AI-based Personalized Feedback in Speech Therapy for People with Aphasia

Eugenia Rykova
University of Eastern Finland, Finland,
Catholic University of Eichstätt-Ingolstadt, Germany
eugenryk@uef.fi

Paper Abstract

In the field of speech and language therapy, artificial intelligence has been used in diagnostics, therapy, and assistive systems for people with aphasia (PWA) (Adikari et al., 2023; Azevedo et al., 2024; Pottinger & Kearns, 2024). AphaDIGITAL project (TDG, 2021) focuses on developing such a mobile application for German-speaking PWA that will provide personalized multilevel feedback with the help of Automatic Speech Recognition (ASR) and further text analysis. To build the corresponding pipeline (Rykova & Walther, 2024), the following questions are addressed.

Which existing ASR solutions are suitable for the task-specific speech of German-speaking PWA?

More than 50 open-source ASR solutions were evaluated with the help of several speech recordings from different corpora (Rykova, Walther, & Zeuner, 2022). Thirteen models were selected and tested with atypical speech, including two small datasets of PWA's speech (Rykova & Walther, in press - a). Based on Character Error Rate (CER), HITS (the number of precisely recognized words) and the number of empty outputs, four open-source ASR models were selected for the pipeline (Fleck, 2022; Grosman, 2022; Guhr, 2022; NVIDIA, 2022). These models are to a greater or lesser extent robust to speaker gender and age. The experiments suggest that for better single-word recognition the audio samples should be not too short and pronounced neither too slowly nor too fast (i.e. intentionally speeded up) (Rykova & Walther, in press - b).

How can selected ASR solutions be improved and/or adapted for the purposes of speech and language therapy?

In the absence of adequate data for ASR models' (re)training, applying the knowledge about non-standard (aphasic and dialect) phonetic features post-hoc to ASR output was attempted. Aphasic features included recognition of syllables as separate words and vowel prolongation. Dialect features were selected from the Thuringia-Upper Saxon dialect group (Wallraff, 2007; Rocholl, 2015; B. Siebenhaar, personal communication, January, 2024). The method combined generating alternative pronunciations based on non-standard patterns (Masmoudi et al., 2014) and using alternatives for evaluation (Ali et al., 2017), and proved to work on the recordings of German aphasia test naming and repetition tasks (Huber, 1993).

How can a combination of selected ASR solutions and existing tools for semantic and grammatical analysis serve for speech production errors analysis?

If the answer of the speaker is not recognized as fully correct or containing phonetic/phonemic errors, it is subject to semantic analysis. It must be compared to the target in terms of their semantic relationship and distance. The current semantic analysis pipeline is built upon GermaNet – a semantic network for the German language (Hamp & Feldweg, 1997). It includes recognition of hyponymy/hypernymy, belonging to the same semantic (sub)category, and different lexical and conceptual relationships, derived from GermaNet. If the answer is not recognized as an existing word, a search for close orthographic matches is performed, and the match that is semantically the closest to the target is subject to the relationship analysis described above. This approach has been tested and described in detail in Rykova & Walther (2023).

Keywords: Aphasia; automatic speech recognition; speech and language therapy; digital health.

REFERENCES

Adikari, A., Hernandez, N., Alahakoon, D., Rose, M. L., & Pierce, J. E. (2024). From concept to practice: A scoping review of the application of AI to aphasia diagnosis and management. *Disability and Rehabilitation*, *46*(7), 1288-1297.

Ali, A., Nakov, P., Bell, P., & Renals, S. (2017, December). WERd: Using social text spelling variants for evaluating dialectal speech recognition. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 141-148). IEEE.

Azevedo, N., Kehayia, E., Jarema, G., Le Dorze, G., Beaujard, C., & Yvon, M. (2024). How artificial intelligence (AI) is used in aphasia rehabilitation: A scoping review. *Aphasiology*, *38*(2), 305-336.

Fleck, M. (2022). *Wav2vec2-large-xls-r-300m-german-with-lm.* Retrieved from https://huggingface.co/mfleck/wav2vec2-large-xls-r-300m-german-with-lm. Accessed September 12, 2022.

