



HAL
open science

IDEA: An Italian Dysarthric Speech Database

Marco Marini, Mauro Viganò, Massimo Corbo, Marina Zettin, Gloria Simoncini,
Bruno Fattori, Clelia d'Anna, Massimiliano Donati, Luca Fanucci

► **To cite this version:**

Marco Marini, Mauro Viganò, Massimo Corbo, Marina Zettin, Gloria Simoncini, et al.. IDEA: An Italian Dysarthric Speech Database. 2021 IEEE Spoken Language Technology Workshop (SLT), Jan 2021, Shenzhen, China. pp.1086-1093, <10.1109/SLT48900.2021.9383467>. <hal-03961538>

HAL Id: hal-03961538

<https://hal.science/hal-03961538v1>

Submitted on 29 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Preprint version of the conference paper:

Marini, M., Viganò, M., Corbo, M., Zettin, M., Simoncini, G., Fattori, B., D'Anna, C., Donati, M., & Fanucci, L. (2021). IDEA: an Italian dysarthric speech database. In 2021 IEEE Spoken Language Technology Workshop (SLT) (pp. 1086-1093). IEEE.

Published paper available at: <https://ieeexplore.ieee.org/document/9383467>

DOI: 10.1109/SLT48900.2021.9383467

The work is subject to the laws of copyright and intellectual property.

DRAFT

IDEA: AN ITALIAN DYSARTHIC SPEECH DATABASE

Marco Marini¹, Mauro Viganò², Massimo Corbo², Marina Zettin³, Gloria Simoncini³,
Bruno Fattori⁴, Clelia D'Anna⁵, Massimiliano Donati¹, Luca Fanucci¹

¹Dept. of Information Engineering, University of Pisa, Pisa, IT

²Dept. of Neurorehabilitation Sciences, Casa Cura Policlinico (CCP), Milan, IT

³Dept. of Psychology, University of Turin, Centro Puzzle, Turin, IT

⁴Dept. of Internal and Experimental Medicine, University of Pisa, Pisa, IT

⁵Audiology and Phoniatrics Unit, AUOP, Pisa, IT

ABSTRACT

This paper describes IDEA a database of Italian dysarthric speech produced by 45 speakers affected by 8 different pathologies. Neurologic diagnoses were collected from the subjects' medical records, while dysarthria assessment was conducted by a speech language pathologist and neurologist. The total number of records is 16794. The speech material consists of 211 isolated common words recorded by a single condenser microphone. The words that refer to an ambient assisted living scenario, have been selected to cover as widely as possible all Italian phonemes.

The recordings, supervised by a speech pathologist, were recorded through the RECORDIA software that was developed specifically for this task. It allows multiple recording procedures depending on the patient severity and it includes an electronic record for storing patients' clinical data. All the recordings in IDEA are annotated with a TextGrid file which defines the boundaries of the speech within the wave file and other types of notes about the record.

This paper also includes preliminary experiments on the recorded data to train an automatic speech recognition system from a baseline Kaldi recipe. We trained HMM and DNN models and the results shows 11.75% and 14.99% of WER respectively.

Index Terms: speech recognition, dysarthria, database, kaldi

1. INTRODUCTION

The evolution of automatic speech recognition (ASR) systems allows us to use smart devices just through our voice. It could be very useful for people with motor impairment or severe problem to handle smart devices. But what happens when people have impaired speech? People affected by motor speech disorders have difficulties using ASR systems according to the level of disorder [1, 2, 3, 4]. One of these types of disorder is the *dysarthria*.

Dysarthria is a neurological acquired motor speech disorder [5] that can impair all processes involved in speech pro-

duction such as respiration, phonation, articulation, resonance and prosody. These manifestations, and imprecise articulation in particular, often result in reduced speech intelligibility, which could negatively impact on the quality of life and the social participation of the speakers [6].

To improve the performance of ASR systems for dysarthric speakers, is fundamental to have enough data to analyse and use in training process. For this reason, during the past years several scientists have developed dysarthric speech databases.

The Whitaker database [7] is a collection of 19275 isolated-word utterances spoken by 6 persons whose speech spans a broad spectrum of dysarthria due to cerebral palsy. It was implemented in 1993.

Nemours database [8] was developed later and contains which contains 814 short nonsense sentences and 74 sentences spoken by each of 11 male speakers with varying degrees of dysarthria.

