ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICASSP55912.2026.11462593

VARIATIONAL LOW-RANK ADAPTATION FOR PERSONALIZED IMPAIRED SPEECH RECOGNITION

Niclas Pokel[1,2,*], *Pehuén Moure*[1,*], *Roman Boehringer*[1,†], *Shih-Chii Liu*[1,†], *Yingqiang Gao*[3,†]

[1]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland
[2]School of Computation, Information and Technology, Technical University of Munich, Germany
[3]Department of Computational Linguistics, University of Zurich, Switzerland
{npokel,pehuen,roman,shih}@ini.ethz.ch, yingqiang.gao@cl.uzh.ch

ABSTRACT

Speech impairments resulting from congenital disorders, such as cerebral palsy, Down syndrome, or Apert syndrome, as well as acquired brain injuries due to stroke, traumatic accidents, or tumors, present major challenges to automatic speech recognition (ASR) systems. Despite recent advancements, state-of-the-art ASR models like Whisper still struggle with non-normative speech due to limited training data availability and high acoustic variability. Moreover, collecting and annotating non-normative speech is burdensome: speaking is effortful for many affected individuals, while laborious annotation often requires caregivers familiar with the speaker. This work introduces a novel ASR personalization method based on Bayesian Low-rank Adaptation for data-efficient fine-tuning. We validate our method on the English *UA-Speech* dataset and a newly collected German speech dataset, *BF-Sprache*, from an individual with structural speech impairment. Both the dataset and the approach are designed to reflect the challenges of low-resource settings that include individuals with speech impairments. Our method significantly improves ASR accuracy for impaired speech while maintaining data and annotation efficiency, offering a practical path toward inclusive ASR.

Index Terms— Automatic speech recognition, personalization, non-normative speech, variational inference, data-efficient learning.

1. INTRODUCTION

Despite having intact cognitive and language abilities, many individuals with speech impairments remain effectively voiceless in a world built around spoken communication. Daily interactions, whether answering a question in class, telling a story, or participating in a group play, can become exhausting, frustrating, or simply impossible. For children, this communication barrier not only isolates them socially but also disrupts their emotional and educational development. The burden extends to families, educators and care providers, who must interpret, advocate and adapt, often without the support of reliable assistive tools [1, 2].

Automatic speech recognition (ASR) systems hold the potential to bridge this gap, but current models are not designed for non-normative speech. Even state-of-the-art models such as Whisper [3] and wav2vec [4, 5] degrade significantly in performance when confronted with atypical articulation, prosodic variation, or inconsistent phoneme production [6].

These challenges are further aggravated for non-English languages due to the lack of representative data, limited linguistic tools, and the need for language-specific adaptation strategies [5, 7]. German remains under-resourced for non-normative speech, particularly for children, with limited publicly available datasets [8, 9, 10].

To address these limitations, researchers have employed various strategies. Fine-tuning large pre-trained models has proven effective for disordered speech, with approaches like those of Shor et al. [11] achieving 60% reductions in word error rate for patients with amyotrophic lateral sclerosis. However, such methods are often prone to overfitting and inefficient parameter usage [12].

Parameter-efficient adaptation methods, including lightweight adapters [12], low-rank adaptation (LoRA) modules [13], and hypernetwork-based speaker tuning [14], improve speech foundation model fine-tuning. However, these methods often rely on English language backbones and assume explicit knowledge of impairment types or sufficient in-domain data [15, 7].

Bayesian Neural Network (BNN)-based approaches offer promising alternatives for improving ASR robustness in low-data, high-variability settings [16], and have been used to maintain robustness across continuous adaptation [17]. BNN approaches may provide key benefits in bridging parameter-efficient fine-tuning methods to the high variability data of impaired speech. Prior works primarily leverage Bayesian low-rank adaptation (LoRA) for efficiency, either via capacity-reducing pruning and quantization [18, 19] or static post-hoc analysis [20, 21], which risks underfitting the complex acoustic features of disordered speech. Conversely, our method employs variational inference (VI) as a dynamic training regularizer that retains full capacity while elastically constraining adaptation to the pre-trained weight structure. To this end, our work proposes:

1. **Variational Low-rank Adaptation framework.** We introduce a Bayesian LoRA method called VI LoRA that captures uncertainty during fine-tuning. This enables robust personalization with significantly less data, while maintaining parameter efficiency, crucial for modeling speech with high acoustic variability.
2. **Data-driven prior estimation.** We develop a prior estimation approach that better captures the multi-modal distribution of layer-wise weight variations in state-of-the-art ASR architectures.
3. **Cross-lingual evaluation.** We validate our method on English and German datasets spanning a range of speech intelligibility. Our results show substantial improvements, especially for speakers with very low intelligibility, demonstrating the framework's

*These authors contributed equally. †These authors also contributed equally. This work was partially funded by the Swiss National Science Foundation project CA-DNNEdge (208227).

