Interspeech 2024
1-5 September 2024, Kos, Greece

Multilingual Speech and Language Analysis for the Assessment of Mild Cognitive Impairment: Outcomes from the Taukadial Challenge

Paula Andrea Perez-Toro[1,2⋆], *Tomás Arias-Vergara*[1,2], *Philipp Klumpp*[1], *Tobias Weise*[1,3], *Maria Schuster*[4], *Elmar Noth*[1], *Juan Rafael Orozco-Arroyave*[1,2], *Andreas Maier*[1]

[1]Pattern Recognition Lab. Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[2]GITA Lab, Faculty of Engineering. Universidad de Antioquia, Medellín, Colombia
[3] Speech & Language Processing Lab. Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[4]Department of Otorhinolaryngology, Head and Neck Surgery. Ludwig-Maximilians University, Munich, Germany

⋆paula.andrea.perez@fau.de

Abstract

Cognitive decline, a hallmark of several neurological conditions, including dementia and Alzheimer's disease, often manifests in noticeable changes in speech patterns and language use. Speech analysis in this context can serve as a valuable tool for early detection and monitoring of cognitive impairment. In this paper, we present the results of our attempts at the TAUKADIAL Challenge for automatically detecting people with mild cognitive impairment and predicting a cognitive score on English and Chinese speakers. In the classification task, we achieved a UAR of 83% using two language-dependent classifiers trained with timing and acoustic features. In the regression task, we obtained an RMSE of 1.87 using English speakers to train the base model with timing, acoustic, and language-dependent features.

Index Terms: Mild Cognitive Impairment, Multilingual, Speech Analysis, Language Analysis

1. Introduction

Cognitive decline is a process often perceived as an inevitable aspect of aging [1], which can manifest in various forms, ranging from normal age-related memory lapses to more severe conditions such as Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). MCI is characterized by noticeable memory deficits, not severe enough to disrupt daily functioning. Unlike AD, which affects a wide range of cognitive skills and compromises independent living, the memory problems associated with MCI are typically minimal to mild, frequently going unnoticed by the individuals themselves.

MCI patients frequently lose personal items, forget appointments or events, or struggle to find the right words. The diagnosis process for MCI combines feedback on memory lapses, family, and friends with a thorough medical evaluation to exclude treatable causes behind the cognitive symptoms (e.g., depression or medication side effects). One of the tests to assess the cognitive function of MCI patients is the Mini-Mental State Examination (MMSE). It consists of a 30-point scale that evaluates aspects such as language production, immediate memory, naming, and spatial attention, with scores above 24 pointing normal cognition.

Crosslingual studies in the context of MCI are crucial for developing universally applicable diagnostic tools. These studies enable researchers to understand how MCI manifests across different languages and cultural backgrounds, providing insights into the universal biomarkers of cognitive decline as well as language-specific characteristics. The TAUKADIAL Challenge [2] seeks to investigate the potential of speech as an indicator of cognitive health worldwide, offering data in two predominant languages: Chinese and English. The challenge's objectives include predicting the MMSE score and detecting subjects with MCI among elderly Chinese and English speakers using speech recordings.

1.1. Related Work

There have been studies in the literature that have addressed the detection of MCI from speech recordings. In general, these studies have emphasized using prosodic measures such as temporal information, intensity, and voice quality [3, 4]. Those measures are related to changes in speech timing, rhythm, and increased pauses that the patients may experience [5].

Language production manifested through cognitive and memory impairments leads to word-finding difficulties, reduced vocabulary, and impaired sentence construction [5, 6]. These changes are categorized into syntactic complexity, lexical richness, informative features, and part-of-speech tagging, highlighting how cognitive impairments affect sentence structure, vocabulary diversity, speech content, and word usage.

In addition, descriptors extracted from Automatic Speech Recognition (ASR)-based methods are also considered for the classification of MCI [7]. Recently, acoustic and linguistic embedding have emerged as non-interpretable but effective features for this task [8].

