Voice Changes in Diabetes Mellitus: a Methodological Commentary

Julia Sidorova[1] and Maria Anisimova[2]

Abstract

Recent research describes the effect of Type 2 diabetes (T2D) on voice, suggesting that it can be diagnosed based on vocal clues. Although these studies have similar experimental designs with respect to the voice data and the analysis methods, the conclusions regarding the voice changes differ substantially and are at times contradictory. This is unexpected, since the mechanism of pathological deterioration behind the observed changes is the same. This year in an article published in J. of Voice it was suggested that vocal changes may be different among ethnicities. Before this hypothesis can be accepted, the study protocols should be improved and unified, to ensure that the empirical evidence is reliable. Additionally, given the recently published data about the temporal voice changes as a result of glucose swings, we propose that the persons in hypo- and hyperglycemic conditions should be excluded from the experiment. Since no study succeeded in diabetes detection, it is timely to mention that there is an alternative methodology for disease detection from voice, which is far more sensitive than the state of the art procedure. We propose a script that is available from the first author on request.

Key words: diabetes, voice, perceptual evaluation, acoustic analysis, pattern recognition, vocal biomarker

1. Introduction

Lately, substantial research effort focussed on diagnosing diabetes with non-invasive methods such as hair analysis [1], facial expression analysis [2], and acoustic analysis of voice [3-5]. The aim is two-fold: (1) building a cost-effective tool for the diagnosis of diabetes mellitus (DM) in the contexts, when standard tests are unaffordable or decided against for a different reason, and (2) describing voice changes in diabetic patients in a way that can provide insights to a physician, as e.g. see [3]. It should be said that non-invasive biomarkers for diabetes are high-risk/ high-reward research with many products being retracted or never reaching the market, e.g. in non-invasive glucose monitoring [8].

With respect to the description of voice changes, Hamdan [3] was the first to describe the pathology of the speech tract caused by DM complications: the expected changes in voice, the group of patients manifesting such changes most clearly, the experimental analysis, and the results that did not align well with the initially expected trends. Since then, two more studies[4, 5] were published. All three studies reported different, at times contradictory trends in voice patterns. As far as building a detection system is concerned, the existing studies have not achieved the recognition of diabetes from voice. Consequently, the contributions of this commentary is three-fold:

I. We review and analyse literature relating to T2D and voice changes focusing on the description and analysis of acoustic changes attributable to DM.

II. We propose to extend the exclusion criteria taking into account the recent literature on the relationship between the glucose swings and voice.

III. We explain an alternative computational analysis, the input to which is a speech file and the output is a prediction about its category (diseased or control). We include a survey of the recent works together with a description of publicly available databases and tools.

The rest of the article is organised as follows, Section 2 describes the literature and analyses the biases in the primary research studies regarding the detection of DM from voice, Section 3 describes a more sensitive computational approach to the detection of the disease from voice, and Section 4 draws the conclusions.

2. Literature Survey on the DM Detection from Voice

A recently literature survey on the topic of T2D and voice [10] left the reader without any conclusive evidence concerning the problem of the description and analysis of acoustic changes attributable to DM, whereby stating "Hamdan *et al* reported no significant differences on acoustic and perceptual measures between people with DM and controls". We have expanded the inclusion criteria from this study with conference proceedings (which traditionally have been the publication venues for important technological breakthroughs in speech technology and computer science) and with the time span after the publication of this survey. The protocol for the literature analysis was as follows.

1. To find for the potentially applicable studies, we have searched across five databases of scientific literature covering publications from 2009 up to the article submission date in 2020, namely: Cochrane library, PubMed, Scopus, Web of Science, and Google Scholar. The logical formula was:

J. Sidorova is with Blekinge Institute of Technology, Vallhallavagän 1, Karlskrona, 37141, Sweden. Julia.a.sidorova@gmail.com.

M. Anisimova is with Zurich University of Applied Sciences, Technikumstrasse, 9, 8400, Winterthur. anis@zhaw.ch

(speech OR voice) AND (diabetes OR glucose OR sugar OR hyperglycemia OR hypoglycemia). The exact search queries are provided in Table IV in the Appendix. The query was kept as broad as possible, so as to retrieve the articles containing the keywords regardless where they appeared. In Google Scholar the search was completed over the titles only, but the formula was extended with synonyms.

2. Once the potentially applicable studies were retrieved, the irrelevant ones were discarded based on the manual scanning of the abstracts and full texts. We took a conservative approach by including and summarizing every piece of research that describes either the construction of a voice biomarker that detects DM from voice or the voice changes attributable to DM.

3. The cited articles and the ones that cite other relevant articles were added to the pool of potentially applicable studies.

4. The research biases in the retrieved studies were accounted for with a Transparency and Reliability Score (TRS). The idea of this score is starting with the initial score 0, to penalise it by 1, each time when some aspect of the study is not described or done in a faulty way. Such a score permits the researcher to give a different weight of consideration to the empirical evidence reported in different sources.

During the search process, we found three primary research articles that describe the static changes in voice attributable to diabetes complications, and/or investigate the possibilities of diabetes detection from voice [3-5] via an acoustic or perceptual analysis and a literature survey on the effect of DM on voice [10].

