feature groups that predict speaking rate most accurately contains LPFed or unfiltered features, indicating LPFed components capture nearly all speaking rate information.

4.2.3. *Case Study*

Figure 5 displays F_2 trajectories for a male subject who utters five repetitions of BBP during each of 13 sessions. The figure displays the trajectories collected from the first and last sessions, spanning 26 months. The [ah-b-iy-ax] phonemes ("**bobby a**") in first of the five BBP repetitions are displayed. During the first session, before the subject's speech declines, the speaking rate is 230 wpm with a normal intelligibility score of 99%; after the decline, during the last session, the speaking rate is reduced to 111 wpm with a severely reduced intelligibility score of 48%.

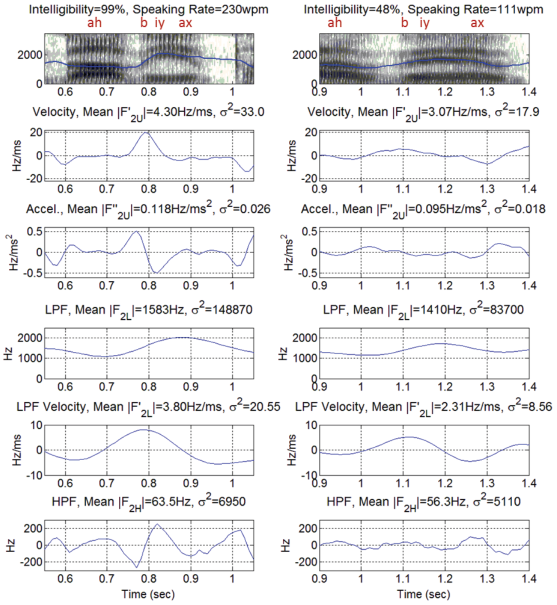

Figure 5: *$F_{2,U}$ trajectory overlaid on spectrogram (top row), $F'_{2,U}$ (2nd row), $F''_{2,U}$ (3rd row), $F_{2,L}$ (4th row), $F'_{2,L}$ (5th row), and $F_{2,H}$ (6th row). Cutoff frequency: 5 Hz. The subject is 54 years old in the first session (left).*

In the top row of Figure 5, the subject's F_2 trajectory has greater dynamic range and variability before the decline than after. In the central region of the diphthong, before the decline, large peaks are present in the velocity, acceleration, LPFed, velocity of LPFed, and HPFed trajectories. These are reduced in the less intelligible, slower speech. The LPFed component appears to capture a cleaner version of the trajectory, and the general shape of the HPFed trajectory appears to be 180^o out of phase with the acceleration trajectory.

5. Discussion and future work

The results confirm previous findings that F_2 mean speed is the individual feature that best predicts intelligibility loss [3–10], and that F_1 contributes little to intelligibility prediction when compared to F_2. We also find that F_2 acceleration aids in speaking rate prediction, and, to a lesser degree than F_2 velocity, intelligibility prediction. The importance of F_2 acceleration features might be due to their ability to capture initiation and termination of articulator motion.

Spearman correlations are higher and predictions are more accurate for speaking rate than intelligibility. This might be due to the observed nonlinearity between the features and intelligibility, and also because of the direct relationship between formant frequencies and articulator movement.

Correlations between the statistics of unfiltered and LPFed formant frequency displacement features (i.e. mean and variance of the 0^{th} derivative) and the assessment metrics are weak or insignificant. However, these features are significant when obtained from the HPFed trajectory. Combined with the resemblance of the F_2 HPFed trajectory to the F_2 acceleration trajectory, this suggests that higher order derivative statistics (more rapidly moving features) are embedded within the HPFed component. The inability of the HPFed component to perform well in prediction suggests the presence of noise.

