Original Project Proposal for the Speech Accessibility Project

The Speech Accessibility Project was originally proposed as the "Consortium Nova Habilitas." The name "Speech Accessibility Project" was the result of brainstorming sessions after the project began, and is universally agreed to be a much better description of the project goals.

This proposal was written on March 6, 2022. This copy of the proposal has been edited slightly, on October 10, 2023, in order to remove images, names, and text references that are the property of organizations other than the University of Illinois. This text, in this form, is distributed with permissions as specified in the standard open-source MIT License (see accompanying file, LICENSE).

Proposal: Consortium Nova Habilitas: Individuals with disabilities helping researchers to improve technology

I. Purpose of the consortium

The purpose of the Consortium Nova Habilitas ("**NOVA H**") is to support the creation of technologies that improve the quality of life for people with disabilities. The consortium will collect data contributed by individuals with disabilities, create metadata and other types of shared infrastructure, and serve as a communication channel by which individuals with disabilities or other diverse technology needs can provide relevant support to researchers and developers interested in serving those needs. The overall concept is to develop and maintain a shared repository built from open-source software that houses de-identified private data to support the proprietary technology goals of our consortium sponsors.

II. Institutional Excellence

The NOVA H team from the University of Illinois Urbana Champaign (UIUC) is drawn from world-renowned campus centers of excellence in speech and hearing science, linguistics, engineering, interdisciplinary and translational research. There are four main reasons why UIUC is the most suitable host for this industry consortium: (1) our commitment to diversity including people with disabilities; (2) our excellent Departments of Speech and Hearing Science and Linguistics; (3) our first-class Departments of Computer Science and of Electrical and Computer Engineering that include groundbreaking work in automatic speech recognition for special populations; and (4) business, technical, and program support at the Beckman Institute for Advanced Science and Technology that leverages the excellence of our campuswide resources.

A. Disability resources and services at the University of Illinois

The University of Illinois is a pioneer in accessibility[1]. The University was the first post-secondary institution to provide a support service program enabling students with disabilities to attend.[2] In 1947, the University of Illinois created a branch campus to support the education of returning WWII veterans, where students started the first university wheelchair basketball team in 1948. In 1949, these programs were moved to the Urbana campus after student wheelchair users demonstrated that they could navigate the physical rigors of the campus and the academic rigors of the state's flagship university. UIUC was a major contributor to the ANSI

[1] https://www.disability.illinois.edu/history-firsts

[2] Commemorative Book Preparation and Publication Committee, "Expanding Horizons: A History of the First 50 Years of the Division of Rehabilitation-Education Services at the University of Illinois," Roxford DT Publishing", Champaign, IL, 1998

A117-A architectural standards for accessibility, initially developed at the University in 1961. Today, the University of Illinois is a training site for the U.S. Paralympics Team,[3] and its office of Disability Resources and Educational Services (DRES) supports a wide array of programs to support students with disabilities, including specialized transportation, housing, athletics, counseling, and other services. Companion research programs take place throughout campus.

B. Speech and Hearing Science and Linguistics at UIUC

Dr. Severina Nelson founded the first speech therapy service at UIUC in 1938 and was Director of the Speech and Hearing Clinic from 1939 to 1959. The Department of Speech and Hearing Science (SHS) was established in 1973. Graduate degrees offered by the department include the M.A.-Clinical, M.A.-Nonclinical, and Ph.D. in Speech and Hearing Science, and the Doctor of Audiology. The M.A.-Clinical degree meets the educational requirements for certification in all 55 U.S. states and jurisdictions. As part of graduate student training, the Speech-Language Pathology clinic[4] was founded in 1975. The current facility serves patients from Champaign-Urbana and around the state and supports research at the University. The UIUC Department of Linguistics is one of the oldest, largest, and most distinguished departments in the country[5]. From its start as an English as a Second Language program in 1946 and the establishment of the department in 1965, today the department continues to support highly-regarded research in phonetics, phonology and language processing.

C. Computer Science and Electrical and Computer Engineering at UIUC

UIUC's Department of Electrical and Computer Engineering (ECE) has hosted groundbreaking research in integrated circuits, transistors, light emitting diodes, antenna technologies and more for decades. UIUC's Department of Computer Science has likewise led research in parallel processing; the National Center for Supercomputing Applications (NCSA) was founded at UIUC in 1986 based on our leadership in this area. UIUC's top-ranked programs in both ECE and CS have attracted outstanding faculty and students for generations. Of particular relevance to this proposal are the groups of faculty members who have focused on automatic speech recognition (ASR). The UASpeech research effort is a collaboration between ECE, Linguistics, SHS and DRES researchers that has flourished since 2008.

The UASpeech corpus[6] was developed so that researchers worldwide could develop ASR for people with dysarthria. Since then, ASR training corpora for people without dysarthria have grown by two orders of magnitude, but UASpeech remains the largest distributable corpus of dysarthric speech. Other than UASpeech, most researchers in assistive speech technology depend on just two widely used dysarthric speech corpora: TORGO, and Nemours. TORGO and Nemours each contain data from fewer contributors than UASpeech (TORGO: 7, Nemours: 11, UASpeech: 16), but more speech per participant (TORGO: 774 isolated words + 634 sentences, Nemours: 814 sentences; UASpeech: 765 isolated words). Metadata distributed with UASpeech include dysarthria type and intelligibility. UASpeech has had a significant impact on the field. Distribution of the UASpeech corpus to researchers on every continent has led to the

[3] https://www.paralympic.org/news/university-illinois-announced-us-paralympic-training-site
[4] https://ahs.illinois.edu/slp-clinic
[5] https://linguistics.illinois.edu/directory/about-department
[6] H Kim, M Hasegawa-Johnson, A Perlman, J Gunderson, T Huang, K Watkin & S Frame, "Dysarthric Speech Database for Universal Access Research," INTERSPEECH 1741-4, 2008

publication of hundreds of papers by scores of research teams. Speech recognition error rates using UASpeech have dropped by almost a factor of three since the corpus was first released.

D. Program management and business and technical support

UIUC provides excellent business and administrative support for complex arrangements funded by industry. UIUC already hosts several industrial consortia, including the Industrial Partners Program of the National Center for Supercomputing Applications (NCSA), started in 1992. At the Beckman Institute for Advanced Science and Technology, complex projects funded by industrial sponsors include the Abbott-funded Center for Nutrition, Learning, and Memory and the GSK Center for Optical Molecular Imaging.

Interdisciplinary research at the Beckman Institute for Advanced Science and Technology is supported by outstanding campuswide information technology infrastructure, and specifically by Beckman's Information Technology Services (ITS), a group of system engineers, programmers, IT specialists, and part-time student staff members under the supervision of Co-PI Erik Hege. Beckman ITS manages on-site and cloud storage for faculty and staff, manages access to the Beckman Institute's computing resources on the NCSA Campus Cluster, and creates software to navigate data access for collaborations between the Beckman Institute and local hospitals.

