Experiences in Designing a Mobile Speech-Based Assessment Tool for Neurological Diseases

Louis Daudet[1(✉)], Christian Poellabauer[1], and Sandra Schneider[2]

[1] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, USA
{ldaudet,cpoellab}@nd.edu

[2] Department of Communicative Sciences and Disorders, Saint Mary's College, Notre Dame, USA
sschneider@saintmarys.edu

http://m-lab.cse.nd.edu/

Abstract. Mobile devices contain an increasing number of sensors, many of which can be used for disease diagnosis and monitoring. Thus along with the ease of access and use of mobile devices there is a trend towards developing neurological tests onto mobile devices. Speech-based approaches have shown particular promise in detection of neurological conditions. However, designing such tools carries a number of challenges, such as how to manage noise, delivering the instructions for the speech based tasks, handling user error, and how to adapt the design to be accessible to specific populations with Parkinson's Disease and Amyotrophic Lateral Sclerosis. This report discusses our experiences in the design of a mobile-based application that assesses and monitors disease progression using speech changes as a biomarker.

Keywords: Speech analysis · Portable diagnostics · Proof of concept · Experience report · Mobile health

1 Introduction

Neurodegeneration is the process through which the neurons and neuronal structures are compromised, hindering their proper functions, or even leading to their death. This neurodegenerative process is the cause of many diseases such as Alzheimer's Disease, Huntington's Disease, Parkinson's Disease (PD) [1] and Amyotrophic Lateral Sclerosis (ALS) [2]. Although there exist some treatments for these diseases aimed at slowing down their progress or helping with their symptoms [3,4], they remain incurable. As these diseases progress, patients struggle with a variety of symptoms such as speech disorders, tremors, difficulties with movement coordination, cognitive decline, and sensory issues [5–7]. An estimated 10 million people live with PD [8], while ALS is thought to impact 6

J. Ye et al. (Eds.): MobiHealth 2020, LNICST 362, pp. 3–17, 2021.
https://doi.org/10.1007/978-3-030-70569-5_1

people out of 100 000. Furthermore, with the aging of the population worldwide, the impact of these diseases is on the rise. For example, the United Nations predicts that, due to aging, the number of people with ALS worldwide is expected to go up 69% between 2015 and 2040, going from about 223,000 to 377,000 people [9]. Besides the tragic effects these diseases have on a human level, they also have a significant financial impact. The worldwide cost of dementia alone, with both PD and ALS being contributing diseases [10,11], is 614 billion dollars, or 1% of the world GDP [12].

There has been a trend towards mobile-based health assessments, as mobile devices are often constantly with their users, but also featuring an increasing array of sensors that can be used to extract valuable health data. In [13], the authors present the various sensors and mobile developments made that can be used by medical professionals to diagnose and monitor conditions such as asthma, hypertension, or diabetes. Specifically, speech has been used in several mobile based health assessment tools. The field of mobile health is finding new applications for all of these developments in mobile technologies, as shown in [14]. When trying to develop mobile health applications, specific challenges need to be taken into account. In [15], the authors list privacy concerns and usability as some of the main difficulties to be addressed. With a traditional test done in a medical setting, the privacy of the patients data is handled by the strict regulations and policies in place. But with a mobile application, the data is being collected from the patients' devices, and needs to be stored and transmitted securely at all times, adding to the complexity of device based assessments. Usability is also complicated by the small screen sizes, the complex inputs, and the sometimes slow interaction speeds of some lower end mobile devices. Similarly, in [16], the authors considered several categories of challenges when designing mobile-based health applications. For the application itself, the two main challenges were the user interface (i.e., how to make sure that the layout of the graphical elements help and not overwhelm the patients), and the design of the task (i.e., how to handle interruptions such as phone calls, how to handle the test being performed in different types of environments). They also noted that several challenges came from the devices' hardware, such as the screen size (i.e. how to be read by different populations on smaller screens, how to account for variations in screen size), the input (i.e. how to handle various types of input scheme), and the network (i.e. how to deal with sometimes spotty or even nonexistent connectivity).

We created a mobile-based application designed to detect the presence of PD and ALS using speech analysis. The application uses speech based tests, adapted from existing speech language pathology tests, to collect speech samples from participants. Using several metrics extracted from these speech samples, we then developed models to identify features that would help with the classification of participants with PD and ALS. As we designed and developed our application however, we met several challenges that we had to address, such as user prompts, noise handling, data safety, speech sample capture, and user error. This paper describes details of our application and the challenges we met when developing it.

