HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 4 (2024)

# Exploring the Application of the Whisper Model in Automatic Aphasia Speech Evaluation

Inpyo Lee[1] and Tracey Salerno[#]

[1]Northern Valley Regional High School at Demarest, USA
[#]Advisor

## ABSTRACT

Aphasia is an everyday communication and speech disorder that impairs the ability of an individual to express through writing and speech. This paper explores the potential of using automatic aphasia speech evaluation models like the Whisper model to evaluate aphasia and potentially other speech impairments. Though effective, traditional methods for aphasia assessment are time-consuming and require specialized clinical expertise. To address these challenges, the study fine-tunes the Whisper using the AphasiaBank dataset to create a more efficient and accessible evaluation tool. The first trial of finetuning focused on the phonemic transcript generation part of the Whisper model and achieved a low accuracy of 56.89%. Minor token prediction errors and word omissions were the major reasons the prediction accuracy was so low. The second trial focused on the model's prediction structure, included prompt and correction tokens, and showed improved accuracy by 70.76%. This indicates that contextual information and correctness tokens can significantly enhance the model's performance. Further research and training of this model should be done on the entire AphasiaBank dataset because only a sample was available for this paper. The results of this paper show that there is a potential for AI models such as the Whisper model to be alternative tools for Aphasia testing and evaluation. This would remedy the scarcity of SLPs and make assessment more accessible to all individuals struggling with communication disorders by deploying an app or tool.

## Introduction

A speech disorder, also known as a communication disorder, is a condition that affects an individual's ability to communicate effectively through speech, language, or voice by having trouble creating the speech sound for communication. At least 5% of U.S. children ages between 3 and 17 have experienced speech disorders (National Institute on Deafness and Other Communication Disorders, 2016). Speech disorder treatment can result in losing billions of dollars of potential production. For example, a teaching job costs nearly three billion dollars annually (Ilapakurti, 2019). Aphasia is a speech disorder that results from damage to the brain, typically caused by a stroke. It widely affects an individual's ability to understand and use language. At a minimum, two million people in the U.S. are experiencing Aphasia (National Aphasia Association, 2016). Aphasia exhibits significant variations based on demographic and socio-cultural factors and the lesion's location, severity, and size. This necessitates an evaluation to develop an effective rehabilitation plan. Aphasia affects an individual's personality and influences relationships with family and others, ultimately impacting their quality of life. A multidisciplinary team consisting of a doctor, physiotherapist, occupational therapist, and language and speech therapist is essential to improve the status of aphasic patients and the patient's social environment (Toğram, 2012). Assessment of aphasia involves a detailed evaluation of language skills in all areas. The assessment should identify areas of deficiency through standardized tests and observation, enabling the prediction of prognosis and planning of therapeutic objectives (Toğram, 2012).

Current Treatment Methods for Communication Disorders

Various standardized Aphasia assessments exist worldwide, including the Boston Diagnostic Aphasia Examination (BDAE), Western Aphasia Battery (WAB), and English Aachen Aphasia Test (EAAT). These assessments evaluate language abilities across various modalities, including comprehension, expression, repetition, naming, reading, and writing (Privitera, 2024). One of them, The Aachen Aphasia Test (AAT), is a standard diagnosis and classification assessment for Germany and worldwide. It evaluates language levels and provides information on syndrome classification and severity. It evaluates spontaneous speech gained by having semi-structured interviews. The interview transcript is evaluated by humans based on a standardized manual. However, there are many challenges. The ATT is time-consuming; one patient takes up to eight hours for data collection and analysis. The feedback also has limitations, as it does not reflect within-individual improvement (Kohlschein, 2007). In addition, there are translation problems due to the society's different linguistic features (Toğram, 2012).

