Automatic Screening for Children with Speech Disorder using Automatic Speech Recognition: Opportunities and Challenges

Dancheng Liu[1], Jason Yang[*1], Ishan Albrecht-Buehler[†*1,2], Helen Qin[†*1,3], Sophie Li[*1,4]
Yuting Hu[1], Amir Nassereldine[1], Jinjun Xiong[1]

[1]University at Buffalo
[2]Nichols School
[3]Thomas Jefferson High School for Science & Technology
[4]Langley High School
[1]{dliu37, jyang86, yhu54, amirnass, jinjun}@buffalo.edu, [2]ishanalbrecht@outlook.com
[3]helen.ty.qin@gmail.com, [4]sophie.zhiran.li@gmail.com

Abstract

Speech is a fundamental aspect of human life, crucial not only for communication but also for cognitive, social, and academic development. Children with speech disorders (SD) face significant challenges that, if unaddressed, can result in lasting negative impacts. Traditionally, speech and language assessments (SLA) have been conducted by skilled speech-language pathologists (SLPs), but there is a growing need for efficient and scalable SLA methods powered by artificial intelligence. This position paper presents a survey of existing techniques suitable for automating SLA pipelines, with an emphasis on adapting automatic speech recognition (ASR) models for children's speech, an overview of current SLAs and their automated counterparts to demonstrate the feasibility of AI-enhanced SLA pipelines, and a discussion of practical considerations, including accessibility and privacy concerns, associated with the deployment of AI-powered SLAs.

Introduction

As arguably one of the most important abilities for human beings, speech plays a vital part in our lives. While speech seems to be used majorly for communication among adults, its significance transcends communication for children, and children with speech disorders face unique challenges that can have enduring consequences if left unaddressed (Bashir and Scavuzzo 1992). Studies have shown that speech is not a mere means of expressing thoughts for children. Rather, it is the cornerstone of effective learning (Young et al. 2002; Tomblin et al. 2000), appropriate social interaction (Redmond and Rice 1998, 2002; Coster et al. 1999) , and mental health (Jerome et al. 2002; BEITCHMAN et al. 2001) during the formative years. Yet, it has been reported (Khan, Freeman, and Druet 2023) that about 9% of children were diagnosed with speech disorders (SD) during assessments prior to the COVID-19 pandemic, and the number increased to 21% in 2022. In total, roughly 1.2 million children in the United States were diagnosed with speech disorders in 2022 (Khan, Freeman, and Druet 2023). The importance of speech and the huge number of affected children together call for immediate and effective screening and treatment.

Speech and language assessments (SLA) have long been essential tools that clinicians and practitioners use to facilitate early treatments for children with SD. SLAs were traditionally carried out by highly skilled speech-language pathologists (SLPs), but the burgeoning demand for efficient and scalable assessment methods has prompted a paradigm shift. Marked by the development of deep learning (DL), we are now able to integrate multiple AI-powered techniques with the traditional SLA pipeline, as shown in Figure 1, promising a revolutionary leap forward in the field of speech and language evaluation.

In this position paper, we bring forth three contributions to the SLA and DL community. First, we outline and survey existing techniques that could be used in this automated SLA pipeline, with a focus on automatic speech recognition (ASR). We will showcase the current progress in adapting ASR models to children's speech. Second, we discuss popular SLAs and their automated prototypes. We will show that the envisioned pipeline in Figure 1 is actually not far from realization. Third, we highlight some of the practical concerns and solutions regarding the accessibility of the AI-powered pipeline and the privacy concerns on user data.

Automatic Speech Recognition

In this section, we explore the fundamental technology that drives the automated SLA: automatic speech recognition (ASR). ASR takes a piece of audio and transcribes it into textual formats that humans visually understand. The output format is majorly in words, but other variations such as phonemes or characters also exist.

[*]Work done as a summer intern at the University at Buffalo.
[†]These authors contributed equally.

