Bridging ASR Gaps for 250,000 People: The DYSTINCT Dataset for German Dysarthric Speech Technology

Anina Klaus[1], Simon Werner[2], Andrew J. Beaton[3], Xenia Klinge[4], Annette Sterr[5], Christian Dohle[5], Dietrich Klakow[1],

[1]Universität des Saarlandes, [2]Technische Universität Berlin, [3]Universität Potsdam, [4]DFKI, [5]P.A.N. Zentrum für Post-Akute Neurorehabilitation

Correspondence: ankl06@dfki.de

Abstract

Dysarthria is a motor speech disorder estimated to affect over 250,000 people in Germany. Automatic Speech Recognition (ASR) could greatly benefit affected populations, but standard ASR models struggle with impaired speech and data resources for specialized training are scarce. To address this limitation, we present a new German dysarthric speech corpus geared towards speech technology development. We collected 8 hours and 16 minutes of both read and spontaneous speech from nine speakers with diverse symptoms of dysarthria and apraxia of speech (AOS). Preliminary ASR evaluation comparing a default and fine-tuned Whisper-medium model in a speaker-dependent and speaker-independent setting shows a diverse range of outcomes. In the speaker-dependent setup, strong word error rate (WER) improvements up to 0.82 points (from 1.01 to 0.19 for a speaker with severe dysarthria) are observed. The speaker-independent version reaches up to a 0.49 point improvement from a WER of 1.20 to 0.71. The results demonstrate the usefulness of read speech in the facilitation of ASR for impaired speech and the benefits of speaker-specific data, but also individual differences in learnability. This highlights the need for more fine-grained analyses in the research on ASR approaches for dysarthric speech.

1 Introduction

Dysarthria is a motor speech disorder caused by neurological injuries and characterized by abnormal neuromuscular execution that affects the speed, strength, range, timing or accuracy of speech movements (Enderby, 2013). The result is highly individual impaired speech. Resulting challenges in communication can negatively impact self-identity and personal relationships (Dickson et al., 2008). It is estimated that this affects over 250,000 people in Germany (Ziegler and Vogel, 2010). Other disorders of speech and language such as aphasia (Trapl-Grundschober et al., 2005) and apraxia of speech (AOS; Takahashi et al., 2025) often co-occur.

This makes dysarthric speech a very challenging case for speech technologies such as automatic speech recognition (ASR; see e.g. Takahashi et al., 2025), excluding a demographic which could greatly benefit from such technology from using it: Besides helping with intelligibility in conversations, ASR could enable voice-controlled applications of home automation or help with written communication (Jaddoh et al., 2023). Many speakers with dysarthria suffer from co-occurring motor impairments which limit movement.

Progress in this field is hindered by the lack of openly available and suitable data, especially in languages other than American English (Bhat and Strik, 2025). To address this gap, we collaborate with the P.A.N. Center for Post-Acute Neuro-Rehabilitation in Berlin to build a dataset of German dysarthric speech: the DYSTINCT database (DYsarthric Speech Technology for INtelligibility and Conversational Translation).

Our contributions are the following: 1) full recordings of nine German speakers suffering from dysarthria or AOS[1], 2) a detailed symptom profile for each speaker and 3) baseline speaker-dependent and speaker-independent ASR results on the collected data.

2 Related Work

Dysarthria affects speech in very diverse ways. A widely-used taxonomy defines six main types of dysarthria (Darley et al., 1969): flaccid, spastic, ataxic, hypokinetic, hyperkinetic, and mixed.

As a first larger collection, the American English Universal Access Speech dataset (UASpeech; Kim et al., 2008) has been widely used in ASR research.

[1]For information regarding access to the dataset, visit github.com/dystinct-project/dystinct-data

The dysarthric section was contributed by 14 speakers with cerebral palsy (CP), each recording 765 isolated words.

A second widely-used American English dataset is the TORGO database (Rudzicz et al., 2012). Dysarthric speech from eight patients with CP or amyotrophic lateral sclerosis (ALS) and healthy speech from seven age- and gender-matched control speakers is included, resulting in 5.5 and 8 hours of single-word and sentence recordings, respectively. Spontaneous speech is included in the form of image descriptions.

