Wednesday, February 16, 2022

- Vocal biomarkers of neurological conditions based on motor timing and coordination
  - T. Quatieri
- Acoustic and kinematic indices of spatiotemporal stability in children's productions of words and nonwords
  - A. Wisler, S. Benham, J. Berlin, J. Wang, L. Goffman
- A Wearable Electromagnetic Articulograph (EMA) for Silent Speech Interface
  - J. Wang, S. Ravi, B. Cao, N. Sebkhi, A. Bhavsar, O. Inan, W. Xu
- Real-world Assessment of DDK Rate in Dysarthric and Healthy Speech
  - P. Kadambi, J. Liss, V. Berisha, S. Hahn
- Collecting remote voice and movement data from people with Parkinson's disease (PD) using multimodal conversational AI: Lessons learned from a national study
  - A. Exner, V. Ramanarayanan, D. Pautler, S. Snyder, H. Kothare, J. Liscombe, S.Sridhar, O. Roesler, W. Burke, M. Neumann, D. Suendermann-Oeft, J. Huber
- A multifactorial facial surface EMG analysis as a novel objective assessment tool for detecting and measuring subclinical speech changes in amyotrophic lateral sclerosis
  - P. Rong, G. Pattee
- Dissociating Dysarthria and Dialect in the Kinematics of Connected Speech
  - J. Berry, Y. Kim
- Automating Vowel Formant Estimations for Disordered Pediatric Speech Samples
  - H. Valentine, G. Carozzi, M. Grigos

6:00-7:30 Poster Session

M. Masapollo, Y. Oh, J. Goel, J. Lowenstein, S. Nittrouer	Electromagnetic articulography is feasible for assessment of speech motor skills in cochlear implant users
A. Fletcher, A. Wisler, E. Gruver, S. Borrie	Beyond Speech Intelligibility: Quantifying Listener Effort and Engagement with Dysarthric Speech
K. Hwang, Y. Chang, V. Berisha, M. McAuliffe, A. Stump, E. Stender, I. Naim, E. Hong, J. Choi, R. Brisman, J. Avobich, E. Levy	Validity of online speech data collection from children with dysarthria due to cerebral palsy
K. Connaghan, J. Green, A. Haenssler, Z. Scheier, M. Keegan, A. Clark, K. Burke, JP. Onnela, J. Berry	What can passive smartphone metrics tell us about communicative participation in ALS?
O. Roesler, W. Burke, H. Kothare, J. Liscombe, M. Neumann, A. Cornish, D. Habberstad, D. Pautler, D. Suendermann-Oeft, V. Ramanarayanan	Multimodal dialog based remote patient monitoring of motor function in Parkinson's Disease and other movement disorder
J. Liscombe, A. Ocampo, H. Kothare, O. Roesler, M. Neumann, A. Cornish, D. Pautler, D. Suendermann-Oeft, V. Ramanarayanan	On the robust automatic computation of speaking and articulation duration in ALS patients versus healthy controls
K. Teplansky, A. Wisler, J. Green, D. Heitzman, S. Austin, J. Wang	Measuring Articulatory Compensation in ALS using a Data-Driven Consonant Distinctiveness Space Approach
T. Talkar, D. Low, D. O'Keeffe, A. Simpkin, S. Ghosh, T. Quatieri	Detection of COVID-19 With Vocal Biomarkers Based on Motor Coordination and Phonetic Approaches.

Vocal biomarkers of neurological conditions based on motor timing and coordination

T. Quatieri

Toward the goal of noninvasive objective means to detect and monitor neurological conditions, MIT Lincoln Laboratory is developing vocal biomarkers that reflect a change in brain functioning as manifested in motor control. Specifically, vocal features are based on timing and coordination of articulatory, phonatory, and respiratory components of vocal expression, motivated by the hypothesis that these relations are associated with neural coordination across different parts of the brain essential in fluent speech. Timing- and coordination-based features are extracted using behavioral measures from the acoustic signal, as well as from associated facial measures during speaking. With the aim of connecting acoustic features to neurophysiology, mappings to articulatory and neurocomputational models of speech production will be discussed. This presentation gives the foundation for extracting our vocal features and illustrates use of these markers by example to three application areas: major depression disorder, autism spectrum disorder and COVID-19. The measurement and modeling framework may provide a common neurophysiological feature basis in detecting and monitoring of neurological conditions from speech and voice, while potentially providing features to distinguish across disorders and to monitor and predict the effect of treatments. The seminar will conclude with the author's perspective on the top unsolved problems in the effective deployment of vocal biomarkers.