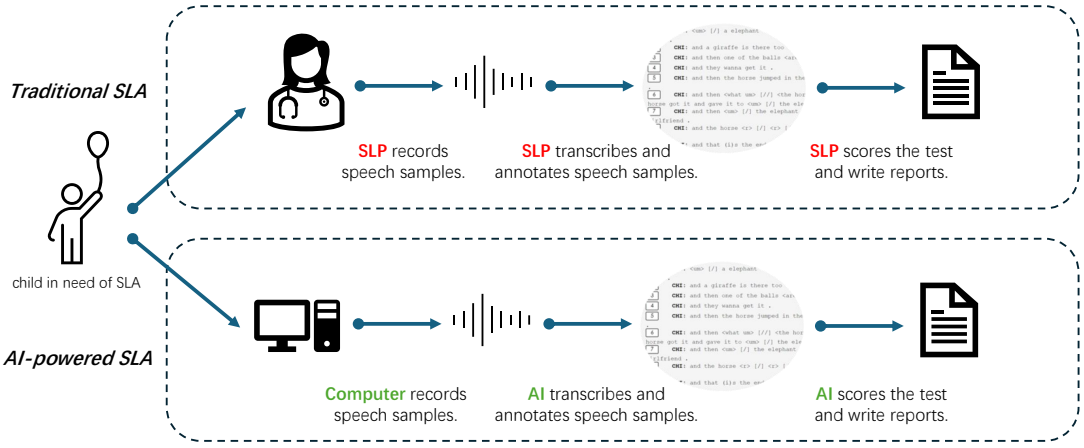

Figure 1: A generic pipeline for automated speech and language assessments in the future. Children in need of SLA will use a local computer to perform all required components of the test which ensures privacy and usability, and such remote SLA will also alleviate the burden of SLPs, helping them focus on the treatment of children in need.

Adult ASR

The majority of works in the field of ASR focus on adult speech, in part due to the relatively broader impacts and easier acquisition of training data. Powered by massive datasets such as CommonVoice (Ardila et al. 2020), ASR models such as Wav2Vec2 (Baevski et al. 2020) and Conformer (Gulati et al. 2020) showcase excellent transcription quality for adults. More recently, the Whisper model (Radford et al. 2022; Bain et al. 2023) has revolutionized the ASR domain with large-scale weak-supervise training and has achieved high robustness towards any adult speech.

Adapting to Children Speech

However, even with the success of the Whisper model, obtaining high-quality ASR models that serve children remains a challenging problem. Children exhibit distinctively different characteristics compared to adults, including rapid changes in pitch, articulation patterns, and vocabulary development, and ASR models that work well with adults fail to excel with children (Bhardwaj et al. 2022). Here, we provide some representative results of evaluating different ASR models on children's speech datasets in Table 1. In particular, we choose the best-performing models – Conformer (Gulati et al. 2020), wav2vec2 (Baevski et al. 2020), Whisper (Radford et al. 2022), and Nvidia NeMo STT (Harper et al. 2024) – on the Librispeech dataset (Panayotov et al. 2015) and evaluate their performance on the two children datasets, ENNI (Liu and Xiong 2024) and My Science Tutor (MyST) (Pradhan, Cole, and Ward 2023). As we can see from these models' word error rate (WER), they suffer from significant accuracy degradation when applied to children's speech. Even the robust Whisper model shows a non-negligible difference that hinders its practical usage.

To prevent accuracy degradation and ensure precise processing and analysis of children's speech for assessments, the adaptation of the ASR models become a necessity. The prevalent adaptation technique employs various fine-tuning methods, with the most popular one being the low-rank approximation (LoRA) (Hu et al. 2021). As shown in Table 2, we can see that the WERs of the ASR models decrease by a significant margin after fine-tuning. However, similar to the results obtained from (Attia et al. 2024), we see that the generalization ability (as measured by the performance fine-tuning on ENNI and testing on the MyST dataset) decreases after fine-tuning.

Training Directly from Children Speech

Recently, some researchers have taken a novel path different from fine-tuning. Similar to Whisper (Bain et al. 2023), they hypothesize that pre-training directly from children's speech corpora will provide more benefits and bring more accurate and robust ASR models to the community. (Li, Hasegawa-

Model	Librispeech	ENNI	MyST
Wav2vec2-conformer	1.81	40.99	22.16
Wav2vec2	2.64	43.89	25.65
Whisper-small	3.30	16.34	17.57
Nemo - STT	1.67	32.41	16.64

Table 1: Performance (WER in percentage) of some recent ASR models on some representative speech datasets.