2 Related Work

There has been extensive research in speech features and using speech as a biomarker to detect neurodegenerative diseases. In [17], the authors showed that variation in the fundamental frequency (F0) could be used to differentiate between healthy and PD patients. Moreover, in [18], the authors found that changes in F0's variability could lead to an early diagnosis of PD. In this longitudinal study, which followed a PD patient for eleven years, including seven years pre-diagnosis, they were able to detect abnormal variability in F0 five years before the diagnosis was made. The work in [19] also identified specific speech metrics that are affected by PD. The authors showed that besides the variability of fundamental frequency already discussed above, breathiness and asthenia (weakness) were the two metrics most impacted by PD. These two metrics were measured by subjective means using the GRBAS scale, an auditory-perceptual evaluation method for hoarseness. The Diadochokinetic (DDK) rate and maximum phonation time, both measured objectively by a computer, were also found to be different (shorter) in PD patients.

Similar to our project, in [20], the authors used a 'quick vocal test' to assess which of the participants in their sample, 46 native speaking Czech, had PD. Their vocal test was made up of three different parts: a sustained phonation task, a DDK task, and a running speech task. Although they were able to get a classification performance of 85%, they used eight metrics from the frequency domain such as jitter, shimmer, and variability of fundamental frequency, to reach that result with only 24 PD patients. This means an average of only three participants per significant metrics, which is below the five to ten recommended to avoid overfitting [21,22].

PD is not the only disease that has been shown to impact the production of speech. In [23], the authors found that ALS affected the speech of the patient by causing abnormal pitch (either too low or too high), limited pitch range, high harmonics-to-noise ratios, and increased nasality, among others symptoms.

From these studies' results, we were encouraged in our hypothesis that different neurodegenerative diseases impact the speech of the patients in specific ways, thereby different speech metrics patterns could assist with the diagnosis of specific neurological diseases. These related efforts on PD detection are different from our system since for one they only rely on a subset of the speech-based tasks contained in our application. They also limited their research to the detection of PD while we have a broader approach that allows for the detection of various neurodegenerative diseases.

3 Application Design

3.1 Overview and Workflow

Developed on iOS, our application is used to collect metadata and speech samples from participants with neurodegenerative diseases PD and ALS, as they have been shown to have a strong impact on speech [17,18,23]. This paper focuses on the design and functionality of the application itself, as well as the challenges involved in them, up to the upload of the data to our servers. The analysis of the data made on the servers is outside of the scope of this paper.

Our application consists of a practitioner questionnaire, which includes an optional feedback form, for research purposes to assess the ease of using the application, a participant questionnaire, and a series of speech-based tasks, and an optional participant feedback form. The workflow of the application can be seen in Fig. 1. The workflow is composed of four main steps, detailed below in different subsections.

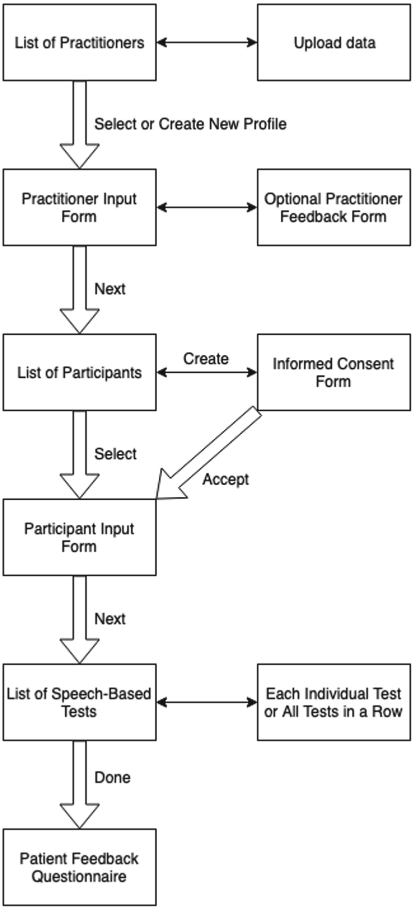

Fig. 1. The workflow of the mobile application

Step 1: The practitioner input/selection. The first screen of the application is the list of first and last names of practitioners registered on the device for selection. It also has two buttons, one to upload the device's data to our cloud servers, and one to create a new practitioner profile. The profile creation page asks for a for basic information (i.e. name, degree, institution), and has an option for the practitioner to submit feedback back to us.