Universal Access speech database (UASpeech) [9] was delivered online in 2008 and consist of dysarthric speech recordings produced by 19 speakers with cerebral palsy. Speech materials consist of 765 isolated words per speaker.

All these databases contain single word or short sentence recordings of dysarthric speakers. Some of those also have video information that enriches the database and helps the researchers. TORGO database [10] store also articulatory data in addition to recordings. Articulatory data is obtained by electromagnetic articulography, which allows the measurement of the tongue and other articulators during speech, and by 3D reconstruction from binocular video sequences.

It is quite clear that several efforts have been made to collect English dysarthric voices for scientific research purposes. In the best of our knowledge, an Italian dysarthric speech database does not exist until today. It is very important to have a database of a specific language because the techniques discovered to improve speech recognition for English speakers may not be good for Italian ones. So, an Italian dysarthric speech database is useful to test the goodness of ASR system for English dysarthric speakers and to make progress in this

topic but in multi-language context.

With the purpose to develop IDEA: the first Italian dysarthric database, we start a collaboration with 3 health facilities: *Azienda Ospedaliero-Universitaria Pisana (AOUP)* [11], *Centro Puzzle di Torino* [12] and *Casa di Cura Privata del Policlinico di Milano* [13]. All these centres are specialized in diagnosis and rehabilitation of speech and language pathologies and host several dysarthric speakers for daily therapy and tests, or for long time hospitalization.

Specially, we have to distinguish our partners in two categories: Long Term Hospitalization (LTH) facilities and Daily Medical Examination (DME). The former provides for treatment of patients for several weeks or also months, so in this kind of facilities there is the possibility to record the same patient's voice for long time and in different days.

In daily medical examination facilities, the specialists have just few hours to evaluate the patient health. In this context there is not the possibility to record a big amount of data from each patient as in the previous case.

Among our partners, Centro Puzzle and Casa di Cura Privata del Policlinico are LTH facilities, while AOUP is DME facility.

Unlike the other databases, IDEA also has a lot of medical information about speakers. This could be very crucial to investigate medical aspects in speech analysis (e.g. dysarthric assessment).

IDEA database is available to anyone willing to use the database for scientific purposes; to access IDEA database is just required to send an email to the authors.

2. METHOD

2.1. Equipment

All the recording procedures have been done in a quiet and free of distractors room inside the facility where the patient has been hosted and respecting all the directives imposed by an ethics commission. The equipment required to start a recording procedure is composed by a computer with screen or a laptop and a condenser microphone [14]. The University of Pisa provided the same microphones to each partner, and a specific software for recording named RECORDIA.

2.2. RECORDIA

RECORDIA is a software based on Java and has two main functionalities: patient characterization and recording procedure. We will describe in details these functionalities in the following sections.

2.2.1. Patient characterization

The user can add a new patient through an Italian Graphical User Interface (GUI) and in order to do that, it is required to insert personal information (name, surname, gender and

age) and medical details. In clinic details the user can choose among 9 pathologies. In case the patient does not have any of those pathologies, the user can select "Other" (OTH) as option and insert some information in a "Note" field.

For each pathology, the user can insert some additional information like date of diagnosis or date of first symptoms, if available. Furthermore, there is the possibility to compile a specific pathology assessment scale according to the pathology. The pairs pathology-scale are showed below:

- Amyotrophic Sclerosis Lateral (ASL) - ASL functional rating scale [15]
- Ataxia (ATX) - Brief Ataxia rating scale [16]
- Huntington's Chorea (HC) - Unified Huntington's disease rating scale [17]
- Multiple Sclerosis (MS) - Kurtzke Expanded Disability Status Scale [18]
- Myotonic Dystrophy (MD) - Muscular Impairment Rating Scale [19]
- Neuropathy (NEU) - Guillain-Barré Syndrome Disability Scale [20]
- Parkinson Disease (PD) - Unified Parkinson's Disease Rating Scale [21]
- Stroke (STR) - National Institutes of Health Stroke Scale [22]
- Traumatic Brain Injury (TBI) - Levels of Cognitive Functioning [23]

In addition to clinical information, the user can classify the type of dysarthria with a specific taxonomy proposed by Duffy [24], evaluate the level of impairment with the Robert-Duffy Dysarthria Profile [25]. The user also can fill out a therapy outcome measure scale [26] and measure the quality of life in the dysarthric speaker [27].

