Received August 22, 2017, accepted October 2, 2017, date of publication October 17, 2017, date of current version November 7, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2762475

Assessment of Speech Intelligibility in Parkinson's Disease Using a Speech-To-Text System

GIOVANNI DIMAURO[1], VINCENZO DI NICOLA[2], VITOANTONIO BEVILACQUA[3], DANILO CAIVANO[1], AND FRANCESCO GIRARDI[4]

[1]Dipartimento di Informatica, Università degli Studi di Bari 'Aldo Moro', 70125 Bari, Italy
[2]Dipartimento di Scienze Mediche di Base, Neuroscienze e Organi di Senso, Università degli Studi di Bari 'Aldo Moro', 70124 Bari, Italy
[3]Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, 70125 Bari, Italy
[4]ASL Bari, 70123 Bari, Italy

Corresponding author: Giovanni Dimauro (giovanni.dimauro@uniba.it)

ABSTRACT Patients with Parkinson's disease (PD) may have difficulties in speaking because of reduced coordination of the muscles that control breathing, phonation, articulation, and prosody. Symptoms that may occur are weakening of the volume of the voice, voice monotony, changes in the quality of the voice, the speed of speech, uncontrolled repetition of words, and difficult speech intelligibility. To date, evaluation of the speech intelligibility is performed based on the unified PD rating scale. Specifically, section 3.1 (eloquence) of the cited scale provides the specialist with some tips to evaluate the patient's speech ability. With the aim of evaluating the speech intelligibility by measuring the variation in parameters in an objective manner, we show that a speech-to-text (STT) system could help specialists to obtain an accurate and objective measure of speech, phrase, and word intelligibility in PD. STT systems are based on methodologies and technologies that enable the recognition and translation of spoken language into text by computers and computerized devices. We decided to base our study on Google STT conversion. We expand Voxtester, a software system for digital assessment of voice and speech changes in PD, in order to perform this study. No previous studies have been presented to address the mentioned challenges based on STT. The experiments here presented are related with detection/classification between pathological speech from patients with PD and regular speech from healthy control group. The results are very interesting and are an important step toward assessing the intelligibility of the speech of PD patients.

INDEX TERMS Parkinson's disease, speech analysis, automatic speech recognition, human voice, speech to text.

I. INTRODUCTION

The reported standardized incidence rates of Parkinson's disease (PD) are 8–18 per 100.000 person-years [1]. The average age of diagnosis for PD is 60 years, but the disease develops at a much younger age in many individuals. As the number of elderly people increases, we must expect a considerable increase in the number of patients. PD is the most relevant of the movement disorders. Patients suffer from incoordination in complex movements. Normal respiration, phonation, and articulation are fundamental for producing well-coordinated speech, while a breakdown in any of these subsystems, or in their coordination, can lead to disordered speech. Parkinson's disease and its common signs (tremor, muscular rigidity, bradykinesia, akinesia, etc.) affect the subsystems of respiration, phonation, and articulation that govern speech motor control. Furthermore, the speech of PD patients reflects the classic physiologic and anatomic changes caused by PD due to disorders of laryngeal, respiratory and articulatory functions [2]–[5]. Disordered oral communication affects most PD patients; a soft voice, monotone, breathiness, hoarse voice quality and imprecise articulation are common symptoms. Voice problems are typically the first to occur, while other disorders, such as prosody, articulation and fluency, appear later and gradually [7]–[9]. The majority of the individuals developing the disease are aged, but the disease sometimes develops in young people. Given that the disease progression is relatively slow, the chance that these individuals could maintain a good quality of life is increased by their improvement in the communication ability. Therefore, the value of an effective treatment for disordered communication is high. The diagnosis of PD is exclusively clinical so there are neither laboratory nor instrumental tests for monitoring the disease evolution and the treatment response. In clinical practice, neurologists widely use the Unified Parkinson's Disease Rating Scale (UPDRS) [10], a scoring system, for the clinical evaluation of PD to follow the progression of their patients' symptoms in a more objective manner. Its 42 items are evaluated through interviews and clinical observation [10], [11].

Speech production can be acoustically measured, and it has been suggested that the study of acoustic parameters of speech may provide an objective and non-invasive measurement of symptomatic changes in PD [12]–[14], [33]. That is why the need for diagnostic tools capable of sensitively and reliably detecting this change is evident. The main acoustic parameters under observation are articulation, voice intensity, frequency spectrum, pressure level and speech intelligibility. Articulation refers to the coordination of the motor subsystems involved in speech production. Imprecise articulation commonly affects speech intelligibility in patients with Parkinson's disease [6], [16]–[21] and can lead to changes in oral diadochokinetic (DDK) production [22], [23]. Voice intensity in people with PD seems to decay faster than that in healthy individuals. Among others, Ho *et al.* examined progressive speech intensity decay using two speech tasks, sustained vowel phonation and sentence reading [24]. It was also shown that the range of articulator movements is reduced in PD, leading to impaired vowel articulation. Vowels are primarily formed by movements of the tongue, lips, and jaw, creating oropharyngeal resonating cavities, which amplify certain frequency bands of the voice spectrum called 'formants'. Different formants characterize different vowels and show different peaks of acoustic energy. Generally, people with PD have a reduced frequency range [21], [24]–[27], [29] that causes monopitch or monotonous speech, which is typically observed in this patient population [7], [8]. Canter and other authors [21], [25], [30] noticed a decrease in the fundamental frequency (F0) range during syllable production or reading tasks in individuals with PD. Moreover, as the severity of the disease increases, a continued measurable decrease in the F0 variability seems evident [27]. Limitations in formant production, from irregular movements of the articulators, decrease normally high-frequency formants and elevate normally low-frequency formants [31]. Therefore, it is important to observe the spectrum of frequencies emitted by patients through the phonation of a vowel. Many studies considering the reduction in the vocal sound pressure level show conflicting results; thus far, there seems to be uncertainty about the fact that Parkinson influences sound pressure [25]–[27]. Fox and Ramig found that the vocal sound pressure level was 2–4 decibels (at 30 cm) lower across a number of speech tasks in people with PD relative to the control group [28]. A 2–4 decibel change is a significant perceptual change in loudness.

