Spoken Grammar Assessment Using LLM

***Abstract*—**
***Index Terms*—**

I. I. INTRODUCTION

II. ARXIV:2410.01579V1 [CS.CL] 2 OCT 2024

III. THE STUDY [12] COMPARED A CASCADED SYSTEM WITH SEPARATE MODULES FOR ASR DISFLUENCY REMOVAL, AND GRAMMAR ERROR CORRECTION, TO AN END-TO-END SYSTEM AND DEMONSTRATED THAT THE PERFORMANCE OF THE LATTER SYSTEM WAS COMPARABLE TO THAT OF THE FORMER. WITH CURRENT ADVANCEMENTS IN ASR TECHNOLOGY, OFTEN IT CAN BE BELIEVED THAT THESE SYSTEMS COULD CAPTURE SPOKEN GRAMMATICAL ERRORS IN THE DECODED TEXT. HOWEVER, THESE SYSTEMS HAVE AN INHERENT BIAS FROM THE LANGUAGE MODEL (LM) TOWARDS THE GRAMMATICALLY CORRECT TEXT. THE STUDY [13] FOUND THAT A DEEP LEARNING-BASED GRAMMATICAL ERROR DETECTION (GED) SYSTEM, FINE-TUNED ON FREE SPEECH DATA, IMPROVED PERFORMANCE ON NON-NATIVE SPOKEN ENGLISH. HOWEVER, CHALLENGES IN ASR AND DISFLUENCY DETECTION LIMITED ACCURATE FEEDBACK. THE WORK [14] EVALUATED THE IMPACT OF ASR ERRORS ON GED USING A DEEP LEARNING-BASED SYSTEM ORIGINALLY TRAINED ON WRITTEN TEXT. ASR CONFIDENCE SCORES WERE INTEGRATED INTO THE GED SYSTEM TO ADDRESS THE GRAMMATICAL ERRORS STEMMING FROM INCORRECT TRANSCRIPTIONS RATHER THAN LEARNER MISTAKES. IN [15], THE AUTHORS EVALUATED ASRBASED METHODS FOR SPOKEN GED, FINDING THAT A SCORE-BASED CLASSIFICATION OUTPERFORMS THE CASCADED APPROACH. THEY ALSO FOUND THAT LM AND N-BEST HYPOTHESES HAD MINIMAL IMPACT ON DECODING-BASED LIKELIHOOD CLASSIFICATION. THE ABOVE TWO STUDIES HIGHLIGHT THE ISSUE WITH THE CURRENT SPOKEN GED SYSTEMS THAT USE SOTA ASR AND THE NEED FOR A SYSTEM USING CUSTOM-BUILT LMS.

In this paper, we introduce an end-to-end SLA system to enable GED or assessment of language grammar from spoken speech. Further, the use of a large language model (LLM) makes the SLA system scalable and practical because no two assessment instances are the same; ensuring that the student cannot be coached for the assessment. The main contribution of the paper is (a) designing a SLA system that can robustly evaluate all aspects of language proficiency, without employing additional WLA tools, thereby significantly reducing the time taken to take the test, (b) proposing a mechanism to incorporate language grammar assessment by exploiting the superior performance of available speech analysis tools on read speech, (c) automatic grammar assessment using a custom-built LM on top of a readily available hybrid ASR system, (d)

IV. The demand for second language (L2) learners to study foreign languages, especially English, leads to the imminent need for the development of language proficiency assessment systems or tools [1], [2]. While several English language assessment tools exist, the assessments are often lengthy because they have separate assessment modules to assess different aspects of language proficiency. The spoken language proficiency assessment is often restricted to assessing the speech articulation of the speaker in terms of pronunciation [3]–[5] and speech delivery in terms of oral fluency [6], [7], which includes speaking rate [8], [9], recognition of pauses, filler words, and analysis of intonation [10] etc. The other important aspects of language like grammar or vocabulary are assessed separately through a written language proficiency as- sessment. Spoken language assessment (SLA) and written lan- guage assessment (WLA) complement each other, providing a comprehensive evaluation of overall language proficiency. Separate SLA and WLA assessments not only extend testing time but may also encourage learners to neglect grammar. In practical settings like call centers and virtual interviews, spoken language communication is important. This highlights the need for a comprehensive SLA system that assesses all aspects of language proficiency. The primary obstacle to integrating grammar assessment into current SLA systems is the limited availability or accuracy of speech analysis tools. Accurate grammar assessment requires precise identification of spoken words by ASR engines, which can be challenging

V. proposing a grammar scoring module that is robust to errors in ASR, and (e) employing LLM to to bring in variations in the test to make the SLA system largely unteachable thus making it scalable and practical. The rest of the paper is organized as follows, we describe the spoken language grammar assessment system in detail in Section II. We conduct experiments in Section III to show the process of automatic generation of paragraphs that can be used in grammar evaluation and show the need for a custom-built LM for speech transcription and we conclude in Section IV.