The Speech Accessibility Project (SAP) dataset is by far the most extensive collection available (Hasegawa-Johnson et al., 2024). In the SAP-240430 release (Zheng et al., 2025), 524 speakers suffering from Parkinson's disease (PD), Down syndrome, ALS, CP or stroke contributed 415 hours of speech in American, Canadian, and Puerto Rican English. Each speaker was asked to read or repeat 350 to 400 phrases. Additionally, free speech was recorded using short open-ended prompts. The database is intended for research on speaker-independent ASR approaches.

Some data collections exist for other languages, such as French, Korean, Cantonese, Dutch, Tamil, Spanish, Czech, Italian and German (see Bhat and Strik, 2025). Many of these, such as the German dataset by Skodda et al. (2011), are not geared towards NLP and are thus very limited in terms of data variety and representativeness. A German dysarthric dataset of 1,272 sentences read by a single speaker is accessible on huggingface[2], but there does not seem to be any publication connected to it or any further information available.

A central goal of research on technology for dysarthric speech is to enable high-quality ASR (e.g. Takahashi et al., 2025). Dysarthria classification serves diagnostic purposes (e.g. Javanmardi et al., 2024), while speech conversion seeks to improve intelligibility (Wang et al., 2020). Parallel to the development of applications for non-impaired speech, the focus in recent years has been on methods of deep learning (Bhat and Strik, 2025). Regarding ASR, the most successful approaches in the recent Interspeech 2025 Shared Task on speaker-independent dysarthric ASR using the SAP dataset worked with the fine-tuning of speech foundation models (Zheng et al., 2025).

[2] https://huggingface.co/datasets/B-Czarnetzki/dysarthric_german

The training data requirements vary by application. While dysarthria classification requires appropriate speech samples annotated with information about the speaker status, ASR training traditionally requires transcribed speech recordings. Usually, dysarthric speech datasets feature this type of data in the form of words or sentences read aloud by the participants. Free, spontaneous speech transcribed afterwards can also occasionally be found, as in the image descriptions of TORGO or the answers to open-ended prompts in the SAP dataset. This type of data is in some regards more representative of conversational speech as a primary use case for dysarthric ASR systems, as it shares the aspect of spontaneity and its effects. Conversational ASR still proves to be a challenging task for modern ASR systems even on healthy speech, due to disfluencies and structural differences from read speech (Maheshwari et al., 2025). The same is true for impaired speech (Tobin et al., 2024), rendering spontaneous speech sections an important component of dysarthric speech datasets aimed at the development and evaluation of ASR approaches.

The impact of accompanying speech impairments such as AOS has not received special attention in previous literature. AOS is characterized by an inability to correctly program sensorimotor commands for the production of speech and is characterized by more unstable error patterns (Cherney and Small, 2009).

3 Corpus Creation

Subjects

We cooperate with the P.A.N. Center for Post-Acute Neuro-Rehabilitation in Berlin for the recruitment and pathological assessment of speakers. The current dataset contains recordings of nine speakers with speech impairments. All speakers are diagnosed with dysarthria or AOS and provided their informed consent to the data collection and use of their pseudonymized speech data for this work along with relevant pathological information.

Methodology

We collect two types of speech data. Firstly, we curated a list of 400 sentences from the Common-Voice dataset (Ardila et al., 2020) to be read aloud by the participants:

100 phrases related to care and daily life, as well as 125 colloquial phrases, aim to provide data which resemble speech that a dysarthric person

could realistically use, and which an ASR system created to assist them in communication would need to be able to handle. 50 automation-related phrases target voice control, a different ASR application which would be very useful to many patients suffering from accompanying motor impairments. 100 additional sentences include minimal pairs that are potentially difficult to distinguish in pronunciation for dysarthric speakers. A final 25 sentences are selected on the basis of biphone diversity to ensure a phonetically diverse collection and highlight potentially difficult sounds missed in the previous groups. Examples are listed in Appendix A.2.

The read sentences provide a starting point for ASR training especially for severely dysarthric speech, which cannot be transcribed with high confidence. Additionally, we collect spontaneous speech. In the pilot recording sessions, image descriptions have proven to be a method that provides intuitive cues for speakers to produce speech. We created two detailed images for this purpose, which depict settings commonly encountered in daily life (see Appendix A.1 for reference). Speakers are asked to describe what they observe in the image. The researcher conducting the recording session can ask follow-up questions on the image if needed.

Recording setup

A streamlit[3]-based recording interface is used for speech collection. This interface can record separate items for different cues such as sentences and images. Typically, a researcher is in the room with the speaker during recording sessions to instruct the speaker and operate the recording tool.