While the research results are promising, some methodological limitations remain, such a as lack of widely accepted protocols and small subject pools. In this paper, we present a study about a new way to measure how much speech is intelligible in Parkinson's disease, basing on a public STT system as Google Speech to Text. This study was performed adding new functions to Voxtester software [35], which will be shortly described in the next section. No previous studies have been presented to address the mentioned challenges. ASR techniques have been used in similar important experiments [39]–[41] but describe studies and experiments much different than ours. In [39] the disability of patients concerns the resection of parts of the oral cavity due to cancer removal. In this case, there is no discussion of anomalies in neuromotor coordination. In fact, the pathology of the oral cavity (and especially the surgical outcomes) causes "peripheral mechanical dyslalia" that have nothing to share with Parkinson's dysarthria and related dysarthrophonia (speech pathology, dysphonia, pathology of the second motor neuron). There is no real dysphonia in the oral cavity pathology. The method in [39] is focused on German language, while a future system based on our study could benefit of the ability of a public STT (Google is just an example) that offer the recognition of almost all languages in the world. Riedhammer *et al.* [40] discuss topics that are covered mostly in [39] and partly much different from our topic. The study presented in [41] is also very appreciable but it shows substantial and important differences from ours. In [41] the authors focus only on consonant mappings because they say that the pronunciation patterns in consonants can play a crucial role in judging the intelligibility of dysarthric speech in Korean language. The plenty literature about Parkinson's Disease reports that voice pronunciation is strongly influenced by the severity of the disease, and many authors consider pronunciation of abnormal vowels as one of the warning signs of the disease itself. So we consider crucial the preservation of the vowels. Furthermore (even if this issue is of particular importance to the study here presented), the Italian language is strongly characterized by the use of vowels that form over 46% of a normal speech, much higher than other languages such as French, English, German, Latin etc.

II. VOXTESTER: DIGITAL ASSESSMENT OF VOICE

There are several software systems that are directly or indirectly used to evaluate some speech disorders in PD, among which we mention Praat [34], a remarkable example of such instruments; however, it seems geared towards use by research centers rather than towards direct use by doctors and speech therapists.

Our project Voxtester [35], a software system simplified and immediately available for use, allows for measuring a small, useful set of parameters as follows:

- Voice spectrum (with comparison);
- F1 to F4 formants;
- acoustic metric of dysarthric speech (FCR, tVSA, qVSA, VAI 3/4);
- diadochokinetic rate;
- decay of voice intensity (emission of a vowel);
- vocal sound pressure level;
- duration of speech.

Voxtester uses a friendly control panel for assessing PDrelated impairment at home, suggesting its use for large clinical trials, especially when early changes or frequent data collection are considered important to document [15]. Additionally, it exchanges data with the Multimedia Electronic Health Record [32] that was already developed by some of the authors of this paper (see http://www.fsem.eu), allowing for a systematic and full recording of patient data.

Further important speech disorders can be evaluated in PD. As a matter of fact, as Parkinson's disease progresses, the speaking rate can either be accelerated or slowed. The speaking rate refers to an advanced state of the disease, and it can mainly be considered to evaluate a modification of the speaking rate with daily therapy. Other deficits in the prosodic characteristics of speech refer, for example, to a decreased naturalness in the melodic line, which can be measured based on changes in the frequency, intensity, rate, and pause. As PD progresses, the intelligibility of speech can be impaired. Voxtester functions are arranged into a single panel (see fig. 1).

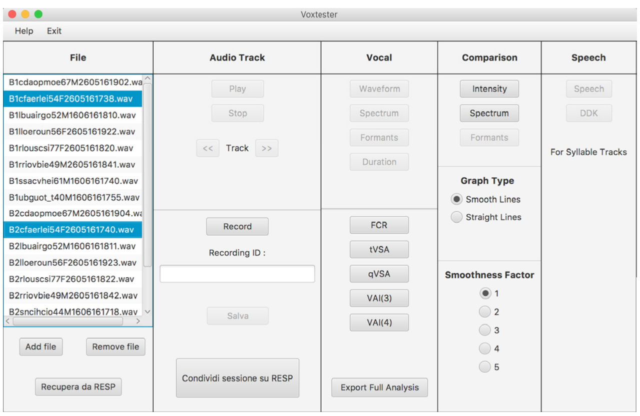

FIGURE 1. **The simplified Voxtester control panel.**

Five columns include all of the following available functions:

- File: file explorer and adding/removing files;
- Audio track: listening, recording and saving new files;
- Vocal: waveform, spectrum, duration of the phonation, and formants;
- Comparison: intensity, spectrum and formants comparison of file selected in the 'file' column, and acoustic metric of dysarthric speech;
- Speech: diadochokinetic rate and speech disorders.

Voxtester can record and process .wav audio files, and we will make it available as free software.

Some functions are automatically activated or deactivated depending on the selected files. One or multiple files can be selected and Voxtester allows for specific analysis that can be performed with the chosen selection. It is based on Java and JavaFX, a Java library that consists of classes and interfaces written in native Java code, which allow one to build network-aware applications that can be deployed across multiple platforms and display information in a high-performance user interface that also features audio and graphics.

III. A PROTOCOL FOR SPEECH ANALYSIS

At present, the evaluation of the speech intelligibility is generally performed based on the UPDRS. Specifically, section 3.1 (*eloquence*) of the cited scale provides the specialist with some tips to evaluate the patient's speech ability, as reported here:

- *Instructions to examiner: listen to the patient's free-flowing speech and engage in conversation, if necessary. Suggested topics: ask about the patient's work, hobbies, exercise, or how he or she got to the doctor's office. Evaluate the volume, modulation (prosody) and clarity, including slurring, palilalia (repetition of syllables) and tachyphemia (rapid speech and running syllables together).*

0: Normal: No speech problems.

1: Slight: Loss of modulation, diction or volume, while all words are easy to understand.

2: Mild: Loss of modulation, diction, or volume with a few unclear words, but the overall sentences are easy to follow.

3: Moderate: Speech is difficult to understand to the point that some, but not most, sentences are poorly understood.

4: Severe: Most speech is difficult to understand or unintelligible.

Further evaluation procedures are reported in [38]. The above metric and scale and the way they are structured cannot be considered a perfectly objective quantitative assessment of the impairment and/or disease. Of note, the ID evaluation using the UPDRS scale is subjective and, to be more reliable, requires at least that a patient is always examined by the same specialist. Ultimately it would be useful to have a system that makes a fast and effective measurement of such voice characteristics. For this reason, we decided to expand Voxtester with a specific task that could make a quantitative and objective measure of the speech impairment. At the same time, we performed significant work with doctors, neurologists, speech therapists and other specialists to define a protocol of analysis of the patient designed specifically to be performed through Voxtester. This protocol requires the vocal production in different modes and execution times, as described in the following list:

a) 2 readings of a phonemically balanced text spaced by a pause (30 sec)

b) execution of the syllable 'pa' (5 sec), pause (20 sec), execution of the syllable 'ta' (5 sec);

c) 2 phonation of the vocal 'a';

d) 2 phonation of the vocal 'e';

e) 2 phonation of the vocal 'i';

f) 2 phonation of the vocal 'o';

g) 2 phonation of the vocal 'u';

h) reading of some phonemically balanced words, pause (1 min), and reading of some phonemically balanced phrases.

