Demonstration of Adapt4Me: An Uncertainty-Aware Authoring Environment for Personalizing Automatic Speech Recognition to Non-normative Speech

Niclas Pokel
Institute of Neuroinformatics
University of Zurich and ETH Zurich
Zurich, Switzerland
ETH AI Center
Zurich, Switzerland
npokel@ethz.ch

Yiming Zhao
ETH Zurich
Zurich, Switzerland
yimizhao@ethz.ch

Pehuen Moure
Institute of Neuroinformatics
University of Zurich and ETH Zurich
Zurich, Switzerland
ETH AI Center
Zurich, Switzerland
pehuen@ini.ethz.ch

Yingqiang Gao*
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
ETH AI Center
Zurich, Switzerland
yingqiang.gao@cl.uzh.ch

Roman Boehringer*
Institute of Neuroinformatics
University of Zurich and ETH Zurich
Zurich, Switzerland
romaboeh@ethz.ch

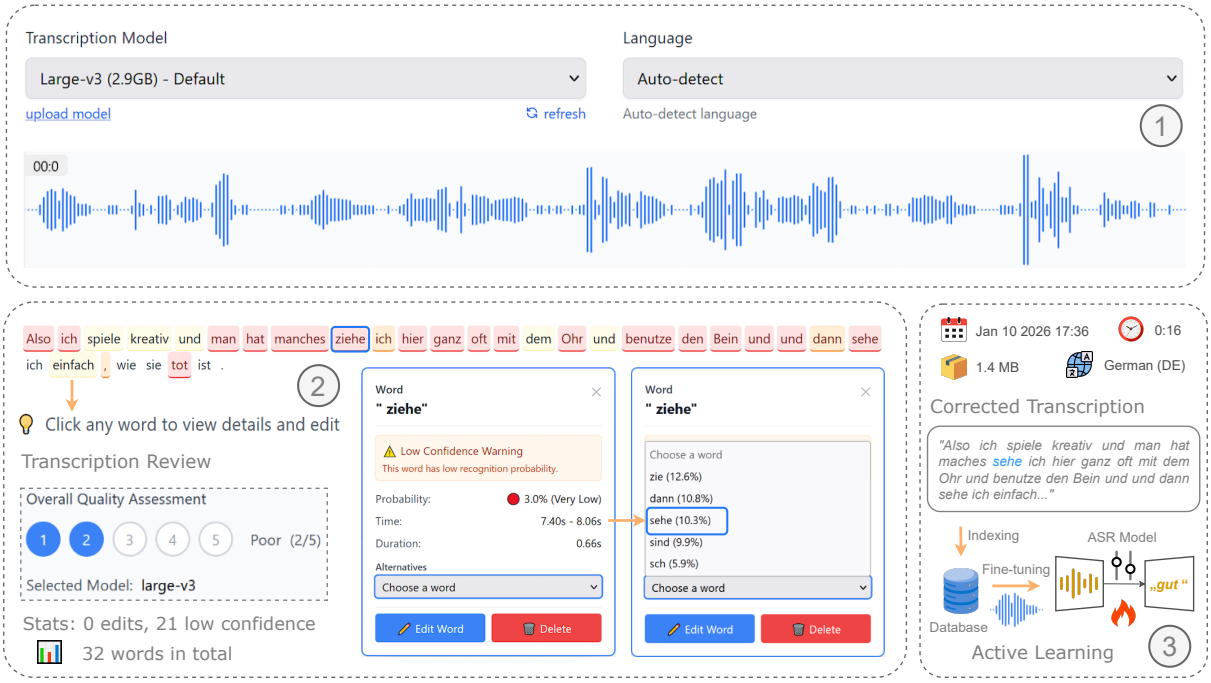

Figure 1: *Adapt4Me*: an end-to-end, human-in-the-loop ASR personalization system for non-normative speech. ① The system prompts the user with high-information sentences for initial recordings; ② Word-level model uncertainty is visualized through color-coding, guiding users directly to likely identify and correct transcribing errors; ③ Users resolve errors via low-effort top-k best corrections powered by a Bayesian backend, with validated edits continuously fed back into active learning and model fine-tuning.

Abstract

Individuals with atypical speech are systematically excluded from Automatic Speech Recognition (ASR) systems, as personalization remains hindered by labor-intensive data collection and technically complex model training. To address these limitations, we propose *Adapt4Me*, a web-based, user-controlled environment that operationalizes Bayesian active learning to enable end-to-end personalization without expert supervision. The app exposes data selection, adaptation, and validation to lay users through a three-stage human-in-the-loop workflow: (1) rapid profiling via greedy phoneme sampling to capture speaker-specific acoustics; (2) backend personalization using Variational Inference Low-Rank Adaptation (VI-LoRA) to enable fast, incremental updates; and (3) continuous improvement, where users guide model refinement by resolving visualized model uncertainty via low-friction top-k corrections. By making epistemic uncertainty explicit, *Adapt4Me* reframes data efficiency as an interactive design feature rather than a purely algorithmic concern. We show how this enables users to personalize robust ASR models, transforming them from passive data sources into active authors of their own assistive technology.

