Evaluation of an Automatic Speech Recognition Platform for Dysarthric Speech

Irene Calvo, Peppino Tropea, Mauro Viganò, Maria Scialla, Agnieszka B. Cavalcante, Monika Grajzer, Marco Gilardone, Massimo Corbo

Preprint version of the paper:

Calvo, I., Tropea, P., Viganò, M., Scialla, M., Cavalcante, A. B., Grajzer, M., Gilardone, M., & Corbo, M. (2021). Evaluation of an automatic speech recognition platform for dysarthric speech. *Folia Phoniatrica et Logopaedica*, 73(5), 432-441.

Published paper available at: https://doi.org/10.1159/000511042

Research Article

An automatic speech recognition platform for dysarthric speech: assessment of performance

Irene Calvo[1+], Peppino Tropea[1+]*, Mauro Viganò[1], Maria Scialla[1], Agnieszka B. Cavalcante[2], Monika Grajzer[2], Marco Gilardone[1] and Massimo Corbo[1]

[1] Department of Neurorehabilitation Sciences, Casa di Cura del Policlinico, Milan, Italy

[2] Gido Labs, Poznan, Poland

[+] These authors contributed equally to this work

Short Title: The evaluation of mPASS platform in terms of speech recognition accuracy and practical applicability in a sample of individuals with and without dysarthria.

*Corresponding Author

Peppino Tropea, PhD,

Casa Cura Policlinico

Department of Neurorehabilitation Sciences

Via Giuseppe Dezza, 48

20144, Milan, Italy

Phone: +39 02 4859 3124 Email: p.tropea@ccppdezza.it

Keywords: Automatic Speech Recognition, dysarthric speech, mPASS platform, neurological disorder

Abstract

Introduction

The use of commercially available Automatic Speech Recognition (ASR) software is challenged when dysarthria accompanies a physical disability. To overcome this issue a Mobile and Personal Speech Assistant (mPASS) platform was developed, using a speaker-dependent ASR software.

Objective.

The aim of this study was to evaluate the performance of proposed platform and to compare mPASS recognition accuracy to a commercial speaker-independent ASR software. In addition, secondary aims were to investigate the relationship between severity of dysarthria and accuracy and to explore dysarthric speech users' perceptions on the proposed platform.

Methods

Fifteen dysarthric speech and twenty normal speech individuals recorded 24 words and 5 sentences in a clinical environment. Differences in recognition accuracy between the two systems were evaluated. In addition, mPASS usability was assessed with a Technology Acceptance Model (TAM) questionnaire.

Results

In both groups, mean accuracy rates were significantly higher with mPASS compared to the commercial ASR for words and for sentences. mPASS reached good levels of usefulness and ease of use according to the TAM questionnaire.

Conclusions

Practical applicability of this technology is realistic: mPASS platform is accurate and it could be easily used by individuals with dysarthria.

Introduction

Neurological disorders such as stroke, brain injuries, motor neuron diseases, and cerebral palsy are known to affect the motor functions of upper and lower limbs. Consequently, the independent and autonomous access to the immediate environment and the use of basic technological devices (e.g., keyboards, mouse, mobile phones, tablets) is often limited.

Automatic Speech Recognition (ASR), a software used to recognize and act upon spoken language, has been introduced as an alternative input method to be used in Environmental Control Systems for people with severe physical disabilities [1, 2]. However, access to ASR application is more challenging when a communication disorder, such as dysarthria, accompanies a physical disability [3].

Dysarthria is a neurological motor speech disorder and is considered one of the most common acquired communication disorders [4]. It can impair all processes involved in speech production including respiration, phonation, articulation, resonance and prosody, resulting in reduced speech intelligibility. In particular, impaired articulation is a common source of degraded intelligibility in these speakers [5].

ASR technology is based on the reproducibility and consistency of the vocal signal that is captured, analyzed and identified by a machine. Therefore, ASR application becomes limited for individuals with dysarthric speech, as their articulation is more imprecise and less consistent in comparison with healthy individuals [6]. Additionally, literature reports that commercially available ASR systems poorly recognize the speech of people with degraded vocal quality [7].

Usually, ASR systems are categorized in two main classes: the speaker-independent systems and the speaker-dependent systems. Commonly available ASR systems created to recognize normal speech belong to the first group (i.e., speaker-independent) [8]. Well-known examples of speaker-independent systems are the Apple and Google virtual assistants. These systems use acoustic models (usually accessed online via specific cloud environment) trained with speech samples of many subjects mostly without speech impairments, so the system requires no training or adaptation to a particular user's speech [8]. Speaker-independent systems are generally considered inadequate for the recognition of dysarthric speech, with recognition accuracy being inversely proportional to dysarthria severity [9, 10].