Relevant campus-level resources available to NOVA H include a system for compensating experimental participants using Amazon ecodes, implemented by our Treasury Operations office. Our campuswide Sponsored Projects Administration (SPA) office has already dedicated negotiators to our team to make NOVA H a reality.

III. NOVA H Management Approach and Management Team

The overall philosophy of our program management approach is a blend of collaborative team science and human-centered design. We will engage two types of collaborators: (1) sponsors to invest and advise, who will be organized into an **Advisory Committee** and a **Technical Advisory Committee**, and (2) partners to support and advocate for contributors to the repository, who will be organized into an **Advocate Advisory Committee** and **a Contributor Community**. The team science approach is a balance of interdisciplinary collaboration, flexible support for research innovations, respect for partners' needs, and rigor in traditional project management approaches and tools. We will rely on continuous engagement with sponsors and partners to iteratively develop and test approaches for the system architecture, tools, outreach, annotation, and delivery of resources to consortium members.

The NOVA H management team includes:

- Principal Investigator **Mark Hasegawa-Johnson**
- Program Management Lead **Patricia Jones**
- Text Annotation Director **Heejin Kim**
- SLP Annotation Director **Clarion Mendes**
- Outreach & Recruitment Director **Meg Dickinson**
- Information Technology & Infrastructure Development Director **Erik Hege**

Principal investigator Dr. Mark Hasegawa-Johnson is a Professor in the Department of Electrical and Computer Engineering, where he currently holds one of four appointments as an

Everitt Faculty Scholar. He is a Fellow of the Acoustical Society of America (for contributions to vocal tract and speech modeling, 2011) and of the IEEE (for contributions to speech processing of under-resourced languages, 2020), was Treasurer of the International Speech Communication Association from 2013-2020 and is currently Senior Area Editor of the IEEE Transactions on Audio, Speech, and Language.

Dr. Patricia Jones is the Associate Director of Research at the Beckman Institute and brings 20 years of program and project management experience to the NOVA H team. She is a former member and chair of the UIUC biomedical Institutional Review Board (IRB) and is highly experienced in human subjects research. She is currently the part-time project manager for the Champaign-Urbana Population Study, a joint study by UIUC and Carle Foundation Hospital, a longitudinal, cross-sectional study of brain health in the community. Dr. Jones is well-positioned to support operational activities of the consortium, ranging from contractual requirements for meeting milestones, project management strategy and tracking for technical, schedule, and budget risks, compensating remote contributors through our campus Treasury Operations office, and assisting the technical and outreach teams in meeting project objectives.

Dr. Heejin Kim is a Research Assistant Professor in the Department of Linguistics. Dr. Kim's research focuses on investigating the acoustic and perceptual characteristics of dysarthric speech and on developing effective communication aids for people with motor speech disorders and their conversation partners. Dr. Kim took full responsibility for acquiring audio-video recordings of speakers in the UA Speech Database as well as for obtaining human listeners' orthographic transcriptions of the dysarthric speech. Dr. Kim is experienced in the linguistic annotations of speech and in conducting phonetic analyses of dysarthric speech. Dr. Kim has successfully supervised undergrad and graduate students and collaborated with other researchers.

Clarion Mendes is a Clinical Assistant Professor and speech-language pathologist certified by the American Speech-Language and Hearing Association (ASHA). She has taught the graduate level course in Motor Speech Disorders at the University of Illinois for the past 8 years. Clarion has been a clinician intervening with populations with acquired neurogenic speech and communication disorders (e.g., Parkinson Disease, CVA). She has received specialty training in two interventions for Parkinson Disease. Clarion has more than a decade's experience in the diagnosis and intervention of acquired communication disorders, and is involved in local Parkinson Disease outreach.

Meg Dickinson is the Director of Communications at the Beckman Institute for Advanced Science and Technology, where she leads outreach, media relations, social media, email marketing, and web initiatives. Meg's support of this project will include working with outreach partners, campus' ad buyer, and guiding production of all related advertisements and communications.

Erik Hege is an IT management professional with over 20 years of experience, including management of personnel, budget, and facilities. Erik was Director of IT Operations at NCS Pearson from January through July of 2002, then joined the University of Illinois. He was the Assistant Dean for Infrastructure at the Graduate School of Library and Information Science from 2008-2019, then became the Director of Information Technology Services at the Beckman Institute.

IV. Outreach and Recruitment of Contributors

Figure 1 schematizes the plan for outreach and recruitment of contributors. Phase 1 of this project will target one etiology; Phase 2 will continue to target the Phase 1 etiology, and will also target four additional etiologies. The outreach plan is different for each etiology, to best support the needs and diversity of each diagnosis and its community. Each outreach plan has been designed with advice from our collaborating etiology experts, as described in the remainder of this section. A special consideration for NOVA H is that three states (IL, TX, WA) have privacy laws restricting the use of "voice print" technologies, and so contributors will not be recruited from those states.