Authorized licensed use limited to: ETH BIBLIOTHEK ZURICH. Downloaded on August 06,2026 at 15:20:28 UTC from IEEE Xplore. Restrictions apply.

effectiveness in low-resource, cross-lingual settings.

Together, these contributions bridge existing research gaps by enabling a personalized, interpretable, and scalable ASR solution tailored for users with atypical speech in multiple languages.

2. METHODOLOGY

2.1. ASR Model and Evaluation Metrics

For all experiments, we used Whisper-Large V3 [3] as the backbone of the impaired speech model, evaluating its performance under both zero-shot and supervised fine-tuning settings, with augmentation via semantic re-chaining (i.e., assembling semantically coherent sentence-level utterances from word-level counterparts [22]). We used word error rate (WER) and character error rate (CER) as metrics to evaluate our experiment outcomes.

2.2. Bayesian Low-rank Adaptation Framework

Our approach addresses the challenges of low data availability, overfitting, and over-parameterization when fine-tuning large models [12] and enhances the standard low-rank adaptation (LoRA) [13], a prominent parameter-efficient fine-tuning (PEFT) technique [23]. LoRA adapts a pre-trained weight matrix $W_0 \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$ by freezing W_0 and introducing a trainable low-rank update ΔW_0. This update is parameterized as the product of two smaller matrices, $B \in \mathbb{R}^{d_{\text{out}} \times r}$ and $A \in \mathbb{R}^{r \times d_{\text{in}}}$, where the rank $r \ll \min(d_{\text{in}}, d_{\text{out}})$. The adapted weight matrix then becomes $W_0 + \frac{\alpha}{r} BA$, with α being a scaling factor.

While LoRA significantly reduces training efforts, in data-sparse scenarios, the matrices A and B can still overfit the limited training data and cause generalization degradation [24]. Drawing inspiration from *Bayes by Backprop* [25], which regularizes neural networks by learning distributions over their weights, we extend LoRA to a Bayesian setting. This allows us to capture uncertainty about the LoRA parameters, which is particularly beneficial for regularization and improving robustness when training data is scarce. Specifically, we perform variational inference (VI) to estimate the posterior distributions of A and B given the training data $\mathcal{D}$. We approximate the true, often intractable, posterior $p(A, B|\mathcal{D})$ with a tractable variational distribution $q_\phi(A, B)$, parameterized by ϕ. Adopting the mean-field approximation, we assume independence between A and B, and further, between individual elements: $q_\phi(A, B) = q_{\phi_A}(A) q_{\phi_B}(B)$, where each factor is a fully diagonal Gaussian:

$$q_{\phi_A}(A) = \prod_{i=1}^{r} \prod_{j=1}^{d_{\text{in}}} \mathcal{N}(A_{ij} | \mu_{A_{ij}}, \sigma^2_{A_{ij}}),$$

$$q_{\phi_B}(B) = \prod_{k=1}^{d_{\text{out}}} \prod_{l=1}^{r} \mathcal{N}(B_{kl} | \mu_{B_{kl}}, \sigma^2_{B_{kl}}).$$

The variational parameters $\phi = \{(\mu_A, \sigma_A), (\mu_B, \sigma_B)\}$ are associated with the LoRA adapter layers (e.g., in query, key, and value projection matrices of multi-head attention) and are learned by minimizing the negative evidence lower bound (ELBO):

$$\begin{aligned}
\phi^* &= \arg\min_\phi \mathrm{KL}[q_\phi(A, B) \,||\, p(A, B|\mathcal{D})] \\
&= \arg\min_\phi \int q_\phi(A, B) \log \frac{q_\phi(A, B)}{p(A, B) p(\mathcal{D}|A, B)} \, d(A, B) \\
&= \arg\min_\phi \underbrace{\mathrm{KL}[q_\phi(A, B) \,||\, p(A, B)] - \mathbb{E}_{q_\phi(A,B)}[\log p(\mathcal{D}|A, B)]}_{-\mathcal{L}_{\text{ELBO}}(\phi)},
\end{aligned}$$

