Signal Analytics Workshop
February 19, 2020

- Sounds of the human vocal instrument
  - Dr. Shri Narayanan
- Assessment of Prosodic and Articulatory Interactions in Autism Spectrum Disorder
  - T. Talkar, J. Williamson, D. Hannon, H. Rao, S. Yuditskaya, D. Sturim, K. Claypool, L. Nowinski, H. Saro, C. Stamm, M. Mody, C. McDougle, T. Quatieri
- A comparison of interpretable acoustic features identified diverse tendencies in several motor speech disorders
  - J. Hlavnička, V. Berisah, J. Liss, J. Rusz
- Evaluating a wearable mechano-acoustic sensor and iPhone application for speech data in individuals with Alzheimer's dementia, mild cognitive impairment, and Parkinson's disease
  - R. Richter, B. Fujii, N. Razin, A. Roberts, J. Rogers, S. Xu
- Evaluating the Validity of Using Speech Analytics for Remotely Tracking Amyotrophic Lateral Sclerosis (ALS) Progress
  - G. Stegmann, S. Hahn, V. Berisha, J. Liss, S. Rutkove, K. Qi, K. Shelton, J. Shefner
- Improving Deep Learning Networks for Automatic Orofacial Assessment across Clinical Populations
  - D. Guarín, B. Taati, A. Bandini, T. Hadlock, Y. Yunusova
- Recognition Of Short Utterances From A Closed Set Produced By Dysarthric Speakers Using A Deep Convolutional Network
  - S. Cai, B. Richburg, M. Maffei, I. Alvarado, J. Cattiau, K. Seaver, J. Green, M. Brenner

K. Smith	Changes in fundamental frequency in spontaneous speech correlate with cognitive function in Parkinson's disease
A. Glotfelty, W. Katz	The Role of Visibility in Silent Speech Tongue Movements: A Kinematic Study of Consonants
A. Wisler, K. Teplansky, J. Green, Y. Yunusova, T. Campbell, D. Heitzman, J. Wang	Analyzing Residuals of Machine Learning based ALS-FRS Speech Score Predications
J. Berry, Y. Kim	Dysarthria and Dialect Affect the Kinematics of Connected Speech
V. Ramanarayanan, D. Suendermann-Oeft	Leveraging Multimodal Dialog Technologies and Machine Learning for Patient Health Diagnosis, Monitoring, and Intervention
A. Wisler, K. Teplansky, J. Green, Y. Yunusova, T. Campbell, D. Heitzman, J. Wang	ALS detection based on onset using correlation structural features from kinematic measurements
K. Teplansky, S. Dutta, B. Cao, J. Wang	Tongue and Lip Motion Patterns of Alaryngeal and Silent Speech
M. Saxon, A. Tripathi, J. Liss, V. Berisha	A new model for objective estimation of hypernasality from dysarthric speech
V. Woisard	Carcinologic Speech Severity Index
K. Cunningham, A. Jacks, K. Haley	Dysarthria in rapid-onset dystonia-parkinsonism: A longitudinal case study

Sounds of the human vocal instrument

Shrikanth (Shri) Narayanan
University of Southern California, Los Angeles, CA
Signal Analysis and Interpretation Laboratory
https://sail.usc.edu/people/shri.html

The vocal tract is the universal human instrument played with great dexterity to produce the elegant acoustic structuring of speech, song and other sounds to communicate intent and emotions. The sounds produced by the vocal instrument also carry crucial information about individual identity and the state of health and wellbeing. A longstanding research challenge has been in improving the understanding of how vocal tract structure and function interact, and notably in illuminating the variant and invariant aspects of speech (and beyond) within and across individuals. The first part of the talk will highlight engineering advances that allow us to perform investigations on the human vocal tract in action-- from capturing the dynamics of vocal production using novel real-time magnetic resonance imaging to machine learning based articulatory-audio modeling--to offer insights about how we produce sounds with the vocal instrument. The second part of the talk will highlight some scientific, technological and clinical applications using such multimodal data driven approaches in the study of the human vocal instrument.