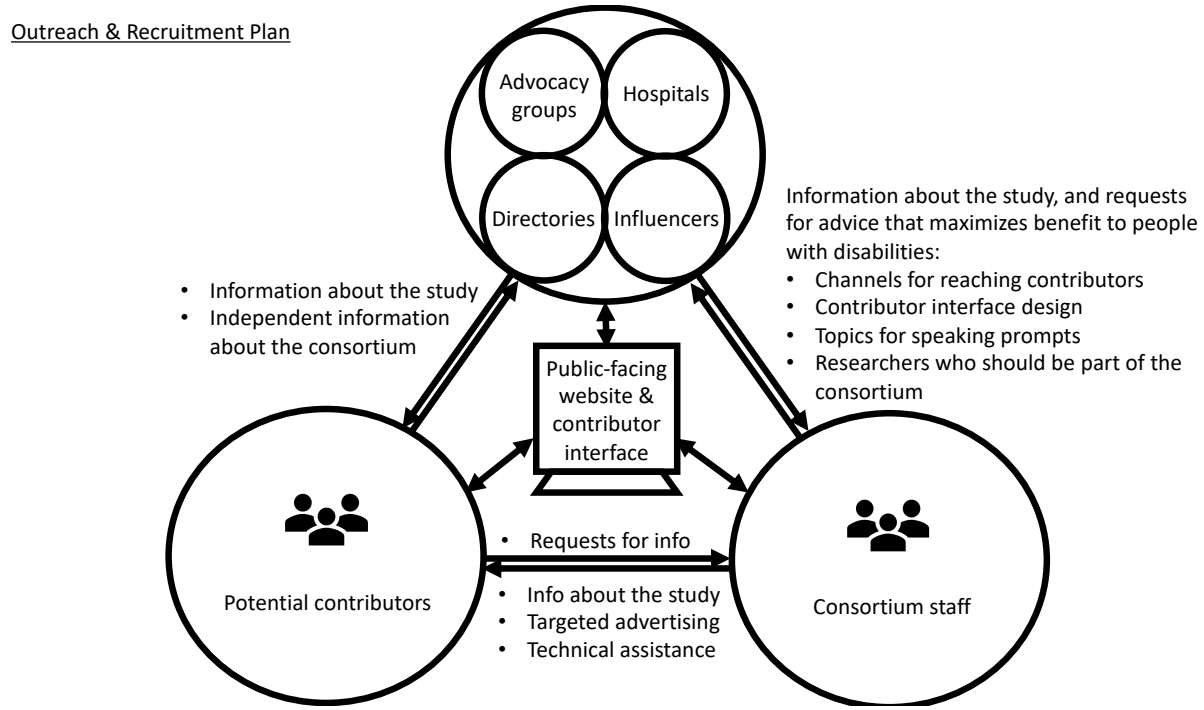

Figure 1. Plan for outreach and recruitment of contributors.

For every etiology, we will work closely with advocacy groups, hospitals, research registries and media. Advocacy groups will be invited to place members on our **Advocate Advisory Committee**, which will be particularly active as we design the contributor interface, so that we can make sure it is successfully adapted to the needs of each etiology. Among the functions envisioned for the Advocate Advisory Committee are: (1) members of the Advocate Advisory Committee will be asked to help us design the prompts for the contributor interface, (2) members of the Advocate Advisory Committee will recommend participants for focus groups that test the contributor interface in months 6 (Phase 1) and 13 (Phase 2) of the consortium, (3) we will work with the Advisory Committee to develop a Data Use Agreement that can be signed by members of the Advocate Advisory Committee, giving them the ability to browse metadata (but not audio files) in the contributed speech database.

A. Phase 1

Phase 1 will focus on recruiting people with Parkinson Disease (PD). PD is commonly cited as the most common motor disorder related to aging[7]: more than 1 million people in the US[8] are affected by PD, 90% of whom eventually suffer from dysarthria.[9]

Co-PI Mendes is certified in LSVT and trained in SPEAK OUT! and LOUD Crowd. She manages a weekly speech intervention group for individuals with PD. She has worked closely with the Parkinson's Disease Support Group of Champaign County since 2015. Although NOVA H will not recruit participants from the State of Illinois, the insights she gains by working with this population have helped us to form connections that will permit robust recruitment across the Midwest and nationwide.

Co-PI Mendes notes several other attributes of PD that make it an ideal etiology for Phase 1 recruitment: (1) In her experience, people with PD are eager to participate in research if they are apprised of updates on the research. (2) People with PD are usually elderly and are therefore "not digital natives:" if speech technology could be made effective for them, it might simplify their interactions with technology and assist with aging in place. (3) More than 90% of people with PD suffer from dysarthria eventually,[10] and their dysarthria causes typical ASR to fail. (4) The speech patterns symptomatic of PD (hypokinetic dysarthria) are less variable across individuals than the speech patterns of some other etiologies: a trained clinician can often immediately diagnose PD upon hearing the patient's speech. Because of the relatively low degree of variability across individuals, it is likely that an ASR trained using a large multi-speaker corpus of hypokinetic dysarthric speech could be deployed with very low error rate for the use of other people with PD.

B. Phase 2

Phase 2 will expand beyond Parkinson's Disease to include four more etiologies. Etiologies in Phase 2 will be selected in consultation with the Advisory Committee. Possible etiologies for Phase 2 include, for example, Multiple Sclerosis (MS), Down Syndrome (DS), Cerebral Palsy (CP), Traumatic Brain Injury (TBI), Amyotrophic Lateral Sclerosis (ALS), Stroke, Deafness, and Stuttering.

MS research in Urbana is facilitated by the Illinois MS Research Collaborative[11], a group of faculty spanning the Colleges of Engineering, Applied Health Sciences, and Liberal Arts and Sciences, with a shared interest in mobility and MS. About 45% of people with MS have mild increasing to severe dysarthria.[12] In order to recruit participants quite broadly from across the United States, we will make use of the NARCOMS registry for Multiple Sclerosis.[13] NARCOMS

---

[7] TR Mhyre, JT Boyd, RW Hamill & KA Maguire-Zeiss, "Parkinson's Disease," Subcell Biochemistry 65:389-455, 2012

[8] R Savica, B Grossardt, WA Rocca & JH Bower, "Parkinson disease with and without Dementia: a prevalence study and future projections," Movement Disorders 33(4):537-43, 2018

[9] G Moya-Galé & ES Levy,"Parkinson's disease-associated dysarthria: prevalence, impact and management strategies," Research and Reviews in Parkinsonism 9:9-16, 2019

[10] https://www.nature.com/articles/s41598-020-68754-0

[11] https://msresearch.illinois.edu/

[12] FL Darley, JR Brown & NP Goldstein, "Dysarthria in multiple sclerosis," J Speech Hearing Rsrch 15(2):229-245, 1972

[13] https://www.narcoms.org/

offers targeted recruitment of experimental participants with MS, at a cost of $1,200 for initial queries plus $5 per individual. A validation study on this research registry[14] had a 20% rate of people who consented to participate; if the same success rate is true for us, we could recruit 200 people by targeting 1,000.

DS is the research focus of Applied Health Sciences faculty including Laura Mattie[15] and Marie Moore Channell[16]. Connections to contributors will be made through their respective research registries, through advocacy organizations including the National Down Syndrome Society[17] and the Canadian Down Syndrome Society[18], and through registries of people with DS who have opted-in to future research studies. One such registry is available through the NIH's DS-Connect using Level 3 access.[19]

CP is an ongoing research focus of Dr. Hasegawa-Johnson and Dr. Kim, and of Applied Health Sciences Faculty including Clitlali Lopez-Ortiz.[20] The primary advocacy organization for people with CP is United Cerebral Palsy (UCP).[21] UCP has a small national staff, serving primarily to organize communication among and on behalf of its regional and local associate organizations. UCP Heartland[22], for example, serves participants in St. Louis and in other cities in Missouri, providing services including adult day habilitation, autism services, employment training & placement, independent residential living, and a directory of services for people with CP.

Research in TBI at UIUC is famous in part because of our long history of serving disabled veterans. The Chez Veterans Center[23] provides academic services, VA services, peer mentoring, career mentoring, a counseling center, a gym, an academic and career center, and residential units to veterans, especially to veterans with disabilities. University of Illinois faculty specializing in TBI include Aron Barbey[24] and Sandraluz Lara-Cinisomo.[25] Information about this study will be communicated to potential contributors with TBI via potential collaborators at VA Hospitals, beginning with the research contacts of Dr. Barbey and Dr. Lara-Cinisomo.