More importantly, these assessments require speech and language pathologists (SLPs) to have clinical and background knowledge of linguistics and culture (Qin, 2016). Standardized tests such as the Western Aphasia Battery, Boston Diagnostic Aphasia Examination, and Comprehensive Aphasia Test evaluate language abilities. In contrast, more straightforward tests like the Token and Copenhagen Cross-Linguistic Naming Test can be used informally at the bedside or with non-native speakers. Neuroimaging techniques like MRI or CT scans are also used to identify underlying brain damage or lesions associated with aphasia. These methods help diagnose aphasia, determine severity and characteristics, and guide individualized treatment planning and intervention strategies.

Treatment options for aphasia focus on improving language abilities through exercises targeting speaking, listening, reading, and writing skills. Speech and language therapy is a basic form of aphasia treatment and may include repetition drills, communication strategies training, and computer-assisted therapy programs. Augmentative and alternative communication (AAC) devices like communication boards or speech-generating devices can aid individuals with severe aphasia in expressing themselves. It is effective as it reflects an individual's progress and evolving needs. Group therapy sessions and family and caregiver involvement are also crucial for supporting individuals with aphasia in their daily communication and rehabilitation efforts. Interdisciplinary collaboration and continuous monitoring and adjustment of treatment plans contribute to optimizing outcomes and enhancing the quality of life for individuals with aphasia (Privitera, 2024).

## Related Works: Automatic Aphasia Speech Evaluation

Speech and Language pathologists (SLP) are considered the most indemand clinicians due to job dissatisfaction and burnout. There is an alarming need for SLPs in schools and post-acute care. Due to the dramatic shortage of Speech and Language Pathologists, automatic speech evaluation has gained interest (Shahin 2020). Automatic speech assessment is significantly more effective than human speech disorder management. Automatic evaluation benefits patients can be done online at an affordable cost (Le & Emily Mower Provost, 2014). With the development of speech recognition models, the accuracy could be more accurate than that of humans (Deka, 2022). Previously, there were many attempts to achieve automatic speech evaluation. Many researches were based on machine learning.

Ilapakurti's study aims to detect pathological voices and classify three disordered categories from waveforms collected on mobile phones. The researchers compared the performance of different machine learning algorithms, including Neural Networks (CNN and RNN). They experimented with various loss functions, such as LeakyRelu and PRelu, and used features like MFCC and Mel-Spectrogram. The best-performing model was a 5-layer CNN trained with MFCC and Mel-Spectrogram, which achieved a sensitivity of 96% and specificity of 18% on test data (2019).

Qin et al. (2016) applied state-of-the-art automatic speech recognition (ASR) DNN-HMM and GMM-HMM to achieve automatic speech assessment for Cantonese-speaking aphasia patients that analyze linguistic and acoustic features. The focus is on attaining reliable time alignments at the word level to analyze supra-segmental duration characteristics and detect content words for linguistic feature analysis. They trained the automatic speech recognition (ASR) model with Cantonese AphasiaBank. The ASR system is evaluated using test speech from 33 aphasia speakers and eight unimpaired speakers in the Cantonese AphasiaBank. For their experiment ASR on aphasia speech, a syllable error rate (SER) of 42.7% was obtained for unimpaired speech, and the SER for aphasia speech was 57.8%, which is lower accuracy and reliability than the human error rate. For the ASR to Automatic Assessment, They achieved reasonable accuracy, with 34.1% syllable error rates (SER).

Bílková et al. focused on mouth, lip, teeth, and tongue movement extraction with convolutional neural networks (CNN) based detection (2020). For lip detection, they extract critical points by using the Dlib library. They were followed by analyzing the state and size of the mouth. For tongue features, they used a model based on the CNN U-net convolutional neural network that was trained on data of different quality and with people of various ages. For teeth detection, they extract features with a mask based on their segmentation and evaluation of the teeth's position and the gap between them. Their lip detection model can correct extract patterns, but tongue and teeth detection evaluation results are still in progress.