Biography of the Speaker:

Shrikanth (Shri) Narayanan the Niki & C. L. Max Nikias Chair in Engineering at the University of Southern California, where he is Professor of Electrical & Computer Engineering, and jointly in Computer Science, Linguistics, Psychology, Neuroscience, Otolaryngology and Pediatrics, Director of the Ming Hsieh Institute and Research Director of the Information Sciences Institute. Prior to USC he was with AT&T Bell Labs and AT&T Research. His research focuses on human-centered information processing and communication technologies. He is a Fellow of the National Academy of Inventors, the Acoustical Society of America, IEEE, ISCA, the American Association for the Advancement of Science (AAAS), the Association for Psychological Science, and the American Institute for Medical and Biological Engineering (AIMBE). He is a recipient of several honors including the 2015 Engineers Council's Distinguished Educator Award, a Mellon award for mentoring excellence, the 2005 and 2009 Best Journal Paper awards from the IEEE Signal Processing Society and serving as its Distinguished Lecturer for 2010-11, a 2018 ISCA Best Journal Paper award, and serving as an ISCA Distinguished Lecturer for 2015-16 and the Willard R. Zemlin Memorial Lecturer for ASHA in 2017. He has published over 800 papers and has been granted seventeen U.S. patents. His research and inventions have led to technology commercialization including through startups he co-founded: Behavioral Signals Technologies focused on the telecommunication services and AI based conversational assistance industry and Lyssn focused on mental health care delivery, treatment and quality assurance.

Assessment of Prosodic and Articulatory Interactions in Autism Spectrum Disorder

T. Talkar, J. Williamson, D. Hannon, H. Rao, S. Yuditskaya, D. Sturim, K. Claypool, L. Nowinski, H. Saro, C. Stamm, M. Mody, C. McDougle, T. Quatieri

We sought to determine if coordination measures (Williamson et al, 2014) (Quatieri et al, 2017) between different speech subsystems aid in discriminating between children with and without autism spectrum disorder (ASD), motivated by findings that children with ASD display higher pitch variability (Bonneh, 2011). 5 children with ASD and 5 controls each read a short passage and repeated a diadochokinetic sequence. From these recordings, pitch, intensity, and formant tracks were extracted to determine eigenvalues from correlation matrices for each subject. The high effect sizes of these eigenvalues highlighted pitch variability and weak ties between intensity and pitch – features which were used to classify ASD subjects with 89% and 100% accuracy respectively. By combining with facial features during speech, we hope to understand the mechanisms of the feedback and feedforward pathways of coordination to develop models of speech and facial motor control in ASD.

A comparison of interpretable acoustic features identified diverse tendencies in several motor speech disorders

J. Hlavnička, V. Berisah, J. Liss, J. Rusz

A majority of studies on acoustic vocal biomarkers explored only a single or limited number of diseases. This study aims to shed light on the diagnostic value of acoustic analysis by comparing diseases associated with various etiologies of motor speech disorders. We analyzed a database of 570 Czech speakers comprising of idiopathic REM sleep behavior disorder (RBD), Parkinson's disease (PD), multiple system atrophy (MSA), progressive supranuclear palsy (PSP), Huntington's disease (HD), cerebellar ataxia (CA), multiple sclerosis (MS), and healthy controls. Each speaker was instructed to perform the sustained vowels /A/ and /I/, rhythm test, reading a passage, monologue, and diadochokinetic task. Total of 49 features was calculated for each speaker, normalized to z-scores to reduce the effect of age and sexual dimorphism, and statistically compared. Majority of features showed significantly increased values for single or multiple disease groups. Despite the considerable overlap of acoustic manifestations, elementary tendencies showed differentiating value especially for recognition of PD and atypical parkinsonian syndromes such as MSA and PSP. Acoustic evidence of reduced amplitude of movements (Hypokinesia) was dominant in PD, PSP, and MSA, but also in RBD group representing subjects under high risk of developing PD. Acoustic evidence of involuntary or exaggerated movements (Hyperkinesia) was frequently observed in MSA, HD, CA, and MS. To conclude, the acoustic analysis can provide valuable insight but requires a thorough interpretation in the context of other findings.

Evaluating a wearable mechano-acoustic sensor and iPhone application for speech data in individuals with Alzheimer's dementia, mild cognitive impairment, and Parkinson's disease

R. Richter, B. Fujii, N. Razin, A. Roberts, J. Rogers, S. Xu

We present feasibility data for a novel, wearable mechano-acoustic sensor for recording speech, gait, and swallowing data from individuals with AD/MCI and PD. Individuals with PD (n = 3, M age = 71; M UPDRS score = 21.3; M Hoehn and Yahr = 2.5; M MoCA = 24.67), AD/MCI (n = 10, M age = 77.5; M Clinical Dementia Rating score = 1.5; M MoCA = 20.1) and their healthy family care partners (n = 9, M age = 61; M MoCA = 27) have completed the study. We compared Praat-analyzed data from the investigational sensor against gold-standard clinical equipment for three structured speech tasks (maximum vowel duration /ah/ - 3 trials, diadochokinetic tasks - AMRs and SMRs, standardized paragraph reading) and a semi-structured conversation task in quiet and noisy conditions using two-tailed, paired t-tests. Preliminary findings suggest the sensor and iOS app-recorded data were consistent with microphone data for all measures except syllable count for /ka/ and mean intensity for sustained vowel phonation.