All this information about the patient are editable over time, so it is possible to update or insert missing details when available.

2.2.2. Registration procedure

As explained in the previous chapters, the registration procedure is supervised by a user, typically a speech-language pathologist or a specialist, who interacts with the GUI and checks that the whole process goes properly.

In the registration procedure, RECORDIA software shows at screen a sequence of words that speaker has to repeat aloud. When a new word appears on the screen, the speaker has a fixed amount of time to pronounce it. This amount of time (2, 3 or 5 seconds) is chosen by the user at the beginning according to the ability of the speaker. If the

speaker gets trouble to do it, the user can try again. If the recording was successful, the user can move to the next one automatically proposed by RECORDIA. The next word is chosen so that the same word does not appear very close in time and as much random as possible. Each word recorded is stored in a single wave file. More repetitions of the same word have different wave files. The total amount of words is 211 chosen by doctors to cover the pronunciation of all Italian phonemes.

As explained in the Introduction, our partners have two different manners to deal with patients. Some have the chance to have the patients at their disposal for long period of time (LTH facilities). The others have just a bunch of minutes to record the patients' voices (DME clinics). For this reason, RECORDIA offers two levels of recording procedures.

The *Level 1* is suited for DME clinics because it requires about 15 minutes to complete the whole procedure. The aim of this level is to acquire a good number of recordings from a speaker despite a limited time at disposal. With good number of recordings, we mean 3 repetitions for 45 words. So, as soon as the speaker has recorded all the 135 wave files, the recording session ends. It is possible to stop the session early for any reason and restart it later from the same point. The Figure 1 shows the list of 45 words that are composed by a prefixed subgroup of 44 words of 211 and the word "occhi". They are chosen in order to have at least one pronunciation of each phoneme. In Figure 2 is shown a screenshot of RECORDIA GUI during Level 1 recording procedure of word "troppo".

The *Level 2* uses all 211 words (listed in Figure 3) and does not have any stop criteria as Level 1. So, the user can record as much as words he wants and stop the session at any time. This is used by LTH facilities because they have a lot of time at disposal to record dysarthric voices as much as possible.

In the first level the amount of data recorded is fixed while in the second level is unlimited. For both the Levels, the recordings are stored in the local PC and then sent to the University of Pisa by the user. Furthermore, they allow the user to collect a comprehensive Italian phoneme inventory, with at least a sample of all the 30 Italian phonemes [28]. The Table 1 shows all the phonemes involved in IDEA database divided by classes. In this table it is also possible see the distribution of occurrence of different phonemes over all 211 words. The *geminate* phonemes are consonants that have different sound duration. Furthermore, both levels include several bi-consonant and tri-consonant clusters but a full examination of clusters was not included in order to limit the number of words and to keep the procedure feasible with impaired speakers.

RECORDIA also gives the possibility to change the contrast of the word showed on screen. This feature is useful for people with sight issues.

Table 1. Phonemes involved in IDEA database divided by classes and with: relative symbol, IPA symbol and distribution of occurrences over 211 words.

Class	Symbol	IPA	Occurrences
plosives (6)	p	p	34
	b	b	17
	t	t	52
	d	d	35
	k	k	39
	g	g	16
fricative (5)	f	f	24
	v	v	21
	s	s	49
	z	z	13
	ʃ	ʃ	2
affricates (4)	ts	ts	5
	dz	dz	2
	tʃ	tʃ	15
	dʒ	dʒ	11
nasals (3)	m	m	38
	n	n	77
	ɲ	ɲ	1
liquids (3)	l	l	53
	r	r	95
	ʎ	ʎ	4
vowels (7)	a	a	151
	e	e	96
	ɛ	ɛ	38
	i	i	91
	o	o	132
	ɔ	ɔ	26
	u	u	27
geminate (19)	ʃʃ	ʃʃ:	3
	ʃʃ	ʃʃ:	1
	bb	b:	3
	dd	d:	1
	ddʒ	ddʒ:	2
	ddz	dz:	1
	ff	f:	4
	gg	g:	1
	kk	k:	6
	ll	l:	3
	mm	m:	1
	nn	n:	5
	pp	p:	1
	rr	r:	4
	ss	s:	9
	tt	t:	10
	ttʃ	tʃ:	2
	tts	ts:	4
	vv	v:	1
approximants (2)	j	j	18
	w	w	10