CCS Concepts

• **Human-centered computing** → **Accessibility technologies**; **Interactive systems and tools**; • **Computing methodologies** → *Active learning settings*; *Speech recognition.*

Keywords

Accessibility, Dysarthria, Active Learning, Human-in-the-Loop, ASR Personalization, Uncertainty Visualization, Low-Friction Annotation

ACM Reference Format:

Niclas Pokel, Yiming Zhao, Pehuen Moure, Yingqiang Gao, and Roman Boehringer. 2026. Demonstration of Adapt4Me: An Uncertainty-Aware Authoring Environment for Personalizing Automatic Speech Recognition to Non-normative Speech. In *ACM Conversational User Interfaces 2026 (CUI '26), July 21–24, 2026, Bremen, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3816046.3816317

1 Introduction

State-of-the-art ASR models like Whisper [32] or Wav2Vec [36] have achieved remarkable performance on normative speech, yet they exclude individuals with motor speech disorders (e.g., dysarthria) [3, 5, 21] or structural impairments of the speech apparatus (e.g., Apert syndrome) [45].[1] While previous work has explored fine-tuning foundation ASR models for non-normative speech, error rates remain substantially higher than those observed for normative speech [17, 20, 38]. This lack of generalization to non-normative speech further reinforces the marginalization of individuals with speech impairments, as their specific communicative needs are rarely addressed by modern speech technologies, thereby limiting their access to the benefits of AI.

Nevertheless, although personalizing ASR models for user-specific transcription is technically feasible, it remains largely inaccessible in practice. Conventional model personalization typically requires hours of voice recordings, which is often exhaustive for individuals with speech impairments and places a substantial burden on family members and caregivers [4, 12, 25, 35]. This process becomes even more burdensome when personalized models must be repeatedly optimized with newly recorded data to adapt to the evolving acoustic characteristics of individuals with speech impairments, who are often everyday ASR users without expert knowledge.

Users often experience ASR systems as "black-box" transcribers, with little insight into which phonemes cause recognition errors[7, 18]. This opacity is largely due to the absence of token-level uncertainty visualizations that reveal the system's transcription difficulties, preventing users from providing targeted and effective feedback in subsequent fine-tuning iterations. Previous work in HCI suggests that well-designed uncertainty visualizations enable users to make better-informed decisions and act as effective supervisors of imperfect AI systems [15].

We argue that effective ASR personalization requires a shift from passive data collection to interactive curating. This shift hinges on exposing the model's epistemic uncertainty, its awareness of what it does not yet model reliably, so users can guide personalization in a principled and efficient way. Recent work in Interactive Machine Learning has shown that users value transparency and agency when training assistive tools [8, 14]. This shift is particularly critical for implementation of ASR personalization in real-world settings, where minimizing user effort is essential for enabling continuous adaptation to newly recorded speech data, but also for ensuring sustained user engagement.

We present *Adapt4Me*, a web-based, accessible, and low-effort environment that operationalizes Bayesian Active Learning (AL) for personalizing ASR for non-normative speech. Compared to existing personalization platforms such as Project Euphonia [39] and Project Relate [22], which rely on centralized data submission to commercial backends, require substantially larger recording efforts, and operate as opaque batch trainers without per-word feedback, *Adapt4Me* exposes word-level uncertainty, supports incremental active adaptation, and keeps the personalization loop under direct user control. By integrating an uncertainty-aware VI-LoRA-based

*Corresponding authors. Visit https://adapt4me.ai for more information.

CUI '26, Bremen, Germany

ACM ISBN 979-8-4007-2741-2/26/07
https://doi.org/10.1145/3816046.3816317

[1] We use *atypical speech* and *non-normative speech* interchangeably to refer to the broad population whose articulation deviates from training-data norms (e.g., dysarthria, structural impairments, age-related changes). *Adapt4Me* is method-agnostic across these subpopulations.

ASR model [31] into a user-friendly interface, *Adapt4Me* offers the following advantages:

- **Efficient "Cold Start"**: A profiling method using Greedy Biphone Coverage [29] that bootstraps personalization in minutes of training data rather than hours;
- **Uncertainty-Guided Feedback**: A visual interface that highlights confused phonemes, guiding users to provide personalization data for active learning through direct human-computer interaction;
- **Longitudinal Resilience**: A lifecycle management feature that, unlike comparable systems, explicitly handles non-linear speech changes (e.g., post-surgery recovery or puberty voice change) by preserving lexical personalization across acoustic re-baselining.