Step 2: The participant input/selection. The participant questionnaire collects the participant's personal and medical information. In order to access the questionnaire, the participant first has to read through and agree to the consent form for our study.

The personal information section consists of asking the participant their name, gender, birth date, and native language. As for the practitioner, we hash the first name, last name and birth date, to create a unique ID for the participant.

The age, gender, and native language are relevant when extracting metrics from the speech samples. The questions about the native language and the strength of the accent serve two purposes. For our initial data collection, it allows us to exclude participants with an accent, as their accents would have been an extra parameter that biased our model. With more data being collected in the future for different accents, we will be able to create specific models for people with different accents. We also ask if the participant has undergone speech therapy of any kind, as it might impact the characteristics of the speech recordings.

The medical information section asks if the participants have an hearing impairment, so that we can assist them if they have difficulties hearing the prompts from the speech based tasks. We also record what cognitive changes if any, have been experienced by the participant, if the participant has been experiencing any unusual movements, or if the participant has felt more emotional or anxious than usual. We also ask if the participant have any problem with their speech. These questions are there for us to see if there is any correlation between the answers obtained from the participants about their self-assess well being, and the metrics extracted from their speech samples.

Finally, we ask what type of disease has the participant has been diagnosed with, and when. For the participants with PD, we also ask when did the participant took the last dose of their treatment. PD having a very regular medication cycle, we want to show a correlated impact on the metrics extracted from the speech samples, by collecting data from the same participants at different point of their medication cycle.

Step 3: The speech-based tasks. The speech-based tasks constitute the core of our application. Fourteen tasks were designed based on de facto standards in the field of speech-language pathology. A summary of the speech based tasks can be seen in the list below.

– Vowel Tests

- Vowel 'Ah' with Timer
- Vowel 'Ah' without Timer
- Vowel 'Eh' with Timer
- Vowel 'Eh' without Timer

Participants with speech impairments will have trouble maintaining a constant pitch or power throughout, or will exhibit vocal fry and breathiness. Two different vowels are being used, 'ah' and 'eh'. We chose these tasks because they have been shown to be a good indicator to detect jitter and shimmer [24,25], and have been used in research to monitor the evolution of PD in patients [26]. Both the time and non timed version of the task are used, in order to see if a difference can be seen in the way control, ALS and PD populations managed running out of air while performing this task.

– DDK Rate

- Repetitions of monosyllabic words. ("Pa"/"Ta"/"Ka") and Repetition of a polysyllabic word. ("PaTaKa") for 5 s intervals

Measure the Alternating Motion Rate (AMR) and Sequential Motion Rate (SMR). The DDK rate is the number of iterations per second a participant is able to produce correctly over a five second window. This permits the assessment of oral motor skills by giving a measure of the participant's ability to make rapid speech movement using different parts of the mouth. This task has been shown to capture differences in control, PD and ALS populations [19,27,28].

– Grandfather Passage

- Reading the grandfather passage

The text of the grandfather's passage was designed to contain almost every phoneme in the English language, thus allowing us to see whether the participants have some difficulties with specific phonemes. As an aside, the history of this text is interesting in its own right [29]. This task has been used in the past to detect acoustic characteristics in speech [30,31], and characteristics specific of PD and Multiple Sclerosis (MS) patients [32].

– Monosyllabic Words & Increasing Syllabic Words

- List of easy words. (mom, Bob, peep, bib, tot, deed, kick, gag, fife, sis, zoos, church, shush, lull, roar)
- Increased syllabic count task. (cat, catnip, catapult, catastrophe/please, pleasing, pleasingly/thick, thicken, thickening)

Here, we test to see at which point, if any, the participants either struggle or become unable to produce the correct word. We look for a "breakdown" in their ability to sequence the words correctly in order to rule out a "motor planning" issue versus a specific motor issue. This task was designed to assess the production of every consonant and vowel in the English language [33–36].

– Picture Description

• Describing a picture presented on the screen

Checking the participant capacity to handle volitional speech, with the extra cognitive load it entails to construct the sentences. A picture is chosen at random from ten possible pictures, and the participants are asked to describe it using any words of their choosing. With this, we are able to both measure the ease of the participants to select and program words on their own, with the extra stress it involves with word finding, semantics, syntax and pragmatic language features. by measuring features such as the rate of word production, the size of the dictionary used (number of different words), and the complexity of the words chosen.