It should be noted that between execution a) and b) and between g) and h), there is a one-minute break. Additionally, before the executions referred to in points c), d), e), f), and g), it is necessary to inhale the largest possible amount of air and continue phonating until the lungs are empty (for the first one of each execution, for the second one the patient performs a 5 second phonation). Between executions c), d), e), f), g) and h), it is appropriate to observe a pause of 30 sec. To ensure the replicability of our study and analyses and to make public our information to other researchers for other analyses, all source data, together with the intermediate and final processed data, have been made available on demand. Here, we report the text, phrases and words used (Italian) [36]:

The phonemically balanced text (Italian language)

IL RAMARRO DELLA ZIA. Il papà (o il babbo come dice il piccolo Dado) era sul letto. Sotto di lui, accanto al lago, sedeva Gigi, detto Ciccio, cocco della mamma e della nonna. Vicino ad un sasso c'era una rosa rosso vivo e lo sciocco, vedendola, la volle per la zia. La zia Lulù cercava zanzare per il suo ramarro, ma dato che era giugno (o luglio non so bene) non ne trovava. Trovò invece una rana che saltando dalla strada finì nel lago con un grande spruzzo. Sai che fifa, la zia! Lo schizzo bagnò il suo completo rosa che divenne giallo come un taxi. Passava di lì un signore cosmopolita di nome Sardanapalo Nabucodonosor che si innamorò della zia e la portò con sé in Afghanistan.

The phonemically balanced phrases (Italian language)

- *Oggi è una bella giornata per sciare.*
- *Voglio una maglia di lana color ocra.*
- *Il motociclista attraversò una strada stretta di montagna.*
- *Patrizia ha pranzato a casa di Fabio.*
- *Questo è il tuo cappello?*
- *Dopo vieni a casa?*
- *La televisione funziona?*
- *Non posso aiutarti?*
- *Marco non è partito.*
- *Il medico non è impegnato.*

Words (Italian)

pipa, buco, topo, dado, casa, gatto, filo, vaso, muro, neve, luna, rete, zero, scia, ciao, giro, sole, uomo, iuta, gnomo, glielo, pozzo, brodo, plagio, treno, classe, grigio, flotta, creta, drago, frate, spesa, stufa, scala, slitta, splende, strada, scrive, spruzzo, sgrido, sfregio, sdraio, sbrigo, prova, calendario, autobiografia, monotono, pericoloso, montagnoso, prestigioso.

While the phonemically balanced text seems meaningless, it contains several interesting features as follows:

- it is sufficiently long and requires the patient to breathe with effort, while it stresses his or her resistance.
- it contains similar and complex phonetics at close range to assess the patient's ability to pronounce very difficult sounds in a short time.
- it requires the patient make changes in expression while reading annexed but isolated phrases or exclamations.

Even balanced phrases and words can stress all muscles involved in voice and speech production to perform rapid and forceful movements for assessing the degree of neuro-control as well as the speech intelligibility.

IV. HOW CAN WE MEASURE SPEECH IMPAIRMENTS ?

As we already said above we wonder whether a speech-to-text system (STT) can help us to obtain an accurate and objective measure of speech, phrase and word intelligibility. STT systems (also called automatic speech recognition, ASR) are based on methodologies and technologies that enable the recognition and translation of spoken language into text by computers and computerized devices. These systems allow a computer to identify the words that a person speaks into a microphone (or a smartphone) and convert it to written text. Having a machine that fluently understands spoken speech has driven speech research for more than 50 years. Although STT systems cannot understand 100% of speech in any acoustic environment, or by any person, they are now fully used in several applications and services. The goal of STT research is to allow a computer to recognize in real-time, with 100% accuracy, all words that are intelligibly spoken by any person, independent of the vocabulary size, noise, speaker characteristics or accent.

In the health care sector, speech recognition is used in the front-end or back-end of the medical documentation process for generating narrative text, such as part of a radiology/pathology interpretation, progress note or discharge summary. The recognized draft document is then routed along with the original voice file to the editor, where the draft is corrected and finalized; deferred speech recognition is widely used in the industry at present. The use of speech recognition systems has shown benefits even to short-term-memory re-strengthening in patients with some neuro-diseases (e.g., brain resection).

Based on the above assumption, we want to test whether a speech-to-text system (STT) can be considered advantageous to estimate the modification of the speech impairment of PD patients. We record the speech of these patients and use STT to generate the narrative text; then, the recognized draft document is routed to a specific word matching software system that we have developed. For this purpose, we could also design and develop a new specific speech recognizer and train it. However, due to the wide availability of free or commercial existing systems, we decided to base our research on the Google Speech to Text conversion. Indeed, at this moment we do not consider it useful to perform the analysis of the available STT systems and the identification of criteria for their choice because Google speech recognition, while being available and easily usable through an Android device and/or any PC or laptop, offers excellent performance in speech recognition (in our case: Italian language). Only if and when our results are proven effective, a thorough investigation of the STT system to be used could help improve the analysis system presented here. At this stage, we will try to estimate the evolution of the speech impairment rather than its absolute intelligibility.

V. GOOGLE CLOUD SPEECH API

Google makes available advanced deep learning neural network algorithms to the user's audio for accurate speech recognition [37]. As you can read on Google Cloud Platform, Google Cloud Speech API enables developers to convert audio to text by applying neural network models in an easy to use API. It enables easy integration of Google speech recognition technologies into developer applications. It sends audio and receives a text transcription from the Cloud Speech API service. It recognizes over 80 languages and variants to support a global user base. The text of users dictating to an application's microphone can be transcribed, enabling command-and-control through voice or transcription of audio files, among many other use cases. Users do not need advanced signal processing or noise cancellation before sending audio to Speech API. The service can successfully handle noisy audio from a variety of environments. Speech recognition can also be tailored to context by providing a separate set of word hints with each API call, and the service tries to associate the text received in the form of audio to a set of words known in advance. In the case under examination in this paper, even the speech specialist knows in advance the text that will be read by the patient and thus instinctively tries to associate the patient's spoken words with those included in the text. Therefore, this software feature must be seen as advantageous.

In our experiment, we agreed with the specialists' recommendation to ignore the short words (<4 characters). We also decided to ignore the accents, as we verified that Google SR sometimes does not correctly recognize the accent of words that have meaning with and without an accent as "trovò" and "trovo" (Engl. "he found it" and "I find"), while it almost always correctly recognizes the accent of words such as *andò* (went), *bambù* (bamboo) or *virtù* (virtue) because these words have no meaning without the accent.