Adapt4Me is a proof-of-concept prototype that demonstrates how advances in ASR models can be converted into tangible solutions to improve communication accessibility for individuals with speech impairments. By bridging the gap between technological advances and real-world deployment, *Adapt4Me* enables efficient and personalized ASR adaptation that can support speech-impaired users over the long term.

2 Motivation

Table 1: Comparison of duration in hours of existing speech datasets (recording hours are reduced by redundant microphones).

Domain	Dataset	Hours
Normative	Common Voice (EN) [2]	~3,800
Normative	LibriSpeech (EN) [26]	~1,000
Dysarthric	SAP (EN) [10]	~415
Dysarthric	UA-Speech (EN) [16]	~9
Dysarthric	BF-Sprache (DE) [29]	~2.5
Dysarthric	TORGO (EN) [34]	~3

The rich diversity of speech impairments cannot be addressed simply by collecting more data, as variability across individuals is substantially higher than in normative speech populations [40, 41, 44]. This high inter-speaker variance poses significant challenges for training personalized ASR models. As shown in Table 1, available non-normative speech resources are orders of magnitude smaller than their normative counterparts. Consequently, this data-scarcity bottleneck limits the applicability of traditional personalization approaches, which are typically data-intensive and depend on extensive expert annotation, making them impractical for deployment in a home settings [27]. In addition, standard personalization approaches typically fine-tune models on small amounts of user data, making them prone to overfitting and catastrophic forgetting [6, 23]. While Parameter-Efficient Fine-Tuning (PEFT [11]) methods such as LoRA [13] partially alleviate these issues, they do not address a more fundamental problem: the lack of an effective and accessible workflow for personalizing ASR models. Without human-in-the-loop (HITL [24, 43]) guidance, users often expend effort recording redundant data, namely utterances that the model already transcribes well, while failing to gather data that is most informative for further reducing recognition errors. This mismatch between user effort and model improvement underscores the necessity of systems that enable data-efficient personalization while minimizing cognitive and physical burden on end users.

To address both data scarcity and the lack of effective personalization workflows in ASR, *Adapt4Me* is designed following the principle of active learning [9, 37]. Rather than treating users as passive data providers, *Adapt4Me* positions them as active participants in the personalization loop, a paradigm aligned with interactive machine learning principles [1]. In this setting, users inspect transcription outputs, judge their quality, and correct erroneous predictions to directly guide model adaptation. This HITL process enables the system to focus data collection on informative samples that contribute most to performance improvement. Through *Adapt4Me*, we demonstrate that effective ASR personalization systems must go beyond algorithmic adaptation alone: they must reduce human effort through targeted interaction and provide actionable feedback that supports iterative model improvement in real-world, home-based settings.

3 System Architecture

The architecture of *Adapt4Me* follows the three-step workflow visualized in Figure 2, starting from speech recording to continuous, uncertainty-guided adaptation.

Step 1: Speech Profiling. To bootstrap personalization for a new user, we employ a *Greedy Biphone Coverage* algorithm [29]. The system first selects a minimal set of initial words from the literature to maximize phonetic variance. These words are used as seeds for an LLM prompt to generate semantically and structurally coherent training samples.

Step 2: End-to-End Personalization. Once the initial audio is captured, the backend personalizes the model using the VI-LoRA algorithm [31]. This parameter-efficient method stabilizes fine-tuning on small datasets and quantifies *epistemic uncertainty* [30]. We aggregate this uncertainty to calculate a *Phoneme Difficulty Score (i.e., PhDScore)*, identifying the user's specific articulatory struggles [30]. This diagnosis drives the *Semantic Re-chaining* engine to synthesize new, targeted sentence prompts rich in these difficult phonemes, constructing a personalized training curriculum for the next phase.

Step 3: Active Learning. Users correct the free speech predictions, guided by word-level uncertainty scores. A key challenge is that standard top-k alternatives from ASR models often lack semantic coherence. To address this, we employ a two-pass decoding strategy: a *coherent pass*, which produces semantically consistent transcriptions, and a *variation pass*, which selectively re-samples only high-uncertainty words while keeping their surrounding context fixed.

4 Human-Computer Interaction Design

Adapt4Me is designed primarily for end users with atypical speech to operate themselves; for users with co-occurring cognitive or motor limitations (e.g., young children, individuals with secondary disabilities), a caregiver can take on the operator role.