- | | |
|----------------|---------------|
| 1) abbassa | 24) maggiore |
| 2) accendi | 25) mezzo |
| 3) accordo | 26) occhi |
| 4) alza | 27) paesi |
| 5) avrebbe | 28) piccolo |
| 6) bagno | 29) pubblico |
| 7) buona | 30) quattro |
| 8) cancello | 31) saranno |
| 9) chiesa | 32) scelta |
| 10) corridoio | 33) scenario |
| 11) difficile | 34) servizio |
| 12) disimpegno | 35) soffitta |
| 13) donna | 36) soggiorno |
| 14) donne | 37) spegni |
| 15) famiglia | 38) sveglia |
| 16) fanno | 39) terrazzo |
| 17) figli | 40) troppo |
| 18) figlio | 41) ufficio |
| 19) forza | 42) uscita |
| 20) giallo | 43) vecchio |
| 21) guerra | 44) veneziane |
| 22) hanno | 45) zero |
| 23) luogo | |

Fig. 1. Level 1 list of words.

3. DESCRIPTION OF DATABASE

IDEA has 16794 total recordings distributed among 45 speakers, 25 males and 20 females. The total amount of time recorded is about 13.72 hours of which about 35% is speech (about 4.1 hours) and 65% is non speech (9.62 hours). For DME speakers, we have recorded 45 words for 3 times in the same day for a total of 135 wave files for each dysarthric speaker. For LTH speakers, our aim was to record at least 3 times all 211 words in different days for a total of 633 wave files for a single speaker. Since a dysarthric speaker typically is affected by other pathologies, a recording process could be very stressful and tiring. Hence, some speakers interrupted the recording procedure, so for them some recordings are missing. Each wave file is encoded in the linear PCM format at 16 bit and 16kHz sample rate.

The Table 2 shows some statistics for each speaker, especially the type of facility, gender, amount of recordings, percentage of recordings done respect to the final aim and pathology. Thus, 21 speakers (46% of total) have complete the entire recording procedure, 8 speakers (17% of total) have recorded less than 50% of the pre-established recordings and 16 speakers have completed the procedure between 50% and 100%.

The Figure 4 shows the distribution of pathologies in the IDEA database. This information can be inferred by the last column of Table 2.



Fig. 2. Screenshot of RECORDIA GUI during Level 1 recording procedure.

3. Annotation

We have analysed all the data in order to detect and annotate any errors or corruptions. This annotation could be useful to organise the data efficiently during ASR training. With this purpose, we decided to classify wave files in 4 categories:

- *Empty*: the file is empty or not playable (due to some RECORDIA failure);
- *Not usable*: the wave file does not contain speech but just noise, or the noise is too loud to understand speech;
- *Without notes*: the wave file is correctly recorded and the speech is pronounced in absence of noise;
- *With notes*: the wave file has some issues that should be tacked into account when we deal with this file.

Each wave file has a *TextGrid* file with the same name. A *TextGrid* file is a *Praat* file [29], which contains some information about wave file like duration and tiers. For instance, for a wave file without notes the associated *TextGrid* file will describe what are the start and stop points of speech inside the recordings and what exactly is pronounced by the speaker within these boundaries.

If the wave file is classified as with notes, in the *TextGrid* file will appear one more tier which describe it. The possible issues are:

- *Truncated*: the speaker speech is truncated because is too close to the beginning or end of record;
- *Substitution*: the speaker pronounced another word rather than the one proposed on screen;
- *Repetition*: the speaker pronounced more times the same word within wave file;
- *Corrupted*: wave file present some microphone saturation or glitch;
- *Background noise*: the wave file presents some background noise, like coughing, but the speech is still understandable;