Grosman, J. (2022). *Fine-tuned XLSR-53 large model for speech recognition in German*. Retrieved from https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-german. Accessed September 12, 2022.

Guhr, O. (2022). *wav2vec2-large-xlsr-53-german-cv9.* Retrieved from https://huggingface.co/oliverguhr/wav2vec2-large-xlsr-53-german-cv9. Accessed September 12, 2022.

Hamp, B., & Feldweg, H. (1997). GermaNet – a lexical-semantic net for german. In *Automatic information extraction and building of lexical semantic resources for NLP applications*.

Huber, W. (1983). *Aachener Aphasie Test (AAT) [Aachen Aphasia Test]*. Verlag für Psychologie Hogrefe, Göttingen, Zürich.

Masmoudi, A., Khmekhem, M. E., Esteve, Y., Belguith, L. H., & Habash, N. (2014, May). A Corpus and Phonetic Dictionary for Tunisian Arabic Speech Recognition. In *LREC* (pp. 306-310).

NVIDIA. (2022). *NVIDIA Conformer-Transducer Large (de)*. Retrieved from https://huggingface.co/nvidia/stt_de_conformer_transducer_large. Accessed September 12, 2022.

Pottinger, G., & Kearns, Á. (2024). Big data and artificial intelligence in post-stroke aphasia: A mapping review. *Advances in Communication and Swallowing*, (Preprint), 1-15.

Rocholl M.J. (2015). *Ostmitteldeutsch – eine moderne Regionalsprache? Eine Untersuchung zu Konstanz und Wandel im thüringisch-obersächsischen Sprachraum [East-Central German – a modern regional language? An investigation into constancy and change in the Thuringian-Upper Saxon language area]*. Hildesheim, Zürich, New York: OLMS.

Rykova, E., & Walther, M. (2023, March). Concept for semantic error analysis in a mobile application for speech and language therapy support. In *Konferenz Elektronische Sprachsignalverarbeitung* (pp. 127-133). TUDpress, Dresden.

Rykova, E., & Walther, M. (2024). AphaDIGITAL – digital speech therapy solution for aphasia patients with automatic feedback provided by a virtual assistant. *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS).*

Rykova, E., & Walther, M. (in press - a). Evaluation of German ASR solutions for speech and language therapy support of people with aphasia. *LOQUENS.*

Rykova, E., & Walther, M. (in press - b). Linguistic and extralinguistic factors in automatic speech recognition of German atypical speech. *Proceedings of the 20th Conference on Natural Language Processing (KONVENS 2024)*.

Rykova, E., Walther, M., & Zeuner, E. (2022, August). *aphaDIGITAL — Avatar-based digital speech therapy solution for aphasia patients: first evaluation*. Poster session presentation at the 35th Fonetiikan Päivät, Joensuu, Finland. Available at https://www.researchgate.net/publication/364676432_aphaDIGITAL_-Avatar-based_digital_speech_therapy_solution_for_aphasia_patients_first_evaluation.

TDG — Translationsregion für digitale Gesundheits-versorgung [Translational Region for Digital Healthcare]. (2021). *AphaDIGITAL: Entwicklung einer digitalen, dezentralen sprachtherapeutischen Versorgung [Development of digital, decentralized speech therapy solutions]*. Retrieved from https://inno-tdg.de/projekte/aphadigital/. Accessed January 25, 2023.

Wallraff, U. (2007). *Ausgewählte phonetische Analysen zur Umgangssprache der Stadt Halle an der Saale [Selected phonetic analyses of the colloquial language of the city of Halle an der Saale]*. [Doctoral Dissertation, Martin-Luther-Universität Halle-Wittenberg, Halle, Germany].

Acknowledgements/Disclaimers

AphaDIGITAL project was sponsored by German Federal Ministry of Education and Research under funding code 03WIR3108A via the TDG innovation ecosystem (Translationsregion für digitale Gesundheitsversorgung [Translational region for digital

healthcare]) and „WIR! – Wandel durch Innovation in der Region” [Change through innovation in the region] program.