here, $p(A, B)$ is the prior distribution over the LoRA matrices. The term $\mathbb{E}_{q_\phi(A,B)}[\log p(\mathcal{D}|A, B)]$ is the expected log-likelihood of the training data, corresponding to the task-specific loss (e.g., cross-entropy for ASR) computed with samples from $q_\phi(A, B)$, typically estimated using Monte Carlo sampling. Let N be the total number of LoRA layers where VI is applied. The KL divergence term in the ELBO, $\mathrm{KL}[q_\phi(A, B)||p(A, B)]$, is theoretically a sum over all N layers. However, numerical instabilities, especially in early training stages before posteriors stabilize, can lead to non-finite KL values for some layers. To ensure robust optimization, our final loss function $\mathcal{L}_{\text{VI}}$ employs an average over layers yielding a finite (non-NaN, non-Inf) KL divergence terms in the current optimization step:

$$\begin{aligned}
\mathcal{L}_{\text{VI}} = &- \mathbb{E}_{q_\phi(A,B)}[\log p(\mathcal{D}|A, B)] + \\
&\beta \underbrace{\Big(\sum \mathrm{KL}[q_\phi(A, B)||p(A, B)] \neq \text{NaN or Inf} \Big)}_{\overline{\mathrm{KL}}[q_\phi||p]},
\end{aligned}$$

where β is a scaling factor. To construct a more informed prior $p(A, B)$, we assume factorization across layers l and between $A^{(l)}$ and $B^{(l)}$, i.e., $p(A, B) = \prod_l p(A^{(l)}) p(B^{(l)})$. For individual elements, we use Gaussian priors: $p(a_{ij}^{(l)}) = \mathcal{N}(a_{ij}^{(l)} | \mu_p, (\sigma_p^{(l)})^2)$ and similarly for $b_{ij}^{(l)}$. A common choice sets $\mu_p = 0$ and employs a single, global prior variance, e.g., $\sigma_p^2 = 1$. However, this assumes that the LoRA updates should operate at a scale comparable to a standard normal distribution, which can be problematic if the pre-trained weights $W_0^{(l)}$ themselves exhibit significantly distinct variances across layers in the pre-trained network parameters, potentially leading to an overly restrictive prior for some layers or an overly loose one for others. For a more informed prior, we first analyze the empirical standard deviations of pre-trained weights within each target layer $W_0^{(l)}$. The final loss is a weighted sum of the standard Whisper loss (90%) and a KL divergence term ($10\% \cdot \mathrm{KL}(q||p)$), which acts as a regularizer. This relative weighting prevents the KL term from dominating when task loss becomes small. For each layer l designated for LoRA updates, we compute its empirical weight standard deviation $\hat{\sigma}_p^{(l)}$ as

$$\hat{\sigma}_p^{(l)} = \sqrt{\frac{1}{|\mathcal{W}^{(l)}| - 1} \sum_{w \in W_0^{(l)}} (w - \overline{w}^{(l)})^2},$$

where $\mathcal{W}^{(l)}$ represents the set of all weights in the original matrix $W_0^{(l)}$, and $\overline{w}^{(l)}$ is their mean. Our empirical analysis of these layer-wise standard deviations, $\{\hat{\sigma}_p^{(l)}\}_{l=1}^N$ (across $N = 288$ target layers in our experiments), reveals a distinct bimodal distribution, as illustrated in Figure 1. A simple Gaussian Mixture Model was used to find the optimal μ for the layer-specific σ prior.

3. EXPERIMENTAL SETUP

We conduct a comprehensive evaluation of our approach across multiple dimensions to assess both personalization effectiveness and generalizability. First, we compare personalization performance against several baselines, including full-parameter fine-tuning, standard low-rank adaptation (LoRA), the high-rank updating technique MoRA [26], and variational inference LoRA (VI LoRA) with single and bimodal priors. Second, we analyze the impact of architectural design choices on recognition accuracy to understand how specific configurations influence performance. Selecting a comparable rank for MoRA

Authorized licensed use limited to: ETH BIBLIOTHEK ZURICH. Downloaded on August 06,2026 at 15:20:28 UTC from IEEE Xplore. Restrictions apply.