1) Aosta	54) destra	107) inglese	160) ruspa
2) Europa	55) devono	108) ingresso	161) sapere
3) abbassa	56) dieci	109) inverno	162) saranno
4) accendi	57) difficile	110) italiano	163) sbarra
5) accordo	58) discreto	111) lacrime	164) sbrana
6) aeroplano	59) disgrazia	112) leggono	165) scale
7) afoso	60) disimpegno	113) letto	166) scelta
8) aiuto	61) diverse	114) libro	167) scenario
9) albero	62) diversi	115) lungo	168) scrive
10) alghe	63) domanda	116) luogo	169) seconda
11) altri	64) donna	117) maggiore	170) sei
12) alza	65) donne	118) mamma	171) sembra
13) androne	66) drago	119) mansarda	172) servizio
14) angelo	67) due	120) marciare	173) sette
15) aprì	68) economico	121) massimo	174) sfrega
16) arancione	69) egregio	122) mattino	175) sgonfio
17) argenteo	70) entrata	123) mezzo	176) siano
18) aspro	71) erano	124) milioni	177) sinistra
19) auto	72) erba	125) molto	178) smalto
20) avrebbe	73) escluso	126) momento	179) soffitta
21) avvolgere	74) esplode	127) morte	180) soggiorno
22) bagno	75) estate	128) mosca	181) soldato
23) bambini	76) famiglia	129) necessario	182) sono
24) barca	77) fanno	130) nero	183) sorpresa
25) benzina	78) fantasmi	131) nessuno	184) spegni
26) bianco	79) fare	132) nostri	185) splende
27) bicicletta	80) farfalla	133) notte	186) spremute
28) blu	81) fatto	134) nuovi	187) stato
29) borotalco	82) figli	135) oasi	188) strada
30) braccio	83) figlio	136) offeso	189) su
31) buona	84) film	137) otto	190) sveglia
32) camera	85) finestra	138) padre	191) temperatura
33) cancello	86) flauto	139) paesi	192) tempi
34) cantina	87) forma	140) particolare	193) terrazzo
35) cento	88) forza	141) passato	194) tigre
36) chiesa	89) freddo	142) pergolato	195) troppo
37) chiudi	90) generale	143) perle	196) ufficio
38) cinque	91) giallo	144) persone	197) umidità
39) cioè	92) giardino	145) piccolo	198) unica
40) classe	93) giorno	146) pinza	199) uomini
41) completo	94) giovane	147) plagio	200) uscita
42) compra	95) giù	148) polvere	201) vecchio
43) comune	96) glicine	149) porta	202) vedere
44) concluso	97) gnocchi	150) prima	203) veneziane
45) confetti	98) gradi	151) pubblico	204) viola
46) consumi	99) grande	152) pulcino	205) visto
47) corpo	100) guerra	153) qualsiasi	206) volpe
48) corridoio	101) hanno	154) quattro	207) volte
49) corte	102) idea	155) realtà	208) volume
50) cravatta	103) ieri	156) resto	209) vuole
51) cuore	104) importante	157) riflesso	210) zero
52) delfino	105) incrocio	158) rinfrescante	211) zuchero
53) dentifricio	106) influenza	159) rosso	

Fig. 3. Level 2 list of words.

- **Word split:** the speaker makes a pause in the middle of the word or tend to pronounce the word in syllabic way;
- **General notes:** is used to annotate some generic issue like if the speaker pronounces the word longer usual (e.g. "donnnna" instead of "donna").

The 70.8% of records are without notes, so it means that they are usable as they are. The 22.8% are with notes, so they need to be used with more attention and just the 6,4% is unusable (5.1% are empty and 1.3% are not usable).

4. BASELINE ASR EXPERIMENTS

4.1. Experimental setup

In the first version of this database (IDEA_1.0), we decided to insert all the speakers who have enough data usable as it is. For this reason, we considered only the recordings without notes for all speakers. Thus, we tried to split these recordings in train and test sets with the following approach: if a word

Table 2. Statistics about speakers: ID, gender, facility type (see *Introduction* for detail), number of records, percentage of completion of total recordings and pathology.

