

# The first Italian Dysarthric Speech Database for improving daily living of severely dysarthric people

*Marco Marini<sup>1</sup>, Mauro Viganò<sup>2</sup>, Massimo Corbo<sup>2</sup>, Marina Zettin<sup>3</sup>, Gloria Simoncini<sup>3</sup>, Bruno Fattori<sup>4</sup>, Clelia D'Anna<sup>5</sup>, Massimiliano Donati<sup>1</sup>, Luca Fanucci<sup>1</sup>*

<sup>1</sup> Dip. di Ingegneria dell'Informazione, Università di Pisa, Pisa, IT

<sup>2</sup> Dip. di Scienze Neuroriabilitative, Casa di Cura Policlinico, Milano, IT

<sup>3</sup> Dip. di Psicologia, Università di Torino, Centro Puzzle, Torino, IT

<sup>4</sup> Dip. di Medicina Clinica e Sperimentale, Università di Pisa, IT

<sup>5</sup> Unità Otorinolaringoiatria audiologia e foniatria, AUOP, Pisa, IT

L'evoluzione dei sistemi di riconoscimento automatico del parlato (RAP) ha permesso l'utilizzo dei dispositivi intelligenti tramite comandi vocali. Questo traguardo è molto importante soprattutto per le persone che soffrono di patologie che ne limitano la mobilità, poiché gli permette di interagire con tali dispositivi senza doverli toccare o manipolare. Ma cosa accade quando la persona ha dei difetti di pronuncia o un linguaggio difficilmente interpretabile?

Le persone affette da disturbi dell'apparato fonatorio, come ad esempio la disartria, hanno difficoltà ad usare i sistemi RAP [1] [2] [3] poiché il loro modo di parlare è molto differente da quello di persone normodotate. Al fine di migliorare le performance di questi sistemi, la comunità scientifica negli anni ha realizzato diversi database di registrazioni di parlato disartrico. Infatti, gli algoritmi e modelli di intelligenza artificiale (IA) hanno bisogno di una grande quantità di dati per poter imparare e lavorare al meglio. Molti di questi database sono in lingua inglese, come ad esempio il Withaker [4] che contiene 19275 parole registrate da 6 persone disartriche dovuta a paralisi cerebrale. Oppure Nemours [5] e Universal Access speech (UASpeech) [6] e Torgo che contengono rispettivamente le registrazioni di singole parole e frasi molto corte di 11, 19 e 7 persone disartriche. Tuttavia, ad oggi non risulta esserci alcun database di parlato disartrico italiano e questo pone grandi limiti ai ricercatori italiani.

Per questo motivo, lo scopo della nostra ricerca è quello di realizzare il primo database di parlato disartrico italiano, utile per realizzare sistemi RAP usabili dai disartrici italiani. Per fare ciò, ci siamo rivolti a diversi enti e strutture ospedaliere per raccogliere più pazienti possibili con un ampio spettro di livelli di disartria. In particolare, le strutture con cui abbiamo collaborato sono 3: Azienda Ospedaliero-Universitaria Pisana (AOUP), il Centro Puzzle di Torino e la Casa di Cura Privata del Policlinico di Milano. Tale collaborazione ha due scopi principali: caratterizzazione del paziente che presta la propria voce (inserimento di dati anagrafici e clinici il più esaustivo possibile); registrare una quantità prefissata di parole per un certo numero di volte per il massimo numero di pazienti possibile.

Per la scelta delle parole da registrare, è stato deciso di affrontare uno scenario realistico e che preveda l'uso di parole singole e non troppo elaborate. Questo perché essendo il primo lavoro su parlato disartrico italiano, si è preferito affrontare una situazione un po' più semplice rispetto ad un riconoscimento automatico del parlato continuo che presenta molteplici criticità come le alterazioni prosodiche. Oltre a parole di uso domestico, si sono aggiunte altre parole al fine di coprire il più possibile tutti i fonemi italiani.

L'obiettivo finale per i pazienti provenienti dalla AOUP è stato quello di registrare 45 parole per 3 volte (non consecutive) per un totale di 135 registrazioni a paziente, mentre per i pazienti provenienti dalle altre strutture, registrare 211 parole per 3 volte (non consecutive) per un totale di 633 registrazioni a paziente. Le 45 parole sono un sottoinsieme delle 211. La differenziazione tra la AOUP e le altre strutture è dovuta al tempo molto limitato che gli operatori sanitari e i medici hanno a disposizione con i loro pazienti e quindi non possono far registrare una grande quantità di dati. Dato che una persona disartrica solitamente è influenzata da altre patologie, un processo di registrazione potrebbe risultare molto stressante e faticoso, per questo motivo alcuni pazienti hanno interrotto la procedura di registrazione, quindi per loro alcune registrazioni sono mancanti. Le persone che sono riuscite a completare interamente la procedura di registrazione sono 21 (46% del totale), 8 persone (17% del totale) hanno registrato meno della metà delle registrazioni prestabilite, mentre 16 persone (35% del totale) ha completato più della metà delle registrazioni. Tutte le registrazioni sono state effettuate con lo stesso microfono a condensatore

[7], usando una codifica lineare in formato PCM a 16 bit e una frequenza di campionamento di 16kHz. Ogni parola registrata è salvata in un singolo file wave.