Acoustic and kinematic indices of spatiotemporal stability in children's productions of words and nonwords

A. Wisler, S. Benham, J. Berlin, J. Wang, L. Goffman

A classic problem in speech motor control concerns the stability of articulatory motion as children or adults produce a linguistic target. While much has been learned about articulatory variability as a function of development and disorder, an ongoing limitation for research and clinical applications is that specialized and expensive equipment is required, which may be overly invasive for some participants. The aim of the present work is to develop a measure of articulatory stability that is based on the acoustic signal, rather than a direct recording of articulatory motion (e.g., Howell et al., 2009). This is a methodological study in which spatiotemporal variability is compared across words and nonwords in 4- and 5-year-old children with typical language and with developmental language disorder (DLD). In general, findings suggest that the acoustic signal may provide information about speech motor variability. Findings from this work will provide new insights into the convergent as well as the divergent relationship between acoustic and kinematic indices across ages, clinical populations, and linguistic contexts.

A Wearable Electromagnetic Articulograph (EMA) for Silent Speech Interface

J. Wang, S. Ravi, B. Cao, N. Sebkhi, A. Bhavsar, O. Inan, W. Xu

Recent work has encouraged the possibility of silent speech interfaces (SSI) as a means of oral communication for patients after laryngectomy or with other voice disorders. Such solutions, however, are currently not feasible for daily operation due to their lack of wearability. We therefore developed a wearable SSI system based on permanent magnet localization with an inertial measurement unit. This novel device generates x, y, z position data along with orientation and magnetic field information of the tracer on the tongue. In this preliminary test, we trained a support vector machine (SVM), deep neural network (DNN), and convolutional neural network (CNN), to classify eight isolated vowels collected from healthy speakers. Data were collected at home while participants were sitting or standing and able to freely move their head and body. The average accuracy was 90.85%, 88.00%, 90.13% for the SVM, DNN, and CNN implementations respectively, which are similar to previously reported results using a commercial EMA (AG500) data for the same classification task. Our preliminary findings demonstrated the potential of this wearable tongue motion device for SSI and other applications.

Real-world Assessment of DDK Rate in Dysarthric and Healthy Speech

P. Kadambi, J. Liss, V. Berisha, S. Hahn

Changes in speech are often harbingers of neurodegenerative diseases such as Parkinson's and amyotrophic lateral sclerosis (ALS) [1,2]. In early disease, maximum performance tasks uncover mild or emerging impairments in speech motor control. One such task is diadochokinesis (DDK), in which a speaker produces a sequence of syllables in rapid succession ("pa ta ka"). The DDK rate, expressed in syllables per second, is used to assess articulatory function. DDK reductions in ALS are detectable prior to decline in speaking rate or intelligibility in early bulbar onset ALS [3]. However, the in-clinic elicitation and manual scoring are barriers to patients with impaired mobility and to SLPs with limited resources. Development of a robust and reliable remote and automated alternative can address this. This paper presents an algorithm that uses an automatic speech recognition (ASR) model [4], for automated DDK rate calculations from recordings of repeated productions of the word "buttercup" elicited via mobile app. Analytical and clinical validation demonstrate the value of the algorithm in reliably assessing DDK rate remotely.

Collecting remote voice and movement data from people with Parkinson's disease (PD) using multimodal conversational AI: Lessons learned from a national study

A. Exner, V. Ramanarayanan, D. Pautler, S. Snyder, H. Kothare, J. Liscombe, S.Sridhar, O. Roesler, W. Burke, M. Neumann, D. Suendermann-Oeft, J. Huber

Telehealth is increasingly gaining recognition in the fields of neurology and speech-pathology (SLP) as a means of addressing accessibility issues experienced by people with PD. We present one potential solution -- a conversational artificial intelligence (AI) agent, Tina, with whom people with PD can interact from the comfort of their homes while she guides them through a customizable assessment. The purpose of this presentation is to discuss lessons learned from deploying and testing this system in an ongoing national study with PD patients. 43 people with PD and 16 age- and sex-matched control participants have been enrolled in the study, with study recruitment continuing. Participants completed four assessment visits, one per week at times that fit their schedule. Issues related to internet connection were anticipated, but were not as impactful as we anticipated. Some issues related to PD were anticipated, including the potential for cognitive impairment to make it difficult for people with PD to execute a session independently with the dialog agent. This was not a large problem in the study. One PD-related issue that was unanticipated was problems with the system voice activity detection routine which occasionally terminated tasks early due to low vocal intensity or long pauses in the speech of people with PD. Throughout the course of the study, we have adjusted criteria for stopping tasks to mitigate these issues. Issues related to caregiver interactions were not anticipated. Caregivers were often involved in helping participants with PD to connect to the system and completing the assessment. Occasionally the web cameras picked up the caregiver and acoustic samples sometimes included speech produced by the caregiver. This study demonstrates the feasibility of collecting acoustic and video data from people with PD for the purposes of speech and motor assessments.