C. Phase 3

We plan to explore the potential for a Phase 3 with our Advisory Committee. Phase 3 represents potential expansion of the project into a number of directions, including possibly the inclusion of speech data in languages other than English, the inclusion of new speaking tasks (e.g., in non-home environments), and the inclusion of non-speech data (e.g., video). Phase 3 may also include the development of new consortium membership frameworks including

---

[14] R.A. Marrie and G. Cutter and T. Tyry and D. Campagnolo and T. Vollmer, "Validation of the NARCOMS registry: diagnosis," Multiple Sclerosis 13(6):770-5, 2007

[15] http://dndlab.shs.illinois.edu/

[16] https://idclresearch.web.illinois.edu/

[17] https://www.ndss.org/

[18] https://cdss.ca/

[19] https://dsconnect.nih.gov/for-professionals/read-the-research-review-policy.html

[20] http://danceneuroscience.kch.illinois.edu/

[21] https://ucp.org/

[22] https://ucpheartland.org/

[23] https://chezveteranscenter.ahs.illinois.edu/

[24] https://en.wikipedia.org/wiki/Aron_K._Barbey

[25] https://ahs.illinois.edu/blog/lara-cinisomo-discusses-veterans-studies-wdws

possibly a multi-tier membership framework that permits universities or smaller companies to join the consortium, with data access and governance appropriately scaled to match the membership fee in each tier.

D. Compensation for contributors

NOVA H will seek 1200 recorded sentences per participant with a staged compensation strategy. Participants will be compensated with up to $200 per contributor: $60 after recording the 400th utterance, an additional $60 after recording their 800th utterance, and a final $80 after recording their 1200th utterance.

E. Content and web presence

The Beckman Institute will create and manage the NOVA H website that includes recruitment information, training materials and a portal to the actual secure system for speech data collection. Beckman's Communications Office provides expertise on web content, advertising copy and design, news story writing, email campaigns, media relations, direct outreach to registries and advocacy groups, and video.

IV. NOVA H Technical Approach

The NOVA H data repository will be built on open-source software and will offer customized services for contributors (our human subjects who provide the speech samples), annotators (our NOVA H team members who will annotate the speech samples), and researchers including our founding sponsors.

B. Contributor interface

The contributor interface is schematized in Figure 2.

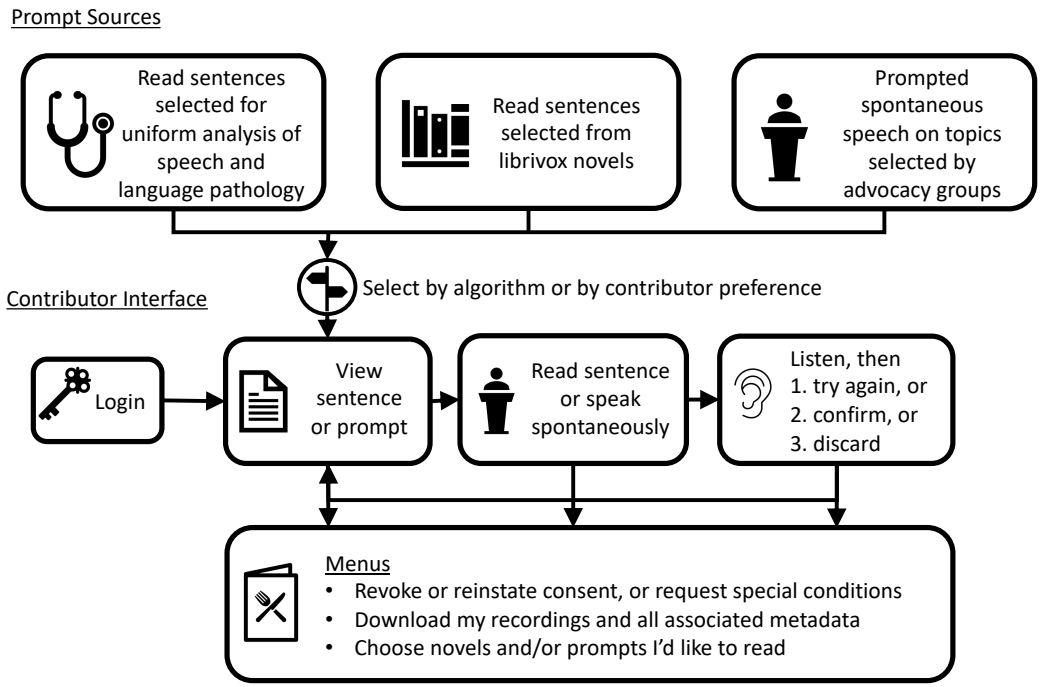

Figure 2. Schematic block diagram of the contributor interface. Prompts are shared sentences, sentences from novels, or spontaneous speech prompts. After logging in, a contributor views a randomly selected prompt, presses a button, reads the prompt, presses another button, listens to the prompt, and then chooses one of three buttons to either try again, confirm the recording, or discard the recording and get a new prompt. Menus are available at any time.

Training materials for the participant interface will be available on the public-facing web page and will be linked from the menu available on every page of the participant interface. People who are interested in contributing will click a "Contribute" button leading to a three-step onboarding process:

1. First, the potential contributor will be given a page with information about the study. If the participant is interested, they will be asked to use dropdown menus to specify their state and/or country of residence, and a broad description of their disability. Their answers to these questions will be automatically screened. If the participant lives in a country or state from which NOVA H is not currently recruiting, or if the participant specifies a disability etiology that NOVA H is not currently recruiting, they will be given an apology page, thanking them for their interest, and asking them to check back later in case future enrollment conditions change. If they live in a country or state from which NOVA H is recruiting, and if their disability etiology is one that NOVA H is currently recruiting, they will be forwarded to the second page.
2. Second, the potential contributor will be given a page with a formal consent form, approved by the University of Illinois IRB. If the potential contributor agrees to the terms of the consent form, they will signify their desire to enroll by entering their e-mail address, choosing a login password, and clicking "consent."
3. Third, the potential contributor will be asked to provide some additional information that will be shared with researchers. The information requested will be designed to

minimize intrusiveness while maximizing utility to researchers and will all be optional. Menu-based responses (e.g., age and gender) will include a "prefer not to specify" option. Optional free-text responses may include information about their disability and a description of their recording environment. Upon completing this page and clicking "submit," contributors will immediately see the main page of the contributor interface.