Furthermore, we decided to verify whether the recognition of some words is affected by any systematic error. As will be shown later, we found that a subset of words was hardly recognized; we will think about the opportunity to preserve these words in the text to read and not consider them in a final analysis protocol of PD patients.

At the end of each speech recognition session, the software system returns the following measures:

- speech duration (time 1/2/3, seconds);
- number of words read;
- characters per second (CPS 1/2/3);
- recognition error rate;
- other measures (formants and others not reported here)

It generates a full report to compare the analysis performed for each HC (Healthy Control) or patient or by grouping them, but as our attention is currently focused on the validation of this approach, we will not mention the specific functions for more detailed analysis.

VI. THE EXPERIMENT

As we stated, the first phase of our experiment is arranged to verify how much the automatic recognition of some words through Google STT is affected by any systematic recognition error. The first group enrolled in this experiment consisted of young people (YHC) because we imagine that their neuromotor characteristics help define a reliable reference lower limit of the 'systematic' error rate for each word. Fifteen healthy people aged 19-29 years were asked to participate in this part of the experiment, including 13 men and 2 women from the Puglia region (Italy), 13 from the Bari area and 2 from the Brindisi area.

They performed the activities specified in section 3 in strict compliance with the rules established in the protocol under a) and the first part of point h), precisely reading the 'balanced text' and 'balanced words'. The reading exercises were performed after a short explanation by the specialist in a quiet room that was echo free while keeping the distance from the microphone 15 to 25 cm. The muscles of the voice system have to be considered outstretched, as the experiment occurred starting from 10 a.m. in a warm room (approximately 22°C) after a brief friendly dialogue. The text and words have been read from a printed sheet (font: Times New Roman, bold, 20).

In tab. 1, the details of the first part of this experiment are reported. Specifically, for each person, these include the sex, age, time (sec) spent reading the balanced text (times 1 and 2) and balanced words (time 3) as well as the characters per second (CPS).

TABLE 1. Data on the YHC - first experimental phase - (time is in seconds).

			TEXT 1^READING		TEXT 2^READING		TEXT 3^READING	
sex	**age**	**from**	**time1**	**CPS1**	**time2**	**CPS2**	**time3**	**CPS3**
M	19	BA	45,61	11,36	54,20	9,56	53,71	5,23
M	21	BR	50,70	10,22	47,20	10,97	59,72	4,71
M	20	BR	40,02	12,94	40,12	12,91	46,71	6,02
M	29	BA	41,61	12,45	42,42	12,21	38,50	7,30
M	22	BA	45,32	11,43	46,52	11,13	53,42	5,26
F	19	BA	40,72	12,72	40,62	12,75	38,82	7,24
M	24	BA	44,12	11,74	43,42	11,93	49,32	5,70
M	19	BA	38,50	13,45	40,20	12,89	42,61	6,59
M	21	BA	44,12	11,74	43,81	11,82	48,82	5,76
M	20	BA	38,32	13,52	39,21	13,21	45,92	6,12
M	20	BA	46,32	11,18	43,21	11,99	37,62	7,47
M	20	BA	41,33	12,53	40,02	12,94	45,82	6,13
M	19	BA	54,52	9,50	46,71	11,09	45,40	6,19
M	19	BA	43,21	11,99	41,40	12,51	36,70	7,66
F	20	BR	46,60	11,12	49,42	10,48	45,40	6,19
		average	**44,07**	**11,86**	**43,90**	**11,89**	**45,90**	**6,24**

In tab. 2, some words from among those that have been read are listed, including the recognition error rate (>20%). Some words have been recognized with a very high error rate, and we will preserve them in the balanced text and words while excluding them in the definition of a regular diagnostic protocol.

The second part of the experiment was performed with a group of Healthy Elderly persons (HEC). Twenty-two healthy persons aged 60-77 years were asked to participate in this experiment, 10 men and 12 women, all from Bari, (Puglia region, Italy). None of the persons reported particular speech or language disorders. They performed the activities specified in section 3 in strict compliance with the rules set out in the protocol under a) and the first part of point h), precisely reading the 'balanced text' and the 'balanced words'.

TABLE 2. Recognition error rate for a subset of words (YHC).

WORDS	not recognized	tot readings	error rate
vaso	15	15	100%
volle	30	30	100%
fifa	24	30	80%
scia	12	15	80%
pozzo	11	15	73%
vedendola	22	30	73%
finì	20	30	67%
divenne	15	30	50%
trovò	13	30	43%
gnomo	6	15	40%
innamorò	12	30	40%
sfregio	5	15	33%
sdraio	4	15	27%
sedeva	8	30	27%
topo	4	15	27%
OTHER WORDS	63	2385	≤20%

The reading exercises were performed after a short explanation by the specialist, in a quiet, echo-free room, keeping the distance from the microphone to 15 to 25 cm. The muscles of the voice system have to be considered outstretched as the experiment occurred in the afternoon (on different days) in a warm room (approximately 22°C) after a brief friendly dialogue. The text and words have been read from a printed sheet.

In tab. 3, the details of the second part of the experiment are reported. Specifically, for each HEC person, the sex, age, time (sec) spent reading the balanced text (times 1 and 2) and balanced words (time 3) as well as the characters per second (CPS) were evaluated. As can be seen in tab. 4, a subset of all words that were read were not always correctly recognized (rec error rate >20%) in the HEC group in this case.

TABLE 3. Data on the HEC – second experimental phase.

		TEXT 1^READING		TEXT 2^READING		TEXT 3^READING	
sex	age	time1	CPS1	time2	CPS2	time3	CPS3
F	69	57,12	9,37	49,99	10,70	47,11	5,96
F	62	100,95	5,30	77,26	6,92	66,4	4,23
F	65	70,87	7,55	57,71	9,27	43,16	6,51
M	68	59,55	8,98	55,08	9,71	43,47	6,46
F	68	55,97	9,56	53,01	10,09	51,98	5,41
M	70	-	-	-	-	64,5	4,36
M	60	60	8,92	54,3	9,85	38,23	7,35
F	60	59,49	8,99	54,97	9,73	43,96	6,39
F	61	66,92	7,99	53,89	9,93	43,77	6,42
M	68	58,92	9,08	56,31	9,50	33,8	8,31
F	63	70,3	7,61	71,26	7,51	44,8	6,27
M	68	58,57	9,13	51,26	10,44	42,02	6,69
F	69	64,45	8,30	54,34	9,85	33,26	8,45
M	76	56,09	9,54	53,15	10,07	49,15	5,72
M	77	67,97	7,87	59,75	8,95	67,33	4,17
F	63	75,88	7,05	62,73	8,53	54,65	5,14
M	69	65,88	8,12	63,09	8,48	38,59	7,28
F	61	59,09	9,05	53,51	10,00	56,81	4,95
F	70	148,83	3,59	104,51	5,12	64,67	4,35
M	62	92,7	5,77	70,11	7,63	49,26	5,70
M	75	68,74	7,78	65,43	8,18	41,19	6,82
F	72	67,37	7,94	67,35	7,94	67,64	4,15
	average	**70,74**	**7,97**	**61,38**	**8,97**	**49,35**	**5,95**