ID	Gend	Facility	# REC	Complete	Pathology
201	F	DME	135	100.00%	OTH
202	F	DME	39	28.89%	HC
203	F	DME	135	100.00%	STR
204	F	DME	41	30.37%	PD
205	M	DME	135	100.00%	ATX
206	M	DME	135	100.00%	ASL
207	M	DME	135	100.00%	STR
208	F	DME	135	100.00%	PD
209	F	DME	135	100.00%	PD
210	M	DME	135	100.00%	MD
211	F	DME	135	100.00%	ASL
212	F	DME	135	100.00%	ASL
213	F	DME	135	100.00%	ASL
214	M	DME	135	100.00%	ASL
215	M	DME	135	100.00%	PD
216	F	DME	135	100.00%	STR
201	M	LTH	630	99.53%	PD
202	M	LTH	630	99.53%	STR
203	M	LTH	630	99.53%	STR
204	F	LTH	14	2.21%	OTH
205	F	LTH	633	100.00%	TBI
206	M	LTH	632	99.84%	OTH
207	F	LTH	630	99.53%	TBI
208	M	LTH	634	100.16%	TBI
209	M	LTH	178	28.12%	TBI
210	F	LTH	630	99.53%	TBI
211	M	LTH	634	100.16%	TBI
212	F	LTH	631	99.68%	TBI
213	M	LTH	576	91.00%	STR
214	M	LTH	630	99.53%	OTH
215	M	LTH	631	99.68%	TBI
216	F	LTH	630	99.53%	TBI
217	F	LTH	445	70.30%	MS
218	M	LTH	217	34.28%	OTH
219	M	LTH	420	66.35%	TBI
220	F	LTH	634	100.16%	TBI
221	M	LTH	633	100.00%	OTH
222	M	LTH	637	100.63%	OTH
223	M	LTH	635	100.32%	TBI
224	M	LTH	254	40.13%	MS
225	M	LTH	589	93.05%	ASL
226	F	LTH	151	23.85%	PD
227	F	LTH	552	87.20%	OTH
228	M	LTH	534	84.36%	STR
229	M	LTH	150	23.70%	ASL

has 3 or more records, 1/3 goes to test and 2/3 go to train. For those words that have less than 3 records, all are used for train. Obviously, for those speakers who have not enough

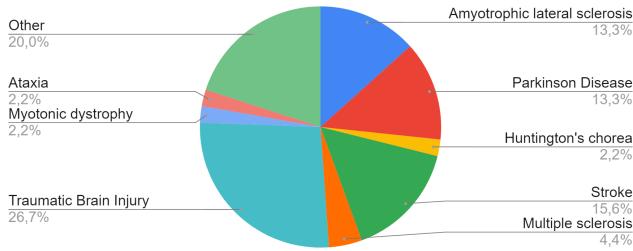


Fig. 4. Percentage of pathologies present in IDEA database.

recordings the test set is empty. The speakers who have at least one records in the test set are 37, so they are included in IDEA_1.0. Finally, IDEA_1.0 includes 11272 recordings without notes (about 67% of whole database). The train and test sets are composed by 9111 and 2161 recordings respectively.

4.2. Results

One of the purposes of developing IDEA database is to create an ASR system for Italian dysarthric speakers. So we tried to run a Kaldi [30] recipe baseline for UA-Speech corpus used in [31], upon IDEA_1.0 database. We decided to use this recipe because UA-Speech database contains single word recordings as IDEA database. Of course, the data preparation of the recipe it is adjusted for IDEA_1.0 database, but the feature extraction and acoustic modelling parts are the same. We evaluated the Word Error Rate (WER) for two models: Gaussian Mixture Model (GMM) based and Deep Neural Network (DNN) based, both combined with Hidden Markov Model (HMM). Table 3 shows that the GMM-HMM ASR system got wrong 254 recordings out of 2161, which means 11.75% of WER. On the other hand, the DNN-HMM ASR system got wrong 324 recordings which means 14.99% of WER. The results for each speaker are shown in Figure 5.

Table 3. Baseline results with numbers of recordings in train and test sets, number of errors and WER for each ASR system.

	Train	Test	Errors	WER
GMM	9111 (7.84h)	2161 (1.89h)	254	11.75%
DNN	9111 (7.84h)	2161 (1.89h)	324	14.99%

The results show that GMM-HMM perform better than DNN-HMM. This differs from results obtained in [31] but it could be explained by the limited amount of data for training. Indeed, IDEA_1.0 has 7.84h for training and 1.89h for testing from 45 speakers, while [31] uses 23.1h for training and 21.7h for test from 15 speakers. So, probably the data is not enough to train all the neural network weights properly.

Despite the WER is very high for some speaker, as shown in Figure 5, the overall WER is quite low for both types of

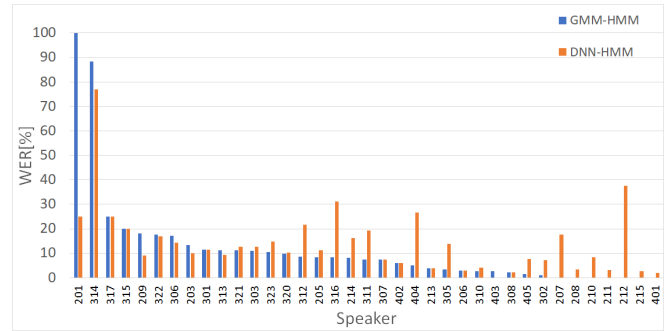


Fig. 5. Baseline results: WER of ASR systems for each speaker.

models. This is because the distribution of recordings in each speaker's test sets is not balanced. For instance, from GMM-HMM model point of view the speaker 201 has 100% of WER but she also has just 4 recordings in test set, so it does not have a big impact on final WER.