All the primary research articles report that there are statistically significant changes attributable to DM at least in some patient groups. Table 1 presents a summary regarding the patient groups, in which the changes were found, whether they were revealed via an acoustic or perceptual analysis, and the characteristics of the corpora. Table 2 specifies which acoustic parameters were affected, the exclusion criteria, the country (as an oblique indicator of ethnicity), and the type of a statistical procedure used to evaluate the significance of the deviations from the norm. The information relevant to our analysis, but which was missing contributed to the decrement of the respective TRS (see explanations in bold face in Tables I-III).

Corpora. From Table I, it can be seen that there is a consensus in the literature [3-5] with respect to the type of voice material, namely, the sound /a/ was recorded (in Hamdan *et al* [3] it was augmented with counting from one to ten). In all the studies, there are around 80 subjects with matching controls, and the recordings were made with a professional microphone in the laboratories of the respective institutions.

Acoustic Analysis. In two studies [4, 5], which confirmed that acoustic differences exist between the diabetics and controls, the acoustic analysis was carried out using CSL (Computerized Speech Lab) to collect the MDVP (Multi-dimensional Voice Program) acoustic parameters: fundamental frequency, jitter, shimmer, amplitude perturbation quotient, noise to harmonic ratio, smoothed amplitude perturbation quotient (sAPR), and relative average perturbation (RAP). Pyniopodjanard *et al* [5] referred to this procedure as a gold standard for acoustic analysis to assess voice pathology. However Hamdan *et al* [3] used a different system and found no significant differences between the patients and controls in the cause of acoustic analysis. Instead, the voice changes in diabetic patients of this study were confirmed via a perceptual evaluation.

Statistical Analysis. The numeric data from the acoustic analysis were subjected to statistical tests to draw conclusions regarding the significance of the detected differences between the diabetics and controls, and linear regression was applied to analyze independent variables associated with diabetes (see Table II for details).

Patient Groups. Table I summarizes the patient groups with voice changes attributable to DM. Some studies [3, 5] did not confirm the presence of static voice changes due to DM in general and stated that these were present only in specific subgroups of patients. Pyniopodjanard *et al* [5] reported that the pathological traits were found only in females. Substantiating this Chitkara and Sharma [4] confirmed that the differences were more prominent in female voices, although male voices also exhibited some statistically significant changes. While Hamdan *et al* [3] found no significant differences between diabetics and controls with respect to acoustic parameters, they reported that diabetic subjects with poor glycemic control and neuropathy had different voice quality which was detected via a perceptual evaluation.

Acoustic changes. As it can be observed from Table II, there are contradicting conclusions (in bold face) regarding the acoustic parameters that are affected and the type of respective trends, i.e., whether they become significantly lower or higher than in controls. More precisely, the discrepancies reported across the studies are as follows:

- lower sAPR [4] vs higher (or unchanged) sAPR [5],
- no changes in RAP and APQ [5] vs lower RAP (for females) and APQ (in both sexes) [4],
- an empirically found decrease in perturbation parameters [4] vs their expected increase [3].

Ethnicity. The voice changes caused by T2D are thought to be attributable to the associated pathologies affecting the muscles and the neurological control of the speech organs [3], i.e. there are typical organic changes, which should suggest that the set of acoustic parameters affected as well as the direction of the change should be the similar. It was hypothesized that voice changes can depend on the subject's ethnicity [5]. Yet, before this is taken as a working hypothesis and, in order to rule out possibly incorrect assumptions coming from a faulty analysis of empirical data, the study protocols need a unification.

Protocols. Let us review the methodological aspects (summarized in Table II):

- the exclusion criteria,
- the descriptive space of the acoustic parameters, and
- the computational methodology.

(1) The exclusion criteria filter out the speakers that have any conditions other than T2D, which are known to affect speech production organs or the neurological mechanisms related to speech.

Recently the results were published regarding the temporal changes in the voices of diabetic patients due to glucose swings [6, 7, 12, 13], which were revealed via a comparison of voice data in diabetic patients in the conditions of hypo- and hyperglycemia. Table III summarizes the literature with regards to the acoustic parameters that were reported to be affected as a result of glucose swings and the ones that form the descriptive space of the voice changes in diabetes that are listed in Section "Acoustic Analysis". In order to study the effects of glucose swings in the voices of diabetics, Czupryniak *at al* [13] applied a statistical analysis to the same gold set of acoustic parameters. (Note that although [13] is an abstract and not a full peer reviewed paper, other studies [6, 19], too, report that glucose swings are non-randomly related to acoustic parameters[6] and, furthermore, that glucose value can be approximated from voice clues on a large corpus of patients [19].) It can be seen that the parameters in the description of static and temporal changes overlap: fundamental frequency, relative average perturbation, shimmer, amplitude perturbation quotient, and noise to harmonic ratio. Therefore, if the aim is to understand the static changes in voice attributable to DM, than the state of hypo- and hyperglycemia should also be added to the exclusion criteria, in order to rule out the variability in the patterns, which can be further detailed in the controlled presence of these two conditions. Therefore in Table II, all the primary research studies received a decrement of 1 for the Exclusion Criteria.