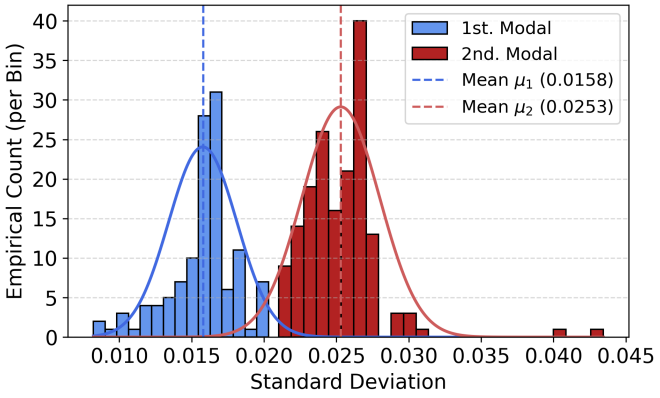

Fig. 1: Histogram of the empirically estimated standard deviations, $\hat{\sigma}_p^{(l)}$, computed individually for each of the $N = 288$ target LoRA layers based on their pre-trained weights $W_0^{(l)}$. Dashed lines indicate the means of the two distinct modes identified by k-means, justifying a layer-type specific prior variance.

is non-trivial, since the rank must divide the Whisper embedding dimension and cannot be directly matched parameter wise. Moreover, Bayesian methods and deterministic methods differ fundamentally in parameterization. To align with the parameter count of VI LoRA, including variance terms, we set $r = 320$ for MoRA. We also evaluated several (VI-)LoRA ranks (r) to optimize the model's configuration. We selected $r = 32$ for our experiments, as it provided a strong balance between performance gains on non-normative speech and a minimal increase in catastrophic forgetting. We observed that higher ranks (e.g., $r = 64$) failed to further improve target domain performance while exacerbating forgetting, a behavior consistent with the "intrinsic rank" hypothesis of LoRA [13].

3.1. Datasets

Developing effective ASR systems for individuals with speech impairments requires datasets that capture the acoustic variability, articulatory challenges, and language-specific characteristics of non-normative speech patterns [27]. We evaluated our method using the following datasets: UA-Speech [28] and BF-Sprache [22], as well as Mozilla Common Voice Dataset [29] for non-impaired speech:

UA-Speech. The English UA-Speech dataset [28] is widely used in dysarthric speech recognition research and has laid the foundation for numerous deep learning-based works [30, 31]. It consists of recordings from 19 speakers with varying degrees of dysarthria, along with recordings from 13 control speakers. We exclusively used the $\approx 66h$ of dysarthic speech. The dataset emphasizes isolated word-level utterances, comprising 455 unique words including digits, letters, and phonetically rich uncommon words.

BF-Sprache. To evaluate our personalization framework across languages, we use the German BF-Sprache [22] dataset. For consistency, the dataset also consists of isolated word-level utterances in the training set, but was tested on spontaneous speech for the most realistic assessment.

The structured design of the two datasets enables controlled experimentation and provides insights into the effectiveness and generalizability of ASR personalization across different intelligibility levels of speech impairment and languages.

Normative evaluation set. To measure the forgetting of the already learned distribution of normal speech we used a split in the respective language (German or English) of the widely known Mozilla Common Voice Dataset [29] for validation and testing.

3.2. Model Evaluation

To ensure robust evaluation, we stratify experiments along two primary dimensions:

1. **Intelligibility levels.** We partition the UA-Speech dataset into subsets based on reported intelligibility levels (i.e., very low, low, medium), enabling a detailed analysis of model performance across speech impairment levels and the calibration of uncertainty estimates throughout the spectrum.
2. **Cross-lingual generalization.** We test our methods on both UA-Speech and BF-Sprache datasets to evaluate how well the personalization framework generalizes across languages.

Table 1: Results of different models on non-normative (BF-Sprache) and normative (CommonVoice) speech. WD refers to Weight Decay, DP to Dual Prior, SP to Single Prior, KL to 10% KL$[q||p]$.