To actively guide user correction, *Adapt4Me* highlights low-confidence words in model transcriptions (e.g., "Pylikan" in Figure 2) using entropy-based uncertainty scores. These uncertainty scores

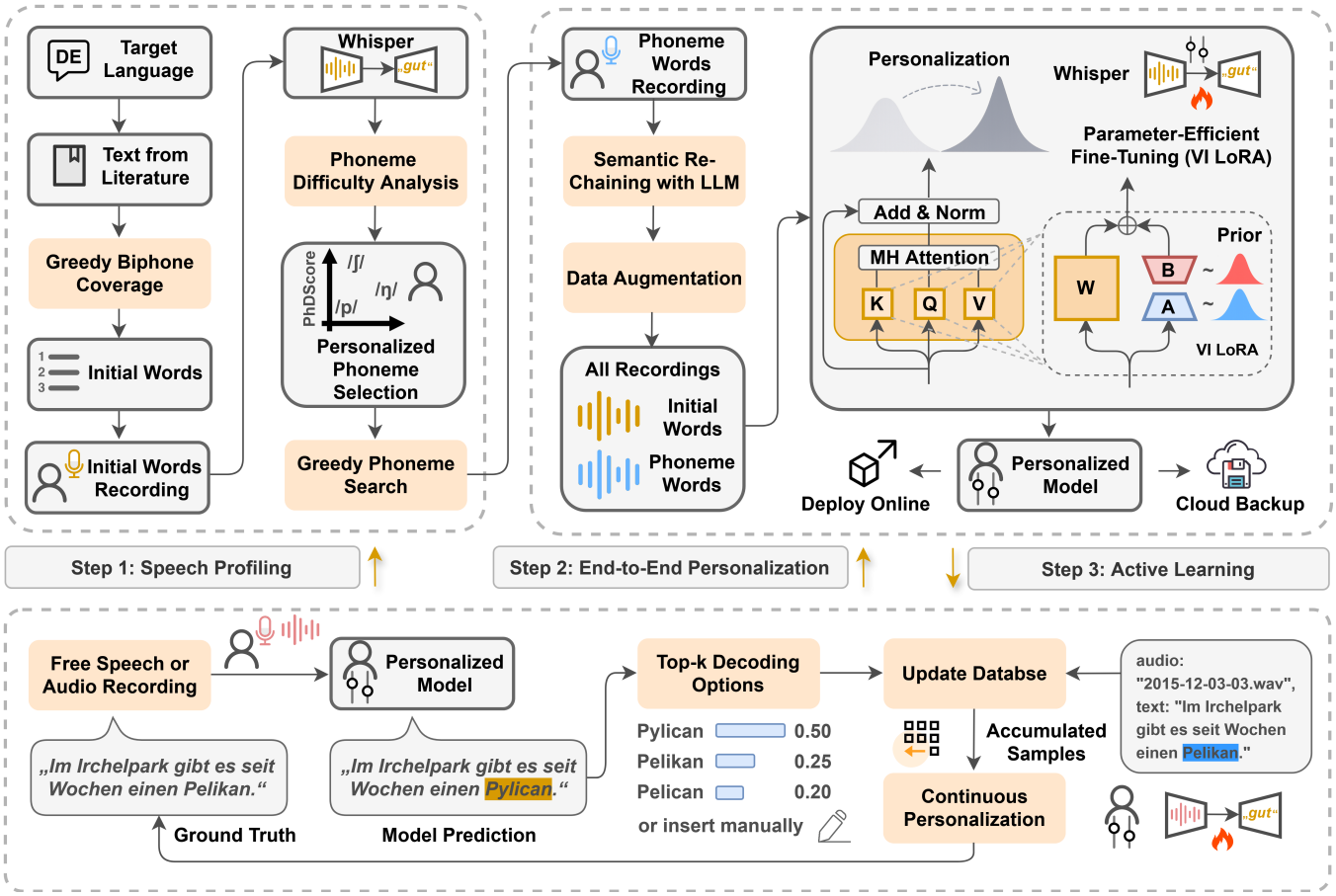

Figure 2: The Adapt4Me Personalization Lifecycle. Step 1 (Speech Recording): The system creates a "Cold Start" profile using a minimized set of 500 words selected via Greedy Biphone Coverage and LLM Prompting [29]. Step 2 (Model Personalization): The backend adapts the model. It uses Semantic Re-chaining (SRC) [29] to generate context, calculates the Phoneme Difficulty Score (PhDScore) to search for highly informative training-samples [30], and fine-tunes the VI-LoRA adapters [31] (Top Right). Step 3 (Active Learning): The continuous interaction loop. The model transcribes free speech, highlights uncertain words (e.g., "Pylikan"), and generates context-aware N-best suggestions to minimize annotation effort.

serve two complementary purposes: (1) as a *speech diagnostic tool*, offering insights into how the model interprets user speech by revealing words or phonemes that consistently cause transcription errors; and (2) as a *task management tool*, directing user attention to the most probable errors, thereby eliminating the need to proofread correctly transcribed segments and reducing cognitive load of the user. Our design choices, word-level color coding paired with top-k selection rather than free-text editing, were informed by consultation with a clinical logopedist and shaped by the motor and cognitive profiles of the target population, though they have not yet been validated in a formal formative study (see Section 7).