A multifactorial facial surface EMG analysis as a novel objective assessment tool for detecting and measuring subclinical speech changes in amyotrophic lateral sclerosis

P. Rong, G. Pattee

A novel computer-assisted multidimensional facial surface electromyographic (EMG) analysis was developed to detect and measure neuromuscular changes related to speech impairment in ALS. Fifty-four linear and nonlinear features were extracted from the surface EMG recordings for the masseter, temporalis, and anterior belly of digastric muscles in 13 patients with ALS and 10 healthy controls, each performed a speech task three times. Through factor analysis, the features were factorized into five internally-consistent, interpretable factors, representing the functioning of masseter, temporalis, digastric, antagonist muscles, and agonist muscles, respectively. These factors explained 40-43% of the variance in the functional speech outcomes and were >= 95% accurate in patient-control classification. Two critical jaw muscular changes were identified in ALS, reflecting (1) decreased motor unit recruitment and synchronization for jaw antagonists, and (2)

increased recruitment for temporalis, attributed to a potential neuromuscular adaptation. The surface EMG analysis therefore shows promise as an automated, non-invasive, objective, and quantifiable means of assessing subclinical neuromuscular changes in the speech motor system related to both upper and lower motor neuron pathologies. Such an analysis may provide a much-needed objective assessment tool for detecting and monitoring speech impairment in ALS, which has important management implications.

Dissociating Dysarthria and Dialect in the Kinematics of Connected Speech

J. Berry, Y. Kim

Given the paucity of standards for collecting and interpreting speech-kinematic data and the limited use of such methods in studying speech-dialectal differences, it is not surprising that almost no bases for dissociating the impact of dialect from dysarthria have been established for speech movements. Our current work analyzes kinematics during connected speech in typical talkers and talkers with dysarthria from two American English dialects (Southern and Midwestern). 45 talkers (15 with dysarthria) read a connected speech passage while movements were registered along the tongue, lips, and jaw. Initial analyses characterizing flesh point range of motion revealed substantive differences between groups, suggesting complex effects of both dialect and dysarthria. Current analyses focus on speed, acceleration, and jerk. The aim of the work is to more effectively dissociate dialectal influences on speech movements from those associated with dysarthria to improve clinical assessment.

Automating Vowel Formant Estimations for Disordered Pediatric Speech Samples

H. Valentine, G. Carozzi, M. Grigos

While acoustic estimation of formants is a long-standing practice for analyzing vowel production, it is rarely applied to disordered pediatric populations. Current best practices are either incredibly time-consuming and inaccessible or use shortcuts that do not readily apply to diverse datasets. This study examines how simple coding within acoustic speech software effectively automates critical steps in formant estimation for disordered pediatric speech sound samples. A custom script is used to manipulate formant extraction settings and identify the steadiest portion of monophthong vowels from single word productions of young children with childhood apraxia of speech. This approach is contrasted against traditional methods of manual software manipulation and vowel midpoint extraction. Data loss from both methods are compared. The feasibility and effectiveness of automated acoustic analysis for disordered pediatric populations are discussed.

Electromagnetic articulography is feasible for assessment of
speech motor skills in cochlear implant users

M. Masapollo, Y. Oh, J. Goel, J. Lowenstein, S. Nittrouer

Cochlear implants (CIs) have led to stunning advances in prospects for individuals with severe-to-profound sensorineural hearing loss to acquire and/or maintain speech production skills, but problems persist. To investigate speech production in individuals with CIs in a comprehensive manner requires a methodology for directly recording speech motor actions, such as electromagnetic articulography (EMA). However, given that both CIs and EMA make use of magnetic fields, concerns of possible cross interference exist that must be addressed if we are to refine our investigations. We here had two distinct goals: first, to test whether CIs affect the recordings obtained from EMA; and second, to test whether EMA interferes with the signal processing of CIs. In an initial experiment, we calibrated a set of EMA sensors in three conditions: (1) with the CI absent; (2) with the CI (and electrodes) in the EMA field, but with the emulator off; and (3) with the CI (and electrodes) in the EMA field and turned on (Experiment 1). We also measured impedances of all CI electrodes when not in the EMA field, and in the EMA field (with EMA off and on). In a subsequent experiment (Experiment 2), speech movements were recorded with simultaneous audio for a NH adult female talker in two conditions: (1) with the CI absent; and (2) with the CI (and electrodes) in the EMA field and turned on. Collectively, our findings indicate that there is no evidence of cross interference between current EMA technology and CIs, and that EMA data can reliably be obtained from CI users.