The sessions are limited to an hour to avoid placing too much strain on the speakers, some of whom are prone to fatigue. Sessions typically begin with a spontaneous speech section with a description of one of the two images. The remaining time is then used for read sentence recordings.

Transcription

The spontaneous speech recordings need to be transcribed. This is done by the researcher who conducted the recordings. In some cases, the speech is very hard to understand, even for familiar listeners. This means that the transcriptions we provide must be understood as noisy and full correctness cannot be guaranteed. Some material had to be discarded because the content could not be transcribed.

[3] https://streamlit.io

Annotation

General speaker information including age, gender, impairment etiology and detailed information on the symptoms of the speech impairment is provided. This information was provided by professional speech-language pathologists who had assessed the speakers individually.

4 Collected Data

ID	Sex	Age	Etiology		Dysarthria		AOS
			Name	ICD-10	Type	Severity	
S1	m	25	TBI	S06.9	flaccid	severe	no
S2	f	36	stroke	I63.1	mixed	severe	no
S3	f	57	stroke	I63.3	-	-	yes
S4	f	36	stroke	I63.3	*INP*	mild	yes
S5	f	50	intracranial hemorrhage	I61.4	ataxic	mild	no
S6	f	46	TBI	T90.5	flaccid	severe	no
S7	f	54	stroke	I63.5	*INP*	medium	yes
S8	m	8	congenital malformation of brain	*INP*	mixed	*INP*	no
S9	m	40	TBI	S06.6, S06.9, S06.5	*INP*	medium	no

Table 1: Overview of the impaired speakers. Missing information is marked as *INP*.

At the moment, the dataset comprises the data of nine speakers with speech impairments. In total, 7 hours and 38 minutes of read and 38 minutes of spontaneous impaired speech were recorded. For two speaker, S7 and S8, no spontaneous speech set is available. Eight of the nine impaired speakers suffer from dysarthria, three from AOS. A table detailing speaker characteristics (anonymized speaker ID, sex, age, etiology name, ICD-10 code, dysarthria type, severity, and AOS diagnosis) is included in Table 1. The first recordings with speakers S1 and S2 served as pilot sessions and therefore partially have lower audio quality and spontaneous speech from image descriptions and open-ended questions, as we tested both types of elicitation methods. Additionally, the full set of items was recorded for one healthy female control speaker.

5 Baseline ASR Experiments

Setup

Whisper is a family of pre-trained models for ASR and speech translation (Radford et al., 2023). In line with the findings of Wang et al. (2025), we leave the decoder frozen and fine-tune using a 70-

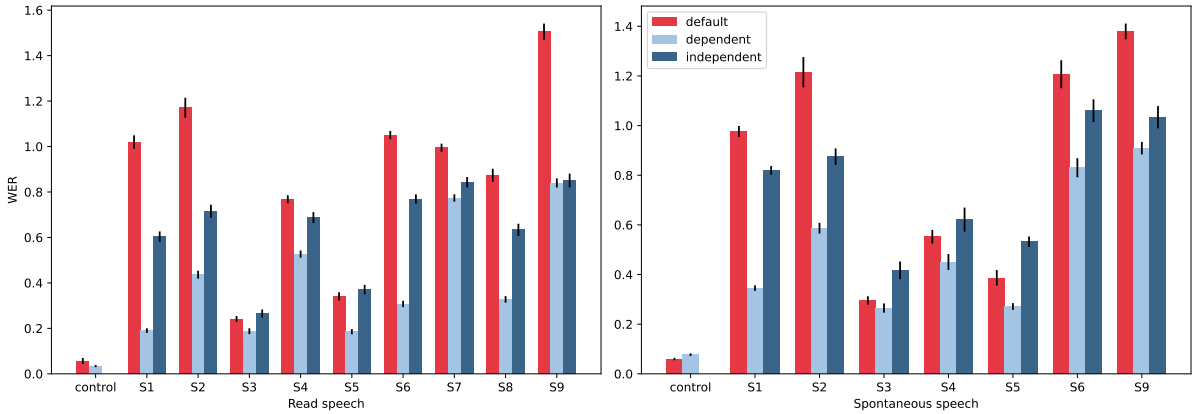

Figure 1: WER outcomes of speaker-dependent and speaker-independent fine-tuning for all speakers

10-20 split for the read data of each speaker, additionally testing on the transcribed spontaneous speech. For the speaker-dependent baseline, training, validation and test data is taken from the same speaker. For the speaker-independent baseline, training and validation data is formed by creating a pool of data from the respective splits of all speakers except the target speaker tested on. The hyperparameter choices are listed in Appendix C; a schematic representation of the data splits can be observed in Appendix B. Results are evaluated using the word error rate (WER), a widely-used metric in ASR research. BERTScore (Zhang et al., 2020) outcomes were also calculated; the results are listed in Appendix D. We tested five randomly sampled splits to account for variance induced by different compositions of the training, validation and test sets. The average and standard error of the WER is reported.