(2) The analysis to identify the descriptive patterns associated with T2D is carried out in the space of the acoustic parameters from the "gold standard". This descriptive space is convenient, as it allows for a comparison. In some studies a bias arises, because the analysis of the full list of acoustic features is not available.

(3) The conclusions about the numeric differences between the diabetics and controls were subjected to a statistical verification. A bias arises in one study, where the statistical analysis was not reported.

3. Alternative Computational Methodology

As discussed in Section 2, the studies on the relation of diabetes and voice [3-5] take a classical statistical approach: once the parameters have been measured, the tests for normality are carried out, in order to decide between the use of parametric and nonparametric methods, and then the groups of diabetics and controls are compared with respect to the values of the few parameters from the gold standard. Given the research objectives, there are two problems with this approach: (1) none of the studies achieved the detection of DM from voice, and (2) the differences in voice quality between the diabetics and controls were confirmed via a perceptual analysis (i.e. undoubtedly heard), and yet they were not statistically detected based on the values of the parameters from the gold set [3]. These challenges can be addressed via the following methodology from the field of vocal biomarkers.

Definition 1 [14]: A vocal biomarker takes speech as an input and evaluates the patient's speech production (quality, competence or other aspects) at either a particular time moment or as a trend during the months of rehabilitation.

Typically its construction has the following two steps.

Step I. A large number of acoustic features is extracted from a voice sample. From this point, a speech sample is represented as a vector with numeric features extracted from the speech file and its class category, for example, patient or control. A feature selection procedure selects a smaller relevant subset from the set of candidate features.

Step II. Construct (or "learn") a classification function from training data, in order to classify between the voices of controls and patients. The discriminative ability of such a function is tested on new data, i.e. the samples that were not used during the training stage.

3.1. Literature Survey on Vocal Biomarkers

Initially this approach of pattern recognition [15-16] based on large number of acoustic features [19] was developed by the community of emotion recognition and then was successfully applied to the detection of diverse pathologies and conditions as was summarized by Sidorova *et al* [14]. Below we provide an account of the research frontier (2019-2020) on vocal biomarkers:

- medical and other conditions for which they recently have been developed,
- databases with voice data, and
- feature extraction systems and tools of computational analysis.

In search for best practices and given the fact that vocal biomarkers are notoriously slow entering an established clinical routine, we have kept the inclusion criteria restrictive and included the articles published only in central journals (1st and 2nd quartile according to Scimargo Journal and country Rank) in all disciplines. In order to obtain the initial pool of articles, in Google Scholar and Scopus, we used the key words "biomarker" and "voice" anywhere in an article. The duplicates were removed. That returned 2370 hits, to which the following filters were applied. (1) Based on the titles and abstracts, articles on the topics other than vocal biomarkers (by Definition 1) were excluded. (2) Non-primary research articles (reviews, opinion papers, and so on) were excluded. (3) Studies on animals, newborns, and singing were excluded. (4) The publication journal ranked within the 1st or 2nd quartile in at least one of the research areas, for which the journal was indexed. (5) Articles that used <20 acoustic features were excluded. (The exception was made for papers [32, 42, 43], because they either formulate a new trend in using tools [32] or open up a new field of applications for vocal biomarkers [42, 43].) (6) A study was excluded, if, based on its description, it likely contained a serious technical problem according to [41] such as that the recognition accuracy was reported on "compromised" data points, namely, any information about the test set was used while choosing the predictive model (e.g. feature selection or, less obviously, the normalization/resampling of the data set was done on the whole data set before cross-validation). (7) The biomarkers were a result of the fusion of diverse technologies other than acoustics (video, transdermal, dialogue structure, MRI, and many others) were not included. Once all the filters were applied, there remained 15 papers [14, 20-32, 42-43] reflecting the last 1,5 years of research advances in the field, which are summarized in Table V with respect to which diagnosis the system aims to automate, the feature extraction system, the computational method, the description of voice data and whether it is publicly available, the details of data recording, and the study's conclusion. It can be seen that the primary research articles are quite similar in experimental designs, which is not a surprise for a mature research field after a decade of active research efforts, in which many authors make their data publicly available (as can be seen from Table V the data from almost half of the described studies are "available on request"), and use open-source libraries for the feature extraction and computational analysis (Table V).

Fields of application: Recently, vocal biomarkers have been successfully constructed for the following applications

a) *for cardiovascular diseases* to predict the outcome of the heart failure, to give a prognosis for survival and a likelihood of hospitalization during a follow-up [20], and to distinguish between diverse measurements reflecting the severity of the pulmonary vascular disease [22],

b) *in neurology,* to accurately with 98%-100% distinguish between the patients with Parkinson's disease and healthy controls [23, 25, 26, 30], to enable an early detection of Parkinson's disease [21, 33] and apathy in older adults [24], to predict whether mild cognitive impairment will develop into the dementia [27], to quantify the severity of cerebellar ataxia [28], to enable an early detection of Alzheimer disease [29], and to monitor the response to diverse treatments for Foreign Accent Syndrome [14], and

c) *attitudes* correlated to androgen-dependent characteristics such as voice: willingness to collaborate on environmental issues in male population [43].

At present, during the pandemic time, vocal biomarkers are being tested for their potential use in the decision support systems for the diagnosing *COVID-19,* e.g. [33, 34].