Most of the current ASR systems for users with dysarthria are speaker-dependent: their potential use is similar to the speaker-independent ones but the systems training and the underlying algorithms are different. They require speaker training prior to the use, to allow the system to be built on samples of the user's own speech. During the training phase, the speaker is required to provide his/her speech samples (words and phrases related to the system's topic scope). Hence, the acoustic

model used is aligned to the particular speech of this user. These systems typically work well only for the person who trains it and usually reach better accuracy rates than speaker-independent software in individuals with severe dysarthria [9, 11]. However, speaker-dependent systems require great training time and effort to reach acceptable accuracy rates, increasing the risk of becoming less feasible and more tiring for people with disabilities.

Previous experiences in building speaker-dependent ASR software yielded accuracy rates ranging between 52 and 99% [10, 12, 13]; however, these solutions were tested on a small number of individuals with severe dysarthria (range 4-8), and the training required an extensive amount of time for each user (up to six weeks with one hour of daily practice) [10, 12, 13]. A long ASR training might be too demanding for people with neurological disabilities because limb motor skills and respiration are often impaired in addition to speech motor functions. Moreover, the application of these approaches was questioned by the end-users in terms of perceived frustration, since these systems often required the repetition of the verbal command and they were susceptible to environmental noise [10, 12, 13].

To overcome these obstacles an innovative concept of a "Mobile and Personal Speech Assistant" (mPASS) platform was developed in order to increase the clinical applicability of ASR systems for individuals with dysarthric speech [14]. The aim was to create a user-friendly speaker-dependent ASR system, easy to use in a clinical/home setting without the necessity of a large amount of training data. The mPASS platform is a set of software tools for building an ASR system, which enable the users to create on their own a customized ASR. A system created this way is tailored to their speech disorders, needs, and capabilities [14]. Most notably, the user can specify the scope of the system (e.g., which words or phrases the system will be able to recognize). The system was tested with eigth individuals (children and young adults) with dysarthria, that were non-homogeneous in term of type and severity of their disorder [15]. Then, a proof-of-concept field trial with a voice-controlled mobile texting application was undertaken recruiting one individual with cerebral palsy, who presented ataxic dysarthria and both upper and lower limbs motility impairments. The mPASS platform showed good accuracy rates in home environment (accuracy=84%), and the individual judged the mPASS application as 49% better than the traditional manual input [16]. A preliminary study involving five persons with Amyotrophic Lateral Sclerosis (pALS) with moderate mixed dysarthria reported an average of 85% accuracy rate for single words [17]. In this preliminary study the perceptual characteristics of pALS speech were rather inhomogeneous but the accuracy of mPASS application was steadily higher compared to that of a commercially available software [17].

Furthermore, to the best of our knowledge, no study has investigated any possible difference between speaker-dependent and speaker-independent systems regarding the recognition accuracy

of non-dysarthric speech. Divergences between these systems' performances with dysarthric and non-dysarthric speakers might influence their application setting and their commercial distribution.

The aim of the present study was to evaluate the performance of the mPASS platform with a larger sample of individuals with and without dysarthria, and to compare mPASS recognition accuracy to a commercial speaker-independent software. In addition, secondary aims were to investigate the relationship between severity of dysarthria and accuracy and to explore dysarthric speech users' perceptions on the mPASS platform.

Materials And Methods

A. Participants

Thirty-five subjects, twenty with normal speech production and fifteen individuals with dysarthric speech were enrolled in this study. Individuals with dysarthria were recruited among the inpatient and outpatient population of Casa di Cura del Policlinico (CCP), a neuro-rehabilitation hospital in Milan, through purposive sampling. Demographic and clinical characteristics of dysarthric speech participants were summarized in Table 1. Individuals with normal speech were recruited among hospital employees, through convenience sampling.

Inclusion criteria for dysarthric speech participants were: stable clinical conditions (ability to undertake a treatment session in a clinical setting) and presence of a motor speech impairment as assessed by an experienced clinician. Exclusion criteria, for both groups, were: presence of cognitive impairment (according to neurologist clinical assessment) that interfere with the ability to willingly understand and give informed consent, inability to read the Italian language, and age<18 years.

Enrolled subjects signed informed consent forms. All research procedures were in accordance with the Declaration of Helsinki and were approved by the Local Ethics Committee (Ethical Committee Milano Area B 133_2017bis).