Beyond Speech Intelligibility: Quantifying Listener Effort and Engagement with Dysarthric Speech

A. Fletcher, A. Wisler, E. Gruver, S. Borrie

This study examines whether listening to dysarthric speech requires greater levels of engagement than listening to healthy speech samples. It also explores whether increased listener engagement is likely to occur when dysarthric speech samples are fully intelligible. To address this question, we focus on a dual-task paradigm. In dual-tasks paradigms, two tasks are performed simultaneously. Reduced performance on a secondary task suggests that the primary task (parsing dysarthric speech) requires more cognitive resources (i.e., greater engagement).

The results of this dual-task experiment will be discussed in conjunction with corresponding listener effort ratings. Implications on communicative participation will be considered, and further analysis of the effects of sentence predictability will also be presented.

Validity of online speech data collection from children with dysarthria due to cerebral palsy

K. Hwang, Y. Chang, V. Berisha, M. McAuliffe, A. Stump, E. Stender, I. Naim,
E. Hong, J. Choi, R. Brisman, J. Avobich, E. Levy

Children with dysarthria due to cerebral palsy often have mobility restrictions and may face barriers to in-person speech-language pathology services and research participation. Therefore, online data collection may be a more effective way to gather their speech data if measures extracted from such data are valid. This study assessed the validity of online speech data collection from children with dysarthria by comparing various acoustic measures extracted from online recordings to those extracted in-person from an audio-recording device. The spectral measure of F2-range of diphthong and temporal measures of fricative-duration difference and articulation rate revealed correlations between recording conditions with r-values greater than .70, whereas F0 variation did not. Our findings indicate that, pending results from further studies, these spectral and temporal measures may lend themselves to valid online speech-data collection from children with dysarthria.

What can passive smartphone metrics tell us about communicative participation in ALS?

K. Connaghan, J. Green, A. Haenssler, Z. Scheier, M. Keegan, A. Clark, K. Burke, JP. Onnela, J. Berry

People with amyotrophic lateral sclerosis (PALS) often experience restrictions to communicative participation. Technological advances provide novel opportunities for monitoring communicative participation through active (speech recordings, surveys) and passive (GPS, communication logs) data streams. The current study explored the utility of passive smartphone metrics to index and monitor communicative participation by PALS. Survey (Communicative Participation Item Bank-short form), GPS, and communication log data were collected using the Beiwe smartphone research platform. Preliminary findings indicate moderate associations between time spent at home and communicative participation, bulbar, and motor function. Communication log data suggest that impaired speech and impaired fine motor function affect call and text behavior respectively, and warrant further investigation. Passive data collection provides a promising approach for addressing communicative participation decline in ALS.

Multimodal dialog based remote patient monitoring of motor function in Parkinson's Disease and other movement disorder

O. Roesler, W. Burke, H. Kothare, J. Liscombe, M. Neumann, A. Cornish,
D. Habberstad, D. Pautler, D. Suendermann-Oeft, V. Ramanarayanan

The need for remote monitoring to support Parkinson's Disease (PD) patients, caregivers, and healthcare professionals in their collaborative efforts for better care has never been greater. We have earlier presented Tina, a virtual agent that is powered by a cloud-based multimodal dialog system designed to conduct automated screening interviews that elicit evidence for detection and progress monitoring of neurological disorders like PD. This work presents our innovations to this system to also allow for assessment of patient motor function via finger-tapping exercises. Most existing contactless RPM systems assess dexterity of finger movements with the help of smartphone apps in which a variant of the finger tapping task is used by asking the user to alternately tap buttons on the screen. In contrast, Tina engages participants in a conversation consisting of structured conversational exercises designed to elicit specific speech, facial, and now limb motor behaviors, thereby, combining the benefits of cost-effective, frequent remote monitoring with the ability for clinicians to review the performance. For the finger tapping exercises, participants are instructed to hold either their left or right hand up to the camera and perform a tapping motion for five seconds, while anatomical landmarks of the participant's hand are derived from the recorded image frames. The finger tapping assessment comprises three tasks that differ based on the instructed goal, i.e. the movement should be as (1) wide, (2) fast, or (3) both wide and fast as possible. Key points of interest during the test are the duration and amplitude of a cycle. If either the duration of the cycle increases or the amplitude decreases, the system will identify this as a positive test result, which should be assessed further. This work demonstrates the feasibility of novel finger-tapping exercises administered through a multimodal dialog agent for assessment of motor function in PD and other movement disorders.