Vocal biomarkers have not become an established clinical routine yet, every paper emphasizes that it is "a new opportunity", "a promising technology", and similar. Beyond being cost-effective, non-invasive, instantaneous and possibly at-a-distance technology, several studies proposed principally novel possibilities: for example, an automated objective system has been constructed to substitute a subjective evaluation of the performance in a range of phonetic tasks that were inherently prone to poor inter-rater reliability [28], to detect AD 25 years before its diagnosis [29], and important and unexpected relations between different pathologies in mental health were discovered [42].

Smartphones in data recording: Smartphones were used for data collection in five out of 13 studies, which caused a drop in recognition accuracies [23] compared to analogous studies but with speech recorded in a sound treated room with a condenser microphone placed by a technician at a fixed distance from the speaker' s mouth.

Feature extraction tools: The vast majority of studies use publicly available feature extraction systems: OpenSmiles [19], praat [35], and a clinical trial application of Vocalis Health [36]. Often, motivated by earlier research on a particular disorder, the specific features were programmed from the primitive ones extractable with open source tools. The options on feature choice range from relying on the prior knowledge about the features that are relevant to the disease in question, e.g. [25, 28], to exploring a large feature space with, for example, multivariate regression to obtain an interpretable quantification of importance for each parameter [22, 24, 30]. Since vocal biomarkers are becoming more and more sophisticated and their field of application is expanding, a new trend is to investigate the advantages of different feature extraction systems: in [32] 19 open source libraries for f0 extraction are compared to see how well they capture subtle phenomena of vocal cord vibration.

Tools of computational analysis: Among the publicly available classification and feature extraction libraries are: Weka [37], scikit learn [38], and R Studio [39]. Note that only one study [29] used deep learning neural networks. This can be explained with a) at times modest sizes of the clinical databases, while deep learning typically requires several thousand data points, b) the reluctance of the medical community to rely on black-box methods, and c) the famous inability of deep neuronal networks to handle noisy data.

4. Discussion and Conclusions

The primary studies unanimously confirmed the changes attributable to DM at least in some patient groups. Such changes were detected either via an acoustic or perceptual analysis. Yet, the discrepancies in the conclusions regarding the acoustic changes concern the patient groups, the acoustic parameters, the direction of their changes. Although that makes a hypothesis of ethnical differences plausible, the literature analysis has revealed that the empirical evidence of such diversity may have the origin in the biases in the primary research studies, such as incomplete exclusion criteria or a lack of statistical verification.

Although no study succeeded in the detection of DM from voice, the most suitable methodology has not yet been applied. To this end, we provide a script that implements the end-to-end computational analysis (available from the first author on request). The script takes a speech file as an input and classifies it into categories ("diseased" or "control"), based on the following:

Step1: A feature extraction to obtain a far richer representation of the speech signal (6,000 parameters based on openSMILES extractor [19] instead of <20 in the gold standard), and

Step 2: Learning and testing of a classification function (the best performing one from the methods listed in Table V not including a deep neuronal network) that assigns a category to a new sample (the input) based on Weka [37].

Other information such as patient's ethnicity, language, sex, glycemic control, neuropathy and other potentially relevant factors can also be naturally integrated as features. (In pattern recognition, many domains have natural taxonomies, for example species, chemicals, etc. cluster into families and subfamilies. Within a taxonomic category, e.g., females or persons with neuropathy, objects have comparable patterns, making it possible to apply methods such as a hierarchical classification guided by a subgroup indicator. For example, see Sidorova and Garcia [17].) This methodology can bypass ethnical specificity (if the hypothesis turns out to be true), namely, ethnicity- and language-independent features may be found from a large set of all considered features. For example, the voice-based diagnostics of the Alzheimer's disease was build from a multilingual corpus from the very beginning [18].

A richer voice description as a result of Step 1 can be used together with the classical statistical analysis in place of the machine learning from Step 2.

Appendix: Query Formulations of the Literature Search for DM and Voice

Table IV lists the queries for the literature search to retrieve the publications on DM and voice.