Setup	Non-Normative		Normative	
	CER	WER	CER	WER
0-shot Inference	40.38 ± 0.00	82.11 ± 0.00	**2.01 ± 0.00**	6.18 ± 0.00
Full Fine-tuning	22.60 ± 1.85	46.43 ± 2.74	2.40 ± 0.34	7.83 ± 0.72
+ WD	22.53 ± 1.55	46.17 ± 2.66	2.38 ± 0.27	7.66 ± 0.49
Standard LoRA	23.85 ± 0.51	46.64 ± 1.47	2.42 ± 0.21	7.11 ± 0.40
+ WD	23.11 ± 0.44	46.18 ± 1.29	2.40 ± 0.19	6.98 ± 0.38
MoRA	25.87 ± 0.66	49.11 ± 1.44	2.54 ± 0.15	7.80 ± 0.23
+ WD	26.43 ± 0.57	48.53 ± 1.47	2.33 ± 0.14	6.97 ± 0.23
DP VI LoRA + KL	**20.09 ± 0.41**	**42.86 ± 1.48**	2.15 ± 0.13	**6.05 ± 0.23**
+ WD	31.42 ± 1.62	55.36 ± 3.51	8.21 ± 0.72	16.82 ± 1.17
SP VI LoRA + KL	21.33 ± 0.51	44.85 ± 1.87	2.02 ± 0.18	6.05 ± 0.27
+WD	26.02 ± 1.06	50.29 ± 2.09	2.33 ± 0.35	7.62 ± 0.65

4. RESULTS AND ANALYSIS

Table 2: Results on the UA-Speech (non-normative) and CommonVoice (normative) dataset for different adaptation methods, including a 0-shot baseline, relative to full fine-tuning (= 100%).

Setup	Speech Type	rel. CER	rel. WER
0-shot	Non-Normative	271.30% ± 0.00%	328.80% ± 0.00%
	Normative	**43.50% ± 0.00%**	**46.94% ± 0.00%**
LoRA	Non-Normative	105.32% ± 1.71%	106.81% ± 2.77%
	Normative	78.55% ± 4.11%	81.21% ± 4.35%
SP VI LoRA	Non-Normative	91.07% ± 2.11%	91.74% ± 2.02%
	Normative	44.17% ± 5.84%	47.29% ± 6.11%
DP VI LoRA	Non-Normative	**88.94% ± 2.36%**	**90.24% ± 1.78%**
	Normative	49.87% ± 6.21%	55.36% ± 5.78%

Table 1 presents a comparative analysis of different adaptation strategies, including standard LoRA, full parameter fine-tuning, and VI LoRA with and without a KL regularization term. All models were fine-tuned on BF-Sprache. The results indicate that VI LoRA, when regularized with a 10% KL$[q||p]$ term, demonstrates a compelling trade-off. This configuration achieves the lowest CER and WER

Authorized licensed use limited to: ETH BIBLIOTHEK ZURICH. Downloaded on August 06,2026 at 15:20:28 UTC from IEEE Xplore. Restrictions apply.

Table 3: CER and WER on BF-Sprache (100% $\approx$ 2h) for different sizes of the training set and adaptation methods.

Train Data	VI LoRA		Full Fine-tuning		LoRA	
	CER	WER	CER	WER	CER	WER
100%	**19.86**	**42.42**	22.28	48.02	23.66	47.55
75%	**22.32**	**44.75**	24.38	49.01	25.91	51.10
50%	**24.77**	**50.40**	28.95	66.04	28.02	58.43
25%	**28.08**	**56.35**	33.07	70.43	31.29	66.94
0%	40.38	82.11	40.38	82.11	40.38	82.11

on the target non-normative speech. Concurrently, it exhibits the least forgetting of normative speech, as evidenced by its leading CER and WER scores, surpassing both standard LoRA and full parameter fine-tuning, which show higher error rates on normative data. These findings generalize across English speakers with varying intelligibility levels in the UA-Speech dataset (Table 2). The table reports relative performance differences from full fine-tuning, aggregated over all dysarthric speakers.

We speculate that the performance of VI LoRA with KL divergence regularization stems from its ability to effectively adapt to non-normative speech characteristics while mitigating catastrophic forgetting. The KL divergence term acts as a regularizer, penalizing significant deviations of the adapted LoRA weights (q) from the original pre-trained weight distribution (p). This constraint likely encourages the model to learn the specific variations of non-normative speech more parsimoniously, preventing overly aggressive updates that could shift the weights too far from their initial state, which is beneficial for normative speech. This allows for controllable adaptation without sacrificing much of the model's generalization ability to the pre-trained normative speech patterns.

Table 3 reports CER/WER on non-normative speech in the BF-Sprache dataset across varying training set sizes. The test set is fixed, while training sets are constructed from randomly selected instances of the remaining data. VI LoRA consistently outperforms all baselines, with its advantage being most pronounced when less data is available. While full fine-tuning underperforms even standard LoRA in low-data settings, it surpasses LoRA once the full training set is used. This trend is particularly evident in WER.