Jothi et al. experimented with neural network-based categorical models and machine learning-based models with various Ensemble Machine Learning classifiers, such as XGBoost, AdaBoost (SVM), AdaBoost, and Bagging and Neural Network Categorical to classify severity levels. They used MFCC (Mel-Frequency Cepstral Coefficient) audio for feature extraction and ANOVA F-value for feature selection. They classified as high-AQ,mild-AQ, and low-AQ using the model. They achieved an accuracy of 0.90 for Adaboost, 0.86 for AdaBoost(SVM), 0.89 for XGBoost, and 0.87 for Bagging Classifier. For the Deep Neutral Network model, they trained with 3 individual datasets. DNN with children speech corpus dataset, AphasiaBank Cantonese, and Speech AphasiaBank EnglishSpeech dataset. They achieved accuracy 83%, 39.35%, and 87.75, respectively(2021)

Using deep learning models, such as 5-layer plain network, 5-layer CNN, and RNN, Ilapakurti et al. created a mobile application for detecting speech disorders. Neural Networks with a 5-layer architecture resulted in 54% after 200000 epochs. 5-layer CNN trained with MFCC and Mel-Spectrogram, which had a sensitivity of 96% and specificity of 18% on the test data. Recurrent Neural Network (RNN) resulted in 97.7% accuracy. The 5-Layer CNN is deployed into the mobile application by freezing the model and converting it into TFLite. (2019).

At least since 2009, there has been an attempt to automatically evaluate therapy's speech using machine learning (Saz 2009). These attempts were based on the Hidden Markov Model (HMM) and/or Gaussian Mixture Model (GMM) with variations of an overall model structure and datasets. Le et al. (2014), Qin et al. (2016), and Kohlschein et al. (2017) use this model in their research. With the introduction of Convolutional Neural Networks (CNN), many researchers used deep learning models without feature extraction. Jothi's and Bílková's words applied CNN to their model.

## Methods

The use of modern machine learning technology has highly affected performance. Many papers use techniques and technology already developed and applied to speech disorder evaluation, so papers use the most recent technology in their research. This study applies the State-of-the-Art ASR model Whisper model to achieve an automatic aphasia classification model. The Whisper achieved robust, high-accuracy speech recognition via large-scale speech data. The research conducted by Jain et al. improved child speech ASR performance, suggesting that adapting the Whisper model could result in better performance in evaluating Aphasia patients (2023). This study uses the LoRA method (Hu, 2021) for efficient fine-tuning in limited hardware resource conditions.

### Overview of the Whisper Model

The Whisper model is a multilingual and multitask speech recognition model that supports more languages than

HIGH SCHOOL EDITION
Journal of Student Research



English. It performs multiple tasks in one model, such as transcription, translation, voice activity detection, alignment, and language identification. The model uses a transformer-based architecture with residual connections and layer normalization to process audio with a combination of convolutional layers, sinusoidal positional encoding, and transformer blocks to process the audio. This model is trained with a dataset of 680,000 hours of labeled audio, including 117,000 hours in 96 languages other than English and 125,000 hours of translation data from other languages to English. (Hu, 2022)

Hu et al. suppose that specifying the task and conditioning information is required to achieve a single model that can be trained to perform the entire speech-processing pipeline. They have done this by formatting all tasks and conditioning information as a sequence of prediction tokens to the decoder. For example, to specify a task, the model can be trained to predict a sequence of tokens such as <|startoftranscript|>, followed by a language token (e.g., <|en|> for English), then a task token (e.g., <|transcribe|>), and finally the output transcript. Additionally, the model can be trained to predict timestamps by including tokens such as <|starttime:10ms|> and <|endtime:30ms|> to specify each caption's start and end times. These tokens are converted by Byte-Pair Encoding (BPE) text tokenizer, which is also used in GPT2 for English, but Whisper changes the vocabulary for the multilingual.