Model	ENNI	MyST
Whisper-small	16.34	**17.57**
Whisper-small (FT)	**7.04**	29.50
Whisper-medium	17.19	**16.67**
Whisper-medium (FT)	**4.88**	20.31
Wav2vec2 Large	43.89	**25.65**
Wav2vec2 Large (FT)	**24.73**	34.30

Table 2: WER% of some recent ASR models after fine-tuning (FT) on the ENNI and MyST speech dataset.

Johnson, and McElwain 2023) has shown promising results in improving speaker diarization and vocalization classification accuracy by pretraining the wav2vec2 model on massive children's home audio. A work-in-progress model that is directly trained from massive children's speech also preliminarily shows high accuracy and robustness in children's speech (Zheng and Hasegawa-Johnson 2024).

Automatic Screening

Traditional Approach of SLA

The conventional approach to SLA, characterized by manual assessments conducted by SLPs, has been a cornerstone in understanding linguistic capabilities. However, the inherent limitations of this approach include time-intensive procedures and resource demands, which delay the service for many children in need. In particular, as illustrated by the generic pipeline in Figure 1, after SLPs obtain speech samples from a child seeking SLA, they need to manually transcribe and annotate the speech samples. It is reported that for every one minute of speech sample, it takes seven to eight minutes for an experienced SLP to convert it to Systematic Analysis of Language Transcripts (SALT) format (Miller, Andriacchi, and Nockerts 2016). Depending on the comprehensiveness of tests issued by SLP, the speech sample duration varies from 7-8 minutes for ENNI (Schneider, Dubé, and Hayward 2005) to 30-45 minutes for core language assessments in CELF-5 (Pearson 2013). The extensive transcription and annotation time introduced in this stage hinders efficient testing for more children, but they could be potentially automated by deep learning models and AI-powered frameworks. Also, after proper annotations of the speech samples, SLPs will need to follow the manual and give out a score or a report on the child's speech ability. The artifact varies across different tests, but some of them are still time-consuming. For example, after the CELF-5 test (Pearson 2013), the SLP will issue a detailed report on the child's performance. While a template has already automated some parts of the report, it still requires some effort.

End-to-End System

Currently, several end-to-end (E2E) evaluation systems have been proposed and implemented to evaluate children's speaking ability automatically. Various frameworks have been proposed to automatically classify children into typical development (TD) and speech-language impairment (SLI) groups, or similar variations. From the early statistical tests like (Gong et al. 2016) to the latest transformer-based models (Johnson et al. 2023; Qin et al. 2024a), the claimed accuracy for identifying children with potential SLI problems has been raised to greater than 95%. Similarly, there have been multiple consumer-level software from some companies that try to achieve similar goals, such as Ambiki (Dias 2023) and Smart Ears (Smart 2023). While all aforementioned works and software show promising potential in automating the SLA and providing accessible remote services, they all either rely on black-box models or lack scientific explanations that hinder the interpretation of the results.

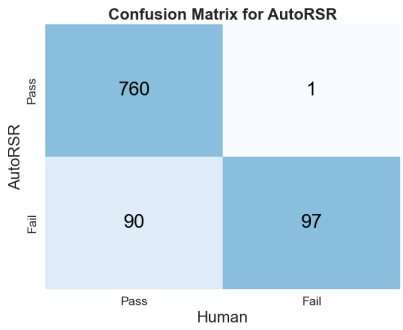

Figure 2: The confusion matrix on AutoRSR vs Human. Pass/Fail indicates whether the child passes/fails the RSR test. Cases when AutoRSR passes and Human fails the child are false negatives.

Explanable System Derived from SLAs

SLA for children is a sensitive area that requires careful attention and strong explanability. The cost of a false negative (that fails to screen a child with speech disorders) is much higher than typical deep-learning tasks. Because of that, frameworks that follow the guidance of SLPs and use tests developed by SLPs are favored over black-box models. As mentioned earlier, SLPs have come up with many forms of SLA tests, each looking at different aspects of children's speech, but all of the tests require extensive human labor, hindering their usage for most people.

We are working towards automating those tests and alleviating the workload of SLPs. Here, we report our progress and provide some preliminary results. In particular, we design and implement the automated Redmond Sentence Recall test (AutoRSR) (Redmond et al. 2019). The AutoRSR follows the same process as the manual version, where it transcribes the children's speech sample with WhisperX (Bain et al. 2023), segments the sample, aligns the transcription with the prompts using BertAlign (Liu and Zhu 2022), applies a modified Levenshtein distance algorithm to count the errors and scores the children's speech sample.