– Multisyllabic Words

• List of complex words. (participate, application, education, difficulty, congratulations, possibility, mathematical, opportunity, statistical analysis, Methodist episcopal church)

Can the participant handle the complex motor patterns required to go from the front to the back of the mouth when saying these words.

– Sentences

• Sentences. (We saw several wild animals, My physician wrote out a prescription, The municipal judge sentenced the criminal, The supermarket chain shut down because of poor management, Much more money must be donated to make this department succeed)

Can the participant program the whole sentence while handling the formation of complex words that compose it. Part of the sentences used in this task were designed by Dr. Julie Liss from Arizona State University. Her goal with these sentences was to determine the type of dysarthria of participants based on rhythmicity of speech while uttering these sentences [37]. These sentences are: In this famous coffee shop they serve the best doughnuts in town, The chairman decided to pave over the shopping center garden, The standards committee met this afternoon in an open meeting.

– Automatic Speech Production

• Iterate the days of the week
• Iterate the months of the year
• Count from 1 to 30

Test the participants' automatic speech production, and their endurance in producing speech. It is considered automatic speech, as opposed to volitional or imitative speech, as the participants do not repeat the words like with the other tasks so far, but do not have to truly think about the words they are saying either, like in the picture description task, since they are part of sequences that are deeply ingrained into the participants' minds for having used them since childhood. The endurance part of the task comes from the length these tasks have, especially the first one. With some diseases, such as ALS, producing speech over such a long list of words in a row can be difficult.

3.2 Challenges in Design

We began data collection with a first version of the application for 28 days in November 2015 before implementing an improved V2 of our application. With this first version, a total of 1260 recordings have been made, corresponding to 103 min recorded, but with unfortunately 34% of these recordings which could not be used. We identified challenges that were addressed in subsequent versions; these challenges being discussed in the remainder of this paper.

User Handicap. The first issue we ran into was the difficulty for some participants, particularly ALS ones, to perform all of the tasks. They often did not have the endurance to go through all of the tests without having to take long pauses to recuperate, In order to deal with this, we added the option to skip tasks, and modified the tasks' order. This order is designed to allow the ALS patients to perform as many tasks as possible before they had to stop the testing. The screen listing the speech based tasks can be seen in Fig. 2.

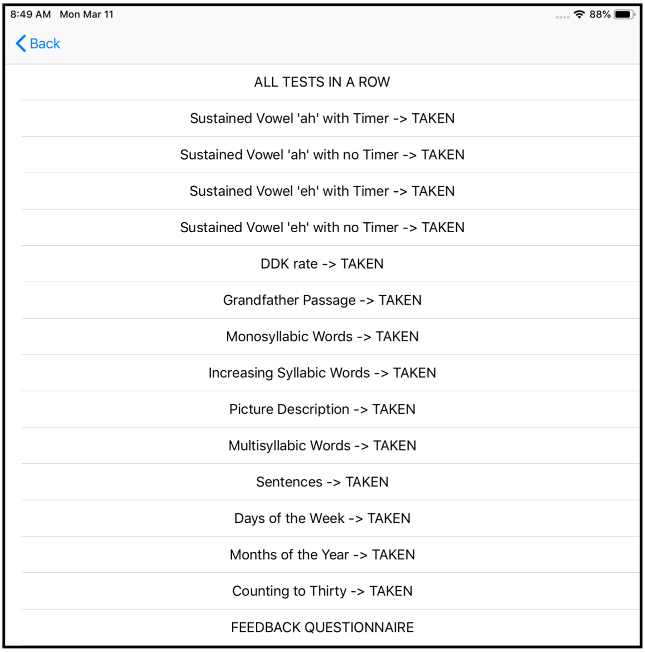

Fig. 2. The list of tasks after completion of a series

User Prompts. Each task needs audible instructions to explain to the participants what they are to do. Using a text-to-speech software avoid introducing any bias for participants that would try to mimic the speech patterns of a human voice. However, text-to-speech voice was reported as confusing for a lot of participants and had a clear negative impact on the application's usage, as participants did not understand the prompts. Instructions provided by a human voice are now used, that we made as even toned and accented as possible.