Evaluating the Validity of Using Speech Analytics for Remotely Tracking Amyotrophic Lateral Sclerosis (ALS) Progress

G. Stegmann, S. Hahn, V. Berisha, J. Liss,
S. Rutkove, K. Qi, K. Shelton, J. Shefner

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder that progressively weakens skeletal muscles, resulting in disability and eventually death. Patients affected by ALS lose their motor function, speech, and the ability to breathe. Disease progression is commonly measured using the ALS functional rating scale (ALSFRS-R), which uses 12 subscales to quantify functional abilities, such as speech, breathing, and gross motor skills, using a 5-point Likert scale. The ALSFRS-R, however, lacks sensitivity to small changes. Here we evaluate the feasibility of tracking participants' neurological health via speech analytics through a custom speech app. We evaluate the test-retest reliability of two measures related to intelligibility, and we evaluate correlation with the ALSFRS-R speech subscale. The reliabilities are high (r = .95-.96) and the correlations with the ALSFRS-R are moderate (r = .64-.79).

Improving Deep Learning Networks for Automatic Orofacial Assessment across Clinical Populations

D. Guarín, B. Taati, A. Bandini, T. Hadlock, Y. Yunusova

The overall goal of this presentation is to review our recent work developing Machine Learning (ML) and computer vision methods for assessment of orofacial abnormalities in neurological diseases and to introduce new techniques that aim to decrease the data requirements of these methods, making them more feasible for clinical implementation.
The methods, algorithms, and software tools that are we are developing represent a paradigm shift in the assessment of orofacial deficits. We are currently translating our tools to be more clinically viable, user-friendly, and convenient so that they become standard components in the clinicians' toolset. In this presentation, we aim to delineate our research program and discuss the advantages and limitations of these technological solutions to important clinical problems.

Recognition of Short Utterances From A Closed Set Produced By Dysarthric Speakers Using A Deep Convolutional Network

S. Cai, B. Richburg, M. Maffei, I. Alvarado, J. Cattiau, K. Seaver, J. Green, M. Brenner

We present a system that learns to recognize a closed set of short utterances produced by speakers with dysarthria, a speech motor disorder caused by neurological conditions such as amyotrophic lateral sclerosis (ALS) and stroke. Our approach tailors recognizers for individual dysarthric speakers by using a small set of training utterances produced by each speaker. It bypasses two challenges for applying large-vocabulary, continuous automatic speech recognition (ASR) to dysarthric speech, namely between-speaker variation in speech symptoms and degradation of intelligibility with increasing sentence length. Powered by TensorFlow.js [4], the system runs in the web browser and performs transfer learning on a base convolutional neural network (convnet) pre-trained on a generic audio dataset of spoken English words [6]. Preliminary results from two individuals with dysarthria caused by ALS show a testing accuracy of approximately 95% over 12 target utterances. Coupled with speech synthesis and other technologies, this recognition system possesses potentials for a self-serve, trainable voice recognizer that can benefit the functional independence and quality of life for individuals with severe dysarthria.

Changes in fundamental frequency in spontaneous speech correlate with cognitive function in Parkinson's disease

K. Smith

Cognitive impairment is common in Parkinson's disease (PD) but better biomarkers are needed. We propose running speech could provide markers of cognitive function in PD. We that hypothesize that acoustic parameters will vary with cognitive demand of the speaking task depending on the degree of cognitive dysfunction.
Fourteen participants with PD performed sustained phonation of /ɑ/ and spontaneous speech while describing a picture of a scene. Fundamental frequency (fo), F1 and F2 were calculated for each task.
The difference in mean fo between the speaking tasks was significantly correlated with MoCA (r=0.63, p=0.02). Poorer cognitive function was associated with a decrease in fo in running speech compared with sustained phonation.
This preliminary work suggests that speech markers may be useful to detect and monitor cognitive function in PD.