5. CONCLUSION

In this paper we have describe IDEA: the first Italian dysarthric speech database. It contains the voices of 45 speakers for a total of 16794 recordings for about 13 hours and 43 minutes. Single word recording was performed for 211 words. Each record has a TextGrid file associated that defines the boundaries of speech in the wave file and adds other important information. Moreover, IDEA store specific medical information about speaker like type of pathologies (there are 9 different options), a specific assessment scale for each type of pathology and other clinical test. We believe that this corpus provides a significant resource for develop assistive technologies since it contains medical and technical information for a good number of users. Some preliminary test has been done to develop an ASR system using Kaldi recipe. This first experiment has been done using about 67% of all IDEA material and shows that ASR systems based on GMM model perform better than DNN model for this amount of data. In the future work we want to use the whole database for training ASR systems and to investigate how to involve clinical information in training process. We are continuing to record other volunteers to further expand our database. IDEA is available free, people interested on IDEA have to contact the author of this paper, who will give the possibility to download it.

6. ACKNOWLEDGMENTS

The authors thank the volunteers for their participation. Much data processing was done with the help of Heidi Christensen, Phil Green and Mauro Nicolao of University of Sheffield (UK).

7. REFERENCES

- [1] Fabio Ballati, Fulvio Corno, and Luigi De Russis, ““hey siri, do you understand me?”: Virtual assistants and dysarthria,” in *Intelligent Environments (Workshops)*, 2018, pp. 557–566.
- [2] Fabio Ballati, Fulvio Corno, and Luigi De Russis, “Assessing virtual assistant capabilities with italian dysarthric speech,” in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, New York, NY, USA, 2018, ASSETS ’18, p. 9, Association for Computing Machinery.
- [3] Davide Mulfari, Gabriele Meoni, Marco Marini, and Luca Fanucci, “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.
- [4] Davide Mulfari, Gabriele Meoni, and Luca Fanucci, “Machine learning in assistive technology: a solution for people with dysarthria,” in *Proceedings of the 11th EAI International Conference on Smart Objects and Technologies for Social Good*, 2018, pp. 308–309.
- [5] Malcolm R McNeil, DA Robin, and RA Schmidt, *Clinical management of sensorimotor speech disorders*, Thieme New York, 1997.
- [6] Pamela Mary Enderby and Joyce Emerson, *Does speech and language therapy work?: A review of the literature*, Whurr London, 1995.
- [7] JR Deller Jr, MS Liu, LJ Ferrier, and P Robichaud, “The whitaker database of dysarthric (cerebral palsy) speech,” *The Journal of the Acoustical Society of America*, vol. 93, no. 6, pp. 3516–3518, 1993.
- [8] Xavier Menendez-Pidal, James B Polikoff, Shirle M Peters, Jennie E Leonzio, and H Timothy Bunnell, “The nemours database of dysarthric speech,” in *Proceedings of Fourth International Conference on Spoken Language Processing. ICSLP’96*. IEEE, 1996, vol. 3, pp. 1962–1965.
- [9] Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas S Huang, Kenneth Watkin, and Simone Frame, “Dysarthric speech database for universal access research,” in *Ninth Annual Conference of the International Speech Communication Association*, 2008.
- [10] Frank Rudzicz, Aravind Kumar Namasivayam, and Talya Wolff, “The torgo database of acoustic and articulatory speech from speakers with dysarthria,” *Language Resources and Evaluation*, vol. 46, no. 4, pp. 523–541, 2012.
- [11] “Azienda ospedaliera universitaria pisana,” <https://www.ao-pisa.toscana.it/>.
- [12] “Centro puzzle di torino,” <https://www.centropuzzle.it/>.
- [13] “Casa di cura privata del policlinico di milano,” <https://www.ccppdezza.it/>.
- [14] “GoMic portable usb condenser microphone,” <http://www.samsontech.com/samson/products/microphones/usb-microphones/gomic/>.
- [15] Hyun Young Kim and Seung Hyun Kim, “Korean version of amyotrophic lateral sclerosis functional rating scale-revised: a pilot study on the reliability and validity,” 2006.
- [16] Sarah Camargos, Francisco Cardoso, Ricardo Maciel, Lucio Huebra, Thiago Roberto Silva, Vilson Geraldo Campos, and Rodrigo Alencar, “Brief ataxia rating scale: a reliable tool to rate ataxia in a short timeframe,” *Movement disorders clinical practice*, vol. 3, no. 6, pp. 621, 2016.
- [17] Sabine Siesling, Jeroen PP Van Vugt, Koos AH Zwinderman, Karl Kiebertz, and Raymund AC Roos, “Unified huntington’s disease rating scale: a follow up,” *Movement disorders: official journal of the Movement Disorder Society*, vol. 13, no. 6, pp. 915–919, 1998.
- [18] Donald E Goodkin, D Cookfair, K Wende, Dennis Bourdette, P Pullicino, B Scherokman, and Ruth Whitham, “Inter-and intrarater scoring agreement using grades 1.0 to 3.5 of the kurtzke expanded disability status scale (edss),” *Neurology*, vol. 42, no. 4, pp. 859–859, 1992.
- [19] J Mathieu, H Boivin, D Meunier, M Gaudreault, and P Begin, “Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy,” *Neurology*, vol. 56, no. 3, pp. 336–340, 2001.
- [20] Christiaan Fokke, Bianca van den Berg, Judith Drenthen, Christa Walgaard, Pieter Antoon van Doorn, and Bart Casper Jacobs, “Diagnosis of guillain-barré syndrome and validation of brighton criteria,” *Brain*, vol. 137, no. 1, pp. 33–43, 2014.
- [21] Movement Disorder Society Task Force on Rating Scales for Parkinson’s Disease, “The unified parkinson’s disease rating scale (updrs): status and recommendations,” *Movement Disorders*, vol. 18, no. 7, pp. 738–750, 2003.
- [22] Patrick Lyden, Mei Lu, Christy Jackson, John Marler, Rashmi Kothari, Thomas Brott, and Justin Zivin, “Underlying structure of the national institutes of health stroke scale: results of a factor analysis,” *Stroke*, vol. 30, no. 11, pp. 2347–2354, 1999.