VI. (a) Block Diagram

VII. II. SPOKEN LANGUAGE GRAMMAR ASSESSMENT

VIII. (b) Functional System (Web Application).

Fig. 1: End to End System for SLA. We only look at the grammar of spoken language. The block diagram of the end-

to-end SLA system is shown in Fig. 1a. It has two parts, the first part, allows for the generation of a paragraph P (example Fig. 4(a)) by prompting an LLM, and the second part takes the audio S(t), spoken by the candidate, corresponding to Pd (example, 4(b)) and assesses for language grammar using Pg (Fig. 4(c)). Unlike traditional SLA systems which take an audio input S(t) and use the output Ps of a standard ASR to automatically compute the pronunciation or oral fluency [16], [17] only, in this paper, we enable grammar assessment on spoken speech. The grammar scoring acts on the output of the ASR, namely, Ps and the gold truth Pg (details mentioned later). This is done by displaying the paragraph Pd generated by an LLM using prompt engineering. We would like to emphasize that we do not focus on oral fluency and pronunciation (red dotted lines in Fig. 1a) which is common in SLA systems in this paper. Further, we do not delve into literature to focus on the proposed SLA system; an implementation is shown in Fig. 1b.

IX. A. Generation of Paragraph

For ¡grammar¿¡correct¿a¡/correct¿/an/the¡/grammar¿ student, ¡grammar¿study/ studied/¡correct¿studying¡/correct¿¡/grammar¿ poetry can be a roller coaster ride. ¡snip¿ can be both vexing and ¡grammar¿¡correct¿demotivating¡/correct¿/motivating/enthusing ¡/grammar¿. (a) A paragraph generated by prompting a LLM (P). A sample P generated by prompting a LLM [18] is shown in Fig. 2a, 5a and 5b. The tags "¡grammar¿ ¡/grammar¿" correspond to the words or phrases that are to be evaluated for grammar. The tag "¡correct¿ ¡/correct¿" shows the correct choice. The correct choice of grammar usage is studying corresponding to study/studied/studying displayed to the student. In practice, both Pd (Fig. 2b) and Pg (Fig. 2c) can be obtained by a simple text parser applied on P (Fig. 2a).

X. B. Spoken Language Grammar Scoring

XI. For (a/an/the) student, (study/studied/studying) poetry can be a roller coaster ride. This journey (is punctuated/punctuates/punctuated) by moments of profound appreciation (with/for/from)simpler pieces and intermittent frustration with more complex works. Some poems (were/have been/are) just plain confusing and no amount of re- reading (seeming/seems/is seeming) to help decipher (the/an/a) in- tended meaning. The puzzlement (that/those/these) results from such (institutions/instances/instigations) can be both vexing and (demoti- vating/motivating/enthusing).

(b) Paragraph displayed to the student (Pd). The student is shown a paragraph Pd on a web interface (Fig. 1b) containing —Pd— words in language L. Of the —Pd— words, a small subset of words Gw (Pd, Pd) help determine the student's grammar proficiency. The student (s) is given time to familiarize themselves with Pd and then reads it into a microphone, generating the audio S(t). The SLA system performs grammar scoring in the following steps.

XII. For a student, studying poetry can be a roller coaster ride. This journey is punctuated by moments of profound appreciation for simpler pieces and intermittent frustration with more complex works. Some poems are just plain confusing and no amount of re-reading seems to help decipher the intended meaning. The puzzlement that results from such instances can be both vexing and demotivating.