Per ogni paziente, prima di registrare qualsiasi parola, è stata compilata una cartella personale contenente dati anagrafici (nome, cognome, genere e data di nascita) e clinici. Tale cartella è stata compilata da un medico competente. I dati clinici comprendono la patologia del paziente (è possibile scegliere la patologia in una lista di 9) ed altre informazioni aggiuntive come la data della diagnosi o la data dei primi sintomi. Inoltre, c'è la possibilità di compilare una scala di valutazione della patologia specifica qualora sia possibile. È altresì possibile specificare il tipo di disartria secondo la tassonomia di Duffy. La lista completa delle patologie selezionabili è la seguente: Sclerosi laterale amiotrofica, Parkinson e parkinsonismi, Corea di Huntington, Sclerosi Multipla, Distrofia Miotonica, Atassia, Ictus, Trauma Cranico, Neuropatia. Se il paziente non presenta alcuna di queste patologie, è possibile scegliere la voce "Altro" ed inserire delle note per spiegare la patologia specifica di quel paziente. Questi dati sono stati inseriti tramite un programma su computer provvisto di una semplice interfaccia grafica, appositamente realizzato dai ricercatori dell'Università di Pisa. Una volta profilato il paziente, il programma permette di iniziare a registrare la sua voce. L'intera procedura di registrazione è completamente assistita da un operatore che interagisce con il programma, il quale proporrà a schermo in modo automatico, la parola che il paziente deve ripetere ad alta voce. Se la registrazione non viene completata in modo corretto (rumore di fondo, il paziente ha pronunciato la parola sbagliata, ecc), l'operatore può far ripetere la registrazione al programma, oppure passare alla parola successiva nel caso in cui la registrazione viene effettuata correttamente. Il programma offre anche la possibilità di cambiare il contrasto della parola mostrata a schermo in modo da renderla più facile da leggere anche ai pazienti ipovedenti. Quando il paziente ha registrato tutte le parole almeno tre volte, il programma notifica il raggiungimento dell'obiettivo e chiude la procedura. Se il paziente vuole continuare a registrare, può farlo liberamente e le parole registrate in più si aggiungono alle altre. Tutti i file di registrazione vengono salvati in locale nel computer dove è installato il programma e successivamente inviate a all'Università di Pisa direttamente dall'operatore.

In totale sono stati arruolati 45 volontari (25 maschi e 20 femmine) con 8 diverse patologie diagnosticate. Al 13,13% dei partecipanti è stata diagnosticata la sclerosi laterale amiotrofica, al 13,3% il Parkinson, al 2,2% la Corea di Huntington, al 15,6% un ictus, al 4,4% la sclerosi multipla, al 2,2% l'atassia e la distrofia miotonica, al 26,7% un trauma cranico, mentre al 20% altre patologie non presenti nell'elenco. In totale sono state registrate 13,72 ore di parlato. Sono già state effettuati degli esperimenti su questo database usando il toolkit Kaldi per la realizzazione di sistemi RAP. Tali risultati verranno presentati in fase di presentazione del progetto.

La nostra università ha finanziato un nuovo progetto di ricerca denominato DESIRE, composto da ingegneri, informatici, medici e umanisti con l'obiettivo di estendere e migliorare il lavoro già fatto, registrando non solo singole parole ma anche intere frasi formate da poche parole.

## Bibliografia

- [1] F. a. C. F. a. D. R. L. Ballati, «Assessing virtual assistant capabilities with italian dysarthric speech,» in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 2018, pp. 93--101.
- [2] D. a. M. G. a. M. M. a. F. L. Mulfari, «Towards a Deep Learning Based ASR System for Users with Dysarthria,» in *International Conference on Computers Helping People with Special Needs*, Springer, 2018, pp. 554--557.
- [3] D. a. M. G. a. F. L. Mulfari, «Machine Learning in Assistive Technology: a Solution for People with Dysarthria,» in *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, 2018, pp. 308--309.
- [4] J. a. L. M. a. F. L. a. R. P. Deller Jr, «The Whitaker database of dysarthric (cerebral palsy) speech,» *The Journal of the Acoustical Society of America*, vol. 93, n. 6, pp. 3516--3518, 1993.
- [5] X. a. P. J. B. a. P. S. M. a. L. J. E. a. B. H. T. Menendez-Pidal, «The Nemours database of dysarthric speech,» in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, 1996.
- [6] H. a. H.-J. M. a. P. A. a. G. J. a. H. T. S. a. W. K. a. F. S. Kim, «Dysarthric speech database for universal access research,» in *Ninth Annual Conference of the International Speech Communication Association*, 2008.
- [7] Samson, «GoMic portable usb condenser microphone,» [Online]. Available: <http://www.samsontech.com/samson/products/microphones/usb-microphones/gomic>.