INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Apkinson – A Mobile Monitoring Solution for Parkinson's Disease

Philipp Klumpp[1], Thomas Janu[1], Tomás Arias-Vergara[2], Juan Camilo Vásquez Correa[2], Juan Rafael Orozco-Arroyave[1,2], Elmar Nöth[1]

[1]Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[2]Universidad de Antioquia, Colombia

philipp-klumpp@live.de, elmar.noeth@fau.de

Abstract

In this paper we want to present our work on a smartphone application which aims to provide a mobile monitoring solution for patients suffering from Parkinson's disease. By unobtrusively analyzing the speech signal during phone calls and with a dedicated speech test, we want to be able to determine the severity and the progression of Parkinson's disease for a patient much more frequently than it would be possible with regular check-ups.

The application consists of four major parts. There is a phone call detection which triggers the whole processing chain. Secondly, there is the phone call recording which has proven to be more challenging than expected. The signal analysis, another crucial component, is still in development for the phone call analysis. Additionally, the application collects several pieces of meta information about the calls to put the results into deeper context.

After describing how the speech signal is affected by Parkinson's disease, we sketch the overall application architecture and explain the four major parts of the current implementation in further detail. We then present the promising results achieved with the first version of a dedicated speech test. In the end, we outline how the project could receive further improvements in the future.

Index Terms: speech monitoring, pathological speech, Parkinson's disease

1. Introduction

Patients who suffer from Parkinson's disease (PD) experience a progressive loss of their motor capabilities throughout the different stages of the disease. The degeneration of dopaminergic cells in the basal ganglia of the human brain causes a lack of the neurotransmitter dopamine [1]. In a healthy person, dopamine greatly affects the way in which the striatum modulates the motor signals that are produced in the frontal cortex. Without this functionality, patient's suffer from a variety of symptoms, such as bradykinesia, postural instability, tremor or impaired speech [2]. Speech impairments are an early and prominent manifestation that can contribute primarily to the diagnosis of PD [3]. The main symptoms of the speech of PD patients include reduced loudness, monopitch, monoloudness, hypotonicity, breathy, hoarse voice quality, and imprecise articulation. These symptoms are typically grouped and called *hypokinetic dysarthria* [4].

The most common treatment of the symptoms of PD is the administration of Levodopa, a pre-stage of dopamine, which is able to pass the blood-brain barrier. Besides the medication, physicians set their focus on a close observation of the disease's progression. Not only does it allow to determine how well a patient responds to the treatment with Levodopa. It also enables to monitor to which extent an affected person is hindered in its everyday life and whether further therapy is required. To determine the severity of PD for a specific patient, a therapist usually has to perform a variety of tests with the person to evaluate the current stage of the disease as well as its progression since the last check-up. This is on the one hand a time consuming task, on the other hand it is impossible to understand how the severity of the symptoms changes over smaller time intervals, like days or even hours. To close this gap in the monitoring of PD, we are implementing the Android smartphone application *Apkinson* which is capable to analyze the everyday speech signal of a patient and come up with a meaningful estimation of the disease's progression.

Several research projects have already worked on mobile monitoring of PD by analyzing the data from wearable accelerometers for tremor detection [5]. Not all of these works rely on smartphone sensors. In a project presented in 2015, a combination of a gesture sensor commonly used in virtual reality applications and a pair of smart glasses was used to track motor deficits developed during PD [6]. In 2011, researchers from the University of California introduced a mobile speech analysis application for the detection of the emotion and stress level of a user [7].

For Parkinson's disease, there are several dedicated speech tests available. A solution like Apkinson, which aims to monitor the degradation of the everyday speech signal of a patient, poses to be a new approach.

The idea behind the application is to record a user's speech signal during a phone call and evaluate it with respect to certain features. Whenever Apkinson detects a decline in the patient's performance with respect to these features, it requests a dedicated speech test from the user. The main goal of this design is to limit the required amount of user interaction to a minimum while still being able to record speech data as frequently as possible. Additionally, Apkinson collects several pieces of meta information about every recorded phone call, such as time and duration, to allow a treating physician to put the results into deeper context. For example, it is possible to evaluate if and how the performance of a patient changes throughout a day.

Apkinson aims to close the monitoring gap between two subsequent check-ups. Therefore, it is not meant to replace the treating physician in any way. This becomes more clear when one takes a closer look at the monitoring workflow shown in Figure 1. After a regular check-up, the patient would start using Apkinson. The application now operates in the background and analyzes the speech signal during phone calls unnoticed by the user. Whenever it detects a decline in the speech quality, it requests a dedicated test. If this test confirms a worse performance, Apkinson suggests an earlier check-up.