As illustrated in Table 4, a qualitative analysis of the transcription outputs reveals a critical distinction between the error patterns of full fine-tuning and our proposed VI LoRA model on out-of-distribution (OOD) phrases, specifically targeting semantically rare or hyper-locally relevant terms (e.g., "Wiedikon") that lack strong language model priors. The fully fine-tuned model consistently exhibits a form of structured hallucination. For instance, it transcribes the Japanese place name "Higashirinkan" as the grammatically plausible but semantically unrelated German sentence "Ein Gassi rennt da." This suggests the model defaults to familiar linguistic patterns when faced with novel acoustic signals, effectively pattern-matching to the closest structure in its learned distribution.

In stark contrast, VI LoRA produces transcriptions that are phonetically much closer to the ground truth, such as "Higashirenpa." While imperfect, this output demonstrates a failure mode that is grounded in the acoustic evidence rather than learned linguistic priors. This highlights a significant limitation of standard metrics like WER and CER, although both models yield high error rates, the nature of VI LoRA's errors is far more interpretable and useful, as it preserves crucial phonetic information.

We speculate that this robustness stems from the stochastic nature of VI LoRA. The inherent variance across multiple forward passes may disrupt the model's tendency for rigid pattern matching. By marginalizing over these varied predictions, the model is forced to find the smallest common ground, which appears to be the underlying phonetics of the input rather than a pre-learned grammatical structure.

Table 4: Qualitative comparison of transcription examples for out-of-distribution phrases. Phonetic transcriptions (IPA) are provided in *italics* to aid interpretation for non-German speakers.

System	Transcription Output	PER/CER
Ground Truth:	"Wiedikon, Enge, Thalwil, Baar."	-
	[ˈviːdɪkɔn, ˈɛŋə, ˈtaːlvɪl, baːr]	-
Full Fine-tuning	"Wie die kann, eine, teilweise, war."	56.0 / 45.7
	[viː diː kan ˈaɪmə ˈtaɪlvaɪzə vaːr]	
VI LoRA (ours)	"Vidikon, Enne, Talwil, Borg."	**20.0 / 25.0**
	[ˈviːdɪkɔn, ˈɛnə, ˈtaːlvɪl, bɔrk]	
Ground Truth:	"Higashirinkan."	-
	[ɕi ga ʃi ɾiŋ kaŋ]	-
Full Fine-tuning	"Ein Gassi rennt da."	86.7 / 63.2
	[aɪn ˈgasi rɛnt da]	
VI LoRA (ours)	"Higashirenpa."	**26.7 / 25.0**
	[ɕi ga ʃi ɾempa]	

5. DISCUSSION AND CONCLUSIONS

In this work, we propose a novel personalization framework using Bayesian LoRA. Our experimental results demonstrate substantial improvements in recognition accuracy for speech-impaired individuals across different intelligibility levels and languages, while maintaining efficient parameter usage. Our experiments revealed several key insights. Our dual prior approach to VI LoRA (Table 1 for BF-Sprache and Table 2 for UA-Speech) showed particularly strong performance, reducing non-normative speech CER to 20.09% (compared to 21.33% for single prior) while maintaining reasonable normative speech performance. This suggests that modeling the bimodal distribution of pre-trained weights significantly improves adaptation capability. The cross-lingual results are especially encouraging, with our framework generalizing effectively from English to German despite the latter's phonological complexity. Furthermore, the observed trade-off between performance on non-normative and normative speech highlights a key design consideration: systems optimized solely for impaired speech may sacrifice generalization, suggesting that multi-objective training strategies could further improve generalizability. While promising, our work has some limitations. Our pipeline assumes independent factorization of $q_\phi(A, B)$ by disentangling it as $q_{\phi_A}(A)q_{\phi_B}(B)$. While computationally efficient, it may not best capture the interactions between LoRA adapter matrices [32, 33]. The main limitation remains the small speaker pool in the BF-Sprache dataset, due to previous resource and ethical approval constraints. With the ethical approval now secured, our immediate future work will focus on expanding this dataset by recruiting a larger, more diverse group of speakers across various conditions and intelligibility levels. Our further work will focus on expanding the speaker base in BF-Sprache over different conditions and intelligibility levels as well as incorporating VI LoRA in an active learning setting for continuous speaker-specific adaptation.