Motor impairments frequently co-occur with speech impairments in individuals with neurodevelopmental disorders [19], which can make fine-grained motor actions such as typing corrections difficult and fatiguing. As a result, manually editing ASR outputs can become a significant barrier for such users [42]. To address this challenge, *Adapt4Me* replaces typing-based corrections with a context-aware top-k selection mechanism, allowing users to correct errors by simply selecting the appropriate word from a short list of alternatives. Manual insertion remains available as a fallback. By transforming error correction from a typing task into a lightweight selection task, the system substantially reduces physical effort and improves accessibility for users with motor limitations.

To support lifelong ASR personalization, *Adapt4Me* is designed to accommodate the dynamic nature of speech impairments. The system incrementally adapts to gradual changes, such as a child's speech development, as users continuously contribute corrected samples. In contrast, major physiological changes, including post-surgical recovery [33] or puberty-related voice changes [28], may render an existing personalized acoustic model less effective. In such cases, users can re-initiate the system from the initial recording stage to establish a new acoustic baseline, while preserving

previously learned semantic and lexical personalization. This design prevents a complete loss of progress and supports long-term, evolving use.

When transcribing with *Adapt4Me*, ten forward passes produce an inference latency of ≈ 2 s on a standard GPU (e.g. NVIDIA GeForce RTX 3090). This is acceptable for the home setting, as *Adapt4Me* was designed primarily to personalize ASR models rather than the real-time user experience. We employ a client-server setup, with a web application (Client) running on edge devices like tablets or smartphones, familiar to children. The computationally intensive personalization is offloaded to a secure cloud backend. This ensures that families do not need high-end GPUs for home deployment of the system.

Unlike sound-treated clinical environments, home settings are often acoustically noisy. *Adapt4Me* therefore incorporates a lightweight Signal-to-Noise Ratio (SNR) check that runs locally in the browser to prevent the submission of low-quality recordings, such as speech contaminated by background television noise, that could degrade system performance. In addition, *Adapt4Me* also supports collaborative use by family members. Prompts are presented in large, simple fonts suited to a speech-impaired child, while the uncertainty visualization and top-k panel are positioned for an adjacent caregiver to oversee corrections. This arrangement supports a dyadic workflow in which the child produces speech and the caregiver validates or selects corrections, distributing cognitive and motor effort across the pair.

5 Envisioned Conference Experience

To effectively demonstrate this home-based, collaborative workflow in a conference setting, we will present *Adapt4Me* via a local web interface deployed on a provided laptop and tablet. Because conference attendees typically possess normative speech, and because live model fine-tuning exceeds feasible booth dwell times, the demonstration is designed as an interactive, simulated authoring scenario.

Attendees will step into the role of the "human-in-the-loop" (e.g., acting as the caregiver or user), tasked with curating pre-recorded, anonymized non-normative speech samples. This workflow allows attendees to directly experience the system's core HCI contributions: interpreting the token-level uncertainty visualizations and utilizing the low-friction top-k selection mechanism to correct transcription errors without typing.

Following the correction phase, attendees will experience a live comparison, contrasting the semantic hallucinations of a baseline Whisper model with the phonetically accurate output of an *Adapt4Me*-personalized model. The physical setup natively supports the dyadic interaction described above, ensuring a fast, engaging 3-minute flow that accommodates high booth throughput while mitigating the challenges of the noisy conference exhibition hall.