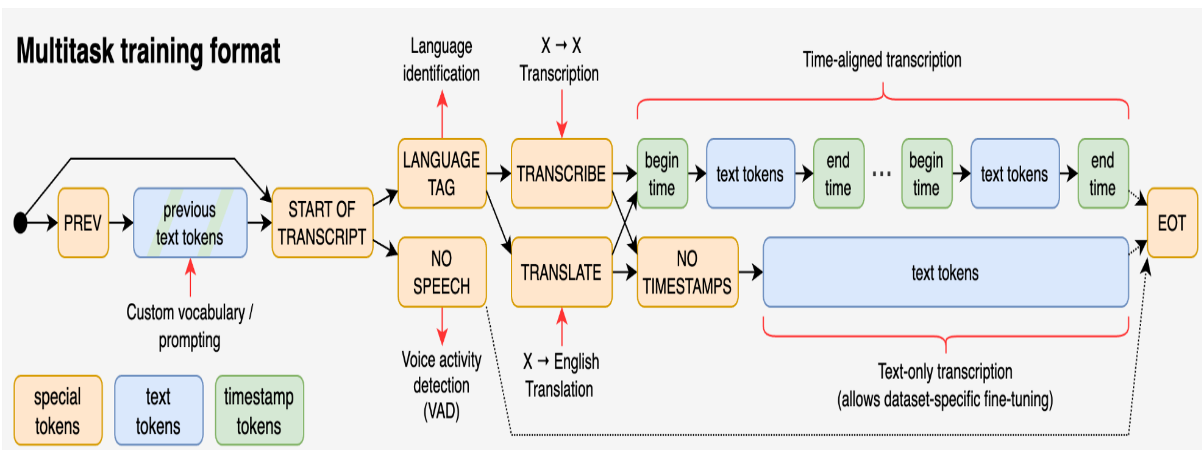


**Figure 1.** The Whisper Model Logic Sequence (Source: https://github.com/openai/whisper )

### AphasiaBank

This paper fine-tunes the Whisper model, in two trials for aphasia patient classification based on the AphasiaBank dataset. It is a free, shared database containing audio and transcripts of around 180 people with aphasia and 140 without (Forbes, 2012). It provides classified train, valid, and test datasets for research purposes, reducing our effort to collect aphasia data. It also provides each audio's ID, session when it was recorded, prompts, correctness, and transcripts. If the speaker correctly speaks the prompt in the audio, that audio is classified as correct. For example, if the prompt is "volcano" (in ARPHABET "V AA L K EY N OW"), but the speaker says "K AA L G EY N OW," this audio is considered incorrect. This research mainly relies on the AphasiaBank data, evaluating whether the speaker correctly pronounces the prompt.

## Results

### First Trial

Whisper generates robust, high-accuracy ASR (Hu, 2022). Using only its own ASR to evaluate speakers could achieve great results. In the first trial, Whisper is fine-tuned to generate ARPABET phoneomon based on AphasiaBank data. The large-v3 Whisper model that OpenAI publicly shares is fine-tuned. It has the largest parameters,1550 M, and accuracy (Hu, 2022). The LoRa fine-tuning method is used for fine-tuning. It is a low-rank breakdown method to reduce the number of trainable parameters while fine-tuning, speeding up speed, and consuming less memory. The large-v3 Whipser model is quantized to 8 bits. Using the generated transcript, this trial compares the ARPANET phonetic transcription codes of the prompt to indicate it is correctly pronounced. This paper uses Hugging's PEET implementation for this fine-tuning (Mangrulkar, 2022).

```
from peft import prepare_model_for_kbit_training
from peft import LoraConfig, PeftModel, LoraModel, LoraConfig, get_peft_model
from transformers import WhisperForConditionalGeneration
from transformers import Seq2SeqTrainingArguments

# Load Pretrained Whipser model and quantize to 8 bit
model = WhisperForConditionalGeneration.from_pretrained(model_name_or_path, load_in_8bit=True, device_map="auto")

model.resize_token_embeddings(len(processor.tokenizer)) # resize model for added tokens
model = prepare_model_for_kbit_training(model) # freeze model for fine-tuning

def make_inputs_require_grad(module, input, output):
    output.requires_grad_(True)
model.model.encoder.conv1.register_forward_hook(make_inputs_require_grad) # make input trainable

# define LoRA config
config = LoraConfig(r=32, lora_alpha=64, target_modules=["q_proj", "v_proj"], lora_dropout=0.05, bias="none")
model = get_peft_model(model, config) # add LoRA layers

training_args = Seq2SeqTrainingArguments( # setup training arguments
    output_dir="reach-vb/test",
    per_device_train_batch_size=8, # batch size as 8
    gradient_accumulation_steps=1,
    learning_rate=1e-3, # set learning rate as 0.001
    warmup_steps=50,
    num_train_epochs=1,
    evaluation_strategy="steps",
    fp16=True, # use fp16 16-bit (mixed) precision training instead of 32-bit training.
    per_device_eval_batch_size=8,
    generation_max_length=128,
    logging_steps=100,
    remove_unused_columns=False,
    label_names=["labels"],
)
```