Now we consider the third part of the experiment performed with a PD patients group. Twenty-eight patients aged 40-80 years were asked to participate in this experiment, 19 men and 9 women, 27 from the Bari area (Puglia region, Italy) and one from Venice. None of the patients reported speech or language disorders unrelated to their PD symptoms prior to this study. All patients were receiving antiparkinsonian treatment. The severity of their disease was classified by the specialists as <4 on the modified Hoehn and Yahr scale, except for 2 patients with stage 4 and 1 patient with stage 5. They performed the activities specified in section 3 in strict compliance with the rules set out in the protocol under a) and the first part of point h), precisely reading the 'balanced text' and the 'balanced words'. Additionally, they performed all exercises a) to h) reported in section 3 (the results of the other exercises will be useful for other current research).

TABLE 4. Recognition error rate for a subset of words (HEC).

WORDS	not recognized	tot readings	error rate
vedendola	37	38	97,37%
vaso	18	20	90,00%
pozzo	16	20	80,00%
volle	29	38	76,32%
scia	14	20	70,00%
sfregio	12	20	60,00%
creta	10	20	50,00%
fifa	18	38	47,37%
gnomo	8	20	40,00%
sdraio	8	20	40,00%
topo	8	20	40,00%
plagio	7	20	35,00%
zero	7	20	35,00%
bagnò	13	38	34,21%
rete	6	20	30,00%
sgrido	6	20	30,00%
giro	5	20	25,00%
glielo	5	20	25,00%
luna	5	20	25,00%
muro	5	20	25,00%
pipa	5	20	25,00%
slitta	5	20	25,00%
trovò	8	38	21,05%
OTHER WORDS	111	3034	≤20%

The reading exercises were performed after a short explanation by the specialist, in a quiet, echo-free room, keeping the distance from the microphone to 15 to 25 cm. The muscles of the voice system have to be considered outstretched as the experiment occurred in the range from 10 a.m. to 2 p.m. (on different days) in a warm room (approximately 22°C) after a brief friendly dialogue. The text and words have been read from a printed sheet.

In tab. 5, the details of the third part of the experiment are reported. Specifically, for each PD person, the sex, UPDRS scale, age, time (sec) spent reading the balanced text (times 1 and 2) and balanced words (time 3) as well as the characters per second (CPS) were evaluated. As can be seen in tab. 6, a subset of all words that were read were not always correctly recognized (rec error rate >20%) in the PD group in this case.

Some preliminary observations can be made on the test execution mode. It should be noted that time 1 for the PD and HEC groups (average time respectively 94,35 and 70,74 sec) is much higher than time 2 (average time respectively 67,50 and 61,38 sec). In the case of YHC, the difference is

TABLE 5. Data on the PD group - third experimental phase.

			TEXT 1^ READING		TEXT 2^ READING		WORDS 1^ READING	
sex	UPDRS	age	time 1	CPS1	time 2	CPS2	time 3	CPS3
F	0	63	//	//	//	//	60,64	4,63
M	2	50	71,73	7,22	53,82	9,62	39,75	7,07
F	1	61	53,40	9,70	51,4	10,08	49,25	5,71
M	2	68	84,05	6,16	63,32	8,18	56,66	4,96
F	0	40	60,92	7,76	52,4	9,89	54,58	5,15
M	1	65	52,40	9,89	50,23	10,31	40,2	6,99
M	1	73	79,35	6,53	71,22	7,27	69,62	4,04
M	0	56	86,81	5,97	66,64	7,77	58,7	4,79
M	1	77	64,75	8,00	60,9	8,51	50,41	5,57
M	0	71	59,84	8,66	56,76	9,13	53,52	5,25
F	1	71	66,22	7,82	48,38	10,71	55,56	5,06
M	0	71	49,95	10,37	45,35	11,42	36,86	7,62
M	1	73	70,98	7,30	65,07	7,96	56,87	4,94
M	2	75	62,20	8,33	56,58	9,16	52,98	5,30
M	2	68	85,37	6,07	63,35	8,18	48,67	5,77
M	0	71	62,33	8,31	52,56	9,86	66,38	4,23
F	2	65	242,50	2,14	180,09	2,88	167,83	1,67
F	1	80	169,29	3,06	//	//	101,19	2,78
M	0	73	66,90	7,74	63,5	8,16	53,04	5,30
M	0	70	65,46	7,91	60,4	8,58	48,25	5,82
F	1	67	79,30	6,53	73,8	7,02	71,67	3,92
F	0	54	55,00	9,42	49,8	10,40	54,7	5,14
F	3	78	163,60	3,17	//	//	108,3	2,59
M	1	72	117,60	4,40	98,8	5,24	87,51	3,21
M	4	65	164,10	3,16	//	//	151,3	1,86
M	4	65	233,00	1,76	//	//	217,3	0,96
M	0	70	112,00	4,63	106,6	4,86	68,31	4,11
M	0	70	68,30	7,58	61,47	8,43	64,5	4,36
		average	94,35	6,65	67,50	8,42	73,02	4,60

TABLE 6. Recognition error rate for a subset of words (PD).

WORDS	not recognized	tot readings	error rate
vaso	27	28	96,43%
fifa	47	50	94,00%
volle	46	50	92,00%
scia	25	28	89,29%
pozzo	25	28	89,29%
vedendola	39	50	78,00%
sfregio	18	28	64,29%
bagnò	26	50	52,00%
sgrido	13	28	46,43%
glielo	13	28	46,43%
finì	23	50	46,00%
sdraio	11	28	39,29%
slitta	11	28	39,29%
giro	11	28	39,29%
iuta	11	28	39,29%
gnomo	10	28	35,71%
trovò	17	50	34,00%
sbrigo	9	27	33,33%
splende	9	28	32,14%
Afghanistan	15	49	30,61%
Nabucodonosor	15	49	30,61%
sole	8	28	28,57%
innamorò	13	49	26,53%
topo	7	28	25,00%
creta	7	28	25,00%
dado	19	77	24,68%
Sardanapalo	12	49	24,49%
sedeva	12	50	24,00%
portò	11	49	22,45%
cosmopolita	11	49	22,45%
divenne	11	50	22,00%
schizzo	11	50	22,00%
neve	6	28	21,43%
grigio	6	28	21,43%
drago	6	28	21,43%
OTHER WORDS	295	3323	≤20%

negligible (indeed in some cases, time 2 is higher than time 1). This result suggests that a more accurate analysis could be achieved by organizing a short exercise session for the patients of the PD and HEC groups before reading the balanced text.