For the sustained vowel task, the challenge is to have the participant understand that the sound has to be sustained for a relatively long time. We displayed a long 'aaaaaaaaah' across the device's screen, and had a small arrow going under its length in 5 s. However, not everyone understood exactly what sound they were expected to make, which was solved by using audible prompts. Also, many participants did not understand that they were to start when the arrow under the text started moving nor were they able to know exactly how long the task was going to run for, and had trouble managing their breath to maintain voicing throughout. So a timer is now used, indicating how long the task is going to run for, and how much of it is left at all times, as can be seen in Fig. 3.

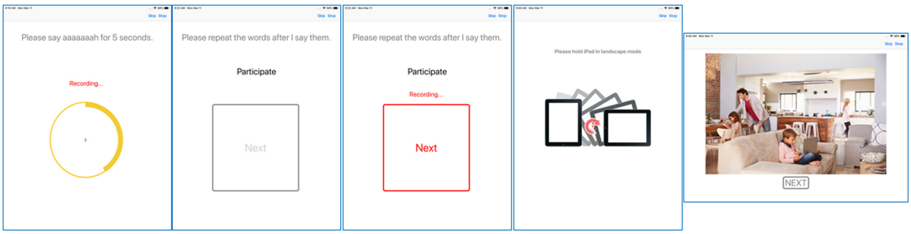

Fig. 3. From left to right: timer for timed tasks (here, the sustained vowel), screen while the participants listen to a word, screen when the participants repeat a word, picture description task while device in portrait mode, picture description task when device has been switched to landscape mode

For the DDK test, special care had to be taken, as here we needed the application to explain what specific sounds to produce, and the fact that these sounds needed to be produced as fast as possible, but not too fast that it hindered the proper production of the expected sounds. The initial design would show the words go across the screen, but for one, people would read 'Pa' several different ways, and, more importantly, the participants would, try to match the speed of the text on the screen rather than reaching their own maximum speed. Like in the sustained vowel task, an audible prompt is now used to indicate the proper pronunciation and a timer is now used to indicate how long remains on the test.

With this grandfather passage, the difficulty is to choose how to display the text. The ability to choose the font, and potentially make the page scrollable would introduce too much variables from participant to participant. So we chose to instead use a fixed font that would be as big as possible as to fit the whole text on the screen. After testing this design, feedback from the practitioners taught us that it would be easier for most participants if the font was a bit smaller and instead the line spacing a bit bigger, so these modifications were integrated into the application.

Clipping of the Recordings. Another challenge with the design of the speech-based tasks was to deal with participants not being timed properly with the application's prompts, talking before the end of the instructions, or going to the next task, or part of a task, as they are still completing the previous one. We thought it would be best to make one recording per word or sentence, making it easier to know what words contained each recordings. After the end of each instruction, a new recording would start, and end when the participant pressed the 'next' button that was on the screen. After some data collection, it became clear that a lot of clipping was happening, from people that would start to talk a little bit before the instruction's recording was done, or tap the next button while they were still talking. This lead to a lot of recordings either too clipped for use, or empty all together. Out of all the recordings that are not usable, 42% of them where due to this issue.

To correct for this, we had a two-fold approach. The first thing we did was to add a color code to the 'next' button making it clear that we were only recording between after the word or sentences had been said aloud by the application, and before the press of the button when the participant is done repeating it. During that time, the button at the center of the screen turns red and a red label indicate that a recording is in progress, as can be seen in Fig. 3. We also changed the way we record, now doing so continuously throughout the task, from beginning to end. We also keep track of the times at which any events happen (end of instruction sound file being played, button being tapped by user, etc.). With this, we are able to know when in the sound file we can find the participant talking. To deal with what clipping still happens in spite of the clearer color coding during the task, we can also crop the sound file for each word or sentences a few 10s of a second before and after the timing information recorded the participant to be talking, insuring that we capture all of the speech sample.

User Error. Another big challenge was handling incorrect inputs from the participants. As seen in Fig. 3, there are two buttons present at the upper right corner of the screen during each task: 'Skip' and 'Stop'. When pressing skip, the practitioner signals that the task was either avoided altogether or that the participant could not complete it. This allows for a task series to be completed even by participant who do not have the capacities to go through all the tasks. When pressing the stop button, the task currently being performed is canceled and the application goes back to the list of tasks. The recordings for that task is not saved, nor are the meta-data about the task, which thus remains as non-taken on the screen with the list of tasks. When doing all the tasks in a row, we added a transition screen in between tasks to redo a specific task without having to stop the series. This screen gives the option, at the end of every task, to either proceed to the next one if all went well, or redo it if the first attempt was not performed properly, without leaving the current series.