This paper gives a short overview over the state of the art in automated monitoring, followed by a description of the major

http://dx.doi.org/10.21437/Interspeech.2017-416

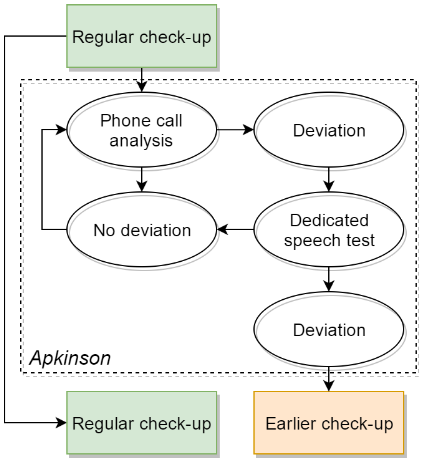

Figure 1: *Schematic diagram of the remote monitoring.*

aspects of Apkinson's current implementation as well as some features that are still in work. Afterwards, the evaluation results for the dedicated speech test will be presented and interpreted. The conclusion will sum up the current progress of the application and outline possible future developments.

2. Implementation

2.1. Architecture

For Apkinson, it is a major goal to reduce the required user interaction to a minimum. PD patients are in most cases elderly people who do not want to change the way they use their smartphone. Ideally, they set up the application once, and then they can forget about it. To make this possible, the required background processing of Apkinson is considerably bigger compared to the active user input.

The application provides four major features which are important for its flawless operation:

- Phone call detection
- Call recording
- Collection of meta information
- Speech signal analysis

All of them are part of the application's background processing chain. The phone call detection runs at any time to be able to detect new calls. It triggers the call recording which handles the process of signal acquisition and initiates both the signal analysis as well as the collection of meta information.

2.2. User interface

The user interface is held very simple. There is only a single Android Activity, which serves as a container for two different Fragments, one for the settings screen, and another one for the results screen. The settings screen gives the user the opportunity to enable or disable a monitoring service, choose a language preference or see some client information. With a swipe gesture, one is able to switch to the result screen, which provides a list of phone calls that were analyzed. Whenever the patient clicks on one of the entries, a dialog pops up which displays all related meta information about the phone call, as well as a small media player to listen to the recorded audio file.

2.3. Phone call detection

The flawless detection of phone calls is very important because all subsequent processes are triggered by this event. A new call can be detected by listening to the current phone state of the device, which then transitions to the *'ongoing'* state. The listener is created inside an Android Service. This is very important, because Services can run in the background, even when the associated application itself is terminated. If the device's memory management system has to destroy the Service to free up memory, it recreates it if the shortage is over. Apkinson also recreates the Service after a reboot of the smartphone. All these provisions enable the application to detect phone calls at any time.

2.4. Call recording

To record the speech of a patient during a phone call, it is very important to first choose the correct audio source. Although Android provides a few audio source options which are designed to capture the voice during a phone call scenario, none of these were applicable for Apkinson, which instead simply makes use of the option *Mic* for the built-in microphone [8]. The reason for this is that the optimized audio sources are all meant to be used for other purposes, such as VoIP applications for example. None of them yielded reliable results during a phone call, when the telephone's dialer application itself also makes use of the microphone.

At this point, it is important to mention that in most countries it is a legal requirement not to record a person's speech without their consent. Therefore, Apkinson must not record the voice of the dialog partner. Unfortunately, this happens from time to time, for example in a noisy environment, when the user increases the volume of the speakerphones. On some devices, this caused the microphone to record not only the speech of the user, but also the speakerphones' output. Android provides an audio pre-processing tool for acoustic echo cancellation, but the test recordings showed that it delivers very poor results. Therefore, it is still a major topic to develop a solution which successfully cancels out this effect. A full list of tested devices is available on our project website on Sourceforge [9].

Speech is recorded in PCM format with 16-bit encoding and mono channel configuration. The preferred sampling rate is 16 kHz, but for Android devices, the only sampling rate that is officially supported by every device is 44.1 kHz. According to this, on some devices it might be necessary to perform a down-sampling after the recording process to match the signal acquisition's required sampling rate of 16 kHz.

It is also important to mention that the call recording thread initializes the microphone instance with a short delay of 3 seconds after detecting an ongoing phone call. This is done to prevent the microphone configuration created by Apkinson to be overridden by the telephone application, which of course also wants to initialize this hardware. This issue could be observed on one of the testing devices and it was solved by adding the short delay.