Results

The results are displayed in Figure 1. The speaker-dependent setting shows improvements of the fine-tuned models for all speakers. The biggest gains are observed for the read speech test set, but performance on spontaneous speech is also notably enhanced. Additionally, individual differences between the speakers can be observed: not only default model performance, but also fine-tuned improvements vary greatly. In the speaker-independent setup, only speakers with a higher WER in the default setup see performance gains. For the control speaker, the default model already performs very strongly – fine-tuning slightly improves the results on the read speech test set, but hurts performance on spontaneous speech.

6 Discussion

Overall, the collected data are very diverse with regard to symptoms and impairment severity. This diversity is also highlighted in the baseline experiments. Generally, it can be observed that the default model performance roughly matches impairment severity. The results of the speaker-dependent fine-tuning show that training on read dysarthric speech only can bring about strong performance gains on both read and spontaneous speech. Notably, the WER outcomes on spontaneous speech are slightly worse, indicating some degree of structural difference of this type of speech. Still, the collection of read speech emerges as a good strategy for the creation of assistive technology, as it is feasible in practice and enables promising results in different types of speech. Considerable differences across speakers exist in learnability, suggesting that dysarthric ASR may not have a one-size-fits-all solution. The strongest improvements can be observed for the speakers S1 and S6, who both suffer from flaccid dysarthria. In contrast, the speakers with the smallest relative improvements are S3, S4 and S7. These are the speakers diagnosed with AOS, whose instability in error patterns may explain the diminished adaptation success. This result points to AOS as a relevant factor in ASR for dysarthric speech, as the two conditions often co-occur.

The speaker-independent results, on the other hand, highlight the difficulty of cross-speaker transfer, which can only be observed for speakers with higher initial WER values and accordingly with more severe impairments. This pattern may be caused by specifics of the dataset composition or characteristics of severely impaired speech.

7 Conclusion

We collected 8 hours and 16 minutes of speech from nine speakers with speech impairments, demonstrating the feasibility of the presented protocol. With this work, we provide the first well-documented German dysarthric speech dataset geared towards the development of speech technology. Baseline experiments highlight the benefit of collecting speaker-specific data and read impaired speech, but also demonstrate notable variability in success rates between speakers. This points to new research directions taking into account individual symptoms and their impact on dysarthric speech technology development. The DYSTINCT dataset provides the basis for the investigation of these questions in the context of German dysarthric speech technology. Future expansions of the data set are planned with new speakers, which substantiate the basis for the analysis of specific factors.

8 Limitations

An important limitation is the size of the dataset. This factor is explained by the costly nature of impaired speech collection. At its current size, it does not allow us to fully disentangle different sources of variance in the performance, such as dysarthria symptoms and gender of the speaker. We do, however, present a dataset which represents diverse dysarthrias. Furthermore, even our current dataset shows promising results in the baseline experiments.

We use image descriptions to collect spontaneous speech. A limitation of this elicitation method is that the type of speech produced is not necessarily representative of conversational speech, as the task is fundamentally different (apart from sharing the aspect of spontaneity). In the pilot recording sessions, we additionally tested open-ended questions similar to the prompts of the SAP dataset. However, the data yield was insufficient and the speakers experienced difficulty answering the questions due to cognitive impairments. The costs and challenges of transcribing severely impaired speech, especially in larger quantities, highlight the need to collect read speech as well.

Additionally, the imperfect nature of transcriptions of severely dysarthric speech has to be kept in mind, especially when interpreting results computed on these items. It is naturally difficult to transcribe spontaneous speech, and even more so for impaired speech. Given our available resources, we provided the best possible transcriptions.

Lastly, the presented experiments are baseline experiments conducted to confirm the usefulness of the collected data. In-depth explorations of architecture choices or speaker-specific factors are out of the scope of this dataset presentation.