After a test series is completed, and all the tasks as marked as taken, the practitioner can still, if needed, redo any of the task that might not have been properly performed by the participant. The data previously collected for that task would then be overwritten by the new recording. This way, the data from tasks that needed redoing are not kept, keeping the data collected as clean as possible. By default, the task series are automatically reset at the end of each day, so that if a task series exists for the selected participant, and it has been started on a day prior to the current one, this task series is closed and a new one is created at the current time and date.

Data Handling. As our application is dealing with medical data, privacy and security are of the utmost importance. It is imperative that the data be kept secured on the device, as well as on the backend servers, and in transit from the former to the later. On the device, the data is kept encrypted by iOS which prevent the data to be accessed by anyone without the device's password.

Initially, our application's data was stored in flat text files. This was easy to implement but made analysing the data complex as dedicated scripts had to be written to query the data. Starting with V2 of the application, we now use a SQLite database through iOS Core Data, allowing the data to be queried using SQL.

To increase the security of the participant's private information, we upload the anonymous information from the participant on a different server than the rest of the data collected. This allows for an extra layer of security: even if one of the server were to be breached, the data of each server would not be useful for an attacker without to data from the other as they would either get access to a list of name with no associated information, or to completely anonymized data. When uploading, the application first separate the anonymous information from the participants (first and last name of the participants), together with the unique ID generated for each participants. The anonymous data is then sent through an encrypted connection to an AWS server. The rest of the data is sent, still through an encrypted connection, to a different server hosted by the Center for Research Computing (CRC) at Notre Dame. Both servers are located behind firewalls to prevent unallowed access.

4 Conclusion and Future Work

Data has been recorded between November 2015 and March 2019. The V2 has been implemented from December 2015 while V3 was used from August 2018. A total of 70 individuals were tested, including control group, ALS suffering individuals and Parkinsons suffering individuals, and team members testing the application. Out of those 70 individuals we excluded all the tests and all bad recordings, which let us with a total of 64 individuals having contributed usable recordings. Each individual has been recorded under a single version of the application, none has tested different versions of the application. The average number of recordings per participant is 26, while the average total recording of an individual is almost 7 min. Of all 64 participants, 56% were men and 44% were female, while the distribution per pathology can be seen in Fig. 4.

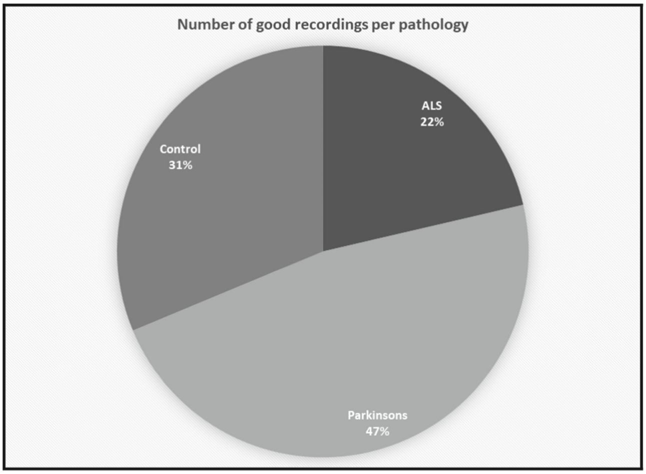

Fig. 4. Distribution of the participants per pathology

A total of 509 min of recordings have been made, out of them 446 min are considered 'good', i.e. can be properly analysed. For most individuals all recording was conducted within a single day, while it was organized in 2 or even 3 d for around 20% of them.

To determine the quality of the recordings, we created a small iOS application that allowed us to efficiently check each of the recordings manually. For each recording, we could set a boolean to indicate if the recording could be used in our analysis, and a comment to indicate why not (no sound, loud ambient noise, participant did not understand the test...). With each version of the application, the percentage of good recordings kept on going higher. Through this data collection process, learning from our errors, we have been able to overcome each of the challenges detailed in this paper. From more than a third of the recordings not fit for analysis, we achieved to go under the 10% threshold with less than 8% of poor quality recordings in the current version of the application, which we consider acceptable, as can be seen in Fig. 5. With each version, the incremental improvements made allowed for the application to perform better. In its current state, it is able to record more accurately, prompting the users clearly and without introducing biases, collecting more metadata for a richer and easier analysis of the recordings.