It should also be highlighted that some words suffer from difficult recognition (top wrong words: *vaso, volle, fifa, scia, pozzo, vedendola*) for all groups. In this case, the age, reading speed and disease do not appear as key factors and are, therefore, likely to consider the misrecognition of the above words as a 'quasi'-systematic error made by Google STT, but we will discuss this result in more detail later.

Some words were correctly recognized with a higher rate in the PD group and this result seems to disprove our hypothesis that the recognition rate related to YHC can be considered a reference rate. This apparent 'anomaly' arises because, in the records of the PD group (with average lower speech rate), words such as *trovò*, *innamorò* and *portò* are recognized as *trovo*, *innamoro* and *porto* in most cases (and we considered them correctly recognized as specified before). In the recordings of the YHC, these words are often confused with *trova*, *innamora* and *porta* (last character differs) and are considered incorrectly recognized. In any case, this 'anomaly' occurs on words that are affected by a significant incorrect recognition rate.

A very interesting result of this experiment can be better observed in fig. 2, where we show the words read by the participants (x-axis) versus the recognition error rate (y-axis). On the x-axis, the words have been ordered with a decreasing recognition error rate of the YHC.

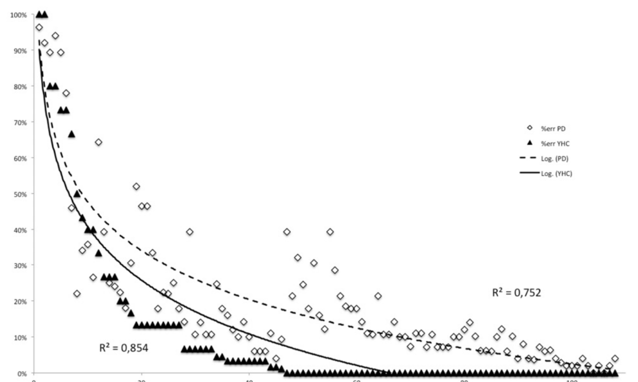

FIGURE 2. Error rate YHC vs. PD with trend lines (y-axis: error rate, x-axis: words ordered in descending error% for YHC).

In fig. 2, we use logarithmic trendlines because they are a best-fit curved line that is most useful when the rate of change in the data quickly increases or decreases and then levels out. A trendline is most reliable when its R-squared value is at or near 1; note the R-squared value calculated on the entire set of words is $R^2 = 0,85$ for the YHC group and $R^2 = 0,75$ for the PD group, which are a relatively good fit of the lines for the data.

To identify a subset of words that is more difficult to recognize in both groups, we report the trendlines in Figure 3 related to the first 15 top incorrectly recognized words (HYC and PD groups). In fig. 4, we report the trendlines

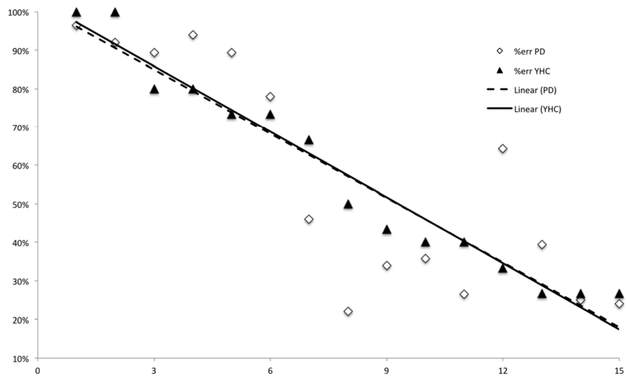

FIGURE 3. Error rate for the 15 top words with trendlines.

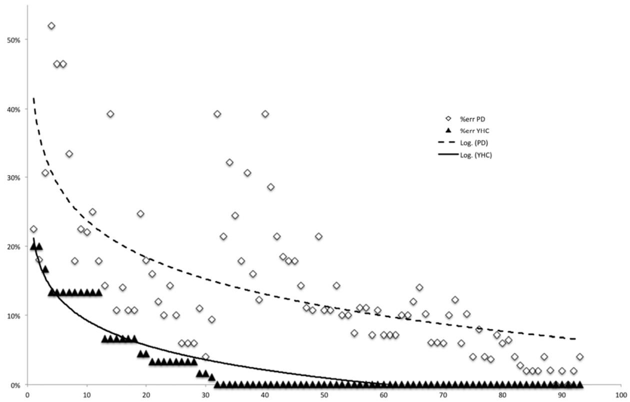

FIGURE 4. Comparison of error rate on 93 words YHC vs PD (y-axis: error rate, x-axis: words ordered in descending error% for YHC).

related to the 93 remaining words (YHC and PD groups). It is evident that the decrease in the error rate is highly correlated between the two groups in Figure 3. In Figure 4, while the decreasing trend remains almost similar, it is evident that there is a strong difference in the recognition error rate.

The profile of the trend lines shows what we have asserted, namely, that a reduced set of words is affected by a similar recognition error in the case of the YHC and PD groups and we consider them as a 'systematic' error due to STT.

It is interesting to note also that the error rate is almost always higher in the case of the PD group (figs 2 and 4). This was a result we were expecting, but it cannot be simply attributed to the disease. This inference needs further analysis because the present study is not focused on the early prediction of the onset of the disease or its assessment in absolute terms, but it is precisely focused on the definition of a specific protocol and the design of a reliable system to evaluate them by studying the variations in the misrecognition of words that we would consider as an assessment of the intelligibility of speech: then specific pre/post treatment assessments could be performed through our system.

We consider also the comparison between the healthy control group of aged-people HEC with PD group.