9 Ethical Considerations

Speakers are provided with extensive information on how their data will be used and their rights according to the GDPR before they provide consent to participating. We ensure the participant's well-being during the recording sessions by frequent check-ins and a limit on the session length of one hour. We are aware that speech in combination with medical information constitutes a dataset with very sensitive data. The speakers' identities are hidden behind pseudonyms and access to the dataset is restricted to research purposes. This publication complies with the DFKI's ethical assessment.

10 Acknowledgements

We would like to thank Lara Pirchner for her help in organizing the recording sessions, as well as Lilly Zühl, Dr. Maja Wiest and Mareike Schrader, who helped with the curation of relevant speaker information. Special thanks go to all the speakers who dedicated their time to improve future ASR technology for patients with speech impairments.

References

Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber. 2020. Common voice: A massively-multilingual speech corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4218–4222, Marseille, France. European Language Resources Association.

Chitralekha Bhat and Helmer Strik. 2025. Speech technology for automatic recognition and assessment of dysarthric speech: An overview. *Journal of Speech, Language, and Hearing Research*, 68(2):547–577.

Leora R. Cherney and Steven L. Small. 2009. Aphasia, apraxia of speech, and dysarthria. *Stroke recovery and rehabilitation*, pages 155–182.

Frederic L. Darley, Arnold E. Aronson, and Joe R. Brown. 1969. Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12(2):246–269.

Sylvia Dickson, Rosaline S Barbour, Marian Brady, Alexander M Clark, and Gillian Paton. 2008. Patients' experiences of disruptions associated with post-stroke dysarthria. *International journal of language & communication disorders*, 43(2):135–153.

Pam Enderby. 2013. Chapter 22 - disorders of communication: dysarthria. In Michael P. Barnes and David C. Good, editors, *Neurological Rehabilitation*, volume 110 of *Handbook of Clinical Neurology*, pages 273–281. Elsevier.

Mark Hasegawa-Johnson, Xiuwen Zheng, Heejin Kim, Clarion Mendes, Meg Dickinson, Erik Hege, Chris Zwilling, Marie Moore Channell, Laura Mattie, Heather Hodges, Lorraine Ramig, Mary Bellard, Mike Shebanek, Leda Sarı, Kaustubh Kalgaonkar, David Frerichs, Jeffrey P. Bigham, Leah Findlater, Colin Lea, and 6 others. 2024. Community-supported shared infrastructure in support of speech accessibility. *Journal of Speech, Language, and Hearing Research*, 67(11):4162–4175.

Aisha Jaddoh, Fernando Loizides, and Omer Rana. 2023. Interaction between people with dysarthria and speech recognition systems: A review. *Assistive Technology*, 35(4):330–338.

Farhad Javanmardi, Sudarsana Reddy Kadiri, and Paavo Alku. 2024. Pre-trained models for detection and severity level classification of dysarthria from speech. *Speech Communication*, 158:article 103047.

Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas S. Huang, Kenneth Watkin, and Simone Frame. 2008. Dysarthric speech database for universal access research. In *Interspeech 2008*, pages 1741–1744.

Gaurav Maheshwari, Dmitry Ivanov, Théo Johannet, and Kevin El Haddad. 2025. ASR benchmarking: Need for a more representative conversational dataset. In *2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine Mcleavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 28492–28518. PMLR.

Frank Rudzicz, Aravind Kumar Namasivayam, and Talya Wolff. 2012. The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation*, 46(4):523–541.

Sabine Skodda, Wenke Grönheit, and Uwe Schlegel. 2011. Intonation and speech rate in parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission. *Journal of Voice*, 25(4):199–205.

Kaito Takahashi, Keigo Hojo, Toshimitsu Sakai, Yukoh Wakabayashi, and Norihide Kitaoka. 2025. Fine-tuning Parakeet-TDT for Dysarthric Speech Recognition in the Speech Accessibility Project Challenge. In *Interspeech 2025*, pages 3304–3308.

Jimmy Tobin, Phillip Nelson, Bob MacDonald, Rus Heywood, Richard Cave, Katie Seaver, Antoine Desjardins, Pan-Pan Jiang, and Jordan R. Green. 2024. Automatic speech recognition of conversational speech in individuals with disordered speech. *Journal of Speech, Language, and Hearing Research*, 67(11):4176–4185.