XIII. (c) The grammatically correct paragraph (Pg).

Fig. 2: A sample P generated using an LLM along with Pd used to display and Pg used for grammar assessment. #1 Building a customized LM (CLM) specific to the paragraph P to enhance the performance of the ASR (ASR-CLM). Let Ps = ASR-CLM(S(t)) be the transcript of S(t). #2 Compute the grammar score (Ss g) For (a/an/the) student, (study/studied/studying) poetry can be a roller coaster ride. (a) Sample sentence displayed to the student a) While maintaining the sequence of the words in Pd and Ps, we create a set p1 = {w Pd — w /Ps} of words that are in Pd but not in Ps. b) Create p2 = {w Gw — w /p1}. c) The grammar score, Ss g = —p2— is the cardinality of the set p2. Note that p2 is a set of all the correctly spoken grammar words by the student. In effect, the SLA of grammar takes S(t), Pd, and Gw as input and produces a score Ss g. Namely, Ss g = G-SCORE(Ps, Pd, Gw) (1) 1) For a student, study poetry can be a roller coaster ride. 2) For an student, study poetry can be a roller coaster ride. 3) For the student, study poetry can be a roller coaster ride. 4) For a student, studied poetry can be a roller coaster ride. 5) For an student, studied poetry can be a roller coaster ride. 6) For the student, studied poetry can be a roller coaster ride. 7) For a student, studying poetry can be a roller coaster ride. 8) For an student, studying poetry can be a roller coaster ride. 9) For the student, studying poetry can be a roller coaster ride. (b) Sentences (correct in italics) expected from the student. Fig. 3: Sample sentence (a) and expected variations (b). where, Ps = ASR-CLM(S(t)). As an example, Gw = {a, studying, punctuated, for, are, seems, the, that, instances, demotivating} for the paragraph shown in Fig. 2 and —Pd— = 61.

XIV. C. Speech to Text (ASR)

The most crucial block is the ASR, which converts the spoken paragraph S(t) into text Ps (see Fig. 1a) because ASR outputs are erroneous [19] leading to an error in grammar assessment. Let *Ps be the true transcript (human transcribed) of S(t). Let s be the error due to ASR, generally captured as the word error rate [20] (WER) between Ps and *Ps, choosing the grammatically correct sentence instead of the spoken wrong sentence. However, a CLM [23] can, easily, be trained to include all possible variations (including the wrong ones) of the sentence to mitigate this. This is the reason for our belief

that an ASR with a custom-built LM (ASR-CLM) can be far more accurate than any state-of-the-art ASR with a general-purpose LM.

XV. s = WER(Ps, *Ps). (2)

Unless s = 0, the audio grammar assessment score Ss g would be different from the true grammar assessment score, *Ss g = G-SCORE(P * s , Pd, Gw). (3) The error in grammar scoring due to an error (s) in ASR is

XVI. III. EXPERIMENTAL ANALYSIS We first describe how to generate a unique assessment para- graph P for each student using ChatGPT. This ensures that the students cannot be coached for the assessment. Subsequently, we experiment with an instance of P to validate the use of an ASR engine equipped with a custom-built LM based on the generated paragraph, namely, ASR-CLM.

g = —Ss g *Ss g—. (4) A. Generating P using ChatGPT We adopt 1-shot learning prompting style for generating new paragraphs (P1, P2, · · ·) as described in Fig. 4. #1 User: """" P """" {Sample P in Fig. 2a.} Generate paragraphs like P. One ¡correct¿¡/correct¿ tag within ¡grammar¿ ¡/grammar¿ tags. Each ¡grammar¿ tag has three options separated by "/". #1 ChatGPT: Thank you for providing the specific format and instructions. The grammar choices are marked within ¡grammar¿, with the correct option indicated using ¡correct¿. #2 User: Generate a paragraph similar to the example shown. #2 ChatGPT: P1 {Generated paragraph (Fig. 5a)} #3 User: Generate use subject "learning physics is easy". #3 ChatGPT: P2 {Generated paragraph shown in Fig. 5b} Fig. 4: 1-shot learning prompting to generate new P.

XVII. We hypothesize that in addition to the way G-SCORE is deter- mined (1), the construction of CLM tightly coupled with the assessment paragraph P performs better than even the state-of-the-art ASR (we use whisper [21] in our experiments). This is due to the fact that a LM plays a significant role in improving the accuracy of an ASR engine. While whisper is trained on extremely large and varied sets of text data, they are likely to lack grammatically incorrect sentences. As an illustration (see Fig. 3) there are three possible options for both the preposition (a/an/the) and the verb (study/studied/studying). Hence, the total number of possible sentences using all options is nine. Most of these (eight of the nine) sentences will rarely occur, in any text databases since they are grammatically incorrect. Hence, text corpora used for training whisper will not include these sentences. Shallow fusion is the most popular approach to combine pre-trained ASR model and LM [22]. Shallow fusion can be expressed mathematically as:

A wide variety of Pn's can be generated using the prompt "Generate just the paragraph. With subject ¡subject¿." This allows for the generation of a completely new paragraph in the desired format; the sample generated P shown in Fig. 5a, and 5b. score(Ps—S(t)) = log (p(Ps—S(t))) + · log(p(Ps)) (5) B. ASR performance We used whisper speech recognition engine and a Kaldi- based ASR with a custom-built LM (ASR-CLM) for compar- ison. The acoustic model of the Kaldi ASR was trained on

XVIII. where Ps is the spoken paragraph, p(Ps—S(t)) is acoustic score, is a scaling factor and p(Ps) is LM score. If Ps is not present in the training text, then p(Ps) = 0, which will make score(Ps—S(t)) very small. This results in the ASR

In ¡grammar¿¡correct¿an¡/correct¿/a/the¡/grammar¿ bustling city, ¡grammar¿exploring/ explored/ ¡correct¿exploration¡/correct¿ ¡/grammar¿ can be an exciting adventure. ¡snip¿ The challenge ¡grammar¿¡correct¿that¡/correct¿/ those/these¡/grammar¿ comes from such ¡grammar¿adventures/¡correct¿explorations¡/correct¿/explorers ¡/grammar¿ can be both thrilling and ¡grammar¿eye-opening/¡correct¿exhausting¡/correct¿/ insightful¡/grammar¿. (a) A paragraph generated by prompting ChatGPT (P1). S(t) It was a late afternoon probably on the 15th of February 2019 my friend and I were walking on the footpath in central Bangalore whisper It was the late afternoon probably on the 15th of February 2019 my friend and I were walking on the footpath in central Bangalore ASR-CLM "It was a late afternoon probably on the 15th of February 2019 my friend and I were walking on the footpath in central

Bangalore". S(t) It am a late afternoon probably on the 15th of February 2019 my friend and I was walking on the footpath into central Bangalore whisper It am a early after noon probably on 15th February 2019 my friend and I was walking on the footpath in central Bangalore ASR-CLM It am a late afternoon probably on the 15th of February 2019 my friend and I was walking on the footpath into central Bangalore TABLE I: Sample S(t). ASR errors, marked in red. For ¡grammar¿¡correct¿an¡/correct¿/a/the¡/grammar¿ physics enthusiast, ¡grammar¿studying/ studied/ ¡correct¿studying¡/correct¿ ¡/grammar¿ physics can be a fascinating journey. ¡snip¿ The understanding ¡grammar¿¡correct¿that¡/correct¿/ those/ these¡/grammar¿ comes from such ¡grammar¿endeavors/¡correct¿pursuits¡/correct¿/ explorations¡/grammar¿ can be both empowering and ¡grammar¿¡correct¿rewarding¡/correct¿/ challenging/exciting¡/grammar¿. compared to g = 3 for a custom-built LM ASR (ASR-CLM). (b) A paragraph generated by prompting ChatGPT (P2). Fig. 5: Paragraph's generated by prompting ChatGPT. Grammar Assessment whisper ASR-CLM *Ss g Ss g(g) Ss g(g) #1 14 (1) 15 (0) 15 #2 11 (1) 11 (1) 10 #3 11 (2) 9 (0) 9 #4 12 (1) 13 (0) 13 #5 12 (1) 12 (1) 13 #6 10 (2) 12 (0) 12 #7 6 (2) 8 (0) 8 #8 15 (3) 12 (0) 12 #10 15 (1) 16 (0) 16 #11 3 (0) 3 (0) 3 #12 6 (2) 8 (0) 8 #13 10 (2) 12 (0) 12 #14 15 (1) 15 (1) 16 #15 14 (1) 15 (0) 15 #16 14 (0) 14 (0) 14 #17 13 (0) 13 (0) 13 Total (20) (3) - TABLE II: Use of whisper and ASR-CLM for grammar assessment. g computed using (4).