The main page of the contributor interface will show a text prompt, chosen at random using an algorithm described later in this section. During Phase 1, prompts will be displayed in the form of visual text, as visual text is the most appropriate presentation modality for almost all contributors with PD. In Phase 2 of the project, presentation modalities will be explored with etiology specialists, and with focus groups from each etiology, to make sure that the presentation modality is appropriate for all (or almost all) potential data contributors.

The only other things on the main page of the contributor interface will be a menu bar or button, and five audio control buttons: "Record," "Stop," "Listen," "Save" and "Skip." In the normal flow of contribution, participants will click the "Record" button, read the prompt sentence, then click "Stop." The "Listen" button will play back their recording to them. If they press "Save," then the recording will be saved to the database, and the existing prompt sentence will be replaced by a new prompt. If they press "Skip," the recording will be deleted, and they will be shown a new prompt. A participant will be free to skip any prompt if desired, by pressing the "Skip" button.

The menu will provide additional control, including the ability to change their metadata, revoke consent, download a zip file containing all their own recordings, change the fraction of their sentence prompts that are drawn from novels vs. sentences about daily life, or choose a novel or a genre of novels from which they would like their prompts to be drawn.

The initial consent form will specify that contributors may revoke consent at any time, but that revocation of consent only affects the data stored on the NOVA H server. Revocation of consent does not affect any copies of their recordings that have already been downloaded by researcher members of the consortium. If a contributor has already been paid for contributing 400, 800, or 1200 utterances, revocation of consent will remove those utterances from the database, but the contributor will not be asked to return the money to us. Revocation of consent will disable the main contributor interface, so that the contributor cannot record any additional audio unless they give consent again.

Each contributor prompt will be selected by first randomly choosing from among three different prompt corpora (shared sentences 45%, non-shared sentences 45%, and spontaneous speech 10%), then by randomly selecting a prompt from within the selected corpus. The contents of all three prompt sets will be optimized with the advice of the Technical Advisory Committee and the Advocate Advisory Committee, based on the following plan: (1) Shared sentences will be read by all participants. Shared sentences will be of two types: (a) sentences created by speech pathologists, for the purposes of facilitating the diagnosis of different types of dysarthria, and (b) command and control sentences, designed to simulate the interactions between a contributor and a smart device or smart home. (2) Non-shared sentences will be selected from English-language public domain audiobooks in the Multilingual LibriSpeech (MLS)

corpus.[26] MLS audiobooks in languages other than English will not be used in Phases 1 or 2 of this project but may be useful in Phase 3. Each contributor will be asked to read from a different MLS novel, which will be selected at random if the contributor does not wish to select a novel for themselves; a menu option in the contributor interface will permit the contributor to choose a novel for themselves. Non-shared sentences will be selected from MLS for the purposes of (a) maximizing lexical diversity across the corpus, and (b) facilitating research that uses automatic voice conversion to expand the utility of the created dataset.[27] (3) Spontaneous speech will be elicited by asking the participant to respond spontaneously to a question or idea. Prompts will be developed with the aid of Advocate Advisory Committee and Technical Advisory Committee members during the first six months of the project. Prompts may refer to activities of daily living, or to other subjects. The total duration of recorded spontaneous speech from any contributor will be limited, to avoid tiring the contributor, and to limit the costs associated with spontaneous speech transcription by human annotators.

When a contributor wishes to stop recording, they may simply close their browser. When they wish to begin recording again, the login URL will ask for their e-mail address and password, then take them immediately to the main contributor interface.

C. Curation and Annotation

Curation and annotation will include validation of the consent form and the speech content, and the creation of several types of metadata, as shown in Figure 3.

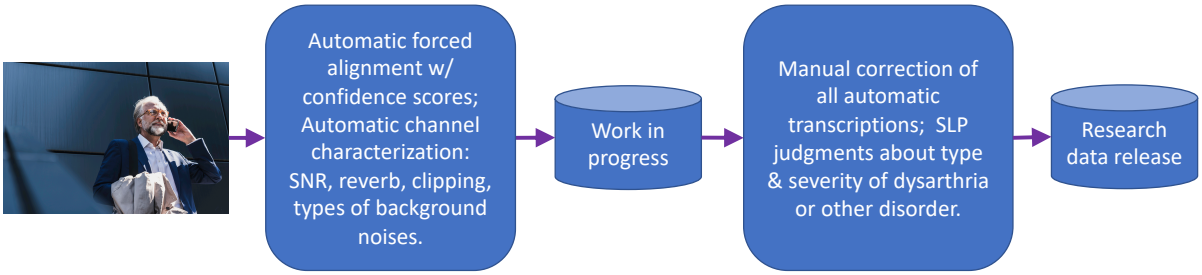

Figure 3. Process chart showing the creation of automatic metadata, text transcripts, and SLP metadata for each speech sample.

After a process to verify consent from each contributor, automatic metadata creation will take place. Every utterance will be validated using an automatic curation process, including automatic channel estimation, automatic speech attribute estimation, and automatic transcription. Automatic channel estimation will estimate the signal to noise ratio and headroom of the recording and will attempt to localize any probable instances of signal dropout, static, and clipping. Automatically estimated speech attributes will include articulation rate, pause durations, and percentage of voicing. ASR will be used to generate an initial transcription of spontaneous utterances that can serve as a basis for corrections made by human annotators. These metadata will be provided to research members clearly marked as automatic annotations.

[26] http://www.openslr.org/94/

[27] J Harvill, D Issa, M Hasegawa-Johnson & CD Yoo, "Synthesis of New Words for Improved Dysarthric Speech Recognition on an Expanded Vocabulary," Proc. ICASSP 2021, pp. 6428-32

Undergraduate RAs under the supervision of Dr. Kim will check every speech recording. In the case of read speech, RAs will create a transcription by correcting the text of the prompt sentence to reflect small errors (e.g., if a word is omitted). Utterances with large errors (e.g., if the contributor said nothing) will be marked as invalid. RAs will also manually transcribe fifteen minutes of spontaneous speech per contributor, beginning with an ASR transcript if they find that doing so improves their transcription speed.

We will also employ speech language pathology (SLP) metadata creation. Speech samples from each contributor (approximately 20 sentences) will be analyzed by two certified speech-language pathologists (CSLPs). CSLPs will each independently decide whether the speech of a contributor appears to have a speech, voice, or language pathology, and if so, to characterize it further to the extent possible. If a contributor is judged to suffer from dysarthria, then a small number of their recorded sentences (approximately 20) will receive two further types of speech pathology annotations. First, the speech sample will be annotated with dysarthria type (flaccid, spastic, ataxic, hypokinetic, hyperkinetic,[28] mixture of any two or more of these, other speech disorder, or no speech disorder). Second, the speech sample will be annotated using the Darley, Aronson & Brown (DAB) differential diagnostic pattern scales.[29] DAB is a set of 46 perceptual attribute scales, including scales related to articulation, intonation, loudness, voice quality, and overall intelligibility.