**Figure 2.** Configuration Code for LoRA fine-tuning. This code above is used to adapt LoRA fine-tuning. We set the learning rate as 8, batch size as 8, and LoRA dropout rate as 0.05. We adapt LoRA in the "q_proj" and "v_proj" module of Whisper.

After 1000 steps, we evaluate the fine-tuned Whisper. It could generate a transcript that is close to the transcript. However, testing the correctness of 342 audios that aren't involved in training results in 56.89% accuracy. Analyzing the ASR results, differences in the token that sounds similar (e.g., the model sometimes predicted AH instead of AA) result in poor accuracy. Also, it sometimes omits some words. For example, when sample audio with the sentence "Heat from fire, fire from heat." is tested, it results "HH IY T F AY R F AY R F R AH M HH IY T". However expected transcript is "HH IY T  F R AH M  F AY ER . F AY ER  F R AH M  HH IY T", which omits "from" (in ARPABET, it is F R AH M).

## Second Trial

In the second trial, the model aims to find correctness independently. Unlike the first trial, which simply outputs the phonemic script, the prediction structure is changed for classification purposes. The main task the model should achieve is to detect a prompt that the speaker says so relevant information is added to the training. At the beginning of the prediction, the prompt is given around “START PROMPT” and “END PROMPT” tokens, indicating the prompt is not part of the transcript. However, in evaluating the fine-tuned model, this information is given to the decoder so that the model can use this information as input. The model aims to generate an ARPABET transcript to reference whether the speaker says the prompt correctly or not, so it includes unique tokens “START OF TRANSCRIPT”, “LANGUAGE TAG”, and “TRANSCRIBE” that Whisper uses. This study has only experimented on English, but language tags can be used for further research on multilingual classification. After the transcript, there is a correctness token. “Incorrect” and “correct” are two possible cases. Figure 2 simplifies this structure. Overall, tokens that indicate “CORRECT”, “INCORRECT”, “PROMPTSTART”, and “PROMPTEND” is added to the tokenizer, resulting in 51870 token embeddings size, which is 4 more than the original Whisper model.

