

# Clinical validation of a novel speech assessment app in amyotrophic lateral sclerosis (ALS)

Leif E. R. Simmatis<sup>1,2</sup>, Jessica Robin<sup>3</sup>, Justin Truong<sup>1</sup>, Timothy Pommée<sup>1,4</sup>, Bharatkumar Koyani<sup>5</sup>, Yana Yunusova<sup>1,2,4</sup>

1. KITE-Univ. health network, Toronto, ON, CAN; 2.Dept. Speech-Lang Path, Univ. Toronto, Toronto, ON, CAN; 3.Winterlight Labs Inc., Toronto, ON, CAN;

4. Sunnybrook Res. Inst., Toronto, ON, CAN; Regeneron Inc., Tarrytown, NY, USA

## Introduction and Aim

ALS often causes speech impairments. We sought to validate a novel digital speech assessment against ALS clinical speech scales using interpretable dimensionality reduction. We addressed challenges of high-dimensional acoustic data and expected confounding effects of sex on acoustic features by using Neural Decomposition (ND) to generate low-dimensional representations of acoustic features ("embeddings") that accounted for sex as a covariate. We asked:

**Is it possible to extract acoustic information related to clinical measures of speech/overall function in ALS? Do specific features correspond strongly to sex? If so, can these effects be mitigated?**

## Hypotheses

(1) Embeddings will correspond strongly (area under receiver operating characteristic curve [AUROC]>0.70) to clinical measures.

(2) Pruning features strongly associated with sex will not strongly affect associations between embeddings and clinical measures.

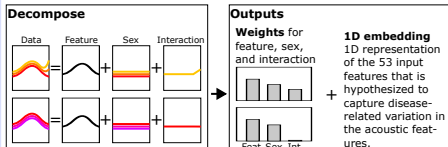
## Methods

Clinical measures were binarized into more-severe/less-severe subgroups using clinical thresholds. Features were extracted from recordings and standardized.

	Threshold	
N (%)	121 (47)	N/A
Yrs. since onset	59 (41)	N/A
SIT intelligibility %	99 (6)	96
SIT rate (WPM)	148 (48)	160
ALFSRS-R total	35 (14)	34
ALFSRS-R bulbar	11 (4)	11
ALFSRS-R resp.	10 (2)	10
FVC (%)	76 (23)	80

All except N(%) are shown as median (IQR).

**Neural Decomposition (ND)** isolated effects of sex on 1D embeddings



**First training:** train ND on full dataset with 53 features - evaluate AUROC of embeddings vs binarized clinical measures.

**Pruning and second training:** train ND on reduced dataset (prune features with bottom quartile of sex effect) - re-evaluate AUROC vs binarized clinical measures.

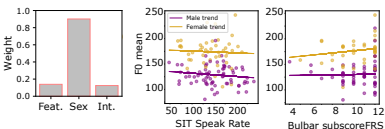
## Results

ND identified groups of features that covaried with sex and those that did not. Retained features were often related to jitter, shimmer, and pause/word duration