In fig.5, we report the trendlines related to the 93 remaining words (PD and HEC groups) after excluding the first 15 top

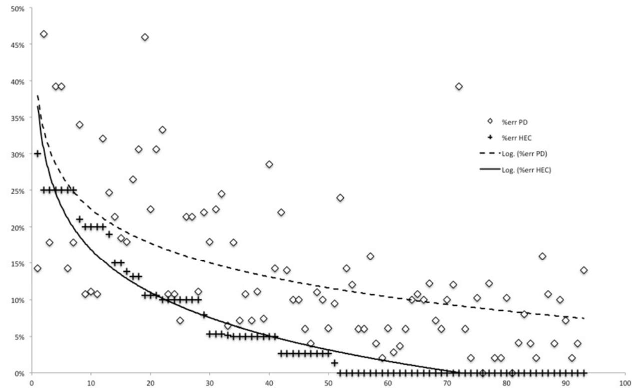

FIGURE 5. Comparison of error rate on 93 words HEC vs PD (y-axis: error rate, x-axis: words ordered in descending error% for HEC).

incorrectly recognized words. It is evident that while the decreasing trend remains almost similar, there is a strong difference in the recognition error rate. The error rate is almost always higher in the case of the PD group also when compared to Healthy Elderly persons, a result we were expecting.

VII. CONCLUSION

The key findings of our study concern the design of a specific speech recording protocol, the use of an effective, simple to use and inexpensive software system, together with a public STT, to obtain some relevant measures in PD treatments, considering it is rather difficult to perform experiments, due to the need for direct interactions with PD patients and the limited number of available patients. We hope that this will confirm the effectiveness of this approach to improve PD treatments. In this paper we deal with the assessment of the intelligibility of speech, a complex task for which the evaluation by humans, albeit subjective, is more reliable than evaluations performed by machines. The results discussed in the paper state that the specific defined protocol and realized software system are useful to show the efficacy of evaluating the intelligibility of speech in Parkinson's Disease by analyzing the variations of the misrecognition of words through a well-known STT public system, such as Google. This approach is an important step toward the automatic assessing of intelligibility of speech in Parkinson's disease. In the next future we will extend experimentation to verify if speech/voice disturbances can be identified basing on the study presented here and through Voxtester software in early parkinsonian patients.

ACKNOWLEDGMENTS

The authors thank Dr. Vito Napoletano and Dr. Marianna Sumerano for their valuable assistance in this study and the "Associazione Parkinson Puglia", which made it possible to perform the experimental phase by involving many patients.

REFERENCES

[1] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *Lancet Neurol.*, vol. 5, no. 6, pp. 525–535, 2006.

[2] K. K. Baker, L. O. Ramig, E. S. Luschei, and M. E. Smith, "Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging," *Neurology*, vol. 51, no. 6, pp. 1592–1598, 1998.
[3] L. O. Ramig, C. Fox, and S. Sapir, "Speech treatment for Parkinson's disease," *Exp. Rev. Neurotherapeutics*, vol. 8, no. 2, pp. 297–309, 2008.
[4] C. A. Moore and R. R. Scudder, "Coordination of jaw muscle activity in parkinsonian movement: description and response to traditional treatment," in *Recent Advances in Clinical Dysarthria*, K. M. Yorkston and D. R. Beukelman, Eds. Boston, MA, USA: College-Hill Press, 1989, pp. 147–163.
[5] K. M. Yorkston, R. M. Miller, and E. A. Strand, *Management of Speech and Swallowing in Degenerative Diseases*. Tucson, AZ, USA: Communication Skill Builders, 1997.
[6] S. Skodda, W. Visser, and U. Schlegel, "Vowel articulation in Parkinson's disease," *J. Voice*, vol. 25, no. 4, pp. 467–472, 2011.
[7] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *Behav. Neurol.*, vol. 11, no. 3, pp. 131–137, 1998.
[8] 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," *J. Speech Hearing Disorders*, vol. 43, no. 1, pp. 47–57, 1978.
[9] S. Sapir, L. O. Ramig, P. Hoyt, C. O'Brien, and M. Hoehn, "Speech loudness and quality 12 months after intensive voice treatment (LSVT) for Parkinson's disease: A comparison with an alternative speech treatment," *Folia Phoniatrica*, vol. 54, no. 6, pp. 296–303, 2002.
[10] C. G. Goetz *et al.*, "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, Nov. 2008.
[11] F. Attivissimo, G. Cavone, A. M. L. Lanzolla, and M. Savino, "Application of hand grip signals for an objective evaluation of Parkinson disease: Analysis and comparison with standard functional clinical tests," *Measurement*, vol. 42, no. 8, pp. 1123–1130, Oct. 2009.
[12] B. T. Harel, M. S. Cannizzaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment," *J. Neurolinguistics*, vol. 17, no. 6, pp. 439–453, Nov. 2004.
[13] H. Cohen, "Disorders of speech and language in Parkinson's disease," in *Mental and Behavioral Dysfunction in Movement Disorders*, M. A. Bédard, Y. Agid, A. D. Korczyn, P. Lesperance, and S. Chouinard, Eds. New York, NY, USA: Humana Press, 2003, pp. 125–134.
[14] R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, "Voice characteristics in the progression of Parkinson's disease," *Int. J. Lang. Commun. Disorders*, vol. 35, pp. 407–418, Jul. 2000.
[15] C. G. Goetz *et al.*, "Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device," *Movement Disorders*, vol. 24, no. 4, pp. 551–556, 2009.
[16] H. Ackermann and W. Ziegler, "Die Dysarthrophonie des Parkinson–Syndroms," *Fortschritte Neurol. Psychiatrie*, vol. 57, no. 4, pp. 149–160, 1989.
[17] A. T. Neel, "Vowel space characteristics and vowel identification accuracy," *J. Speech, Lang. Hearing Res.*, vol. 51, pp. 574–585, Jun. 2008.
[18] A. R. Bradlow, G. M. Torretta, and D. B. Pisoni, "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.*, vol. 20, no. 3, pp. 255–272, 1996.
[19] K. Forrest, G. Weismer, and G. S. Turner, "Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults," *J. Acoust. Soc. Amer.*, vol. 85, no. 6, pp. 2608–2622, 1989.
[20] H. Ackermann, I. Hertrich, and T. Hehr, "Oral diadochokinesis in neurological dysarthrias," *Folia Phoniatrica Logopaedica*, vol. 47, pp. 15–23, Feb. 1995.
[21] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: III. Articulation, diadochokinesis, and over-all speech adequacy," *J. Speech Hearing Disorders*, vol. 30, pp. 217–224, Aug. 1965.
[22] H. Ackermann, J. Konczak, and I. Hertrich, "The temporal control of repetitive articulatory movements in Parkinson's disease," *Brain Lang.*, vol. 57, no. 2, pp. 312–319, 1997.
[23] S. Fletcher, "Time-by-count measurement of diadochokinetic syllable rate," *J. Speech Hearing Res.*, vol. 15, pp. 763–770, Dec. 1972.
[24] A. K. Ho, R. Iansek, and J. L. Bradshaw, "Motor instability in Parkinsonian speech intensity," *Neuropsychiatry Neuropsychol. Behav. Neurol.*, vol. 14, no. 2, pp. 109–116, 2001.
[25] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: I. Intensity, pitch, and duration," *J. Speech Hearing Disorders*, vol. 28, no. 3, pp. 221–229, 1963.
[26] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: II. Physiological support for speech," *J. Speech Hearing Disorders*, vol. 30, pp. 44–49, Feb. 1965.
[27] E. J. Metter and W. R. Hanson, "Clinical and acoustical variability in hypokinetic dysarthria," *J. Commun. Disorders*, vol. 19, no. 5, pp. 347–366, 1986.
[28] C. M. Fox and L. O. Ramig, "Vocal sound pressure level and self-perception of speech and voice in men and women with idiopathic Parkinson disease," *Amer. J. Speech Lang. Pathol.*, vol. 6, pp. 85–94, May 1997.
[29] V. Fraïle and H. Cohen, "Prosody in Parkinson's disease: Relations among duration, intensity and fundamental frequency range," in *Proc. 23rd Annu. Meet. Int. Neuropsychol. Soc.*, Seattle, WA, USA, Feb. 1995, doi: 10.1017/S1355617700001843.
[30] A. J. Flint, S. E. Black, I. Campbell-Taylor, G. F. Gailey, and C. Levinton, "Acoustic analysis in the differentiation of Parkinson's disease and major depression," *J. Psycholinguistic Res.*, vol. 21, no. 5, pp. 383–399, 1992.
[31] N. Roy, S. L. Nissen, C. Dromey, and S. Sapir, "Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy," *J. Commun. Disorders*, vol. 42, no. 2, pp. 124–135, 2009.
[32] G. Dimauro, D. Caivano, F. Girardi, and M. M. Ciccone, "The patient centered electronic multimedia health fascicle—EMHF," in *Proc. IEEE Workshop Biometric Meas. Syst. Secur. Med. Appl. (BIOMS)*, Oct. 2014, pp. 61–66.
[33] G. Defazio, M. Guerrieri, D. Liuzzi, A. F. Gigante, and V. di Nicola, "Assessment of voice and speech symptoms in early Parkinson's disease by the Robertson dysarthria profile," *Neurological Sciences*, vol. 37, no. 3, pp. 443–449, Nov. 2015, doi: 10.1007/s10072-015-2422-8.
[34] P. Boersma and D. Weenink. (Feb. 2017). *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.25*. [Online]. Available: http://www.praat.org/
[35] G. Dimauro, D. Caivano, V. Bevilacqua, F. Girardi, and V. Napoletano, "VoxTester, software for digital evaluation of speech changes in Parkinson disease," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Benevento, Italy, May 2016, pp. 1–6, doi: 10.1109/MeMeA.2016.7533761.
[36] F. Farace *et al.*, "Free anterolateral thigh flap versus free forearm flap: Functional results in oral reconstruction," *J. Plastic, Reconstruct. Aesthetic Surgery*, vol. 60, no. 6, pp. 583–587, 2007.
[37] *Google Cloud Platform—Google Cloud Speech API*. [Online]. Available: https://cloud.google.com/speech/
[38] *National Parkinson Foundation—Evaluation*. [Online]. Available: http://www.toolkit.parkinson.org/content/evaluation
[39] A. Maier, M. Schuster, A. Batliner, E. Nöth, and E. Nkenke, "Automatic scoring of the intelligibility in patients with cancer of the oral cavity," in *Proc. INTERSPEECH*, Aug. 2007, pp. 1206–1209.
[40] K. Riedhammer *et al.*, "Towards robust automatic evaluation of pathologic telephone speech," in *Proc. IEEE ASRU*, Dec. 2007, pp. 717–722.
[41] M. J. Kim, Y. Kim, and H. Kim, "Automatic intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 4, pp. 694–704, Apr. 2015.