2.5. Meta information

For every recorded phone call, Apkinson collects a series of meta information:

- Date and time
- Duration of call / user's part of call
- Dialog partner in list of contacts (yes/no)

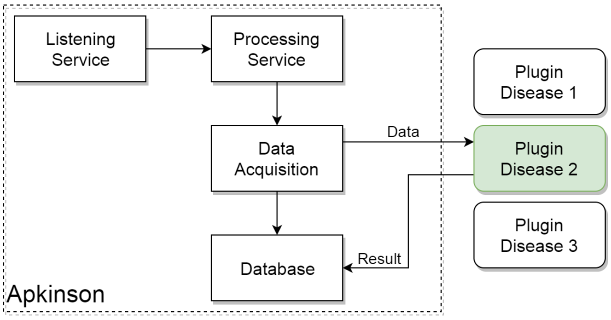

Figure 2: *Plugin architecture for the speech signal analysis with respect to different diseases.*

- Incoming or outgoing call

This information could enable a treating physician to identify certain trends in the performance of a patient. Knowledge about the exact time of a phone call can be useful to determine how effective a medication is. Furthermore, it could be interesting to evaluate how the patient performs with dialog partners from the list of contacts compared to others, with whom the user is less familiar. The results from phone calls with familiar persons could be weighted higher, because factors like nervousness can be excluded. In later implementations, knowledge about the user's speaking time could help to analyze how many words are produced per minute for example. It could already be shown that people who suffer from a dysarthric speech impairment have a decreasing speaking rate [10]. For example, it could be possible to determine that along with the severity of PD, the ratio between incoming and outgoing phone calls heavily shifts to one side. At the moment, Apkinson only collects these pieces of meta information to allow for a manual evaluation.

2.6. Speech signal analysis

In future releases, Apkinson should support the monitoring of many different diseases which lead to an impaired speech. To reduce the overall size of the application, the speech signal analysis is meant to be implemented in a separate plugin. This way, every device only has to supply the speech signal analysis for the particular disease of the user. Furthermore, it allows other developers to create new speech analysis plugins without applying any major changes to the main application. The communication between client and plugin is realized via an AIDL (Android Interface Definition Language) interface and the concept is shown in Figure 2.

A first version of the dedicated PD test was already implemented using the open-source speech recognizer library *Pocketsphinx* from the CMU Sphinx project [1] with a modified configuration. The patient has to perform a Diadochokinetic (DDK) task, a speech exercise in which the person has to constantly produce syllables of subsequent consonant-vowel combinations [11]. For a short time period, the patient has to produce the keyword *'Pa-ta-ka'* as fast and as clearly as possible. Apkinson runs two separate recognizers. The first counts the number of produced keywords, resulting in a text string of several recognized *'Pa-ta-ka'*. The second recognizer is able to detect each of the possible syllables *'Pa'*, *'Ta'*, *'Ka'*, *'Ba'*, *'Da'* and *'Ga'*. The detected results of two subsequent syllables are independent of each other, thus every syllable combination is possible (e.g. *Pa-ga-ka-ta*. The Levenshtein distance, a metric which evaluates the similarity of two text strings by computing the required insertions, deletions and substitutions to transform one string to the other, is computed for the two recognizer results. This value can be normalized with respect to the number of spoken keywords. The result is the syllable error rate (SER), which ranges from 0 for a very good to 3 for a very poor performance.

3. Evaluation

The speech signal analysis for the recorded phone calls is still in development. In this paper we want to describe the results achieved with the evaluation of the first dedicated speech test implementation. To evaluate its reliability, Apkinson had to analyze audio data recorded for a similar study [12] from PD patients as well as from a healthy control group. For convenience, files were read from memory instead of manually recording them with the built-in microphone. This is possible because the acoustic conditions of both scenarios are quite similar and thus yield similar results [13].

The group of participants was composed of 50 patients suffering from PD, as well as 50 health controls. In each group, there was an equal amount of male (25) and female (25) participants. The age of the PD patients ranged from 33 to 81 with an average of 61 years, and from 31 to 86 with and average of 61 years for the healthy controls respectively. All the participants were native Colombian Spanish speakers. For the healthy control group, the available audio data covered a duration of 3 minutes and 23 seconds. For the PD patients, 3 minutes and 54 seconds of recorded audio were available.

The Parkinson's disease group was composed of patients who had reached different stages of the disease. For every patient, the PD severity had been classified with respect to the Unified Parkinson's Disease Rating Scale (UPDRS). A higher score indicates a higher PD severity [14]. Table 1 shows the distribution of patients with respect to certain intervals on the UPDR Scale.