B. Robertson Dysarthria Profile

All individuals with dysarthria were assessed with the Italian version of the Robertson Dysarthria Profile (RDP) [18, 19] by an experienced speech and language therapist, as it represents the current clinical practice in the hospital. RDP is a scale used for the assessment of the clinical features of dysarthria. It is comprised of 71 items divided into different categories: respiration, phonation, facial muscles, diadochokinesis, reflexes, articulation, intelligibility and prosody. The scoring ranges from 0 to 284 ("severe" and "within normal limits", respectively). For the purpose of the present study, the analysis was focused on the total scoring of the RDP (RDP_{TOT}) and on the partial score of two specific

categories: articulation (RDP_{ART}) and intelligibility (RDP_{INT}). Articulation category is based on the clinician evaluation of speech-sound accuracy during words and sentences repetition tasks. Intelligibility category is assessed during reading and spontaneous speech tasks through the qualitative ratings of the clinician, a caregiver and an external listener.

C. mPASS platform

The mPASS platform is an online web-based application developed by Gido Labs (Gido Labs sp. z o.o., Poznan, Poland, available at: www.mpass.gidolabs.eu). The platform can be accessed from any device with a Google Chrome web browser and it allows the creation of individual, personalized speech recognition systems by people with speech disorders without assistance. The necessary parts of such a system are: (1) acoustic model which represents the relationship between an audio signal of a given person's voice and chosen linguistic units of speech (usually phonemes), (2) dictionary which defines words that can be recognized by a given ASR system, and (3) language model which conveys the information on how those words can be arranged into recognized sentences. The mPASS platform creates all 3 components based on a personal user input. The overview of system architecture has already been described in a previous study [15] and is recapped here. The mPASS platform is designed for non-technical users and it does not require any special hardware. Therefore, it can be used in different clinical and home environments. After creating an online account, a software toolchain guides the user in the process of creating his/hers personalized ASR system. Firstly, the user defines the vocabulary needed; then, with the help of visual and acoustic feedback cues, the user records his/her speech samples for ASR training. Users have full control over the duration of each recording session and the amount of data collected for the targeted ASR system. This allows to overcome problems previously reported in the literature [8, 12, 20, 21] and keeps the users engaged in the process. In addition, the mPASS platform provides tools for semi-automated development of the dictionary and the language model (represented by rule-based grammar or a statistical n-gram language model). ASR system training is performed automatically – the acoustic model for each given user is trained based on the collected speech samples. After the training phase, the user can test the newly created ASR system by recording a series of phrases/sentences proposed by mPASS platform (within the expected scope of the system). The recorded samples are recognized by the developed ASR system and, based on the results, the recognition accuracy is provided to the user. In addition, the user has the option to collect more training data and re-train the whole system in an attempt to improve its accuracy. Finally, users can export the ASR system created with mPASS to a desired speech-based application. Although the envisioned target device is a mobile device (the system was tested on Android and iOS smartphone operative systems) the tests were also run on Linux PC and Mac laptop. Users can then export the ASR system created with mPASS to a desired

speech-based mobile application. As mPASS works with the users' own vocabulary, acoustic and language models, the platform is language-agnostic and can be used with any language.

The mPASS platform allows to create different types of ASR systems with different levels of complexity, ranging from small-vocabulary, command-based systems, to dictation-based systems with different vocabulary sizes for the recognition of sentences and phrases [15, 16].

Although recently deep neural networks are being used to address speech recognition challenges, their proper training requires large amount of data. Collecting the required amount of data from a given speaker with speech impairments would be impractical and too tiring for the user. Therefore, in the present study, the mPASS platform was used to create a speaker-dependent ASR based on Hidden Markov models (HMMs). The system was used for the recognition of single discrete words and sentences. Phonemes were used as a basic recognition unit that is represented by a single HMM with three states and four or eight mixtures per state.

Mel Frequency Cepstral Coefficients (MFCC) were used for calculating feature vectors of the speech signal. The dimension of each vector was 39 (12 spectral components, energy, 12 delta components, delta energy, 12 delta-delta components, and delta-delta energy).

*D. **Experimental setup***

Dysarthric speech and normal speech participants utilized the mPASS online platform for the recording and testing of 24 words and 5 sentences in a clinical environment. The words, listed in Table 2, comprised the principal phonemes and consonant clusters of the Italian language. The sentences (Table 3) were short expressions with a maximum of three words combination, describing states of being, commands and small requests.