GIOVANNI DIMAURO was born in Taranto, Italy, in 1964. He received the Laurea degree in computer sciences from the Università degli Studi di Bari in 1987. He is currently an Associate Professor of computer programming, multimedia systems, cooperative systems with the Università degli Studi di Bari. His research activities, described in over 150 papers published in scientific journals, proceedings and books, are currently on e-health, multimedia systems, pattern recognition with applications in medicine such as new diagnosis technology for anemia and in Parkinson disease.

VINCENZO DI NICOLA was born in Bari, Italy, in 1957. He received the degrees in medicine and surgery and specialization in ear, nose, and throat from the University of Bari, in 1982 and 1985, respectively. He is currently an Associate Professor of phoniatrics with the University of Bari. His research activities, described in over 180 papers published in scientific journals, proceedings and books, are currently on laryngeal oncology, voice/speech pathology and acoustic analysis of voice with applications in medicine, such as new evaluation technology for disease.

VITOANTONIO BEVILACQUA received the Laurea degree in electronic engineering and the Ph.D. degree in electrical engineering from the Polytechnic University of Bari, Italy. Since 1996, he has been involving and investigating in the field of computer vision and image processing, human–machine interaction, bioengineering, machine learning, and soft computing (neural networks, evolutionary algorithms, with main applications in medicine, biometry, bioinformatics, and bioengineering. In 2017, he was qualified as an Associate Professor of bioengineering and has been the Head of the Industrial Informatics Laboratory, Department of Electrical and Information Engineering, Polytechnic University of Bari, since 2017. He is currently a Tenured Professor of human–computer interaction with the Department of Electrical and Information Engineering, Polytechnic University of Bari. He has authored over 140 papers in refereed journals, book, and international conferences proceedings.

DANILO CAIVANO received the Ph.D. degree from the University of Bari Aldo Moro in 2002. He carries out his research in the Software Engineering Laboratory, Department of Informatics. His research and teaching activities focus on topics related to Software Engineering with emphasis on project and process management in collocated and distributed contexts and on software development, maintenance, and testing. Since 2007, he has been a Chief Executive Officer of SER&Practices, a spin off company of the University of Bari that he has contributed to start up. He is actively involved in the Project Management Institute, Southern Italy Chapter, and the International Software Engineering Research Network. He is currently an Assistant Professor with the University of Bari Aldo Moro.

FRANCESCO GIRARDI was born in Bari, Italy, in 1962. He received the degree in medicine and surgery from the Università degli Studi di Bari in 1986 and the bachelor's degree in computer science in 2013. Three years later, he specialized in general haematology at the Università degli Studi di Bari and he is currently with the Azienda Sanitaria Locale in Bari, Italy. Since 2013, he has been cooperating with the Computer Science Department, University of Bari, and the Department of Electrical and Information Engineering, Polytechnic University of Bari. His research activities are currently on the e-health field.