Table 1: *Number of patients with particular UPDRS score.*

UPDRS interval	Patients
$0-20$	10
$21-40$	22
$41-60$	12
$61-80$	5
$81-100$	1

Two values are of interest for analyzing the reliability of the dedicated speech exercise. First of all, it is necessary for this task to correctly detect the number of produced keywords. The actual number was determined before by manual evaluation of the files. In the next step, it is important to precisely identify the syllables which were spoken by the user to allow for a flawless and reliable computation of the syllable error rate.

4. Results

Both the recognizer for counting the keyword phrase as well as the syllable recognizer produced usable results for most of the recordings. Only for two recordings of the healthy control group, they were unable to detect any keyword and therefore failed to determine a syllable error rate. The analysis of the

[1] http://cmusphinx.sourceforge.net/

healthy control data set took 6 minutes and 8 seconds, whilst the analysis of the PD data required 6 minutes and 22 seconds. Thus, it took the recognizer on average 1.72 seconds to analyze 1 second of speech. The results of the keyword detection are presented in detail in Figure 3.

For the recordings of the healthy group, the recognizer correctly counted the number of keywords in 76 % of the cases. Six out of the 50 samples were assigned with a keyword count which deviated by more than 1 from the real count. For the group of PD patients, the number of correctly assessed records was slightly lower with 72 %, but the number of records which yielded a deviation from the real count of more than 1 remained at 6. On average, the deviation for the healthy control group was 0.64, and 0.62 for the PD patients respectively.

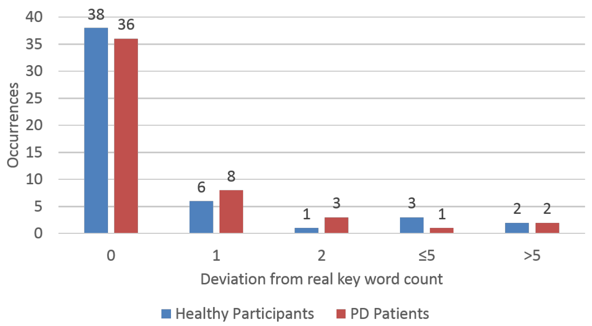

Figure 3: *Deviation between detected and real keyword count.*

Figure 4 summarizes the results of the syllable error rate determination. It shows different results for the two groups. For the healthy controls, the mean SER was 1.34. Additionally, it can be seen that 26 % of the healthy participants achieved an SER of 0.0, indicating a flawless performance. The group of PD patients achieved a mean syllable error rate of 1.67. This value is almost 20 % higher compared to the result of the healthy controls. The number of participants which achieved a low error rate of no more than 0.5 is clearly smaller with only 7, compared to the 13 for the healthy controls. On the other side, 28 % of all Parkinson patients ended up with a very high error rate greater than 2.5.

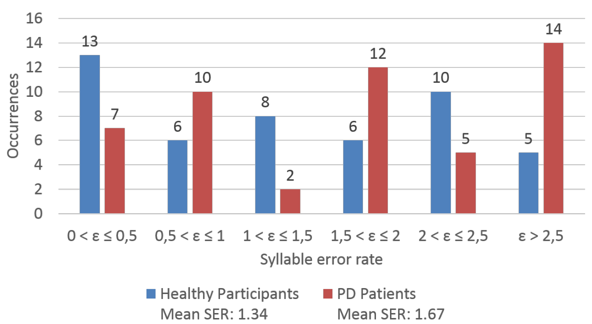

Figure 4: *Detected SER for healthy speakers and PD patients.*

5. Discussion

The keyword recognition of the first version of Apkinson already produced good results. For 88 % of all participants, the keyword count deviated by not more than 1 from the real value. This value is particularly interesting as it remains the same for both the healthy as well as the PD group. This indicates that even for people who might already suffer from an impaired speech which is less intelligible, the system is still able to correctly detect the keyword.

The determination of the syllable error rate indicated that there are differences between both groups. Especially in the boundary regions, these differences are very pronounced. However, in the regions closer to the center, the distribution looks more random. The mean values of 1.34 (healthy) and 1.67 (PD) show once more that the overall performance of the two groups varied. Unfortunately, this variation is not as pronounced as it could be observed in other studies [15], where the mean value of the PD group was more than three times higher compared to this of the healthy participants.

A likely explanation for the weaker performance of the SER detection of Apkinson was the usage of an acoustic model trained for native English speakers. The participants of the study were all native Colombian Spanish speakers though. However, considering that we only made use of a regular speech recognizer which was not designed for the evaluation of a DDK task, the results were still really good and we could measure a difference between the two groups.