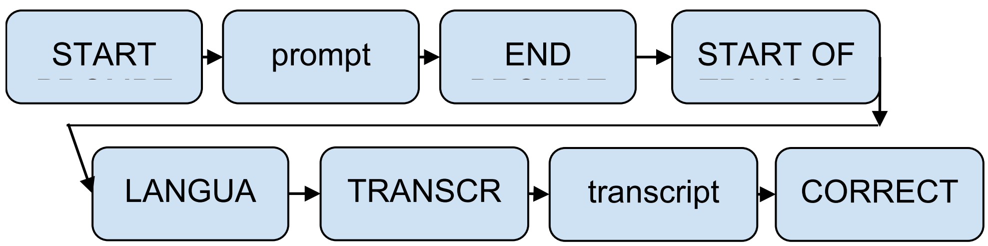


**Figure 3.** Structure of prediction sequence. Capitalized sections are a special token.

```python
from g2p import make_g2p
transducer = make_g2p('eng', 'eng-arpabet') #load g2p that converts English to ARPABET

def prepare_dataset(batch):
    # load and resample audio data from 48 to 16kHz
    audio = batch['filename']

    # compute log-Mel input features from input audio array
    batch["input_features"] = feature_extractor(audio["array"], sampling_rate=audio["sampling_rate"]).input_features[0]

    # encode target text to label ids for Aphasia classfication purpose
    if batch['correctness']:
      correctness = '<|correct|>'
    else:
      correctness = '<|incorrect|>'

    phonetic_prompt = transducer(batch['prompt']).output_string
    promptpart = '<|promptstart|>' + phonetic_prompt + '<|promptend|>'
    promptpart_token = processor.tokenizer(promptpart,add_special_tokens=False).input_ids
    transcript_token = processor.tokenizer(batch['transcript']).input_ids
    correctness_token = processor.tokenizer(correctness,add_special_tokens=False).input_ids

    newlabels = promptpart_token + transcript_token + correctness_token

    # it result expected output sequence in tokenized form, e.g. ['<|promptstart|>', 'HH', ' AW', ' S', ' ', '<|promptend|>',
    # '<|startoftranscript|>', '<|en|>', '<|transcribe|>', 'HH', ' AW', ' S', ' ', '<|correct|>] in tokenized form

    batch["labels"] = newlabels
    return batch

buff = common_voice.map(prepare_dataset, num_proc=1)
common_voice = buff
```

**Figure 4.** The Code for Preparing a Dataset. It generates a prediction sequence in the format mentioned in this paper.

The second trial is also fine-tuned and evaluated under the same conditions as the first trial. However, prompts such as “START PROMPT” and “END PROMPT” tokens and ARPHABET converted prompts are forced in the decoder to give a prompt. This conversation has been done through the `Gi2Pi` library (Pine, 2022). Predicted correctness is based on a "CORRECTNESS" unique token that fine-tuned Whipser predicted. It results in 70.76% accuracy, which is reasonable and better than the first trial.

## Conclusion

In conclusion, speech disorders, particularly aphasia, significantly impact an individual's ability to communicate, affecting both their personal and social lives. Traditional assessment methods, though effective, have limitations such as time consumption, the necessity of specialized clinical knowledge, and challenges in translation and cultural adaptation. The advent of modern technology, especially automatic speech recognition (ASR) and machine learning models, offers promising solutions to these challenges by enabling more efficient, accessible, and accurate assessments. This paper investigated the Whisper model, a state-of-the-art ASR model, for the its potential in evaluating and classifying aphasia. The Whisper model contains a robust architecture, multilingual capabilities and large-scale training data that make it a good candidate for automatic aphasia speech evaluation. LoRA fine-tuning methods were applied to show that the model is adaptable to other practical real-world scenarios and deployments.

The result of the two fine-tuning trials showed the potential and limitations of ASR. The first trial worked on phonemic transcription and revealed that the model can produce transcripts close to the expected output. However the overall accuracy showed that the model was detecting incorrect pronunciations. Therefore the second trial improved the classification approach by integrating prompt information and correctness tokens into the model's prediction structure. Now with both the input and prediction structures modified, there was an improved accuracy of 70.76%. The second trial with further refinements show that the prediction accuracy can increase by some modifications to the structure. The limitation of this study is that the whole AphasiaBank training dataset was not available therefore the model was only trained on the sample set which is a smaller data set than what is available. Future studies can do further refinements and fine-tuning on the whole AphasiaBank to create a deployable Whisper model for automatic speech evaluation.

In conclusion, the integration of ASR models, like the Whisper Model, can be a huge advancement in the field. It can remediate the need for more SLPs and perform a more objective evaluation of Aphasia. This paper shows a good start to improving the accuracy and reliability of these models, and therefore, Aphasia and overall speech evaluation and maybe even recognition can be more accessible to individuals suffering with communication disorders through apps or online.

## Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

## References

A. Ilapakurti, S. Kedari, J. S. Vuppalapati, S. Kedari and C. Vuppalapati, "Artificial Intelligent (AI) Clinical Edge for Voice disorder Detection," 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 2019, pp. 340-345, doi: 10.1109/BigDataService.2019.00060.