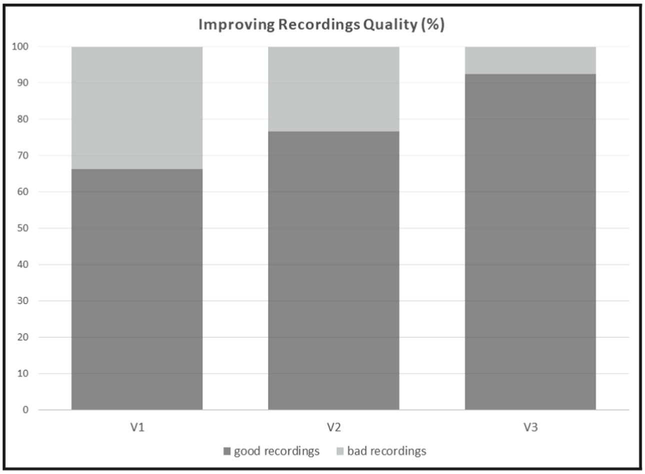

Fig. 5. Proportion of bad recordings per version of the application

With the data that we have now collected, we are working on building statistical and machine learning models to classify the recordings with high accuracy. We will first work on extracting metrics from each of the recordings, both from the time and frequency domain. In the time domain, these could be the number of utterance per second for the DDK test, or the number of words per second, total time per sentences, amount of time in between each words for the sentences tests, and grandfather passage. These can be measured by using Sphinx [38] to measure the start and end of each words in the tests. For the frequency domain, a large array of metrics will be extracted using python and praat [39], such as the shimmer, jitter, average pitch, variance in intensity, breathiness... All these metrics will then serve as the basis for our modelling work to classify each recording as control, PD or ALS. The work is currently in progress and will be presented as part of a future paper.

Acknowledgment. This research was supported in part by GE Health and the National Football League through the GE/NFL Head Health Challenge. The research was further supported in part by the National Science Foundation under Grant Number IIS-1450349.

The authors wish to thanks Dr. Julie Stierwalt from the Mayo clinic, Rochester, MI, Dr. Kim Winter from the Hospital for Special Care, New Britain, CT, and the Rock Steady Boxing program participants and leaders, especially Don Sheliga and Morgan Books.

References

1. Poewe, W., et al.: Parkinson disease. Nat. Rev. Dis. Primers **3**(1), 1–21 (2017)
2. Kiernan, M.C., et al.: Amyotrophic lateral sclerosis. Lancet **377**(9769), 942–955 (2011)

3. Connolly, B.S., Lang, A.E.: Pharmacological treatment of parkinson disease: a review. JAMA **311**(16), 1670–1683 (2014)
4. Vucic, S., Rothstein, J.D., Kiernan, M.C.: Advances in treating amyotrophic lateral sclerosis: insights from pathophysiological studies. Trends Neurosci. **37**(8), 433–442 (2014)
5. Tysnes, O.-B., Storstein, A.: Epidemiology of Parkinson's disease. J. Neural Transm. **124**(8), 901–905 (2017)
6. Sveinbjornsdottir, S.: The clinical symptoms of Parkinson's disease. J. Neurochem. **139**, 318–324 (2016)
7. Boillée, S., Velde, C.V., Cleveland, D.W.: ALS: a disease of motor neurons and their nonneuronal neighbors. Neuron **52**(1), 39–59 (2006)
8. Vos, T., et al.: Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet **388**(10053), 1545 (2016)
9. Arthur, K.C., Calvo, A., Price, T.R., Geiger, J.T., Chio, A., Traynor, B.J.: Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat. Commun. **7**, 12408 (2016)
10. Ringholz, G., Appel, S.H., Bradshaw, M., Cooke, N., Mosnik, D., Schulz, P.: Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology **65**(4), 586–590 (2005)
11. Strong, M.J., et al.: Consensus criteria for the diagnosis of frontotemporal cognitive and behavioural syndromes in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis **10**(3), 131–146 (2009)
12. World Health Organization, et al.: Dementia: A Public Health Priority. World Health Organization (2012)
13. Sim, I.: Mobile devices and health. New Engl. J. Med. **381**(10), 956–968 (2019)
14. Silva, B.M., Rodrigues, J.J., de la Torre Díez, I., ópez-Coronado, M.L., Saleem, K.: Mobile-health: a review of current state in 2015. J. Biomed. Inf. **56**, 265–272 (2015)
15. Zhang, C., Zhang, X., Halstead-Nussloch, R.: Assessment metrics, challenges and strategies for mobile health apps. Issues Inf. Syst. **15**(2) (2014)
16. Al-Saadi, T.A., Aljarrah, T.M., Alhashemi, A.M., Hussain, A.: A systematic review of usability challenges and testing in mobile health. Int. J. Acc. Financ. Rep. **5**(2), 1–14 (2015)
17. Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J. Acoust. Soc. Am. **129**(1), 350–367 (2011)
18. Harel, B., Cannizzaro, M., Snyder, P.J.: Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cogn. **56**(1), 24–29 (2004)
19. Midi, I., Dogan, M., Koseoglu, M., Can, G., Sehitoglu, M., Gunal, D.: Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. Acta Neurol. Scand. **117**(1), 26–34 (2008)
20. Rusz, J., et al.: Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. Mov. Disord. **26**(10), 1951–1952 (2011)
21. Peduzzi, P., Concato, J., Kemper, E., Holford, T.R., Feinstein, A.R.: A simulation study of the number of events per variable in logistic regression analysis. J. Clin. Epidemiol. **49**(12), 1373–1379 (1996)
22. Vittinghoff, E., McCulloch, C.E.: Relaxing the rule of ten events per variable in logistic and cox regression. Am. J. Epidemiol. **165**(6), 710–718 (2007)