6 Evaluation and Clinical Evidence

We deployed *Adapt4Me* in a home setting for a teenage user with structural speech impairment. As shown in Figure 3, only 75 minutes of cumulative interaction (recording and correction) reduced

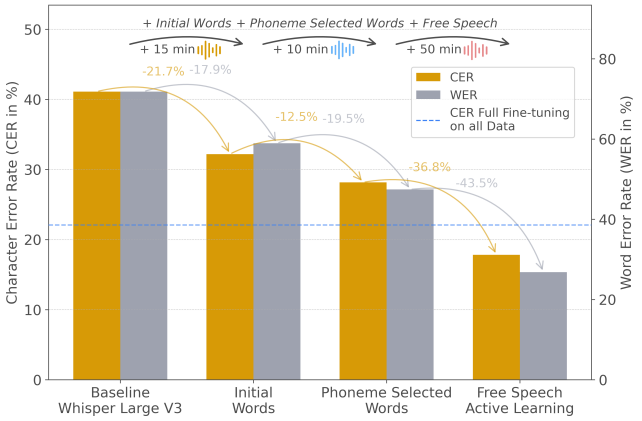

Figure 3: Longitudinal user study with a teenage child with structural speech impairment. The system reduces word error rate (WER) from a non-functional ~70% to usable ~25% within under 90 minutes of total interaction. Uncertainty-aware active learning (orange/grey bars) eventually outperforms the full fine-tuning baseline (blue dashed line), demonstrating the effectiveness of VI-LoRA in data-scarce regimes.

the Word Error Rate (WER) from a non-functional 70% of a Whisper Large baseline [32] to a usable level of approximately 25%. The proposed active learning pipeline with VI-LoRA fine-tuning outperformed a brute-force full-parameter fine-tuning baseline (blue dashed line) while using substantially less data, providing preliminary evidence that targeting high-uncertainty phonemes is an efficient personalization strategy. We note that these findings come from a single longitudinal deployment and treat them as illustrative rather than conclusive; broader evaluation across speakers is ongoing. Qualitative analysis further shows that personalization restores communicative intent. Rather than producing semantic hallucinations, the adapted model makes phonetically plausible errors that remain intelligible to humans. For example, when dictating Swiss train stations (*"Wiedikon, Enge, Thalwil, Baar"*), the baseline hallucinated unrelated sentences, whereas *Adapt4Me* produced *"Vidikon, Enne, Talwil, Borg"*. Despite high WER, these phonetic spellings are understandable in context, making the system useful. Because residual errors remain phonetically grounded in the user's intent, downstream conversational use (e.g., dictation, voice-agent commands) continues to convey meaning even when WER alone would suggest unusability, a qualitative pattern we have observed repeatedly in live testing. Prior work [30] validated model-predicted phoneme difficulties against clinical logopedic assessments. The strong correlation between the model's internal uncertainty estimates and therapists' evaluations confirms that the visual feedback provided by *Adapt4Me* is clinically meaningful rather than a mere statistical artifact.

7 Future Work

Adapt4Me offers a novel platform for large-scale speech science. By aggregating anonymized model states, specifically the trajectories of uncertainty heatmaps, researchers could perform meta-analysis

on speech patterns without manual annotation. This could reveal latent population-level clusters, such as distinguishing between age-related phonetic drift and pathology-driven changes, effectively turning the ASR model into a diagnostic sensor.

We also plan a formative study with end users and caregivers to validate the uncertainty visualization and top-k correction interface, examining whether the chosen visual encoding and interaction granularity match users' actual needs. Closely related, we are interested in the social dynamics of dyadic personalization, how parent-child or caregiver-user pairs negotiate authority over corrections and how their differing goals (accuracy, speed, engagement) shape the personalization trajectory.

Finally, sustained engagement, especially for child users, will likely require gamification of the correction loop. Casting prompt recording and top-k selection as a reward-driven game could turn a repetitive curation task into a motivating activity, an avenue we view as essential for real-world long-term deployment.

Acknowledgments

We would like to thank Dr. Corinne Mathys Zulauf for her logopedic consulting, Dr. Samra Hamzic for valuable discussion and Philipp Guldimann for the initial data collection and annotation.

References

[1] Saleema Amershi, Maya Cakmak, W. Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (2014), 105–120. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1609/aimag.v35i4.2513 doi:10.1609/aimag.v35i4.2513

[2] 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*, 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.). European Language Resources Association, Marseille, France, 4218–4222. https://aclanthology.org/2020.lrec-1.520/

[3] Fabio Ballati, Fulvio Corno, and Luigi De Russis. 2018. Assessing virtual assistant capabilities with Italian dysarthric speech. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 93–101.

[4] Paula Cronin, Rebecca Reeve, Patricia McCabe, Rosalie Viney, and Stephen Goodall. 2020. Academic Achievement and Productivity Losses Associated with Speech, Language and Communication Needs. *International Journal of Language & Communication Disorders* 55, 5 (2020), 734–750. doi:10.1111/1460-6984.12558

[5] Joseph R. Duffy. 2005. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (2nd ed.). Elsevier Mosby, St. Louis, MO. Includes bibliographical references and index..