Bílková, Z. (2020). Human-computer interface based on tongue and lips movements and its application for speech therapy system. *Electronic Imaging*. https://doi.org/10.2352/issn.2470-1173.2020.1.vda-389

Deka, C., Shrivastava, A., Nautiyal, S., & Chauhan, P. (2022). *AI-Based Automated Speech Therapy Tools for persons with Speech Sound Disorders: A Systematic Literature Review*. https://arxiv.org/pdf/2204.10325.pdf

Forbes, M. M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language*, *33*(3), 217–222. https://doi.org/10.1055/s-0032-1320041

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Jain, R., Barcovschi, A., Yiwere, M., Corcoran, P., & Cucu, H. (2023). Adaptation of Whisper models to child speech recognition. *arXiv preprint arXiv:2307.13008*.

Jothi, D. K. R., Yawalkar, P., & Mamatha, V. L. (2021). Automatic Speech Assessment System for Aphasia Speech Disorder. *Annals of the Romanian Society for Cell Biology*, *25*(5), 5382–5392. http://www.annalsofrscb.ro/index.php/journal/article/view/6425

Kohlschein, C., Schmitt, M., Schuller, B., Jeschke, S., & Werner, C. J. (2017). A machine learning based system for the automatic evaluation of aphasia speech. *OPUS (Augsburg University)*. https://doi.org/10.1109/healthcom.2017.8210766

Le, D., & Emily Mower Provost. (2014, September 14). Modeling pronunciation, rhythm, and intonation for automatic assessment of speech quality in aphasia rehabilitation. *15th Annual Conference of the ISCA (INTERSPEECH),*. https://doi.org/10.21437/interspeech.2014-373

Mangrulkar, S., Gugger, S., Debut, L., Belkada, Y., Paul, S., & Bossan, B. (2022). Peft: State-of-the-art parameter-efficient fine-tuning methods. *URL: https://github. com/huggingface/peft*.Chicago

National Aphasia Association. (2016). *Aphasia Statistics*. The National Aphasia Association. https://aphasia.org/aphasia-resources/aphasia-statistics

National Institute on Deafness and Other Communication Disorders. (2016, May 19). *Quick Statistics About Voice, Speech, Language*. NIDCD. https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language

Pine, A., Littell, P., Joanis, E., Huggins-Daines, D., Cox, C., Davis, F., Antonio Santos, E., Srikanth, S., Torkornoo, D., & Yu, S. (2022). G$_i$2P$_i$ Rule-based, index-preserving grapheme-to-phoneme transformations. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 52–60). Association for Computational Linguistics.

Privitera, A. J., Ng, S. H. S., Kong, A. P., & Weekes, B. S. (2024). AI and Aphasia in the Digital Age: A Critical Review. *Brain sciences*, *14*(4), 383. https://doi.org/10.3390/brainsci14040383

Qin, Y., Lee, T., Pak, A., & Sam Po Law. (2016). *Towards automatic assessment of aphasia speech using automatic speech recognition techniques*. https://doi.org/10.1109/iscslp.2016.7918445

Saz, O., Yin, S.-C., Lleida, E., Rose, R., Vaquero, C., & Rodríguez, W. R. (2009). Tools and tech-nologies for Computer-Aided Speech and Language Therapy. *Speech Communication*, *10*, 51. https://doi.org/10.1016/j.specom.2009.04.006ï

Shahin, M., Zafar, U., & Ahmed, B. (2020). The Automatic Detection of Speech Disorders in Children: Challenges, Opportunities, and Preliminary Results. *IEEE Journal of Selected Topics in Signal Processing*, *14*(2), 400–412. https://doi.org/10.1109/jstsp.2019.2959393

Toğram, B., & Maviş, İ. (2012). Validity, reliability and standardization study of the language assessment test for aphasia. *Turkish Journal of Neurology*, *18*(3), 096-103.