The Role of Visibility in Silent Speech Tongue Movements: A Kinematic Study of Consonants

A. Glotfelty, W. Katz

This study focused on the spatiotemporal patterns of tongue movement produced during silent and audible speech for five consonants of American English, selected to represent a continuum of tongue visibility (/g/, /w/, /ɹ/, /l/, and /ð/). Lip aperture data were also analyzed, as lip opening affects tongue visualization during face-to-face conversation. Electromagnetic articulography (EMA) was used to measure 11 native speakers' productions of consonant-vowel (CV) syllables in two vowel contexts (/ɑ/ and /i/). Results indicated increased consonant duration and lip aperture in the silent condition across all five consonants. Spatial analyses revealed that the overall area of tongue movement did not differ between the conditions. However, tongue movements occurred more anteriorly in silent speech than in audible speech for all consonants except /g/, the only consonant without a visible tongue component. These anterior-shifted patterns for visible lingual consonants suggest that degree of tongue visibility is a component of speech feedforward articulatory maps, and that tongue visibility is included in the feedback processes used by talkers.

Analyzing Residuals of Machine Learning based ALS-FRS Speech Score Predictions

A. Wisler, K. Teplansky, J. Green, Y. Yunusova, T. Campbell, D. Heitzman, J. Wang

In this study, we consider the speech component of the Amyotrophic Lateral Sclerosis (ALS) functional rating scale (FRS). The ALSFRS is an instrument for measuring motor function in individuals with ALS. The speech component measures function along a 5-point scale ranging from 4 (assigned to patients with intact intelligibility) to a score of 0 (assigned to patients who have completely lost the ability to produce intelligible speech). By breaking motor function into five levels, the ALSFRS can create an objective set of criteria for each level. Although this allows for more reliable evaluation, the coarse nature of this rating scale inherently limits the precision of the measurement.
In this study, we develop a supervised learning regression model designed to map acoustic features to the speech score of the ALSFRS. We then investigate the residuals of the model's predictions. The purpose of examining residuals is to determine whether these "error terms" contain meaningful information about speech motor control that is not captured by the coarse ALSFRS assessment. This would suggest that the supervised learning models can provide more fine-grained information on motor speech control than the labels they are trained on.
To assess this, we use the clinical measure of intelligible speaking rate (ISR) as a proxy measure of speech motor function (Yorkston & Breukelman, 1981). We then perform a regression analysis to examine the relationship between our models' residuals and ISR while controlling for each patient's true ALSFRS speech score.

Dysarthria and Dialect Affect the Kinematics of Connected Speech

J. Berry, Y. Kim

Methods for measuring articulatory movements during speech are becoming increasingly viable in the clinical setting. Few standards for acquiring and analyzing articulatory-kinematic data have been established, and little work has examined the disambiguation of typical articulatory differences and differences associated with speech disorder. We examine speech kinematics during connected speech in typical talkers and talkers with dysarthria from two distinct dialects of American English. Electromagnetic articulography data were collected during connected speech, using a standard reading passage, from typical talkers (n=30) and talkers with dysarthria (n=19). Participants are divided among upper Midwestern and Southern American English dialects. Current results suggest changes in articulatory range-of-motion resulting from differences in dialect and dysarthria. Ensuing analyses will examine dynamic aspects of articulation and metrics of kinematic stability.

Leveraging Multimodal Dialog Technologies and Machine Learning for Patient Health Diagnosis, Monitoring, and Intervention

V. Ramanarayanan, D. Suendermann-Oeft

Using conversational speech and video technologies in conjunction with signal processing and machine learning algorithms can allow patients to interact with an engaging virtual agent in fun and simple tasks on demand in the comfort of their homes, and also obtain various automatically-computed, clinically-validated measures of disease onset and progression that can help those patients and caregivers in diagnosis, monitoring and understanding. In this vein, we present NEMSI, an interactive technology-based diagnosis, monitoring and intervention solution for patient mobile health applications. So far the NEMSI system has collected multiple interactions with depression and ALS patients with positive results and encouraging feedback from patients and clinicians. We are continuing to explore other domains of potential applicability, such as for kids and adults with Autism Spectrum Disorder (ASD) as part of future work.

ALS detection based on onset using correlation structural features from kinematic measurements

A. Wisler, K. Teplansky, J. Green, Y. Yunusova, T. Campbell, D. Heitzman, J. Wang

Amyotrophic Lateral Sclerosis (ALS) is a fatal neurodegenerative disease that inhibits the ability of the brain to control muscle movements. Although there is currently no cure for ALS, early detection of the disease can facilitate more effective clinical interventions, increasing both life expectancy and quality of life (Kiernan et al., 2011). Recently several researchers have explored the idea of using speech data to distinguish individuals with ALS from healthy controls (Norel et al., 2018; Wang, Kothalkar, Cao, & Heitzman, 2016). Although this approach to early detection has shown promise, speech is not always the initial area affected by ALS. In fact, bulbar-onset (BO), which primarily effects speech and swallowing, represents only about 25% of ALS patients, whereas spinal-onset (SO) accounts for approximately 70% of ALS patients (Kiernan et al., 2011). Prior research has disregarded the degree to which the region of onset affects our ability to detect ALS. In this study, we design a supervised learning classifier to discriminate between individuals with ALS and healthy controls, in order to examine how the classification performance varies between the BO and SO participants.