We evaluate the performance of the AutoRSR software on the 948 samples with children's ages provided by the original authors of RSR, and the results are summarized in the confusion matrix in Figure 2. On the available samples, the AutoRSR achieved 90.4% accuracy with only 1 false negative. Although the accuracy is high, we also observed that the Whisper model scores are harsher than those of SLPs. On average, Whisper scores are three points lower than human scores, and we provide a detailed distribution of score differences in Figure 3. After consulting SLPs, we conclude this difference to be attributed to two factors: 1. Whisper sometimes makes mistakes and transcribes incorrectly; 2. SLPs transcribe with more context, causing them to guess some words when the child is pronouncing unclearly. Both factors would be alleviated by fine-tuning the model with RSR data (but as discussed above, at a potential cost of generalization).

However, despite the success of the automated RSR framework, it is noteworthy that generic automation of SLP-

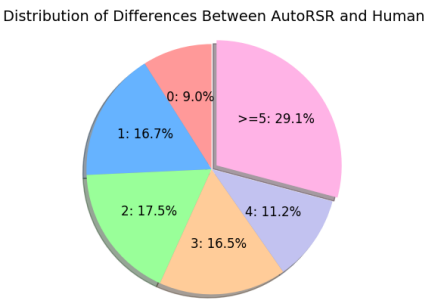

Figure 3: The distribution of differences between the automated RSR and human scorers. We can see that there is still a subtle difference between the scores even though the final classification is mostly correct.

guided tests still faces numerous challenges. The biggest challenge among them is the need for an accurate phonetic ASR model that captures phoneme-level mistakes that children tend to make. While the RSR test does not require careful annotations on those phonetic mistakes, other tests such as the Edmonton Narrative Norms Instrument (ENNI) test require SALT annotations and emphasize the correct language regularization (Schneider, Dubé, and Hayward 2005). As a result, current state-of-the-art Whisper-based models are no longer suitable for automating the ENNI test, because Whisper often infers words, leading to an auto-correct regularization that deviates from the intended purpose.

Report Generation

The final step in some of the tests such as CELF (Pearson 2013) requires the SLP to issue a report. As mentioned earlier, such a process is very time-consuming with little expertise required. Given the development of LLMs, we believe that they could help make this process fully automated. While there have not been works investigating the potential of LLMs generating SLP reports, similar works have been successful in asking LLMs to generate financial reports (Zhao et al. 2024), extracted clinical notes (Li et al. 2024), and SOAP notes (Hu et al. 2021). We believe that SLP reports could also be achieved in similar ways with an adequate knowledge base and appropriate usage of RAG.

Practical Concerns on Availability and Privacy of Automated SLA

In this section, we discuss some of the practical concerns that limit the usage of automated SLAs under resource-constrained settings. Those concerns and their solutions are particularly important for the deployment of AI-powered SLAs in less developed regions.

Availability of DL Models due to Privacy Concerns

Children's speech is highly sensitive data that requires a high level of privacy. Due to the uncontrollable nature of cloud services (Qin et al. 2024b) and the unacceptable cost of secure computation (Knott et al. 2021; Xu et al. 2024a), the most favorable method for the automated SLA is to deploy ASR models and other necessary models onto edge devices. However, edge ASR presents a unique set of challenges that need to be addressed, including the RAM limitations, the tradeoff between model quantization and performance, and computation overhead and delays, all of which will affect the practical performance of the automated SLA.

Numerous works have tried to address different individual challenges. For example, PI-Whisper provides fine-grained fine-tuning and inference via individualized profiling according to additional characteristics of the speakers and archives high accuracy using smaller models (Nassereldine et al. 2024). Whisper-KDQ provides a working implementation of a distilled and quantized Whisper model that fits into edge devices while even increasing the accuracy (Shao et al. 2023). Transformer accelerations that reduce inference time, such as flash attention (Dao et al. 2022), ONNX runtime with process-in-memory (Kim et al. 2022), and even optimized edge transformer architectures (Bergen, O'Donnell, and Bahdanau 2021; Xu et al. 2024b), could also help the automated SLA on the edge. However, these works do not integrate with each other, and a unified fast, accurate, and low-resource ASR model that serves the need is yet to be explored.