[6] Robert M. French. 1999. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences* 3, 4 (1999), 128–135. doi:10.1016/S1364-6613(99)01294-2

[7] Neta Glazer, Yael Segal-Feldman, Hilit Segev, Aviv Shamsian, Asaf Buchnick, Gill Hetz, Ethan Fetaya, Joseph Keshet, and Aviv Navon. 2025. Beyond Transcription: Mechanistic Interpretability in ASR. *arXiv preprint arXiv:2508.15882* (2025).

[8] Steven M Goodman, Emma J McDonnell, Jon E Froehlich, and Leah Findlater. 2025. SPECTRA: Personalizable Sound Recognition for Deaf and Hard of Hearing Users through Interactive Machine Learning. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*.

[9] Dilek Hakkani-Tür, Giuseppe Riccardi, and Allen Gorin. 2002. Active Learning for Automatic Speech Recognition. In *2002 IEEE international conference on acoustics, speech, and signal processing*, Vol. 4. IEEE, IV–3904.

[10] Mark Hasegawa-Johnson, Xiuwen Zheng, Heejin Kim, Clarion Mendes, Meg Dickinson, Erik Hege, Chris Zwilling, Marie Channell, Laura Mattie, Heather Hodges, Lorraine Ramig, Mary Bellard, Mike Shebanek, Leda Sarı, Kaustubh Kalgaonkar, David Frerichs, Jeffrey Bigham, Leah Findlater, Colin Lea, and Bob MacDonald. 2024. Community-Supported Shared Infrastructure in Support of Speech Accessibility. *Journal of Speech, Language, and Hearing Research* 67 (09 2024), 4162–4175. doi:10.1044/2024_JSLHR-24-00122

[11] Junxian He, Chunting Zhou, Xuezhe Ma, Taylor Berg-Kirkpatrick, and Graham Neubig. 2022. Towards a Unified View of Parameter-Efficient Transfer Learning. In *Proceedings of the Tenth International Conference on Learning Representations*.

[12] Elaine R. Hitchcock, Daphna Harel, and Tara McAllister Byun. 2015. Social, Emotional, and Academic Impact of Residual Speech Errors in School-Aged Children: A Survey Study. *Seminars in Speech and Language* 36, 4 (2015), 283–294. doi:10.1055/s-0035-1562911

[13] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. LoRA: Low-rank Adaptation of Large Language Models. *ICLR* 1, 2 (2022), 3.

[14] Hernisa Kacorri, Kris M Kitani, Jeffrey P Bigham, and Chieko Asakawa. 2017. People with visual impairment training personal object recognizers: Feasibility and challenges. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5839–5849.

[15] Matthew Kay, Tara Kola, Jessica R. Hullman, and Sean A. Munson. 2016. When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5092–5103. doi:10.1145/2858036.2858558

[16] 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*. 1741–1744. doi:10.21437/Interspeech.2008-480

[17] Minkyu Kim and Jeong Eon Sung. 2023. Exploring the Efficacy of OpenAI Whisper for the Recognition of Dysarthric Speech. *Applied Sciences* 13, 12 (2023), 7092. doi:10.3390/app13127092

[18] Korbinian Kuhn, Verena Kersken, and Gottfried Zimmermann. 2025. Evaluating ASR Confidence Scores for Automated Error Detection in User-Assisted Correction Interfaces. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[19] Giulio E Lancioni, Jorge Navarro, Nirbhay N Singh, Mark F O'Reilly, Jeff Sigafoos, Antonella Mellino, Pietro Arcuri, Gloria Alberti, and Valeria Chiariello. 2025. People with Neuro-Motor Impairment, Lack of Speech, and General Passivity Can Engage in Basic Forms of Activity and Communication with Technology Support. *Advances in Neurodevelopmental Disorders* 9, 1 (2025), 105–114.

[20] Shansong Liu, Mengzhe Geng, Shoukang Hu, Xurong Xie, Mingyu Cui, Jianwei Yu, Xunying Liu, and Helen Meng. 2022. Recent Progress in the CUHK Dysarthric Speech Recognition System. *arXiv preprint arXiv:2201.05845* (2022).

[21] Ben Maassen. 2007. *Speech Motor Control: In Normal and Disordered Speech*. Oxford University Press.

[22] Alicia Martin, Robert L MacDonald, Pan-Pan Jiang, Marilyn Ladewig, Julie Cattiau, Rus Heywood, Richard Cave, Jimmy Tobin, Philip C Nelson, and Katrin Tomanek. 2025. Project Euphonia: Advancing Inclusive Speech Recognition through Expanded Data Collection and Evaluation. *Frontiers in Language Sciences* 4 (2025), 1569448.