Participants were sitting in front of a laptop, where the online platform of the mPASS software was running; at least one clinician was present to facilitate and supervise the use of the platform. Participants were asked to read aloud the word/sentence appearing on the screen, cued by a changing color button. Every participant would record the 24 words and 5 sentences for 5 times. Then, mPASS platform used these recordings to create his/her personal ASR system.

Afterward, the recognition accuracy trial was performed in real time with the participant recording again all the words and the sentences, appearing in random order.

The microphone used for the recording was the standard built-in laptop microphone. The recording and testing of the mPASS platform were performed in a room at the CCP hospital clinic. The acoustic environment resembled a quiet home environment. No special measures were taken to prevent noise (i.e., no echo-cancellation chambers where used).

Participants completed the training and testing of the mPASS platform during different 45'-sessions, accordingly to the individuals' clinical conditions and fatigue level. The first session included signing of the informed consent.

E. Usability assessment

At the end of the testing dysarthric speech participants filled in an ad-hoc questionnaire built accordingly to the Technology Acceptance Model (TAM) [22, 23].

In the Information Systems community, TAM is the most popular model among those proposed to explain and predict the acceptance of a system. TAM questionnaire provides a basis with which one traces how external variables influence belief, attitude, and intention to use. Two cognitive beliefs are posited by TAM: perceived usefulness and perceived ease of use.

In the framework of the present study the developed TAM questionnaire comprised ten questions (Table 4) with a 4-point Likert scale (1 = strongly disagree; 4 = strongly agree) analyzing users' perceptions on mPASS platform's ease of use and usefulness.

F. Commercial ASR software

To test the concurrent validity of mPASS the speech samples collected during the recognition trial were used at Gido Labs premise with an open access commercially available speaker-independent ASR software, with the release of Italian language models: Pocketsphinx (Pocketsphinx, CMUSphinx, Carnegie Mellon University) [24]. To make the comparison fair, we ensured that both Pocketsphinx and mPASS ASR were using the same set-up grammar during decoding process. More specifically, parameters controlling VAD (Voice Activity Detection) and feature extraction had the same values. The language model, dictionary, and the decoding algorithms were equivalent. The only crucial difference was in the acoustic model: mPASS was using its own model trained for specific user, while Pocketsphinx utilized online available speaker independent Italian acoustic model designed for this platform.

G. Data and Statistical analysis

Accuracy rates (as percentage over the total [%]) in recognizing dysarthric and normal speech were measured as the average of words and sentences correctly identified by the two tested ASR systems (i.e., mPASS and commercial ASR).

All variables were analyzed using descriptive statistics. Variables were not normally distributed (Kolmogorov- Smirnov test). Therefore, a Wilcoxon test was used to determine statistical significance of the accuracy using both ASR systems (i.e., mPASS and commercial ASR) for both groups (i.e., normal and dysarthric speech). Spearman's correlation was used to explore the association between

RDP scores (i.e., RDP_{TOT}, RDP_{ART}, RDP_{INT}) and accuracy rates for dysarthric speech participants.

Data were processed by using custom routines developed under Matlab environment (Mathworks Inc., Natick, MA, USA). For all statistical tests, the significance was set at $\alpha=0.05$.

Results

A. Participants' characteristics

All enrolled participants completed the entire experimental procedure. No adverse events were reported during the study. One individual with dysarthria (participant #5), who completed the mPASS data collection, was however excluded from the final analysis as Italian was not his mother tongue. Dysarthric speech participants training time varied from one to four 45'-sessions, with a mean of 2.3±1 sessions. All normal speech individuals completed data collection in one 45'-session.

Most of the enrolled individuals with dysarthria showed articulatory imprecision, pneumophonic coordination deficit and harsh voice. The group was slightly heterogeneous, including individuals with different type of dysarthria and with both increased and decreased speech rate.

The control group subjects had a mean age of 39.5±10.8 year (range 26-67), 60% were females. None of the participants was a trained speaker and Italian regional varieties were fairly represented..

B. Accuracy rates

Figure 1 shows words recognition accuracy rates (mean and one standard deviation). The left panel shows values for individuals with dysarthric speech using both the mPASS (ACC_{mPASS}) and the commercial ASR software (ACC_{comm}): 88.7±12.2% (range 58.3-100%) and 63.7±21.0% (range 8.3-87.5%) respectively. The right panel shows words recognition accuracy rates for normal speech subjects with both ASR platforms: mPASS reached 98.3±2.5% (range 91.6-100%) and commercial ASR 86.3±8.1% (range 75.0-95.8%).

Statistical analysis showed that the difference in words recognition accuracy between mPASS and commercial ASR reached statistical significance in the dysarthric speech group ($p<0.001$) and also in the normal speech group ($p<0.0001$).