Tongue and Lip Motion Patterns of Alaryngeal and Silent Speech

K. Teplansky, S. Dutta, B. Cao, J. Wang

Laryngectomy is the surgical removal of a larynx due to oral or laryngeal cancer (Bailey, Johnson, & Newlands, 2006). After the larynx is removed, the articulatory control system cannot modify the length of the vocal tract based on larynx height and the patients lose the ability to control pitch (Sorokin, Olshansky, & Kozhanov, 1998). Current available postoperative treatments include the following voice prosthesis options: 1) tracheoesophageal (TEP) speech, 2) the use of an electrolarynx, and 3) esophageal speech (Liu & Ng, 2007; Mignano et al., 1993). While these methods provide a means of communicating, voice options for alaryngeal communication typically result in poor intelligibility, robotic sounding speech, require extensive training, and are cumbersome to use (Liu & Ng, 2007). Consequently, the use of current alaryngeal communication devices often results in social isolation and a decreased quality of life (Mertl et al., 2018).

A new model for objective estimation of hypernasality from dysarthric speech

M. Saxon, A. Tripathi, J. Liss, V. Berisha

Hypernasality is a common symptom across many motor-speech disorders. For voiced sounds, hypernasality introduces an additional resonance in the lower frequencies and, for unvoiced sounds, there is reduced articulatory precision due to air escaping through the nasal cavity. However, the acoustic manifestation of these symptoms is highly variable, making hypernasality estimation very challenging, both for human specialists and automated systems. Previous work in this area relies on either engineered features based on statistical signal processing or machine learning models trained on clinical ratings. Engineered features often fail to capture the complex acoustic patterns associated with hypernasality, whereas metrics based on machine learning are prone to overfitting to the small disease-specific speech datasets on which they are trained. Here we propose a new set of acoustic features that capture these complementary dimensions. The features are based on two acoustic models trained on a large corpus of healthy speech. The first acoustic model aims to measure nasal resonance from voiced sounds, whereas the second acoustic model aims to measure articulatory imprecision from unvoiced sounds. To demonstrate that the features derived from these acoustic models are specific to hypernasal speech, we evaluate them across different dysarthria corpora. Our results show that the features generalize even when training on hypernasal speech from one disease and evaluating on hypernasal speech from another disease (e.g. training on Parkinson's disease, evaluation on Huntington's disease).

Carcinologic Speech Severity Index

V. Woisard

Speech disorders affect the quality of life of patients treated for cancer of the oral cavity or oropharynx. The lack of reliability and validated measure of severity explains the difficulty to determine how to decrease the sequalae and justify the development of automatic speech processing.
The aim of this study is to determine an automatic measurement of speech severity for oral and oropharyngeal cancer.
87 patients treated for cancer of the oral cavity or oropharynx and 42 controls performed different tasks of speech production, specifically targeting vocal production, prosody, comprehensibility, acoustic-phonetic decoding and intelligibility. The audio recordings of these productions were then subjected to a human perceptive evaluation, and automatic processing in order to determine different scores and study the correlations.
The best correlation between human perceived severity and automatic processing was obtained by the combination of some acoustics extracted parameters and likelihood score obtained by automatic speech recognition (r=0.86).
Automatic speech processing provides valid, reliable and reproducible parameters that can be used as a reference for follow-up of patients treated for oral cavity or oropharynx cancer.

Dysarthria in rapid-onset dystonia-parkinsonism: A longitudinal case study

K. Cunningham, A. Jacks, K. Haley

Rapid-onset dystonia-parkinsonism (RDP) is a rare neurodegenerative disease caused by a mutation in the ATP1A3 gene with a characteristic sudden onset of motor symptoms in a period of hours to days. Though important to the diagnosis, the profile, onset, and progression of the presumed dysarthria have not been characterized in detail to our knowledge.

The purpose of the current study was to characterize the speech of two people with RDP, asking:
1. What is the dysarthria profile in RDP?
2. What is the pattern of onset and progression of dysarthria; does it follow the rapid onset pattern of motor symptoms?