[23] Michael McCloskey and Neal J. Cohen. 1989. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. Psychology of Learning and Motivation, Vol. 24. Academic Press, 109–165. doi:10.1016/S0079-7421(08)60536-8

[24] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. 2023. Human-in-the-Loop Machine Learning: A State of the Art. *Artificial Intelligence Review* 56, 4 (2023), 3005–3054.

[25] Allyson D Page and Kathryn M Yorkston. 2022. Communicative Participation in Dysarthria: Perspectives for Management. *Brain sciences* 12, 4 (2022), 420.

[26] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. 2015. Librispeech: An ASR corpus based on public domain audio books. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 5206–5210. doi:10.1109/ICASSP.2015.7178964

[27] Joon Sung Park, Danielle Bragg, Ece Kamar, and Meredith Ringel Morris. 2021. Designing an Online Infrastructure for Collecting AI Data From People With Disabilities. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 52–63. doi:10.1145/3442188.3445870

[28] Ana P Pinheiro, Jean-Julien Aucouturier, and Sonja A Kotz. 2024. Neural Adaptation to Changes in Self-voice during Puberty. *Trends in Neurosciences* (2024).

[29] Niclas Pokel, Pehuén Moure, Roman Boehringer, and Yingqiang Gao. 2025. 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)*. 82–86. doi:10.21437/DiSS.2025-17

[30] Niclas Pokel, Pehuén Moure, Roman Boehringer, and Yingqiang Gao. 2025. Data-Efficient ASR Personalization for Non-Normative Speech Using an Uncertainty-Based Phoneme Difficulty Score for Guided Sampling. arXiv:2509.20396 [eess.AS] https://arxiv.org/abs/2509.20396

[31] Niclas Pokel, Pehuén Moure, Roman Boehringer, Shih-Chii Liu, and Yingqiang Gao. 2026. Variational Low-rank Adaptation for Personalized Impaired Speech Recognition. In *ICASSP 2026-2026 IEEE International Conference on Acoustics,*

Speech and Signal Processing (ICASSP). IEEE, 18457–18461.

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

[33] Srinivasa Naidu Brahma Reddy and Adi Narayana Mallikarjuna Reddy. 2023. Automatic Speech Recognition for Cleft Lip and Cleft Palate Patients Using Deep Learning Techniques. In *Speech and Language Processing for Human-Machine Communications*. CRC Press. doi:10.1201/9781003433941-2

[34] 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 (2012), 523–541. doi:10.1007/s10579-011-9145-0

[35] Zhanna Sarsenbayeva and et al. 2022. Methodological standards in accessibility research on motor impairments: a survey. *ACM Computing Surveys (CSUR)* 55, 7 (2022), 1–35.

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

[37] Burr Settles. 2009. *Active Learning Literature Survey*. Technical Report 1648. University of Wisconsin-Madison, Department of Computer Sciences.

[38] Amrit Singh, Ayush Goyal, Anil Kumar, and Rohit Kumar. 2024. An Investigation of Whisper ASR for Dysarthric Speech Recognition. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 11711–11715. doi:10.1109/ICASSP48485.2024.10446973

[39] 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 (2024), 4176–4185.

[40] Johannes Tröger, Felix Dörr, Louisa Schwed, Nicklas Linz, Alexandra König, Tabea Thies, Juan Rafael Orozco-Arroyave, and Jan Rusz. 2024. An Automatic Measure for Speech Intelligibility in Dysarthrias—validation Across Multiple Languages and Neurological Disorders. *Frontiers in Digital Health* 6 (2024), 1440986.

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

[42] Keith Vertanen and Per Ola Kristensson. 2009. Parakeet: a continuous speech recognition system for mobile touch-screen devices. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (Sanibel Island, Florida, USA) *(IUI '09)*. Association for Computing Machinery, New York, NY, USA, 237–246. doi:10.1145/1502650.1502685

[43] Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. 2022. A Survey of Human-in-the-Loop for Machine Learning. *Future Generation Computer Systems* 135 (2022), 364–381.

[44] Fang Xiong, Jon Barker, and Heidi Christensen. 2019. Phonetic Analysis of Dysarthric Speech Tempo and Applications to Robust Personalised Dysarthric Speech Recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 6465–6469. doi:10.1109/ICASSP.2019.8683134

[45] C. E. Zimmerman, J. Sun, A. M. Wes, G. H. Vu, C. L. Kalmar, L. S. Humphries, S. P. Bartlett, M. A. Cohen, J. W. Swanson, and J. A. Taylor. 2020. Long Term Speech Outcomes Following Midface Advancement in Syndromic Craniosynostosis. *The Journal of Craniofacial Surgery* 31, 6 (2020), 1775–1779. doi:10.1097/SCS.0000000000006581