Concerning sentences accuracy rates for both groups using both ASR systems (i.e., mPASS platform and commercial ASR software), the dysarthric speech group presented a mean recognition accuracy of 98.6±5.3% (range 80.0-100%) with mPASS and a mean of 88.6±23.1% (range 20.0-100%) with commercial ASR. The normal speech group presented a mean of 100% recognition accuracy for both ASR systems.

C. *Correlation between RDP scores and accuracy rates*

All Spearman's rho had a magnitude below 0.5 (range: 0.02-0.42). All three RDP scores (RDP_{TOT}, RDP_{ART}, RDP_{INT}) showed a trend of positive correlation with both ACC_{mPASS} and ACC_{comm}, indicating higher RDP scores are related to higher accuracy rates (Figure 2).

RDP_{TOT} presented a weak correlation with both ASR systems accuracy (rho_{mPASS}=0.10, rho_{comm}=0.02) RDP_{INT} scores showed a weak correlation with ACC_{mPASS} (rho_{mPASS}=0.23) and a moderate correlation with ACC_{comm} (rho_{comm}=0.41). RDP_{ART} scores showed a moderate correlation with ACC_{mPASS} (rho_{mPASS}=0.42) and a weak correlation with ACC_{comm} (rho_{comm}=0.16).

D. *Usability assessment*

The TAM questionnaire results showed a mean score of 17.3 (± 3.07) for perceived usefulness (20 means strongly agree) and a mean score of 16.4 (± 2.71) for perceived ease of use (20 means strongly agree). In table 4, for each item of the TAM questionnaire, are reported the mean score and standard deviation.

Discussion

This study analyzes the performance of mPASS, a novel tool for developing speaker-dependent ASR system, in recognizing normal and dysarthric speech compared to a commercial speaker-independent ASR software.

To the best of our knowledge, this is the first study directly comparing the performance of a speaker dependent and a speaker independent software in the same sample of dysarthric and normal speech individuals. Previous small-scale studies investigated the accuracy and use of different ASR adaptive, independent and dependent software with dysarthric speech [10-13]. A previous case study reported the accuracy rate of three different speaker-adaptive systems with one individual with dysarthria and one normal speech individual [24]. One study reported on the accuracy of a speaker-independent software used with typical speech and dysarthric speech individuals [25]. Sample size in the present study is larger than in previous ones whose performed analysis using four [10] and eight [12] subjects. Furthermore, inclusion criteria in this study, where individuals presented different degrees of dysarthria severity ranging from mild to severe, might have been larger than in previous studies that included people with moderate and severe dysarthria [11].

In this study, mPASS achieved significantly higher accuracy rates than commercial ASR (Pocketsphinx) when used by individuals with dysarthria. This finding confirms that ASR performance is better with speaker-dependent software than speaker-independent ones for the recognition of dysarthric

speech [8, 9, 11, 20]. Moreover, a statistically significant difference between mPASS and Pocketpshinx for single words recognition was found even in a group of individuals without any speech disorder. Therefore, mPASS might represent a valuable software for the recognition of both dysarthric and non-dysarthric speech.

In this study the mPASS recognition accuracy rate for dysarthric speech (i.e., mean words was 88.6%) is consistent with the single case study reported by Cavalcante and Grajzer [15]. Moreover, this result is comparable to accuracy rates obtained by other speaker dependent ASR software used with people with dysarthria, which showed a range varying between 64-100% [9, 11-13].

Sentence recognition accuracy for dysarthric speech in the present study was very high (98.5%). However, this result might not indicate real accuracy for sentences, as the number of target stimuli was very small (five sentences) and recognition accuracy with speaker-dependent software usually decreases as the vocabulary increases.

The number of word target stimuli (i.e., 24) is larger than most previous studies, reporting accuracy for a list of a word varying from 12 to 47 [13]. Moreover, the number of repetitions for each stimulus with mPASS was only five, while most of other studies recorded the stimulus for more than 20 times. Therefore, contrary to previous studies [12, 13] the time needed to train mPASS was considerably reduced, with a maximum total training time of four 45'-sessions. As accuracy rate in speaker-dependent ASR software increases with the number of repetition, mPASS might represent a good option for reaching good accuracy rate with minimum effort and energy required to the individual. This is important as people with dysarthria often tire easily when requested to speak for long period of time. Moreover, in the case of neurodegenerative disorders on-going reassessments of communication needs and adjustment to the supports are required at short time intervals [26].