D. Researcher and Member Admin interfaces

The researcher and member admin interfaces will include three key functions, as shown in

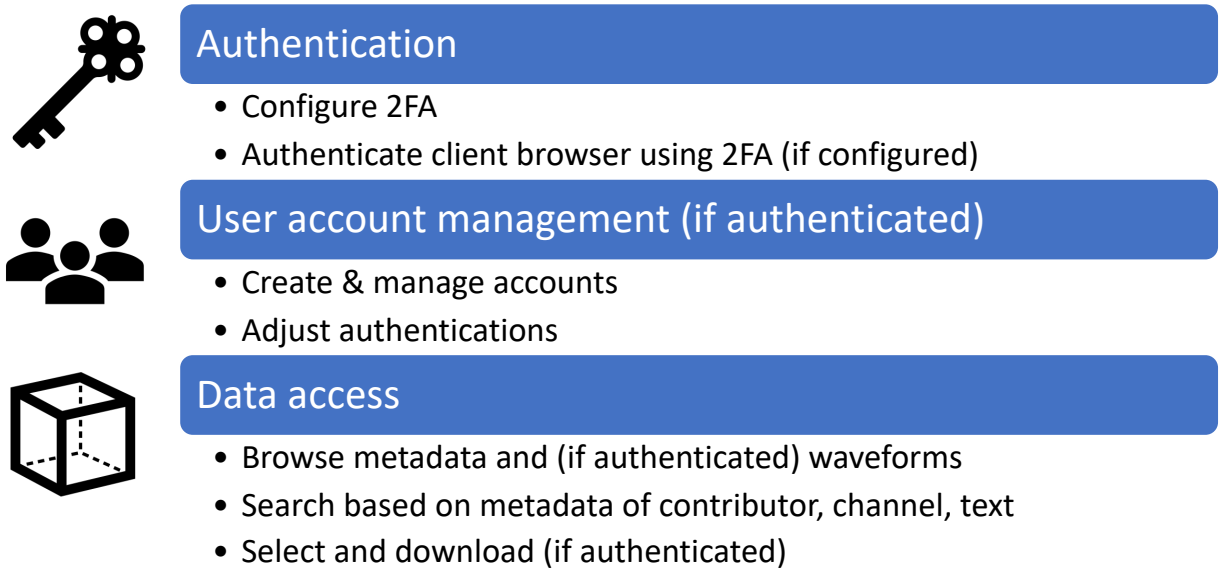

Figure: authentication, user account management, and data access.

[28] FL Darley, AE Aronson & JR Brown, "Differential patterns of dysarthria," J Speech & Hearing Rsrch 12:246-69, 1969

[29] FL Darley, AE Aronson & JR Brown, Motor Speech Disorders. W.B. Saunders, Philadelphia, PA, 1975

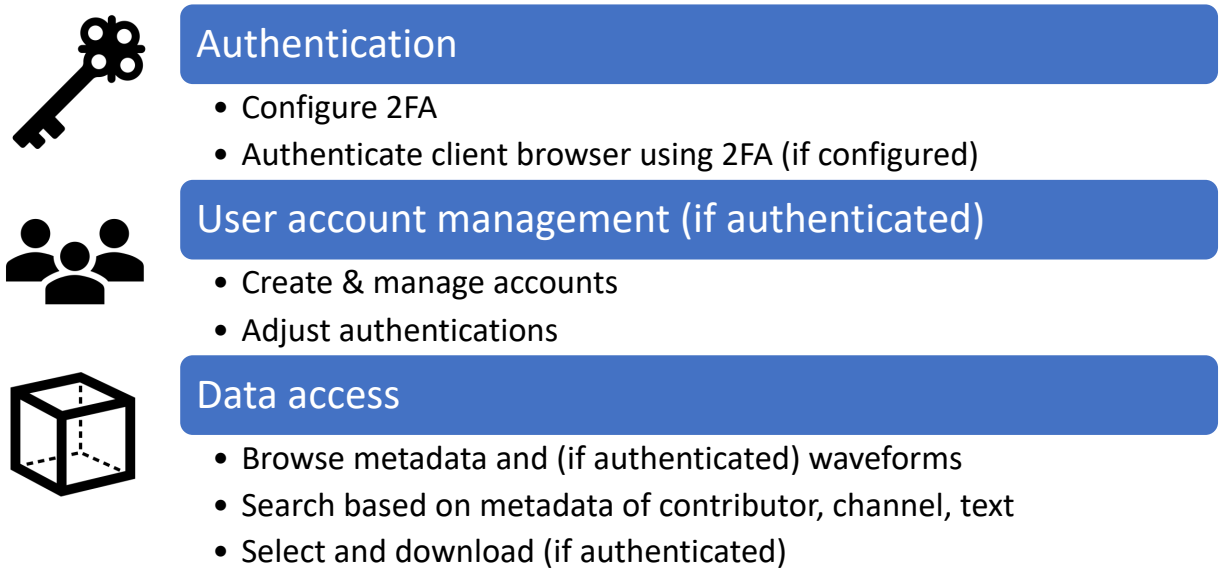

Figure 4. The researcher and member admin interfaces will include three key components: authentication, user account management, and data access.

For authentication, each new researcher account will be initialized with a one-time password. Upon presenting the one-time password, the researcher will be permitted to set up two-factor authentication, which will be required for their future logins.

The data access dashboard will include search tools (searching on any of the corpus metadata fields), filtered browsing (through lists of data returned by any prior search), and download. Alternatively, member organizations can request periodic automatic or manual pushes of data to their own cloud infrastructure (e.g., through Globus). Every metadata field will have an "last modified" field, so that researchers can optionally choose to download only the speech samples and metadata that have been modified since their own most recent download.

E. Architecture

Phase 1: First year – Contributor interface, annotator interface, researcher interface, and admin interface. Designed for ~200 contributing users, and 5 companies (~12 users each) of downloading users.

Phase 2: Second year - Additional capacity to accommodate additional ~1000 contributing users with different types of disabilities.

Architectural and Design requirements

Phase 1:

We are estimating dozens of users each day (50-60), each logged on for 30-60 minutes. The interface options available to a user will depend on their role. The application anticipates multiple user roles:

- Public – People looking for information about the project – this is a public website.