Authorized licensed use limited to: ETH BIBLIOTHEK ZURICH. Downloaded on August 06,2026 at 15:20:28 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Allyson D Page and Kathryn M Yorkston, "Communicative Participation in Dysarthria: Perspectives for Management," *Brain Sciences*, vol. 12, no. 4, pp. 420, 2022.

[2] Loes van Bemmel, Chiara Pesenti, Xue Wei, and Helmer Strik, "Automatic Assessments of Dysarthric Speech: the Usability of Acoustic-Phonetic Features," in *Proc. Interspeech 2023*, 2023, pp. 141–145.

[3] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever, "Robust Speech Recognition via Large-scale Weak Supervision," in *Proc. International Conference on Machine Learning*. PMLR, 2023, pp. 28492–28518.

[4] Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli, "Wav2vec: Unsupervised Pre-Training for Speech Recognition," in *Proc. Interspeech 2019*, 2019, pp. 3465–3469.

[5] Murali Karthick Baskar, Tim Herzig, Diana Nguyen, Mireia Diez, Tim Polzehl, Lukas Burget, and Jan Černockỳ, "Speaker Adaptation for Wav2vec2 Based Dysarthric ASR," in *Proc. Interspeech 2022*, 2022, pp. 3403–3407.

[6] Hannah P. Rowe, Sarah E. Gutz, Marc F. Maffei, Katrin Tomanek, and Jordan R. Green, "Characterizing Dysarthria Diversity for Automatic Speech Recognition: A Tutorial From the Clinical Perspective," *Frontiers in Computer Science*, vol. 4, pp. 770210, 2022.

[7] Jimmy Tobin, Phillip Nelson, Bob MacDonald, Rus Heywood, Richard Cave, Katie Seaver, Antoine Desjardins, Pan-Pan Jiang, and Jordan R Green, "Automatic Speech Recognition of Conversational Speech in Individuals with Disordered Speech," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 11, pp. 4176–4185, 2024.

[8] Lars Rumberg, Christopher Gebauer, Hanna Ehlert, Maren Wallbaum, Lena Bornholt, Jörn Ostermann, and Ulrike Lüdtke, "kidsTALC: A Corpus of 3- to 11-year-old German Children's Connected Natural Speech," in *Proc. Interspeech 2022*, 2022, pp. 5160–5163.

[9] Bence M. Halpern, *Making Speech Technology Accessible for Pathological Speakers*, Ph.D. thesis, University of Amsterdam, 2022.

[10] Philipp Lukas Guldimann, "Speech Recognition for German-Speaking Children with Congenital Disorders: Current Limitations and Dataset Challenges," M.S. thesis, ETH Zürich, 2024.

[11] Joel Shor, Dotan Emanuel, Oran Lang, Omry Tuval, Michael Brenner, Julie Cattiau, Fernando Vieira, Maeve McNally, Taylor Charbonneau, Melissa Nollstadt, Avinatan Hassidim, and Yossi Matias, "Personalizing ASR for Dysarthric and Accented Speech with Limited Data," in *Proc. Interspeech*, 2019, pp. 784–788.

[12] Jinzi Qi and Hugo Van Hamme, "Parameter-efficient Dysarthric Speech Recognition Using Adapter Fusion and Householder Transformation," in *Proc. Interspeech*, 2023, pp. 151–155.

[13] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al., "LoRA: Low-rank Adaptation of Large Language Models," *ICLR*, vol. 1, no. 2, pp. 3, 2022.

[14] Max Müller-Eberstein, Dianna Yee, Karren Yang, Gautam Varma Mantena, and Colin Lea, "Hypernetworks for Personalizing ASR to Atypical Speech," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 1182–1196, 2024.

[15] Enno Hermann and Mathew Magimai Doss, "Few-shot Dysarthric Speech Recognition with Text-to-speech Data Augmentation," in *Proc. Interspeech 2023*, 2023, pp. 156–160.

[16] Pehuen Moure, Longbiao Cheng, Joachim Ott, Zuowen Wang, and Shih-Chii Liu, "Regularized Parameter Uncertainty for Improving Generalization in Reinforcement Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23805–23814.

[17] Sayna Ebrahimi, Mohamed Elhoseiny, Trevor Darrell, and Marcus Rohrbach, "Uncertainty-guided Continual Learning with Bayesian Neural Networks," in *International Conference on Learning Representations*, 2020.

[18] Bai Cong, Nico Daheim, Yuesong Shen, Rio Yokota, Mohammad Emtiyaz Khan, and Thomas Möllenhoff, "Improving LoRA with Variational Learning," *arXiv preprint arXiv:2506.14280*, 2025.