Fairness towards Marginalized Group

Moreover, the fairness of ASR models and the SLA framework needs to be researched, especially when facing marginalized groups. While the bias of ASR models has been researched for adults, in particular, African American accents (Martin and Wright 2022; Mengesha et al. 2021; Johnson et al. 2024), there has been little research involving the speech of children. At the same time, how the fine-tuning process affects each group of speakers is seldomly studied, with only PI-Whisper showing some results, indicating that profile-based fine-tuning helps the fairness of ASR (Nassereldine et al. 2024).

Conclusion

In this paper, we have explored the intricacies of ASR techniques tailored for children, discussed the development and implementation of an automated speech-language assessment pipeline, and addressed the critical privacy concerns associated with edge ASR technology. Our high-level discussions and practical implementations highlight the promising advancements in these areas and their potential to benefit children significantly.

We have demonstrated that ASR systems can be effectively adapted to account for the unique speech patterns and linguistic behaviors of children **on word-level**, but the **generalizability** of children ASR models and **phonetic** ASR models remain an open challenge in the area. Similarly, with the current ASR models, we can prototype some (simple) automated speech-language assessments, but more complex SLAs require stronger ASR models to achieve desirable practicality. Furthermore, the deployment of private ASR models on the edge deserves more research focus, particularly on the direction of integrating multiple optimization methods and fine-tuning more fair and accessible models.

Acknowledgements

This material is based upon work supported under the AI Research Institutes program by National Science Foundation and the Institute of Education Sciences, U.S. Department of Education through Award # 2229873 - National AI Institute for Exceptional Education. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education. The work is also supported in part by the NSF Grants # 2235364 and # 2329704.

References

Ardila, R.; Branson, M.; Davis, K.; Kohler, M.; Meyer, J.; Henretty, M.; Morais, R.; Saunders, L.; Tyers, F.; and Weber, G. 2020. Common Voice: A Massively-Multilingual Speech Corpus. In Calzolari, N.; Béchet, F.; Blache, P.; Choukri, K.; Cieri, C.; Declerck, T.; Goggi, S.; Isahara, H.; Maegaard, B.; Mariani, J.; Mazo, H.; Moreno, A.; Odijk, J.; and Piperidis, S., eds., *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 4218–4222. Marseille, France: European Language Resources Association. ISBN 979-10-95546-34-4.

Attia, A. A.; Liu, J.; Ai, W.; Demszky, D.; and Espy-Wilson, C. 2024. Kid-Whisper: Towards Bridging the Performance Gap in Automatic Speech Recognition for Children VS. Adults. arXiv:2309.07927.

Baevski, A.; Zhou, Y.; Mohamed, A.; and Auli, M. 2020. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 12449–12460. Curran Associates, Inc.

Bain, M.; Huh, J.; Han, T.; and Zisserman, A. 2023. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. *INTERSPEECH 2023*.

Bashir, A. S.; and Scavuzzo, A. 1992. Children with language disorders: natural history and academic success. *Journal of Learning Disabilities*, 25(1): 53–65.

BEITCHMAN, J. H.; WILSON, B.; JOHNSON, C. J.; ATKINSON, L.; YOUNG, A.; ADLAF, E.; ESCOBAR, M.; and DOUGLAS, L. 2001. Fourteen-year follow-up of speech/language-impaired and control children: Psychiatric outcome. *Journal of the American Academy of Child; Adolescent Psychiatry*, 40(1): 75–82.

Bergen, L.; O'Donnell, T. J.; and Bahdanau, D. 2021. Systematic Generalization with Edge Transformers. In Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*.

Bhardwaj, V.; Ben Othman, M. T.; Kukreja, V.; Belkhier, Y.; Bajaj, M.; Goud, B. S.; Rehman, A. U.; Shafiq, M.; and Hamam, H. 2022. Automatic Speech Recognition (ASR) Systems for Children: A Systematic Literature Review. *Applied Sciences*, 12(9).

Coster, F.; Goorhuis-Brouwer, S.; Nakken, H.; and Spelberg, H. L. 1999. Specific language impairments and behavioural problems. *Folia Phoniatrica et Logopaedica*, 51(3): 99–107.