XIX. IV. CONCLUSIONS

XX. 960 HOURS OF SPEECH DATA FROM LIBRISPEECH DATABASE [24]. THE CUSTOM LM WAS TRAINED ON THE TEXT COMPRISING ALL POSSIBLE VARIATIONS OF THE GIVEN SENTENCES (EXAMPLE FIG. 3B). WE RECORDED SPEECH CORRESPONDING TO ALL VARIATIONS OF THE BELOW SENTENCE, "IT (WAS/IS/AM) A LATE AFTERNOON PROBABLY (ON/IN/OF) THE 15TH OF FEBRUARY, 2019. (I AND MY FRIEND/MY FRIEND AND I) (WAS/WERE/WILL BE) WALKING ON THE FOOTPATH (IN/INSIDE/INTO) CENTRAL BANGALORE." NAMELY, 3(WAS/IS/AM) × 3(ON/IN/OF) × 2(I AND MY FRIEND/MY FRIEND AND I) × 3(WAS/WERE/WILL BE) × 3(IN/INSIDE/INTO) = 162 UTTERANCES. WE FOUND THAT ASR-CLM WAS ABLE TO EXACTLY TRANSCRIBE THE UTTERANCE (EVEN WHEN THERE WAS AN ERROR IN GRAMMAR) WHILE WHISPER "CORRECTED" THE GRAMMATICAL ERROR. TABLE I SHOWS TWO EXAMPLES WHERE ASR- CLM ACCURATELY RECOGNIZES THE SPOKEN WORDS, REGARDLESS OF GRAMMATICAL CORRECTNESS, WHILE WHISPER FALLS SHORT. IN THE FIRST EXAMPLE (TABLE I) THE ARTICLE "A" WAS REPLACED BY "THE" BY WHISPER WHILE IN EXAMPLE TWO, THE ARTICLE "A" WAS NOT RECOGNIZED BY WHISPER. OVERALL, THE ABILITY OF ASR-CLM TO RECOGNIZE WHAT WAS SPOKEN IS 84.7% WHILE THAT OF WHISPER WAS 46%. THE PERFORMANCE WAS COMPUTED ON 137 UTTERANCES; 25 OF THE 162 UTTERANCES WERE DISCARDED BECAUSE OF NOISE. THE POOR ACCURACY OF THE SOTA ASR HIGHLIGHTS THE NEED FOR A CLM-ASR FOR THE PURPOSE OF SLA OF GRAMMAR.

To the best of our knowledge, a standard speech dataset for spoken grammar assessment with manual annotations of grammatical errors in conversational or read speech is currently unavailable. To evaluate our SLA system, we used an in-house dataset consisting of audio recordings from 17 students speaking a generated paragraph, which was manually assessed by a linguist to mark the grammar score (*Ss g). We used both whisper and ASR-CLM to convert the spoken paragraph to text and compute Ss g. The error in assessment is captured in parenthesis for each student in Table II. Larger grammar assessment errors (g = 20) due to whisper are observed Language proficiency assessment is a common requirement for L2 speakers of English. There exist several SLA tools to assess pronunciation and oral fluency but none of them venture into assessing language grammar, instead, they depend on WLA systems. We designed and implemented a practical, scalable and robust SLA system to assess grammar. The design, to display the paragraph with options, made sure the audio obtained for assessment had no spontaneous speech characteristics like filler words, or repetitions and resembled "read" speech thereby enhancing the ASR performance. Additionally, the use of a custom LM in ASR-CLM leads to improved ASR performance, resulting in robustness in grammar assessment. The use of LLM enables the generation of paragraphs that are largely non-repetitive thereby making the proposed system