- Contributors – Users who record and upload samples. To create a Contributor account, a user will have to click a button agreeing to the terms of an electronic informed consent document approved by the IRB.
- Member Admins - Each consortium member organization will have a Member Admin account with the authority to create Researcher accounts. Member Admin accounts will be created by the Program Management Lead upon confirmation that an organization has signed the consortium member agreement.
- Researchers – Users who search, browse, and download data.
- Limited access researchers (e.g., members of the Advocate Advisory Committee) may be created for disability advocates and others if they first sign a data use agreement that has been approved by all members of the Advisory Committee. Limited access researchers will be permitted to browse the metadata, but not the speech audio.
- Annotators – People that will be adding metadata to the speech, ensuring the data is clean, and ensuring the data is saved in the desired format. Annotator accounts will be created by Internal Admins.
- Internal Admins - Internal Admins will include the PI and Co-PIs. Internal Admins will have authority to create or delete user accounts and to assign user roles. Internal Admins will also use the interface to gather statistics about the data, spot-check data quality, and perform other high-level needs.

Contributors will login with email and password. All other non-Public user roles will require 2-factor authentication.

The data will be stored in two databases: a Speech Database that is available to Researchers, Member Admins, and Annotators, and a Private Database that is only available to Internal Admins (more detail about these databases is provided below). While they could be stored on the same server, we plan to store them on separate servers for an added level of security. This is how we handle MRI data currently. **NOTE:** There will be 3 versions of each of these databases as we will have development, test, and production environments. There will also be both on-site and off-site copies of the data. File storage requirements in Phase 1 are estimated to include about 1400 utterances from each Contributor:

- 200 x 1400 = 280,000 utterances
- Each about 5 seconds of audio or about 160K each
- Two copies of each Speech Dataset: one original, one normalized to standard file format and sampling rate
- 0.5TB of storage will be allocated for the datasets and metadata

The **Speech Database** will have two tables:

- The contributor shareable metadata table will include metadata about each contributor that can be shared with any researcher who has signed a consortium member agreement or an approved data use agreement. In this table, the contributor key will be a unique identifier (UUID). Each contributor is associated with the following metadata fields. Each

metadata field will each have associated “last modified” fields, specifying date and time in a standard format (e.g., ISO 8601).
    - Age or birth year
    - Gender
    - Contributor identification of their own native language, if not English
    - Disability category (drop-down menu)
    - Contributor description of their own disability (free text)
    - Contributor description of their recording environment
    - Dysarthria or other speech pathology type, as annotated by SLPs based on audio samples the Contributor has uploaded
    - Perceptual attributes of dysarthria, including the 46 integer scales of the DAB annotation system, plus additional attributes selected by SLPs as being relevant to each etiology
- The audio table will include the original and post-processed audio recordings (as blobs or links), and metadata associated with the audio file. The audio table key will be an utterance UUID. Each metadata field will have an associated "last modified" field. Fields in this table will include:
    - UUID of the contributor
    - Original uploaded audio file as it was uploaded by the contributor (blob or link)
    - Post-processed audio file, e.g., normalized to a common file format and sampling rate, also possibly normalized to a standard level (blob or link)
    - Text of the prompt that the contributor was reading
    - Text of the utterance as transcribed by ASR
    - Automatically estimated channel conditions (SNR, clipping frequency, etc.)
    - Text of the utterance as transcribed or corrected by a human annotator
    - Is-Valid: this field will be checked when a human annotator has verified that the audio file is valid.

The **Private Information Database** (only available to Internal Admins) is organized as follows:
- Unique identifier (UUID) that ties it back to the individual in the Speech Database
- Metadata:
    - State/Country of residence
    - Email address
    - Cumulative number of contributions, so that human subject payments can be made appropriately
    - Future items to be determined

Contributors may upload, remove, and download their own audio files.

Training materials, FAQ, and initial triage of the tech support issue queue will be provided by engineering RAs (one graduate, one undergraduate) hired for this purpose. Data contributors will be able to easily request tech support by pressing a "Support" button that provides a template for trouble ticket submission. RAs will be trained to answer tech support

issues that they know how to answer, and to route the remaining issues to the Program Management Lead, to the ITS team, or to other appropriate consortium personnel.

Phase 2: Potentially additional development to allow for Contributors with different types of disabilities to participate. Additional storage to go from ~200 to ~1200 Contributors. Targeting about 1400 utterances per Contributor

- 1200 x 1400 = 1680000 utterances
- Each about 5 seconds of audio or about 160K each
- 2 Terabytes of storage will be allocated for the datasets and metadata

User Access / Security

For initial setup, use "local" account management. Use email account as login name, encrypted passwords, use 2FA, support password reset function, not reliant on UIUC credentials, and note that account holders may have multiple roles.

Other resources

1. Accessibility Expert: Campus has multiple options (Disability Resources and Educational Services, Office of the Vice Chancellor of Diversity Equity and Inclusion, Tech Services, AITS, etc.) for us to consult with to ensure the application and associated tools are all accessible.
2. Campus Security has resources to check application for security
3. Software to perform recordings such as selected JavaScript libraries or HTML 5's MediaStream API.
4. Development in a Windows environment – this is what we currently use for most of our applications. Other options are possible such as .Net Core, C#, Windows server, MS SQL server.
5. Virtual Machines: Recommend 3 web servers for development, test, and production for the applications. 6 VMs for storing the Audio and Private Information Databases on separate servers. (Dev, test, and prod for each of the databases on their own server). 3 VMs for dev, test, and production if the Audio and Private Information databases are stored on the same server. Total VMs is either 6 or 9 – We will be using existing campus infrastructure for the VMs. Specs: 4 CPU, 8GB RAM, 60GB Fast Storage, 60 GB Bulk Storage.
6. SSL certificates – free for the University
7. Beckman Isilon storage (existing) for storing the sound files, metadata, consent forms. Storage includes production and Disaster Recovery instances
8. DNS registration

V. Schedule and Key Milestones

A work breakdown structure with associated milestones and responsible points of contact is shown in Table 1 below. EOM ("End of Month") is used to express the relative due dates of milestones, where the precise date will depend on the start date of the period of performance of the NOVA H contract.

WBS #	WBS Element Name	Due Date	POC
1.	**Project Management**		P. Jones
1.1.	**Governance**		P. Jones
1.1.1.	Advisory Committee of founding industrial sponsors established	EOM 1	M. Hasegawa-Johnson P. Jones
1.1.2.	Technical Advisory Committee of experts from industrial sponsors established	EOM 1	M. Hasegawa-Johnson P. Jones
1.1.3.	Advocate Advisory Committee from patient/advocacy groups established	EOM 2 Expanded membership EOM 12	M. Hasegawa-Johnson M. Dickinson
1.1.4.	Institutional Review Board	EOM 3	P. Jones
1.1.5.	Github organization created, and repositories initialized	EOM 3	M. Hasegawa-Johnson
1.1.6.	Consortium Member Agreement	EOM 4	P. Jones
1.1.7.	Selection of Etiologies for Phase 2 by a consensus of the Advisory Committee	EOM 9	M. Hasegawa-Johnson P. Jones
1.1.8.	Consortium Member Agreement revised to include terms for non-founder members, who will be recruited for the beginning of Phase 3	EOM 20	P. Jones
1.2.	**Documentation**		P. Jones
1.2.1.	Annual Report	EOM 12 and EOM 24	P. Jones
1.2.2.	Quarterly Status Reports	Each Quarter	P. Jones
1.2.3.	Monthly Status Reports	Each EOM	P. Jones
1.2.4.	Notes from Committee meetings	As needed	P. Jones
1.2.5.	IRB protocol documentation: initial protocol plus amendments and renewals as needed	EOM 3 with amendments as needed	P. Jones. Product of task 1.1.4 above.
1.2.6.	Github initial project documentation	EOM 4	M. Hasegawa-Johnson
2.	**NOVA H Repository**		E. Hege