Dao, T.; Fu, D. Y.; Ermon, S.; Rudra, A.; and Ré, C. 2022. FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. arXiv:2205.14135.

Dias, K. 2023. Ai-generated speech therapy soap notes.

Gong, J. J.; Gong, M.; Levy-Lambert, D.; Green, J. R.; Hogan, T. P.; and Guttag, J. V. 2016. Towards an Automated Screening Tool for Developmental Speech and Language Impairments. In *Proc. Interspeech 2016*, 112–116.

Gulati, A.; Qin, J.; Chiu, C.-C.; Parmar, N.; Zhang, Y.; Yu, J.; Han, W.; Wang, S.; Zhang, Z.; Wu, Y.; and Pang, R. 2020. Conformer: Convolution-augmented Transformer for Speech Recognition. arXiv:2005.08100.

Harper, E.; Majumdar, S.; Kuchaiev, O.; Jason, L.; Zhang, Y.; Bakhturina, E.; Noroozi, V.; Subramanian, S.; Nithin, K.; Jocelyn, H.; Jia, F.; Balam, J.; Yang, X.; Livne, M.; Dong, Y.; Naren, S.; and Ginsburg, B. 2024. NeMo: a toolkit for Conversational AI and Large Language Models.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2021. LoRA: Low-Rank Adaptation of Large Language Models. arXiv:2106.09685.

Jerome, A. C.; Fujiki, M.; Brinton, B.; and James, S. L. 2002. Self-esteem in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 45(4): 700–714.

Johnson, A.; Shankar, N. B.; Ostendorf, M.; and Alwan, A. 2024. An exploratory study on dialect density estimation for children and adult's African American Englisha). *The Journal of the Acoustical Society of America*, 155(4): 2836–2848.

Johnson, A.; Veeramani, H.; Balaji Shankar, N.; and Alwan, A. 2023. An Equitable Framework for Automatically Assessing Children's Oral Narrative Language Abilities. In *Proc. INTERSPEECH 2023*, 4608–4612.

Khan, T.; Freeman, R.; and Druet, A. 2023. Louder Than Words: Pediatric Speech Disorders Skyrocket Throughout Pandemic. Komodo Health:https://www.komodohealth.com/hubfs/2023/Speech_Pathology_Research_Brief.pdf.

Kim, S. Y.; Lee, J.; Kim, C. H.; Lee, W. J.; and Kim, S. W. 2022. Extending the ONNX Runtime Framework for the Processing-in-Memory Execution. In *2022 International Conference on Electronics, Information, and Communication (ICEIC)*, 1–4.

Knott, B.; Venkataraman, S.; Hannun, A.; Sengupta, S.; Ibrahim, M.; and van der Maaten, L. 2021. CrypTen: Secure Multi-Party Computation Meets Machine Learning. In *arXiv 2109.00984*.

Li, D.; Kadav, A.; Gao, A.; Li, R.; and Bourgon, R. 2024. Automated Clinical Data Extraction with Knowledge Conditioned LLMs. arXiv:2406.18027.

Li, J.; Hasegawa-Johnson, M.; and McElwain, N. L. 2023. Towards Robust Family-Infant Audio Analysis Based on Unsupervised Pretraining of Wav2vec 2.0 on Large-Scale Unlabeled Family Audio. In *INTERSPEECH 2023*. ISCA.

Liu, D.; and Xiong, J. 2024. FASA: a Flexible and Automatic Speech Aligner for Extracting High-quality Aligned Children Speech Data. arXiv:2406.17926.

Liu, L.; and Zhu, M. 2022. Bertalign: Improved word embedding-based sentence alignment for Chinese–English parallel corpora of literary texts. *Digital Scholarship in the Humanities*, 38(2): 621–634.

Martin, J. L.; and Wright, K. E. 2022. Bias in Automatic Speech Recognition: The Case of African American Language. *Applied Linguistics*, 44(4): 613–630.

Mengesha, Z.; Heldreth, C.; Lahav, M.; Sublewski, J.; and Tuennerman, E. 2021. "I don't think these devices are very culturally sensitive."-impact of automated speech recognition errors on African Americans.

Miller, J. F.; Andriacchi, K.; and Nockerts, A. 2016. Using language sample analysis to assess spoken language production in adolescents. *Language, Speech, and Hearing Services in Schools*, 47(2): 99–112.