The relationship between severity and characteristics of dysarthria and accuracy has not been widely investigated in the past. Previous literature reported that mild and moderate dysarthric speech provides better accuracy rates than severe dysarthria [25]. However, in this study population, RDP total score showed a weak correlation with percentage of recognition accuracy. RDP total score is a perceptual measure that marks the grade of speech impairment and includes different items that might not directly influence speech production (i.e., reflexes). Furthermore, RDP intelligibility score and RDP articulation score showed a weak and moderate positive correlation with accuracy rates, respectively. Previous studies reported correlations of intelligibility measures and recognition success for dysarthric speech with speaker-adapted software in a single case study [27] and in ten individuals with chronic spastic dysarthria [28]. The heterogeneous nature of the dysarthric impairment in the sample may account for weak/moderate correlations. Moreover, it is possible that, in a speaker-dependent software like mPASS, intelligibility is not critical in determine accuracy, as long as words

articulation is consistent. Greater variability in the articulation of speech sounds would account for poorer performance in recognition accuracy.

Users' positive perceptions and satisfactions represent an important driving force in planning and implementing technology devices aimed at supporting individuals' social interactions and communication abilities. In this study, TAM questionnaire results showed individuals with dysarthria supposed the mPASS platform to be a useful and easy to use tool. However, during the experimental setup, a clinician was always present and acting as a moderator between the individual and mPASS interface, as most patients were not used to technological devices. Substantially, the clinician was not a person with an IT or ASR background.

Contrary to previous studies [13] mPASS accuracy rates seem large enough to allow for using the ASR software in real-life applications. Practical applicability of this technology is realistic: mPASS platform is simple and easy to use, and the system could be re-trained at home – adding more speech samples for training that would allow to further increase accuracy rates.

This study presents some limitations. Firstly, the small sample number limits the generalization of these findings to a wider population of individuals with dysarthric speech, even though the number of individuals recruited in the present study is larger than in previous research. Secondly, due to the exploratory nature of the present study the mPASS ASR system was tested in a clinical environment; therefore, mPASS accuracy rates reported in his study might not replicate in other and noisier environments (e.g., outdoors or at home). However, to resemble "real-word" environment, recordings were not performed with professional equipment in a soundproof room, probably leading to less precise but more lifelike measurements. Moreover, the commercial ASR software used as a comparison (Pocketsphinx) is not considered the benchmark of speaker-independent ASR software. Using different speaker-independent systems might provide different results.

Future studies would need to investigate the mPASS ASR performance in recognizing dysarthric speech in a larger group of individuals, and further examine the correlation between accuracy and dysarthria severity and characteristics. Concerning this latter point, future investigation will also take into account the use of non-linear model and multivariate statistics in order to better define these associations. In addition, future studies are needed to explore the use of mPASS software with the creation of individualized applications tailored to the users' needs.

Acknowledgement

The authors wish to thank all the participants enrolled in the study for their time and contributions.

Statement of Ethics

Enrolled subjects signed written informed consent forms.

All research procedures were conducted ethically in accordance with the World Medical Association Declaration of Helsinki.

All research procedures were approved by the Local Ethics Committee (Ethical Committee Milano Area B 133_2017bis).

Disclosure Statement

A. B. Cavalcante and M. Grajzer have financial interests in a company that may be affected by the research reported in the enclosed paper.

No potential competing interests are reported by other authors.

Funding Sources

This work was supported in part by the “SLACIAMOCI non-profit organization”.

Author Contributions

I.C., M.Gi. and M.C. designed the study based on the platform developed and set up by A.B.C. and M.Gr. I.C., M.V. and M.S. performed patients' enrolment and data collection. I.C. and P.T. performed data analysis and interpreted the results. M.V., M.S., A.B.C., M.Gr., M.Gi. and M.C. contributed in interpreting the results, I.C., P.T. and M.V. took the lead in writing the manuscript. All authors provided critical feedback and helped shape the manuscript. All authors read and approved the final version of the manuscript.

References

1. Derosier R, Farber RS. Speech recognition software as an assistive device: a pilot study of user satisfaction and psychosocial impact. Work. 2005;25(2):125-34.

2. Koester HH. Usage, performance, and satisfaction outcomes for experienced users of automatic speech recognition. Journal of rehabilitation research and development. 2004 Sep;41(5):739-54.

3. Kim M, Kim Y, Yoo J, Wang J, Kim H. Regularized Speaker Adaptation of KL-HMM for Dysarthric Speech Recognition. IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society. 2017 Sep;25(9):1581-91.