On the robust automatic computation of speaking and articulation duration in ALS patients versus healthy controls

J. Liscombe, A. Ocampo, H. Kothare, O. Roesler, M. Neumann, A. Cornish,
D. Pautler, D. Suendermann-Oeft, V. Ramanarayanan

Speaking and articulation duration have been shown to be important biomarkers of disorder severity in ALS and other motor speech disorders. While signal processing algorithms afford significant advantages over manual annotation in that they allow for scalability and automatic computation of these and other biomarkers at scale, we must ensure that they are accurate and reliable. This work investigates the various parameters of 2 automated algorithms used to compute speaking and articulation duration for speech data collected via a conversational AI agent and estimates, via simulated tuning experiments, optimal settings for bulbar symptomatic amyotrophic lateral sclerosis (ALS) patients in comparison to healthy controls. We also uncover non-intuitive differences in optimal parameter settings required for robust computation of articulation duration versus speaking duration. We found that optimal settings for the automatic prediction of articulation and speaking duration were dependent on both task and cohort type.

Measuring Articulatory Compensation in ALS using a Data-Driven Consonant Distinctiveness Space Approach

K. Teplansky, A. Wisler, J. Green, D. Heitzman, S. Austin, J. Wang

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease that causes muscular weakness and atrophy. In early ALS, patients often show a reduced speaking rate yet maintain high speech intelligibility until some unknown critical point in disease progression. It remains unknown if the initial ability to preserve speech intelligibility is due to an articulatory compensation strategy. Using the NDI Wave System, tongue and lip movement patterns from speakers with ALS (n = 19) were tracked during production of 10 consonants (/b/, /g/, /w/, /v/, /d/, /z/, /l/, /ɹ/, /dʒ/, /j/). Participants were

stratified into three groups (pre-symptomatic, compensation, post-compensation) based on their speech intelligibility and speaking rate scores. A comprehensive articulatory consonant distinctiveness space (ACDS) approach was used to analyze the distinctness of consonants. We also used a support vector machine learning model to classify the consonants in each group. Results show a larger ACDS area in the compensatory group than the pre-symptomatic and post-compensation group. Findings of higher classification accuracy in the pre-symptomatic group (91.58%) than the compensatory (86.73%) and the post-compensatory (85.36%) did not correspond to an increase in consonant distinctiveness. The increase in ACDS area may reflect a potential compensation mechanism in early ALS that is activated in response to neuropathological changes.

Detection of COVID-19 With Vocal Biomarkers Based on Motor Coordination and Phonetic Approaches.

T. Talkar, D. Low, D. O'Keeffe, A. Simpkin, S. Ghosh, T. Quatieri

We sought to determine if measures of neuromotor coordination and phoneme-based features derived from recordings of simple speech tasks could aid in detecting the presence of COVID-19 and characterizing the effect of COVID-19 on speech production subsystems. In this pilot analysis, we utilized recordings from a dataset with 12 COVID-19 positive patients and 15 COVID-19 negative patients at University Hospital Galway performing read speech, free speech, counting, sustained vowel, and diadochokinetic sequence tasks. From these recordings, we extracted out a set of acoustic time series to determine eigenvalues from correlation matrices for each subject. We additionally extracted phoneme-based features utilizing the KALDI AspIRE chain model. Both of these features were used to create machine learning models to discriminate between the COVID-19 positive and COVID-19 negative groups. Eigenvalues from the harmonic-to-noise ratio (HNR) time series derived from read speech were able to discriminate between the two groups with an accuracy of 0.95, and suggested that HNR was more constant over time for individuals with COVID-19 as compared to healthy individuals. Additionally, the analysis identified that the velocity of fundamental frequency and speech envelope was more erratic in individuals with COVID-19 as compared to healthy individuals. Phoneme-based features, comprising of rate and duration based features, were able to discriminate between the two groups with an accuracy of 0.90, suggesting decreased rate of articulation and increased durations of phonemes in individuals with COVID-19. The results from this analysis highlight the promise of using nonintrusive sensing through simple speech recordings for early warning and tracking of COVID-19.