The results from the intelligibility and speaking rate prediction experiments suggest that the LPFed component below 5 Hz is associated with both intelligibility and speaking rate decline. This association might be due to at least one of three factors: (1) the LPFed component is a less noisy measurement of the formant trajectory, and therefore more accurately represents the actual formant trajectory; (2) changes in the formant frequencies across time are slow, below 5 Hz; and/or (3) the unfiltered signal is affected by the motions of several vocal tract structures with varying degrees of impairment, while the LPFed component might represent the motion of a particular vocal tract structure that is actively declining. For example, suppose that when transitioning from a low-back to high-front vowel, a patient is unable to rapidly constrict the vocal tract because of slowed tongue movements. While the area of the constriction remains large, formant frequencies would be expected to remain relatively constant despite the tongue advancing forward to approximate the low vowel [25]. Determining the articulatory mechanisms of the abnormal formant features will require additional research.

Another area for future work involves exploring different frequency regions by applying various cutoff frequencies and bandpass filters. In this study, we assume that 5 Hz represents syllabic rate, but syllabic rate will vary during each session. Other cutoff frequencies and band-passed frequency regions may provide additional acoustic information, and might also be associated with physiological differences exhibited by patients with ALS. Future work also involves understanding the formant features that contribute to intelligibility loss versus those that contribute to speaking rate decline, as intelligibility and speaking rate are correlated ($\rho=0.48$, $p=3.38e-9$).

6. Conclusions

We have presented preliminary results exploring automatically extracted formant features that predict speech deterioration in ALS. Our main novel findings are that acceleration features derived from the frequency trajectory of F_2 predict speaking rate decline and aid in prediction of intelligibility decline, and that applying an LPF with a 5 Hz cutoff frequency to F_2 velocity and acceleration trajectories yields prediction results comparable to those achieved without LPFing. Since the majority of the information appears to be present in the LPFed frequency components, a reduced representation of the formant frequency trajectories might be sufficient to capture important trends in intelligibility and speaking rate decline. These results can be applied toward rapid automatic monitoring of ALS progression.

7. Acknowledgements

We would like to thank Nicolas Malyska and Brian Helfer from MIT Lincoln Laboratory, and Brian Richburg from Mass. General Hospital Institute of Health Professions.

8. References

[1] J. R. Green, Y. Yunusova, M. S. Kuruvilla, J. Wang, G. L. Pattee, L. Synhorst, L. Zinman, and J. D. Berry, “Bulbar and speech motor assessment in ALS: Challenges and future directions,” *Amyotroph. Lateral Scler. Front. Degener.*, vol. 14, no. 7–8, pp. 494–500, 2013.

[2] C. Armon and D. Moses, “Linear estimates of rates of disease progression as predictors of survival in patients with ALS entering clinical trials,” *J. Neurol. Sci.*, vol. 160, pp. S37–S41, 1998.

[3] R. D. Kent, J. F. Kent, G. Weismer, R. E. Martin, R. L. Sufit, and B. R. Brooks, “Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects,” *Clin. Linguist. Phon.*, vol. 3, no. 4, pp. 347–358, 1989.

[4] R. D. Kent, R. L. Sufit, J. C. Rosenbek, J. F. Kent, G. Weismer, R. E. Martin, and B. R. Brooks, “Speech Deterioration in Amyotrophic Lateral Sclerosis: A Case Study,” *J. Speech Lang. Hear. Res.*, vol. 34, no. 6, pp. 1269–1275, Dec. 1991.

[5] J. F. Kent, R. D. Kent, J. C. Rosenbek, G. Weismer, R. E. Martin, R. L. Sufit, and B. R. Brooks, “Quantitative Description of the Dysarthria in Women with Amyotrophic Lateral Sclerosis,” *J. Speech Hear. Res.*, vol. 35, pp. 723–733, Aug. 1992.

[6] G. Weismer, R. Martin, R. D. Kent, and J. F. Kent, “Formant trajectory characteristics of males with amyotrophic lateral sclerosis,” *J. Acoust. Soc. Am.*, vol. 91, no. 2, pp. 1085–1098, Feb. 1992.

[7] M. Mulligan, J. Carpenter, J. Riddel, M. K. Delaney, G. Badger, and P. Krusinski Rup Tandan, “Intelligibility and the Acoustic Characteristics of Speech in Amyotrophic Lateral Sclerosis (ALS),” *J. Speech Hear. Res.*, vol. 37, Jun. 1994.

[8] G. Weismer, J.-Y. Jeng, and J. S. Laures, “Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders,” *Folia Phoniatr. Logop.*, vol. 53, pp. 1–18, 2001.