Nassereldine, A.; Liu, D.; Xu, C.; and Xiong, J. 2024. PI-Whisper: An Adaptive and Incremental ASR Framework for Diverse and Evolving Speaker Characteristics. arXiv:2406.15668.

Panayotov, V.; Chen, G.; Povey, D.; and Khudanpur, S. 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.

Pearson. 2013. Clinical Evaluation of Language Fundamentals — Fifth Edition CELF-5.

Pradhan, S. S.; Cole, R. A.; and Ward, W. H. 2023. My Science Tutor (MyST) – A Large Corpus of Children's Conversational Speech. arXiv:2309.13347.

Qin, R.; Cook, R.; Yang, K.; Abbasi, A.; Dobolyi, D.; Seyedi, S.; Griner, E.; Kwon, H.; Cotes, R.; Jiang, Z.; and Clifford, G. 2024a. Language Models for Online Depression Detection: A Review and Benchmark Analysis on Remote Interviews. *ACM Trans. Manage. Inf. Syst.* Just Accepted.

Qin, R.; Liu, D.; Yan, Z.; Tan, Z.; Pan, Z.; Jia, Z.; Jiang, M.; Abbasi, A.; Xiong, J.; and Shi, Y. 2024b. Empirical Guidelines for Deploying LLMs onto Resource-constrained Edge Devices. arXiv:2406.03777.

Radford, A.; Kim, J. W.; Xu, T.; Brockman, G.; McLeavey, C.; and Sutskever, I. 2022. Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356.

Redmond, S. M.; Ash, A. C.; Christopulos, T. T.; and Pfaff, T. 2019. Diagnostic accuracy of sentence recall and past tense measures for identifying children's language impairments.

Redmond, S. M.; and Rice, M. L. 1998. The socioemotional behaviors of children with SLI: Social Adaptation or Social Deviance? *Journal of Speech, Language, and Hearing Research*, 41(3): 688–700.

Redmond, S. M.; and Rice, M. L. 2002. Stability of behavioral ratings of children with SLI. *Journal of Speech, Language, and Hearing Research*, 45(1): 190–201.

Schneider, P.; Dubé, R. V.; and Hayward, D. 2005. The Edmonton Narrative Norms Instrument. University of Alberta Faculty of Rehabilitation Medicine website:http://www.rehabmed.ualberta.ca/spa/enni.

Shao, H.; Wang, W.; Liu, B.; Gong, X.; Wang, H.; and Qian, Y. 2023. Whisper-KDQ: A Lightweight Whisper via Guided Knowledge Distillation and Quantization for Efficient ASR. arXiv:2305.10788.

Smart, E. 2023. Revolutionizing Spanish Articulation Therapy: How Smarty Ears Online Empowers Spanish-Speaking SLPs And Students.

Tomblin, J. B.; Zhang, X.; Buckwalter, P.; and Catts, H. 2000. The Association of Reading Disability, Behavioral Disorders, and language impairment among second-grade children. *Journal of Child Psychology and Psychiatry*, 41(4): 473–482.

Xu, C.; Yu, F.; Xu, Z.; Inkawhich, N.; and Chen, X. 2024a. Out-of-Distribution Detection via Deep Multi-Comprehension Ensemble. arXiv:2403.16260.

Xu, C.; Yu, F.; Xu, Z.; Liu, C.; Xiong, J.; and Chen, X. 2024b. QuadraNet: Improving High-Order Neural Interaction Efficiency with Hardware-Aware Quadratic Neural Networks. In *2024 29th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 19–25.

Young, A. R.; Beitchman, J. H.; Johnson, C.; Douglas, L.; Atkinson, L.; Escobar, M.; and Wilson, B. 2002. Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry*, 43(5): 635–645.

Zhao, H.; Liu, Z.; Wu, Z.; Li, Y.; Yang, T.; Shu, P.; Xu, S.; Dai, H.; Zhao, L.; Mai, G.; Liu, N.; and Liu, T. 2024. Revolutionizing Finance with LLMs: An Overview of Applications and Insights. arXiv:2401.11641.

Zheng, H.; and Hasegawa-Johnson, M. 2024. Combining self-supervised and self-training for child's automatic speech recognition. Restricted Internal Document.