The authors have no competing interests to declare

References

1. H. Huang, W. Hu, Z. Han et al, "Hybrid progressive algorithm to recognize type II diabetic based on hair mineral element content", Conf. Proc. IEEE Eng Med Biol Soc., vol. 5, pp. 4716-4718, 2005.
2. B. Zhang, B.V. Vijaya Kumar, D. Zhang, "Noninvasive diabetes mellitus detection using facial block color with a sparse representation classifier", IEEE Trans. Biomed. Eng., vol. 61, pp. 1027-1033, 2014.
3. A. Hamdan, J. Jabbour, J. Nassar, I. Dahouk, S. Azar, "Vocal characteristics in patients with type 2 diabetes mellitus", European Archives of Otorhinolaryngology, vol. 269, pp. 1489-1495, 2012.
4. D. Chitkara, R.K. Sharma, "Voice based detection of type 2 diabetes mellitus", in Int. Conf. on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics, Chennai, India, 2016, pp. 83-87.
5. S. Pyniopodjanard, P. Suppakitjanusant, P. Lomprew, N. Kasekomkosin, L. Chailurkit, B. Ongphiphadhanakul, "Instrumental Acoustic Voice Characteristics in Adults with Type 2 Diabetes", Journal of Voice, 2019.
6. C. Tschope, F. Duckhorn, M. Wollf, G. Saeltzer, "Estimating blood sugar from voice samples: a preliminary study", in Int. Conf. on Computational Science and Computational Intelligence, Las Vegas, USA, 2015, pp. 804-805.
7. Y. Ulanovsky, A. Frolov, A. Kozlova, "Method of non-invasive determination of glucose concentration in blood and device for the implementation of thereof", patent WO2014/049438.
8. T. Lin, A. Gal, Y. Mayzel, K. Horman, K. Bahartan, "Non-invasive glucose monitoring: a review of challenges and recent advances", Current Trends in Biomedical Engineering and Biosciences, vol 6, no 5, 2017.
9. Sidorova J., "New advances in glucose level estimation from voice", in preparation.
10. R. Ravi, D. Gunjawate, "Effect of diabetes mellitus on voice: a systematic review", Practical Diabetes, vol. 36, no. 5, pp. 177-180, 2019.
11. B. W. Schuller, "Speech Emotion Recognition, Two decades in a nutshell, Benchmarks, and Ongoing Trends", Communications of the ACM, vol. 61, no. 5, pp.90-99, 2018.
12. P. R. Michaelis, "Detection of Extreme Hypoglycemia and Hyperglycemia based on automatic analysis of speech patterns", US patent US 7,925,508 B1, 2011.
13. L. Czupryniak., E. Sielska-Badurek, A. Niebisz, M. Sobol, M. Kniecik, K. Jedra, E. Ezymanska-Garbacz, K. Niemczyk, "378-P: Human voice is modulated by hypoglycemia and hyperglycemia in Type 1 diabetes", poster presentation, American Diabetes Association, San Francisco, California, 2019.
14. J. Sidorova, S. Carlsson, O. Rosander, I. Moreno-Torres, M. Berthier, "Towards disorder-independent automatic assessment of emotional competence in neurological patients with a classical emotion recognition system: application in foreign accent syndrome", IEEE Transactions on Affective Computing, In press.
15. J. Sidorova, T. Badia, "ESEDA: Tool for enhanced speech emotion detection and analysis", In the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution, Italy, Florence, 2008, 17-19.
16. J. Sidorova, T. Badia , "Syntactic learning for ESEDA.1, tool for enhanced speech emotion detection and analysis", In Proc. Of Internet Technology and Secured Transactions Conference, UK, London, pp. 1-6, 2009.
17. J. Sidorova, J. Garcia, "Bridging from syntactic to statistical methods: classification with automatically segmented features from sequences", Pattern Recognition, Vol 48, pp 3749-3756, 2015.
18. K. Lopez-de-Ipiña, J.B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C.M. Travieso, M. Ecay-Torres, P. Martinez-Lage, H. Eguiraun, "On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature", Cognitive Computation, vol. 7, no. 1, pp. 44-55, 2015.
19. F. Eyben, M. Wöllmer, F. Gross, B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor", in MM' 13, Barcelona, Spain, 2013, pp. 835--838.
20. E. Maor, D. Perry, D. Mevorach, N. Taiblum, Y. Luz, I. Mazin, A. Lerman, G. Koren, V. Shalev, "Vocal Biomarker is associated with hospitalization and mortality among heart failure patients", Journal of Americal Heart Association, vol. 9, no 7, 2020.
21. J.M. Tracy, Y. Ozkanca, D. C. Atkins, R. Hosseni Ghomi, "Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease", Journal of Biomedical Informatics, vol.104, 2020.
22. J. D. S. Sara, E. Maor, B. Borlaug, B. R. Lewis, D. Orbelo, L. O. Lerman, A. Lerman, "Non-invasive vocal biomarker is associated with pulmonary hypertension", Plos One, vol. 15, no. 4, 2020.
23. S. Arora, L. Baghai-Ravary, A. Tsanas, "Developing a large scale population screening tool for the assessment of Parkinson's disease using a telephone-quality data", The Journal of the Acoustical Society of America , vol.145, no. 5, pp. 2871-2884, 2019.
24. A. Konig, N. Linz, R. Zeghari, X. Klinge, J. Troger, J. Alexandersson, P. Robert, "Detecting Apathy in Oldr Adults with Cognitive Disorders Using Automatic Speech Analysis", Journal of Alzheimer's Disease, vol. 69, no. 4, 2019, pp. 1183-1193.
25. L. Ali, C. Zhu, M. Zhou, Y. Liu, "Early Diagnosis of Parkinson's disease from Multiple Voice Recordings by Simultaneous Sample and Feature Selection", Expert Systems with Applications, 137, pp.22-28, 2019.
26. A. U. Haq, J.P. Li, M.H. Memon, , A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, M. Shahid, "Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings", IEEE Access, 7, pp.37718-37734, 2019.
27. J.J. Meilán, F. Martínez-Sánchez, I. Martínez-Nicolás, T.E. Llorente, and J. Carro, "Changes in the Rhythm of Speech Difference between People with Nondegenerative Mild Cognitive Impairment and with Preclinical Dementia", Behavioural Neurology, 2020.
28. B. Kashyap, P.N. Pathirana, M. Horne, L. Power, D. Szmulewicz, "Quantitative Assessment of Speech in Cerebellar Ataxia Using Magnitude and Phase Based Cepstrum", Annals of biomedical engineering, vol. 48, no. 4, pp.1322-1336, 2020.
29. S. de la Fuente Garcia, C.W. Ritchie, S. Luz, "Protocol for a conversation-based analysis study: PREVENT-ED investigates dialogue features that may help predict

dementia onset in later life”, BMJ open, vol. 9, no. 3, p.e026254, 2019.