4. Enderby PM, Emerson J. Does speech and language therapy work? : a review of the literature. London: Whurr Publishers; 1995.

5. Rong P, Yunusova Y, Wang J, Zinman L, Pattee GL, Berry JD, et al. Predicting Speech Intelligibility Decline in Amyotrophic Lateral Sclerosis Based on the Deterioration of Individual Speech Subsystems. PloS one. 2016;11(5):e0154971.

6. Rudzicz F. Using articulatory likelihoods in the recognition of dysarthric speech. Speech Communication. 2012 2012/03/01/;54(3):430-44.

7. Muhammad G, Mesallam TA, Malki KH, Farahat M, Alsulaiman M, Bukhari M. Formant analysis in dysphonic patients and automatic Arabic digit speech recognition. Biomedical engineering online. 2011 May 30;10:41.

8. Rosen K, Yampolsky S. Automatic speech recognition and a review of its functioning with dysarthric speech. Augmentative and Alternative Communication. 2000;16(1):48-60.

9. Young V, Mihailidis A. Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: a literature review. Assistive technology : the official journal of RESNA. 2010 Summer;22(2):99-112; quiz 13-4.

10. Hamidi F, Baljko M, Livingston N, Spalteholz L. CanSpeak: a customizable speech interface for people with dysarthric speech. International Conference on Computers for Handicapped Persons: Springer; 2010. p. 605-12.

11. Fager SK, Beukelman DR, Jakobs T, Hosom JP. Evaluation of a speech recognition prototype for speakers with moderate and severe dysarthria: a preliminary report. Augment Altern Commun. 2010 Dec;26(4):267-77.

12. Hawley MS, Enderby P, Green P, Cunningham S, Brownsell S, Carmichael J, et al. A speech-controlled environmental control system for people with severe dysarthria. Medical engineering & physics. 2007 Jun;29(5):586-93.

13. Hawley MS, Cunningham SP, Green PD, Enderby P, Palmer R, Sehgal S, et al. A voice-input voice-output communication aid for people with severe speech impairment. IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society. 2013 Jan;21(1):23-31.

14. Cavalcante AB, Lorens L. Use case: a mobile speech assistant for people with speech disorders. Proceedings of the 7th Language & Technology Conference2015. p. 192-97.

15. Cavalcante AB, Grajzer M. Proof-of-concept Evaluation of the Mobile and Personal Speech Assistant for the Recognition of Disordered Speech. 2016;9:589.

16. Cavalcante AB, Grajzer M. Mobile and Personal Speech Assistant for the Recognition of Disordered Speech. Proceedings of the Second International Conference on Smart Portable, Wearable, Implantable and Disability-oriented Devices and Systems (SPWID 2016)2016. p. 6-10.

17. Calvo I, Tropea, Scialla, Cavalcante AB, Grajzer M, Gilardone M, et al. An automatic speech recognition platform for dysarthric speech: assessment of accuracy. Proceedings of the Sixth National Conference on Bioengeenering, (GNB 2018). 2018.

18. Robertson SJ. Dysarthria profile. 1982.

19. Fussi F, Cantagallo A. Profilo di valutazione della disartria. Omega Edizioni, Torino. 1999.

20. Raghavendra P, Rosengren E, Hunnicutt S. An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition systems. Augmentative and Alternative Communication. 2001;17(4):265-75.

21. Havstam C, Buchholz M, Hartelius L. Speech recognition and dysarthria: a single subject study of two individuals with profound impairment of speech and motor control. Logopedics, phoniatrics,

vocology. 2003;28(2):81-90.

22. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly. 1989:319-40.

23. Chuttur MY. Overview of the technology acceptance model: Origins, developments and future directions. Working Papers on Information Systems. 2009;9(37):9-37.

24. Hux K, Rankin-Erickson J, Manasse N, Lauritzen E. Accuracy of three speech recognition systems: Case study of dysarthric speech. Augmentative and Alternative Communication. 2000;16(3):186-96.

25. Fager SK, Burnfield JM. Speech Recognition for Environmental Control: Effect of Microphone Type, Dysarthria, and Severity on Recognition Results. Assistive Technology. 2015;27(4):199-207.

26. Hanson EK, Fager SK. Communication supports for people with motor speech disorders. Topics in Language Disorders. 2017;37(4):375-88.

27. Kotler A-L, Thomas-Stonell N. Effects of speech training on the accuracy of speech recognition for an individual with a speech impairment. Augmentative and Alternative Communication. 1997;13(2):71-80.

28. Ferrier L, Shane H, Ballard H, Carpenter T, Benoit A. Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. Augmentative and Alternative Communication. 1995;11(3):165-75.