hard to be memorized by students. We can observe that the grammar scoring mechanism, by design, is not affected by ASR mis-recognition of non Gw words. [1] L. Jin and H. Zhu, "Developing standardized speech and language assessment tools in Mandarin Chinese: A context for improving reading and writing," Journal of Chinese Writing Systems, vol. 7, no. 3, pp. 150– 160, 2023. [2] H. Franco, H. Bratt, R. Rossier, V. Rao Gadde, E. Shriberg, V. Abrash, and K. Precoda, "Eduspeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications," Lan- guage Testing, vol. 27, no. 3, pp. 401–418, 2010. [3] K. Sheoran, A. Bajgoti, R. Gupta, N. Jatana, G. Dhand, C. Gupta, P. Dadheech, U. Yahya, and N. Aneja, "Pronunciation Scoring With Goodness of Pronunciation and Dynamic Time Warping," IEEE Access, vol. 11, pp. 15485–15495, 2023. [4] H. Pei, H. Fang, X. Luo, and X. Xu, "Gradformer: A Framework for Multi-Aspect Multi-Granularity Pronunciation Assessment," IEEE ACM Trans. Audio Speech Lang. Process., vol. 32, pp. 554–563, 2024. [5] B. Lin and L. Wang, "Exploiting Information From Native Data for Non- Native Automatic Pronunciation Assessment," in 2022 IEEE Spoken Language Technology Workshop (SLT), pp. 708–714, 2023. [6] A. Preciado-Grijalva and R. F. Brena, "Speaker fluency level classification using machine learning techniques," arXiv preprint arXiv:1808.10556, 2018. [7] S. P. Dubagunta, E. Moneta, E. Theocharopoulos, and M. Magimai Doss, "Towards Automatic Prediction of Non-Expert Perceived Speech Flu- ency Ratings," in Companion Publication of the 2022 International Conference on Multimodal Interaction, pp. 7–11, 2022. [8] A. Imran, M. Pandharipande, and S. K. Kopparapu, "Speakrite: Monitor- ing speaking rate in real time on a mobile phone," International Journal of Mobile Human Computer Interaction (IJMHCI), vol. 5, no. 1, pp. 62– 69, 2013. [9] S. K. Kopparapu, Non-linguistic analysis of call center conversations. Springer, 2015. [10] J. P. Arias, N. B. Yoma, and H. Vivanco, "Automatic intonation assess- ment for computer aided language learning," Speech Communication, vol. 52, no. 3, pp. 254–267, 2010. [11] K. Mukherji, M. Pandharipande, and S. K. Kopparapu, "Improved Language Models for ASR using Written Language Text," in 2022 National Conference on Communications (NCC), pp. 362–366, 2022. [12] S. Bannò, R. Ma, M. Qian, K. M. Knill, and M. J. F. Gales, "Towards End-to-End Spoken Grammatical Error Correction," in ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 10791–10795, 2024. [13] K. Knill, M. Gales, P. Manakul, and A. Caines, "Automatic grammatical error detection of non-native spoken learner english," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8127–8131, 2019. [14] Y. Lu, M. J. F. Gales, K. Knill, P. Manakul, L. Wang, and Y. Wang, "Impact of ASR Performance on Spoken Grammatical Error Detection," in Interspeech, 2019. [15] C. Venkata Thirumala Kumar, M. Sirigiraju, R. Vaideeswaran, P. K. Ghosh, and C. Yarra, "Can the decoded text from automatic speech recognition effectively detect spoken grammar errors?," in 9th Workshop on Speech and Language Technology in Education (SLaTE), pp. 41–45, 2023. [16] A. Panda, R. Acharya, and S. K. Kopparapu, "Oral Fluency Classi- fication for Speech Assessment," in 31st European Signal Processing Conference, EUSIPCO 2023, Helsinki, Finland, September 4-8, 2023, pp. 231–235, IEEE, 2023. [17] L. Fontan, M. L. Coz, and S. Detey, "Automatically measuring L2 speech fluency without the need of ASR: a proof-of-concept study with Japanese learners of French," in INTERSPEECH, 2018. [18] OpenAI, "GPT-3.5: OpenAI's Generative Pre-trained Transformer 3.5." https://platform.openai.com, 2023. Accessed: 2023-06-26. [19] C. Anantaram, S. K. Kopparapu, C. Patel, and A. Mittal, "Repairing General-Purpose ASR Output to Improve Accuracy of Spoken Sentences in Specific Domains Using Artificial Development Approach," in Pro- ceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI'16, p. 4234–4235, AAAI Press, 2016. [20] J.-C. Junqua and J.-P. Haton, Robustness in Automatic Speech Recog- nition: Fundamentals and Applications. Kluwer Academic Publishers, 1996. [21] OpenAI, "Whisper: OpenAI's Automatic Speech Recognition (ASR) System." https://github.com/openai/whisper, 2022. Model: Whisper Tiny.en, Accessed: 2023-06-26. [22] R. Prabhavalkar, T. Hori, T. N. Sainath, R. Schlüter, and S. Watanabe, "End-to-End Speech Recognition: A Survey," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 325–351, 2024. [23] K. Heafield, "KenLM: Faster and Smaller Language Model Queries," in Proceedings of the Sixth Workshop on Statistical Machine Translation, (Edinburgh, Scotland, United Kingdom), pp. 187–197, Association for Computational Linguistics, 2011. [24] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5206–5210, 2015.

REFERENCES