30. S. Yücelbaş, “Simple Logistic Hybrid System Based on Greedy Stepwise Algorithm for Feature Analysis to Diagnose Parkinson’s Disease According to Gender”, Arabian Journal for Science and Engineering, vol. 45, no. 3, pp.2001-2016, 2020.
31. F. Karlsson, L. Hartelius, “How Well Does Diadochokinetic Task Performance Predict Articulatory Imprecision? Differentiating Individuals with Parkinson’s Disease from Control Subjects”, Folia Phoniatrica et Logopaedica, vol. 71, no. 5-6, pp. 251-260, 2019.
32. J. Hlavnička, R. Čmejla, J. Klempí, E. Růžička, Rusz, , 2019. Acoustic Tracking of Pitch, Modal, and Subharmonic Vibrations of Vocal Folds in Parkinson’s Disease and Parkinsonism. IEEE Access, 7, pp.150339-150354.
33. “The Audio Test for Potential Coronavirus Infection Built by Voice Tech Startups Has a New Home, retrieved 12/05/2020 at ”https://voicebot.ai/2020/04/02/the-audio-test-for-potential-coronavirus-infection-built-by-voice-tech-startups-has-a-new-home/
34. “Diagnosing coronavirus by listening to your voice”, , retrieved 12/05/2020 at https://www.israel21c.org/diagnosing-coronavirus-by-listening-to-your-voice/
35. P. Boersma, D. Weenink, “Praat: Doing Phonetics by computer” [computer software], Version 5.3.84.
36. Y. Levanon, LL-S interventor method and system for diagnosing pathological phenomenon using a voice signal. US patent. 2008.
37. E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. Witten, “WEKA: A machine learning workbench for data mining”, in Data mining and knowledge discovery handbook, O. Maimon & L. Rokach (eds.), US, Springer, 2005 , pp. 1265-1277.
38. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blonde, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, “Scikit-learn: Machine learning in Python”, the Journal of machine Learning research, vol. 12, pp. 2825-2530, 2011.
39. J. Allaire, J., 2012. RStudio: integrated development environment for R. Boston, MA.
40. B.M. Bot, C. Suver, E.C. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap , J. Wilbanks, E.R. Dorsey, S.H. Friend, ”The mPower study, Parkinson disease mobile data collected using ResearchKit”, Scientific data, vol. 3, no. 1, pp. 1-9, 2016.
41. J. Friedman, T. Hastie, R. Tibshirani. The elements of statistical learning. New York: Springer series in statistics; 2001. Page 245 for a wrong way to do cross-validation.
42. A. S. Cohen, T. L. Fedechko, E. K. Schwartz, T. P. Le, P. W. Foltz, J. Bernstein, ... ,B. Elvevåg, “Ambulatory vocal acoustics, temporal dynamics, and serious mental illness”, Journal of abnormal psychology, vol. 128, no 2, 2019.
43. N. Landry, J.E. Desrochers, C. Hodges-Simeon, S. Arnocky, “Testosterone, facial and vocal masculinization and low environmentalism in men”, Journal of environmental psychology, 64, pp.107-112, 2019.
44. V. Hozjan, Z. Kacic, A. Moreno, A. Bonafonte, A. Nogueiras, “Interface Databases: Design and Collection of a Multilingual Emotional Speech Database”, in LREC, Las Palmas, Spain, 2002, pp. 2024 -- 2028.

TABLE 1: Summary of the published conclusions regarding the voice changes in T2D subjects.

Study:	Voice changes detected in subgroups:	Method:	Corpus:	Number Patients:	Transparency and Reliability Score
Hamdan et al., 2012 [2]	diabetics with poor glycemic control and/or neuropathy	Via perceptual evaluation	Vowel /a/ and counting from one to ten	82 patients and matching controls	0
Chitkara and Sharma, 2016 [3]	both sexes, though gender differences exist.	Via acoustic analysis	Vowel /a/	Unspecified, 177 voice samples	-1
Pinyopodjanard et al, 2019 [4]	females, subjects with disease duration >10 ys, poor glycemic control, neuropathy	Via acoustic analysis	Vowel /a/	83 patients and 70 controls	0

TABLE II: Acoustic parameters found to be affected.