[9] Y. Kim, R. D. Kent, and G. Weismer, “An Acoustic Study of the Relationships Among Neurologic Disease, Dysarthria Type, and Severity of Dysarthria,” *J. Speech Lang. Hear. Res.*, vol. 54, no. 2, p. 417, Apr. 2011.

[10] Y. Yunusova, J. R. Green, L. Greenwood, J. Wang, G. L. Pattee, and L. Zinman, “Tongue Movements and Their Acoustic Consequences in Amyotrophic Lateral Sclerosis,” *Folia Phoniatr. Logop.*, vol. 64, no. 2, pp. 94–102, 2012.

[11] Y. Yunusova, G. Weismer, J. R. Westbury, and M. J. Lindstrom, “Articulatory Movements During Vowels in Speakers With Dysarthria and Healthy Controls,” *J. Speech Lang. Hear. Res.*, vol. 51, no. 3, p. 596, Jun. 2008.

[12] Y. Yunusova, J. R. Green, M. J. Lindstrom, L. J. Ball, G. L. Pattee, and L. Zinman, “Kinematics of disease progression in bulbar ALS,” *J. Commun. Disord.*, vol. 43, no. 1, pp. 6–20, Jan. 2010.

[13] R. DePaul and B. R. Brooks, “Multiple Orofacial Indices in Amyotrophic Lateral Sclerosis,” *J. Speech Hear. Res.*, vol. 36, pp. 1158–1167, Dec. 1993.

[14] Y. Yunusova, J. R. Green, J. Wang, G. L. Pattee, and L. Zinman, “A protocol for comprehensive assessment of bulbar dysfunction in ALS,” *J. Vis. Exp.*, vol. 48, 2011.

[15] P. Rong, Y. Yunusova, J. Wang, and J. R. Green, “Predicting Early Bulbar Decline in Amyotrophic Lateral Sclerosis: A Speech Subsystem Approach,” *Behav. Neurol.*, vol. 2015, pp. 1–11, 2015.

[16] K. Yorkston, D. Beukelman, M. Hakel, and M. Dorsey, *Sentence Intelligibility Test*. Communication Disorders Software, 2007.

[17] P. Rong, Y. Yunusova, and J. R. Green, “Speech Intelligibility Decline in Individuals with Fast and Slow Rates of ALS Progression,” in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[18] L. J. Ball, D. R. Beukelman, and G. L. Pattee, “Timing of Speech Deterioration in People with Amyotrophic Lateral Sclerosis,” *J. Med. Speech-Lang. Pathol.*, vol. 10, no. 4, 2002.

[19] K. M. Yorkston, E. Strand, R. Miller, A. Hillel, and K. Smith, “Speech Deterioration in Amyotrophic Lateral Sclerosis: Implications for the Timing of Intervention,” *J. Med. Speech-Lang. Pathol.*, vol. 1, no. 1, pp. 35–46, 1993.

[20] T. F. Quatieri and R. B. Dunn, “Speech enhancement based on auditory spectral change,” in *Acoustics, Speech and Signal Processing (ICASSP), 2002 IEEE International Conference on*, 2002, vol. 1, pp. 257–360.

[21] T. F. Quatieri and R. A. Baxter, “Noise reduction based on spectral change,” in *Applications of Signal Processing to Audio and Acoustics*, 1997.

[22] D. D. Mehta, D. Rudoy, and P. J. Wolfe, “Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking,” *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1732–1746, Sep. 2012.

[23] B. S. Helfer, T. F. Quatieri, J. R. Williamson, D. D. Mehta, R. Horwitz, and B. Yu, “Classification of depression state based on articulatory precision.,” presented at the Interspeech, Lyon, France, 2013, pp. 2172–2176.

[24] F. Pellegrino, C. Coupé, and E. Marsico, “Across-Language Perspective on Speech Information Rate,” *Language*, vol. 87, no. 3, pp. 539–558, 2011.

[25] K. N. Stevens and A. S. House, “Development of a Quantitative Description of Vowel Articulation,” *J. Acoust. Soc. Am.*, vol. 27, no. 3, May 1955.