- [23] WD Gouvier, PD Blanton, KK LaPorte, and C Nepomuceno, "Reliability and validity of the disability rating scale and the levels of cognitive functioning scale in monitoring recovery from severe head injury.," *Archives of physical medicine and rehabilitation*, vol. 68, no. 2, pp. 94–97, 1987.
- [24] Joseph R Duffy, *Motor speech disorders-e-book: Substrates, differential diagnosis, and management*, Elsevier Health Sciences, 2013.
- [25] Giovanni Defazio, Marta Guerrieri, Daniele Liuzzi, Angelo Fabio Gigante, and Vincenzo 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, 2016.
- [26] Pamela Enderby, Alexandra John, and Brian Petheram, *Therapy outcome measures for rehabilitation professionals: speech and language therapy, physiotherapy, occupational therapy*, John Wiley & Sons, 2013.
- [27] Valentina Piacentini, Annalisa Zuin, Davide Cattaneo, and Antonio Schindler, "Reliability and validity of an instrument to measure quality of life in the dysarthric speaker," *Folia Phoniatrica et Logopaedica*, vol. 63, no. 6, pp. 289–295, 2011.
- [28] Martina Tresoldi, Maria Rosaria Barillari, Federico Ambrogi, Elena Sai, Umberto Barillari, Elvira Tozzi, Letizia Scarponi, and Antonio Schindler, "Normative and validation data of an articulation test for italian-speaking children," *International Journal of Pediatric Otorhinolaryngology*, vol. 110, pp. 81–86, 2018.
- [29] Paul Boersma, "Praat: doing phonetics by computer," <http://www.praat.org/>, 2006.
- [30] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, L. Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Roman Silovsky, Georg Stemmer, and Karel Vesely, "The kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. Dec. 2011, IEEE Signal Processing Society, IEEE Catalog No.: CFP11SRW-USB.
- [31] Feifei Xiong, Jon Barker, and Heidi Christensen, "Deep learning of articulatory-based representations and applications for improving dysarthric speech recognition," in *Speech Communication; 13th ITG-Symposium*. VDE 2018, pp. 1–5.