Figure Legends

Fig. 1. Mean and one standard deviation (one side error band) of accuracy rates for dysarthric individuals (on left) and normal speech subjects (on right), using mPASS ASR (black bars) and commercial ASR (light gray bars). The labels * and ** indicate a statistical difference ($p<0.001$ and $p<0.0001$, respectively) among groups.

Fig. 2. Scatter-plot illustrating the relationship of accuracy rates against the RDP scores (RDP_{TOT}, RDP_{ART}, RDP_{INT}). Dysarthric individuals using mPASS and Commercial ASR are represented with squares and circles, respectively. Regression lines are included with solid lines for mPASS, dashed lines for Commercial ASR.

Table Legends

Table 1: Demographic and clinical characteristics of dysarthric speech participants. The highest possible RDP scores (i.e., "within normal limits") are reported in brackets in the first line.

Table 2: The 24 words used for the recording and testing. For each word are reported the principal phonemes and consonant clusters of the Italian language.

Table 3: The five sentences used for the recording and testing.

Table 4: The TAM questionnaire. Mean and standard deviation score for each item.

TABLES

Table 1: Demographic and clinical characteristics of dysarthric speech participants. The highest possible RDP scores (i.e., "within normal limits") are reported in brackets in the first line.

#	Age	Gender	Diagnosis	RDP_{TOT} (284)	RDP_{INT} (24)	RDP_{ART} (20)
1	40	F	Tetraparesis in bleeding of pons	198	18	18
2	79	M	Left hemorrhagic stroke	156	12	8
3	86	M	Left ischemic stroke	243	18	20
4	75	F	ALS	184	18	13
5	48	M	ALS	168	18	15
6	70	F	ALS	193	19	20
7	88	F	Left ischemic stroke	179	13	12
8	72	F	MSA	190	20	19
9	70	M	Parkinson's Disease	220	22	18
10	70	F	PLS	186	21	19
11	85	F	Cerebellar Syndrome	199	15	12
12	39	F	Cerebral Palsy	166	16	16
13	82	M	Left ischemic stroke	210	15	14
14	62	M	ALS	229	24	20
15	49	M	ALS	149	9	7

(ALS: Amyotrophic Lateral Sclerosis; MSA: Multiple Systems Atrophy; PLS: Primary Lateral Sclerosis)

464 Table 2: The 24 words used for the recording and testing. For each word are reported the principal
465 phonemes and consonant clusters of the Italian language.

Word	Phonemes and Consonant Clusters	Word	Phonemes and Consonant Clusters
Bicchiere	/b/ ; /jɛ/; /e/	Scimmia	/ʃ/
Pipa	/p/; /a/	Lettera	/l/
Mucca	/m/ ; /u/	Rosa	/r/; /z/
Telefono	/t/ ; /o/	Ciao	/t͡ʃ/
Dente	/d/; n+t	Gelato	/d͡ʒ/
Natale	/n/	Tazza	/ts/
Caramella	/k/	Zero	/dz/
Gatto	/g/	Treno	t+r
Bagno	/ɲ/	Poltrona	l+t+r
Famiglia	/f/; /ʎ/	Completo	m+p+l
Vino	/v/; /i/	Dentifricio	n+t; f+r
Sapone	/s/	Stufa	s+t

466

467 Table 3: The five sentences used for the recording and testing.

Sentence (Italian Language)	Sentence (English Language)
Sono stanco	I'm tired
Ho fame	I'm hungry
Mangio il gelato	I eat ice-cream
Voglio bere	I want to drink
Chiudi la porta	Close the door

468

469 Table 4: The TAM questionnaire. Mean and standard deviation score for each item.

Construct	Measurement instrument	Mean (STD)
Perceived ease of use	I find mPASS system easy to use	2.9 (0.92)
	The directions are clear and understandable	3.6 (0.50)
	Learning how to use mPASS system is easy for me	3.3 (0.83)
	Using mPASS system is funny for me	3.1 (1.03)
	Interacting with mPASS system does not require a lot of my mental effort	3.4 (0.74)
		TOTAL 16.4 (2.71)
Perceived usefulness	The mPASS system is useful to me	3.4 (0.76)
	The mPASS system could increase my communicability efficiency	3.4 (0.76)
	The mPASS system could help me in ADL	3.2 (0.80)
	The mPASS system could enhance my effectiveness in ADL	3.6 (0.76)
	I would suggest mPASS system to people with my same deficit	3.6 (0.63)
		TOTAL 17.3 (3.07)

470

471