2.1.	**Data storage** in secure infrastructure	V1 due EOM 5 V2 due EOM 12	E. Hege
2.2.	**Data structure and overall architecture and tools**	V1 due EOM 5 V2 due EOM 12	E. Hege
2.2.1.	Creation of a protocol for routine periodic evaluation of tools, and for identification and resolution of software revision priorities	EOM 5	E. Hege
2.2.2.	Technical support queue implemented and staffed	EOM 5	M. Hasegawa-Johnson E. Hege
2.3.	**Metadata:** pipeline created to store metadata provided by Contributors and Annotators	V1 due EOM 5 V2 due EOM 12	E. Hege
2.3.1.	Pipeline created for automatic metadata creation	V1 due EOM 5 V2 due EOM 12	M. Hasegawa-Johnson
2.3.2.	Protocol created for speech-language pathology metadata	V1 due EOM 5 V2 due EOM 12	C. Mendes
2.3.3.	Protocol created for verifying and correcting audio transcripts	V1 due EOM 5 V2 due EOM 12	H. Kim
2.4.	**User interfaces / dashboard(s)**		E. Hege
2.4.1.	Contributor UI/dashboard.	V1 due EOM 5 V2 due EOM 12	E. Hege
2.4.2.	Member Admin UI/dashboard	V1 due EOM 5 V2 due EOM 12	E. Hege
2.4.3.	Researcher UI/dashboard	V1 due EOM 5 V2 due EOM 12	E. Hege
2.4.4.	Annotator UI/dashboard	V1 due EOM 5 V2 due EOM 12	E. Hege
2.5.	**Online training interfaces / support**		M. Hasegawa-Johnson
2.5.1.	Contributor training documentation and FAQ	V1 due EOM 5 V2 due EOM 12	M. Hasegawa-Johnson
2.5.2.	Researcher training documentation and FAQ	V1 due EOM 5 V2 due EOM 12	M. Hasegawa-Johnson
3.	**Community Outreach and Engagement**		M. Hasegawa-Johnson M. Dickinson H. Kim C. Mendes P. Jones
3.1.	**Awareness campaign and Recruitment to participate**		M. Dickinson

3.1.1.	Acquire research participant directory (at least one) relevant to etiology 1	EOM 4	M. Dickinson
3.1.2.	Write promotional material relevant to etiology 1	EOM 4	M. Dickinson
3.1.3.	Subcontracts/recruitment agreements established with at least two advocacy organizations affiliated with etiology 1	EOM 5	M. Dickinson
3.1.4.	Subcontracts/recruitment agreements established with at least one hospital with a patient population covering etiology 1	EOM 6	M. Hasegawa-Johnson
3.1.5.	Initiate advertising campaign relevant to etiology 1	EOM 6	M. Dickinson
3.1.6.	Acquire research participant directories relevant to etiologies 2-5	EOM 11	M. Dickinson
3.1.7.	Revise or rewrite promotional materials to be relevant to etiologies 2-5	EOM 11	M. Dickinson
3.1.8.	Subcontracts/recruitment agreements established with advocacy organizations affiliated with etiologies 2-5	EOM 13	M. Dickinson
3.1.9.	Subcontracts/recruitment agreements established with hospitals affiliated with etiologies 2-5	EOM 14	M. Hasegawa-Johnson
3.1.10.	Initiate advertising campaigns relevant to etiologies 2-5	EOM 14	M. Dickinson
3.2.	**Consent to participate**	Ongoing	P. Jones
3.3.	**Focus group prototyping of Contributor UI/dashboard**		M. Dickinson E. Hege C. Mendes
3.3.1.	First etiology focus group test of Contributor UI/dashboard	EOM 6	M. Dickinson E. Hege C. Mendes
3.3.2.	Second and third etiology focus group tests of Contributor UI/dashboard	EOM 13	M. Dickinson E. Hege C. Mendes
3.4.	**Data collection of speech samples from contributors**	Ongoing.	M. Hasegawa-Johnson M. Dickinson

		Etiology 1 due EOM 11. Etiologies 2-5: due EOM 23	H. Kim C. Mendes E. Hege P. Jones
3.5.	**Annotation and release to researchers of speech samples**		M. Hasegawa-Johnson C. Mendes H. Kim
3.5.1.	Generation of automatic metadata for speech samples	Etiology 1: EOM 12. Etiologies 2-5: EOM 24	M. Hasegawa-Johnson
3.5.2.	Generation of speech-language pathology metadata for speech samples	Etiology 1: EOM 12. Etiologies 2-5: EOM 24	C. Mendes
3.5.3.	Verification or correction of transcriptions of speech samples	Etiology 1: EOM 12. Etiologies 2-5: EOM 24	H. Kim

VI. Addenda

A. Member benefits

The Advisory Committee will consist of six members: one representative of the University of Illinois, and one representative of each Founding Member of the Consortium. The Advisory Committee will meet once per three months.

B. IP ownership

All Coalition IP (IP developed in a project funded in whole or in part by coalition membership fees) will be owned by the University of Illinois. The University of Illinois will not file patent applications covering any Coalition IP. No governmental funding will be used to support any Coalition Project without prior written approval from all Coalition Members.

C. Licenses

Software that is Coalition IP will be released under the MIT/X11 license. Speech Signal Data that is Coalition IP will be licensed to members under non-exclusive, perpetual, royalty-free license terms.

D. Liability

Liability terms will need to be reviewed by University of Illinois legal counsel.

E. Privacy

Coalition IP that is Speech Signal Data will be made available to Coalition Members who have agreed to the privacy terms which can be briefly summarized as: (3.1) the data user will not make data available to other organizations, except that it may make data available to its own representatives if they have agreed to the same terms, (3.2) the data user will use reasonable safeguards to prevent unauthorized access, and (3.3) the data user will not attempt to identify any natural person based on anything present in the data.

F. Consent

Contributors will agree to consent terms approved by the University of Illinois IRB. The University of Illinois will draft consent forms in a manner that will enable commercial use of the Speech Signal Data.