6. Conclusions

Apkinson aims to provide a unique monitoring solution for PD patients which operates mainly in the background by recording phone calls. It is one of the major goals to close the monitoring gap between two check-ups, without drawing any user attention. The current design of Apkinson already accomplishes several of the requirements. The signal acquisition procedure requires minor improvements to prevent recording the dialog partner, but the overall signal quality is sufficient to perform an evaluation as soon as the phone call analysis was successfully implemented. The collection of meta information can be very helpful as it enables a treating physician to better understand acquired results.

The first implementation of the dedicated speech exercise was able to detect a measurable difference in the syllable error rate between healthy users and patients suffering from PD. Additionally, the reliability of the keyword recognition worked particularly good. Other studies showcased that the difference in the SER can be even more pronounced, and with further improvements to the speech recognizer, it should be possible to increase the quality of Apkinson's evaluation results.

The next big step for Apkinson will be the addition of the online speech analysis for phone calls, which is currently in development. It will allow the application to fully work in the background, perform an evaluation of a patient's performance, and trigger the dedicated speech test whenever necessary.

The code of the application is published as an open-source project on Sourceforge [2], allowing other developers to share their ideas, knowledge and implementations to push Apkinson to the next level.

7. Acknowledgements

The work reported here was carried out during the 2016 Jelinek Memorial Summer Workshop on Speech and Language Technologies, which was supported by Johns Hopkins University via DARPA LORELEI Contract No HR0011-15-2-0027, and gifts from Microsoft, Amazon, Google, and Facebook.

[2] `sourceforge.net/projects/apkinson`

8. References

[1] J. M. Fearnley and A. J. Lees, “Ageing and Parkinson’s disease: substantia nigra regional selectivity,” *Brain*, vol. 114, no. 5, pp. 2283–2301, 1991.

[2] J. Jankovic, “Parkinsons disease: clinical features and diagnosis,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[3] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, “Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinsons disease,” *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 350–367, 2011.

[4] J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, “Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients,” *Journal of Speech and Hearing Disorders*, vol. 43, no. 1, pp. 47–57, 1978.

[5] F. M. Khan, M. Barnathan, M. Montgomery, S. Myers, L. Côté, and S. Loftus, “A wearable accelerometer system for unobtrusive monitoring of Parkinson’s diease motor symptoms,” in *Proceedings of the IEEE international conference on Bioinformatics and Bioengineering (BIBE)*, 2014, pp. 120–125.

[6] J. Güttler, R. Shah, C. Georgoulas, and T. Bock, “Unobtrusive tremor detection and measurement via human-machine interaction,” *Procedia Computer Science*, vol. 63, pp. 467–474, 2015.

[7] K.-h. Chang, D. Fisher, J. Canny, and B. Hartmann, “How’s my mood and stress?: An efficient speech analysis library for unobtrusive monitoring on mobile phones,” in *Proceedings of the 6th International Conference on Body Area Networks*, 2011, pp. 71–77.

[8] P. Klumpp, “Implementation of a mobile monitoring application for patients with Parkinsons disease,” Master’s thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2017, available on Sourceforge: https://sourceforge.net/projects/apkinson/files/Documentation/Master%20Thesis.pdf/download.

[9] *List of tested devices*, available on Sourceforge: https://sourceforge.net/projects/apkinson/files/Documentation/Tested%20Devices.txt/download.

[10] S. Niimi, Masaki Nishio, “Speaking rate and its components in dysarthric speakers,” *Clinical Linguistics & Phonetics*, vol. 15, no. 4, pp. 309–317, 2001.

[11] M. Asgari and I. Shafran, “Predicting severity of Parkinson’s disease from speech,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology (EMBC)*, 2010, pp. 5201–5204.

[12] J. R. Orozco-Arroyave, J. D. Arias-londoño, J. F. Vargas-Bonilla, M. C. González-Rátiva, and E. Nöth, “New Spanish speech corpus database for the analysis of people suffering from Parkinsons disease,” in *Proceedings of the 9th international Conference on Language Resources and Evaluation (LREC)*, 2014, pp. 1–210.

[13] J. C. Vásquez-Correa, J. Serrà, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, and E. Nöth, “Effect of Acoustic Conditions on Algorithms to Detect Parkinsons Disease from Speech,” in *Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[14] C. G. Goetz, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, G. T. Stebbins, M. B. Stern, B. C. Tilley, R. Dodel, B. Dubois *et al.*, “Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan,” *Movement Disorders*, vol. 22, no. 1, pp. 41–47, 2007.

[15] E. Nöth *et al.*, “Remote Monitoring of Neurodegeneration through Speech,” Final Presentation of the Third Frederick Jelinek Memorial Summer Workshop, August 2016.