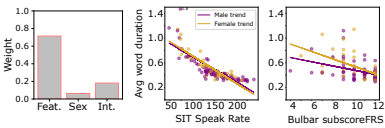
	Feature	Sex	Interaction	Kept
F0_max	0.21	0.23	0.25	
F0_mean	0.14	0.9	0.12	
F0_median	0.18	0.94	0.15	
F0_min	0.16	0.82	0.13	
F0_range	0.21	0.22	0.26	
F0_variance	0.12	0.079	0.13	
hnr_ac_max	0.65	0.16	0.28	
hnr_ac_mean	0.96	0.46	0.16	
hnr_ac_median	0.97	0.6	0.16	
hnr_ac_min	0.23	0.18	0.13	
hnr_ac_range	0.2	0.17	0.3	
hnr_ac_variance	0.096	0.2	0.19	
hnr_cc_max	0.37	0.24	0.32	
hnr_cc_mean	0.93	0.54	0.16	
hnr_cc_median	0.88	0.48	0.28	
hnr_cc_min	0.2	0.18	0.18	
hnr_cc_range	0.27	0.35	0.26	
hnr_cc_variance	0.14	0.31	0.18	
int_max	0.22	0.96	0.11	
int_mean_db	0.26	0.43	0.29	
int_mean_energy	0.25	0.39	0.31	
int_mean_sones	0.19	0.46	0.29	
int_median	0.16	0.36	0.21	
int_min	0.13	0.13	0.25	
int_range	0.088	0.083	0.08	
int_variance	0.89	0.092	0.18	
jitter_ddp	0.94	0.083	0.12	
jitter_local	0.93	0.19	0.16	
jitter_local_abs	0.97	0.19	0.075	
jitter_rap	0.83	0.1	0.28	
long_pause_ct_norm	0.094	0.1	0.16	
long_pause_dur	0.11	0.085	0.12	
long_pause_ratio	0.12	0.13	0.17	
med_pause_ct_norm	0.081	0.15	0.097	
med_pause_dur	0.11	0.095	0.11	
med_pause_ratio	0.13	0.12	0.17	
phonation_rate	0.14	0.46	0.13	
shimmer_ap01	0.99	0.16	0.12	
shimmer_ap02	0.99	0.19	0.06	
shimmer_ap03	0.99	0.082	0.11	
shimmer_d01	0.99	0.13	0.11	
shimmer_d02	0.99	0.14	0.07	
shimmer_local_db	0.99	0.12	0.06	
short_pause_ct_norm	0.11	0.14	0.14	
short_pause_dur	0.099	0.12	0.099	
zcr_kurtosis	0.18	0.18	0.21	
zcr_mean	0.087	0.07	0.091	
zcr_skewness	0.16	0.49	0.17	
zcr_var	0.079	0.075	0.12	
articulation_rate	0.7	0.25	0.25	
avg_word_dur	0.7	0.11	0.14	
speech_rate	0.72	0.21	0.18	

Feature, sex, and interaction weights captured interpretable patterns in the raw data.

**F0 mean** showed a strong sex effect and so was **discarded** from the final feature set



**Average word duration** showed a minimal sex effect and so was **retained** in the final feature set



Bamboo data 1D embedding was strongly associated with clinical features (AUROC >0.70). Removal of features with high sex influence minimally impacted these associations.

	AUROC		
	Full (53)	Reduced (14)	Change
Sex	0.652	0.555	-0.097
SIT intelligibility (%)	<b>0.703</b>	<b>0.727</b>	<b>+0.024</b>
SIT rate (WPM)	<b>0.714</b>	<b>0.687</b>	<b>-0.027</b>
ALFSRS-R total	0.542	0.534	-0.008
ALFSRS-R bulbar	<b>0.694</b>	<b>0.686</b>	<b>-0.008</b>
ALFSRS-R resp.	0.592	0.591	-0.001
FVC (%)	0.503	0.503	n/c

## Conclusions

Low-dimensional (1D) embeddings of acoustic features showed strong associations with clinical measures (AUROC >0.70).

Removal of features confounded by sex had a small impact on associations with clinical measures.

These results show the promise of using acoustic features to capture disease-related phenomena in a highly interpretable way.

## References

- Traxinger K., Kelly C., Johnson BA, et al. (2013). Prognosis and epidemiology of amyotrophic lateral sclerosis. *Neurology Clinical Practice* 3:13-320.
- Tomik B. and Gullof, RJ. (2016) Dysarthria in amyotrophic lateral sclerosis: A review. *Amyotrophic Lateral Sclerosis*, 11:1-2, 4-15
- Martens K and Yau C (2020) Neural decomposition: functional ANOVA with Variational Autoencoders. *ArXiv*: 2006.14293v2

## Acknowledgements

We would like to sincerely thank Dr. Madhura Kulkarni for her assistance in organizing the data that was analyzed in this project.