	Hamdan et al 2012 [3]	Chitkara, Sharma, 2016 [4]	Pinyopojanard et al, 2019 [5]
Acoustic parameters affected	**No differences in acoustic parameters**	In both sexes, **lower** shimmer, APR, NHR,	In females lower f0, **higher sAPR**. No

	(yet the changes were confirmed via a perceptual analysis). Expects increase in the perturbation parameters.	**sAPR**, and additionally, specifically in females, lower jitter, RAP. Statistics for f0 was not reported.	**changes detected in RAP, APQ, and sAPQ.**
Exclusion criteria	conditions that are known to affect speech production organs. Neurological conditions and glucose swings were not excluded.	unspecified	neurological and other conditions that are known to affect voice patterns, and some other. Glucose swings were not excluded
Country	Lebanon	India	Thailand
Statistical analysis	Standard statistical tests to verify that values in the diseased group are significantly different from the controls.	unspecified	Tests as in [3] and linear regression to analysed independent variables associated with diabetes.
Transparency and Reliability Score	-2	-3	-1

TABLE III: STATIC AND TEMPORAL CHANGES IN VOICE [9]

Temporal changes triggered by glucose swings	Energy, amplitude of **fundamental frequency**, indicator of voice/phonation probability, formant frequencies of F1, F2, F4, residual to harmonic ratio, harmonic to all energy ratio, **relative average perturbation (RAP)**, number of fundamental periods, time of fundamental periods, simple voice quality, **shimmer**, **amplitude perturbation quotient (APQ),** unharmonic to harmonic ratio, **noise to harmonic ratio** (NHR) [13].
Static changes due to diabetes reported	**Fundamental frequency** [5,] jitter, **RAP, shimmer, APQ**, smoothed APQ, **NHR** [4].

TABLE IV: QUERIES FOR THE LITERATURE SEARCH

Database	Query	#new hits (not listed, if were retrieved by previous searches)	#manually selected as relevant and relevant papers that cite the hit
Cochrane Library	#1 speech MeSH #2 glucose OR hypoglycemia OR sugar OR hypoglycemia OR diabetes OR SMBG OR self monitoring of glucose #3 #1 AND #2	223	N/A
PubMed	(("speech"[MeSH Terms] OR "speech"[All Fields])	956 (more	N/A

	OR ("voice"[MeSH Terms] OR "voice"[All Fields])) AND (("glucose"[MeSH Terms] OR "glucose"[All Fields]) OR ("hyperglycaemia"[All Fields] OR "hyperglycemia"[MeSH Terms] OR "hyperglycemia"[All Fields]) OR ("hypoglycaemia"[All Fields] OR "hypoglycemia"[MeSH Terms] OR "hypoglycemia"[All Fields]) OR ("sugars"[MeSH Terms] OR "sugars"[All Fields] OR "sugar"[All Fields]) OR ("diabetes mellitus"[MeSH Terms] OR ("diabetes"[All Fields] AND "mellitus"[All Fields]) OR "diabetes mellitus"[All Fields] OR "diabetes"[All Fields] OR "diabetes insipidus"[MeSH Terms] OR ("diabetes"[All Fields] AND "insipidus"[All Fields]) OR "diabetes insipidus"[All Fields]) OR SMBG[All Fields] OR (("ego"[MeSH Terms] OR "ego"[All Fields] OR "self"[All Fields]) AND monitoring[All Fields] AND ("glucose"[MeSH Terms] OR "glucose"[All Fields])))	recent than 1999)	
Web of Science	(speech OR voice) AND (diabetes OR glucose OR sugar OR hyperglycemia OR hypoglycemia)	808	[22, 23]
Scopus	(“speech” OR “voice”) and (“diabetes” OR “glucose” OR “sugar” OR “hyperglycemia” OR “hypoglycemia”)	886 hits (more recent than 2014)	[21]
Google Scholar	allintitle: diabetes OR glucose OR hypoglycemia OR hyperglycemia OR sugar vocal OR acoustic OR perceptual OR speech OR voice; time span of 2009-2019 in papers and patents.	221	[24]

TABLE V: VOCAL BIOMARKERS 2019-MID 2020.

Ref	Features	Computational Approach	Diagnosis	DB Availability	Voice Data	Way of Recording	Conclusion
[20]	600, Vocalis Health	Machine learning (ML), cross-validation (CV)	Overall survival in Heart Failure (HF) and hospitalization during follow up.	Available on request	20 s length voice samples, 2267 patients	call centre	A new opportunity to home telemedical monitoring for cardiovascular diseases.
[14]	6,000,	ML, CV	Following the	Available on	2 patients, 2	Phonetic lab,	Automation of

	OpenSmiles		response in patients diagnosed with Foreign Accent Syndrome to diverse experimental therapies	request	controls, 10 phrases in the reading test, 5 time points separated by months of treatment. 200 phrases from reading test and a large emotional corpus with acted emotional speech by healthy actors [44].	condenser microphone.	the assessment of the skills in paralinguistic communication in neurological patients for the disorders with hard to characterise voice pathology patterns.
[21]	2268 features corresponding to Geneva Minimalistic Acoustic Parameter Set extracted with Open Smiles	ML, CV	Early detection of Parkinson's Disease (PD)	Available from m-Power study [40]	10 s recording per participant vocalising the /a/ phoneme, 2,289 individuals	Smartphones	A technology for early detection of PD.
[22]	223 features, Vocalis	Stat. analysis	Pulmonary hypertension: moderate or greater than normal, and other measurements reflecting the severity of pulmonary vascular disease	Available as supplementary material with [22]	83 patients, 3 voice recordings of 30 s duration. Recording 1: reading a pre-specified text, Recording 2: describing a positive emotional experience; Recording 3: describing a negative emotional experience.	smartphone	1) The potential for identifying at-risk patients with heart failure (HF) using voice analysis. 2) Using voice analysis to detect hemodynamic changes in patients with established HF or pulmonary hypertension.
[23]	307 features summarised in the paper, yet the extraction tools are unspecified	ML (Random Forest), CV	PD vs healthy controls	unspecified	2759 recordings from 1483 PD and 15321 from 8300 controls	smartphone (a significant drop down in accuracies compared to controlled voice recordings)	A step forward toward assessing the development of a reliable cost-effective and practical clinical decision support tool with smartphones.
[24]	27 features, extracted with praat	ML (Logistic Regression	Apathy in older adults	unspecified	60 patients aged 65 or older with neurodegenerativ	With a tablet computer's internal	The first study to investigate whether