23. Green, J.R., et al.: Bulbar and speech motor assessment in ALS: challenges and future directions. Amyotrophic Lateral Sclerosis Frontotemporal Degeneration **14**(7–8), 494–500 (2013)
24. Brockmann, M., Drinnan, M.J., Storck, C., Carding, P.N.: Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. J. Voice **25**(1), 44–53 (2011)
25. Horii, Y.: Jitter and shimmer differences among sustained vowel phonations. J. Speech, Lang. Hear. Res. **25**(1), 12–14 (1982)
26. Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O.: Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans. Biomed. Eng. **57**(4), 884–893 (2010)
27. Canter, G.J.: Speech characteristics of patients with Parkinson's disease: III. articulation, diadochokinesis, and over-all speech adequacy. J. Speech Hear. Disord. **30**(3), 217–224 (1965)
28. Tjaden, K., Watling, E.: Characteristics of diadochokinesis in multiple sclerosis and Parkinson's disease. Folia Phoniatrica et Logopaedica **55**(5), 241–259 (2003)
29. Reilly, J., Fisher, J.L.: Sherlock holmes and the strange case of the missing attribution: a historical note on "the grandfather passage". J. Lang. Hear. Res. (2012)
30. Vogel, A.P., Maruff, P., Snyder, P.J., Mundt, J.C.: Standardization of pitch-range settings in voice acoustic analysis. Behav. Res. Methods **41**(2), 318–324 (2009)
31. Zraick, R.I., Wendel, K., Smith-Olinde, L.: The effect of speaking task on perceptual judgment of the severity of dysphonic voice. J. Voice **19**(4), 574–581 (2005)
32. Tjaden, K., Sussman, J.E., Wilding, G.E.: Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. J. Speech, Lang. Hear. Res. **57**(3), 779–792 (2014)
33. Duffy, J.R., et al.: Temporal acoustic measures distinguish primary progressive apraxia of speech from primary progressive aphasia. Brain Lang. **168**, 84–94 (2017)
34. Strand, E.A., McNeil, M.R.: Effects of length and linguistic complexity on temporal acoustic measures in apraxia of speech. J. Speech, Lang. Hear. Res. **39**(5), 1018–1033 (1996)
35. Odell, K., McNeil, M.R., Rosenbek, J.C., Hunter, L.: Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers. J. Speech, Lang. Hear. Res. **34**(1), 67–80 (1991)
36. Haley, K.L., Overton, H.B.: Word length and vowel duration in apraxia of speech: the use of relative measures. Brain Lang. **79**(3), 397–406 (2001)
37. Liss, J.M., LeGendre, S., Lotto, A.J.: Discriminating dysarthria type from envelope modulation spectra. J. Speech, Lang. Hear. Res. **53**(5), 1246–1255 (2010)
38. Huggins-Daines, D., Kumar, M., Chan, A., Black, A.W., Ravishankar, M., Rudnicky, A.I.: PocketSphinx: a free, real-time continuous speech recognition system for hand-held devices. In: 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings, vol. 1, p. I. IEEE (2006)
39. Styler, W.: Using praat for linguistic research. University Colorado at Boulder Phonetics Lab (2017)