[19] Cristian Meo, Ksenia Sycheva, Anirudh Goyal, and Justin Dauwels, "Bayesian-LoRA: LoRA based Parameter Efficient Fine-Tuning using Optimal Quantization levels and Rank Values trough Differentiable Bayesian Gates," in *2nd Workshop on Advancing Neural Network Training: Computational Efficiency, Scalability, and Resource Optimization (WANT@ICML 2024)*, 2024.

[20] Adam X. Yang, Maxime Robeyns, Xi Wang, and Laurence Aitchison, "Bayesian Low-rank Adaptation for Large Language Models," in *The Twelfth International Conference on Learning Representations*, 2024.

[21] Vaibhav Seth, Ayan Sengupta, Arinjay Pathak, Aastha A K. Verma, Natraj Raman, Sriram Gopalakrishnan, Niladri Chatterjee, and Tanmoy Chakraborty, "Robust and Efficient Fine-tuning of LLMs with Bayesian Reparameterization of Low-Rank Adaptation," *Trans. Mach. Learn. Res.*, vol. 2025, 2025.

[22] Niclas Pokel, Pehuén Moure, Roman Boehringer, Yingqiang Gao, "Adapting Foundation Speech Recognition Models to Impaired Speech: A Semantic Re-chaining Approach for Personalization of German Speech," in *12th edition of the Disfluency in Spontaneous Speech Workshop (DiSS 2025)*, 2025, pp. 82–86.

[23] Vladislav Lialin, Vijeta Deshpande, Xiaowei Yao, and Anna Rumshisky, "Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning," *arXiv preprint arXiv:2303.15647*, 2024.

[24] Yang Lin, Xinyu Ma, Xu Chu, Yujie Jin, Zhibang Yang, Yasha Wang, and Hong Mei, "LoRA Dropout as a Sparsity Regularizer for Overfitting Control," *arXiv preprint arXiv:2404.09610*, 2024.

[25] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra, "Weight Uncertainty in Neural Network," in *International Conference on Machine Learning*. PMLR, 2015, pp. 1613–1622.

[26] Ting Jiang, Shaohan Huang, Shengyue Luo, Zihan Zhang, Haizhen Huang, Furu Wei, Weiwei Deng, Feng Sun, Qi Zhang, Deqing Wang, and Fuzhen Zhuang, "MoRA: High-Rank Updating for Parameter-Efficient Fine-Tuning," *arXiv preprint arXiv:2405.12130*, 2024.

[27] Katherine C. Hustad, Ashley Sakash, Aimee Teo Broman, and Paul J. Rathouz, "Differentiating Typical From Atypical Speech Production in 5-Year-Old Children With Cerebral Palsy: A Comparative Analysis," *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 7, pp. 2585–2603, 2019.

[28] 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 *Interspeech 2008*, 2008, pp. 1741–1744.

[29] Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber, "Common Voice: A Massively-Multilingual Speech Corpus," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, Eds., Marseille, France, May 2020, pp. 4218–4222, European Language Resources Association.

[30] Guilherme Schu, Parvaneh Janbakhshi, and Ina Kodrasi, "On Using the UA-Speech and TORGO Databases to Validate Automatic Dysarthric Speech Classification Approaches," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[31] Xavier F Cadet, Ranya Aloufi, Sara Ahmadi-Abhari, and Hamed Haddadi, "A Study on the Impact of Self-supervised Learning on Automatic Dysarthric Speech Assessment," in *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, 2024, pp. 630–634.

[32] Jiacheng Zhu, Kristjan Greenewald, Kimia Nadjahi, Haitz Sáez De Ocáriz Borde, Rickard Brüel Gabrielsson, Leshem Choshen, Marzyeh Ghassemi, Mikhail Yurochkin, and Justin Solomon, "Asymmetry in Low-rank Adapters of Foundation Models," in *Proceedings of the 41st International Conference on Machine Learning*, 2024, pp. 62369–62385.

[33] Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen, "DoRA: Weight-decomposed Low-rank Adaptation," in *Proceedings of the 41st International Conference on Machine Learning*, 2024, pp. 32100–32121.

Authorized licensed use limited to: ETH BIBLIOTHEK ZURICH. Downloaded on August 06,2026 at 15:20:28 UTC from IEEE Xplore. Restrictions apply.