	and OpenSmiles	), CV			e disorder: 30 with apathy and 30 without. Recording 1: talking about a positive event. Recording 2: talking about a negative event.	microphone.	automatic speech analysis can be used to characterise and detect apathy. The findings reinforce the usability of speech as a reliable biomarker.
[25]	26 dysphonic features defined in previous research and extracted with praat	Deep neuronal network	PD	unspecified	Training set contains 1040 speech samples: 20 PD, 20 healthy controls. 26 voice samples from each subject: vowels, numbers, words, sentences. Disjoint test set: 28 PD, /a/, /o/ three times	Controlled settings.	100% accurate speaker-independent discrimination between the subjects with PD and controls.
[26]	23 numeric features already extracted by the providers of the database	ML (Support Vector Machine), CV	PD	Available from the UC Irvine machine learning repository	31 persons (including 23 with PD), 195 recordings, from every subject on average 6 vowel phonations were obtained	Sound treated booth	99-100% accurate
[27]	45 features extracted with praat (12 of them were significantly different in the two groups)	Statistical analysis (Mann-Whitney)	Older people with Mild Cognitive Impairment (MCI) and a high probability to develop dementia vs those with MCI that will not do so with a high probability.	Available on request	Reading test by 86 individuals: 73 MCI, 13 pre-AD. Only the vowel nucleus were used for analysis from paragraphs of text.	A sound proof room, a head mounted condenser microphone placed 14 cm away from the speaker's mouth.	Variations in rhythm rate and intensity distinguish between preAD and MCI with a low probability to develop dementia.
[28]	23 features selected from 1) Mel-frequency cepstral coefficients, 2) magnitude based cepstral features. The feature choice was	ML (Support Vector Machine), CV	Cerebellar ataxia (CA)	unspecified	42 patients, 23 controls, 3 times say "British constitution" (a classical phrase for eliciting the features of ataxic speech).	A quiet room, a condenser microphone 10 cm away from the subject's mouth recorded with Android phone	An automated objective system to quantify CA severity and to substitute a subjective evaluation of the performance of a range of motor and phonetic tasks inherently

	motivated by earlier research on discriminating between ataxic speech and normal or other speech disorders.						prone to poor inter-rater reliability, i.e. designing a CA diagnostic decision support.
[30]	752 features	ML (Simple Logistic Classifier), CV	PD	Available from UCI learning repository	252 subjects: 188 PD, 64 healthy. /a/ three times.	Unspecified, but via email the authors commented that they used a condenser microphone in a quiet room.	Automatic recognition of PD with 89% with gender-specific models.
[31]	23 features, the extraction system is implied to be developed by the authors.	ML (Logistic Regression), CV	PD	unspecified	38 PD and 38 controls, repeated syllables /pa/, /ta/, /ka/ produced for as many times as the subject could in one breath.	unspecified	The suitability of the type of speech material to classify PD vs controls and to predict how well they will be assessed for a reading test.
[32]	Considers only f0 detection with 19 available pitch detectors	Statistical analysis	Differentiating between PD and Parkinsonian syndromes (Multiple System Atrophy and Progressive Supranuclear Palsy), which differ from PD by more wide-spread neurodegeneration, rapid progression, and poor response to dopaminergic medication.	unspecified	22 patients with PD, 21 patients with Multiple System Atrophy, 18 patients with Progressive Supranuclear Palsy.	Quiet room, headset with a condenser microphone.	Some freely available feature extraction systems are more suitable for specific purposes.
[42]	Macroscopic speech measures for clinical and physiological science, which are 5 features constructab	statistics	Serious mental illness: a broad range of psychiatric symptoms. 16 persons with schizophrenia, 8 persons with major depressive disorder, and one	unspecified	25 subjects responded a structured but open-ended probe, e.g. “give a step-by-step explanation of how you boil an egg”. On average each participant		

	le with open source libraries		subject with bipolar disorder.		completed 4,5 such sessions.		
[43]	7 features extracted with praat. Then the measurements were combined into a masculinity score.	Statistical nalysis	Whether androgen-dependent physiological characteristics (incl. voice) predict reduced environmental attitudes.	unspecified	162 males /eh/, /ee/, /ah/, /oh/, /oo/	Quiet room, controlled settings, microphone 20 cm away from the participant’s mouth.	More masculine voices were found to be with less favourable environmental attitudes and decreased willingness to preserve the environment.