Michaela Trapl-Grundschober, Raoul Eckhardt, Peter Bosak, and Michael Brainin. 2005. Early recognition of speech and speech-associated disorders after acute stroke. *Wiener medizinische Wochenschrift (1946)*, 154:571–6.

Disong Wang, Jianwei Yu, Xixin Wu, Songxiang Liu, Lifa Sun, Xunying Liu, and Helen Meng. 2020. End-to-end voice conversion via cross-modal knowledge distillation for dysarthric speech reconstruction. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7744–7748.

Qianli Wang, Zihan Zhong, Satwinder Singh, Clarion Mendes, Mark Hasegawa-Johnson, Waleed Abdulla, and Seyed Reza Shahamiri. 2025. Dysarthric speech conformer: Adaptation for sequence-to-sequence dysarthric speech recognition. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. *Preprint*, arXiv:1904.09675.

Xiuwen Zheng, Bornali Phukon, Jonghwan Na, Ed Cutrell, Kyu J. Han, Mark Hasegawa-Johnson, Pan-Pan Jiang, Aadhrik Kuila, Colin Lea, Bob MacDonald, Gautam Mantena, Venkatesh Ravichandran, Leda Sari, Katrin Tomanek, Chang D. Yoo, and Chris Zwilling. 2025. The Interspeech 2025 Speech Accessibility Project Challenge. In *Interspeech 2025*, pages 3269–3273.

W. Ziegler and M. Vogel. 2010. *Dysarthrie verstehen - untersuchen - behandeln*. Forum Logopädie. Thieme.

A Example Items

The following items are examples from the described interface to elicit speech, first by describing images, second by reading sentences.

A.1 Images

A.2 Sentences

Guten Tag!
Kannst du mal nach den Kindern sehen?
Hilfe!
Und ab geht die Post.
Oh ja!
Wie lautet die Diagnose?

B Data splits

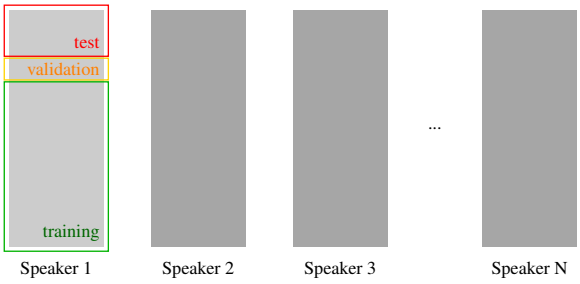

Figure 2: Speaker-dependent fine-tuning for speaker 1

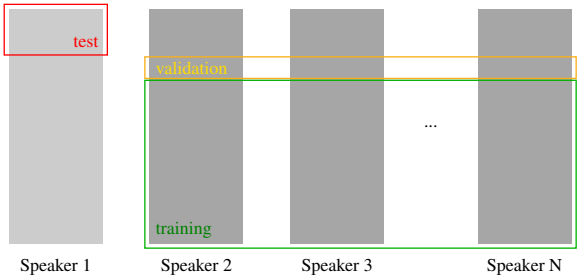

Figure 3: Speaker-independent fine-tuning for testing on target speaker 1

C Hyperparameters

Parameter	Value
Optimizer	AdamW
Batchsize	8
Learning rate	1×10^{-5}
Warmup steps	10
Epochs	10

D BERTScore results

Besides WER, BERTScore results of the ASR experiments were calculated. This metric captures semantic closeness of the prediction and the ground truth transcription. Results are shown in Figure 2 for the read speech test set and 3 for the spontaneous speech test set.

Speaker	default	dependent	independent
control	0.99	0.99	-
S1	0.74	0.94	0.83
S2	0.75	0.88	0.8
S4	0.78	0.85	0.81
S3	0.93	0.95	0.92
S7	0.72	0.78	0.75
S6	0.71	0.91	0.77
S5	0.91	0.95	0.89
S8	0.78	0.9	0.81
S9	0.71	0.78	0.78

Table 2: BERTScore results on read speech

Speaker	default	dependent	independent
control	0.98	0.98	-
S1	0.73	0.9	0.75
S2	0.76	0.84	0.76
S4	0.83	0.88	0.84
S3	0.9	0.92	0.89
S6	0.71	0.81	0.73
S5	0.9	0.91	0.83
S9	0.7	0.77	0.75

Table 3: BERTScore results on spontaneous speech