There are few studies in the literature where they considered a cross-lingual dataset for cognitive impairment assessment. In the context of AD detection, last year Multilingual Alzheimer's Dementia Recognition through Spontaneous Speech challenge [9] considered a crosslingual corpora composed of English and Greek speakers. The challenge included two tasks; classification of dementia with the highest accuracy reported of 87% and prediction of the MMSE score with lowest Root Mean Squared Error (RMSE) of 3.72.

Some studies have explored the feasibility of cross-linguistic AD detection in English, German, and Spanish [10]. The authors evaluated from a classical acoustic and prosodic features to those based on neural embedding, reaching accuracies of up to 70% while training and testing in different languages.

For MCI classification, multilingual word embeddings were used in [11] to analyze Cookie Theft narratives in English and Swedish, forming clusters to generate multilingual topics. The

 10.21437/Interspeech.2024-2115

authors reported accuracies of up to 72% on classifying MCI vs. Healthy Control (HC) subjects. Similarly, in [12], word embeddings were employed to assess MCI in verbal fluency tasks. By analyzing semantic similarities semantic similarities in these word sequences, area under the ROC curve of up to 0.68 were achieved.

1.2. Contributions

Our contributions are:

- Performing a multidimensional approach that combines classical and state-of-the-art methods to investigate speech and language cues that reveal the presence of MCI from audio recordings.
- Improvement over baseline results for the MCI classification and MMSE prediction tasks of the Taukadial 2024 Challenge [2] using acoustic and linguistic embeddings together with speech timing features.
- Evaluation of linguistic and acoustic features, as well as multi-modal fusion approaches for the classification of AD, using ASR and speaker diarization.

The rest of the paper is organized as follows: Chapter 2 includes a description of the challenge data and the preprocessing steps that we followed. In Chapter 3 we describe the methods used for characterization of the speech signals, automatic detection of MCI, and prediction of the MMSE score. In Chapter 4 we reported the obtained results for the two tasks of the challenge. Chapter 5 includes the conclusions of our participation in this challenge.

Table 1: *Metadata of the training set*

MCI	
Age	73.36 (SD 6.14, range 61-87)
Men	39.2% (n = 87)
Women	60.8% (n = 135)
MMSE	25.84 (SD 3.73, range 13-30)
Duration	58.92 (SD 36.61, range 12.7-240.9)
HC	
Age	71.85 (SD 6.65, range 61-87)
Men	38.2% (n = 63)
Women	61.8% (n = 102)
MMSE	29.07 (SD 1.08, range 25-30)
Duration	63.07 (SD 33.85, range 10.2-209.6)

Table 2: *Metadata of the test set*

MCI	
Age	77.90 (SD 9.15, range 59-91)
Men	52.4% (n = 33)
Women	47.6% (n = 30)
MMSE	25.86 (SD 3.27, range 18-30)
Duration	72.33 (SD 46.94, range 20.46-257.65)
HC	
Age	67.68 (SD 4.71, range 62-82)
Men	36.8% (n = 21)
Women	63.2% (n = 36)
MMSE	29.05 (SD 1.06, range 26-30)
Duration	63.47 (SD 48.55, range 10.7-253.471)

2. Data Preprocessing

The dataset in this work was created by the organizers of the Interspeech Taukadial 2024 challenge [2], which comprises English and Chinese recordings from MCI patients and NC subjects. The training dataset consists of 387 recordings from 129 participants (MCI, HC). The test dataset consists of 120 recordings from 40 participants. Participants were asked to describe the three different images. Those images do not match between languages. The recordings are matched for age and gender and have been acoustically enhanced and normalized.

For this study, we, obtained transcriptions using a commercial state-of-the-art ASR service[1]. Despite removing most of the interviewer parts, we found that a significant part of the English recordings contained part of the interviewer's voice. We used speaker diarization to identify how many subjects were present in a single recording. Additionally, the language of the recordings was automatically identified using the same service.

The information of the training and test sets is reported in Tables 1 and 2.

3. Methods

Our methodology considers state-of-the-art linguistic models, such as BERT and RoBERTa, known for their ability of encoding and understanding contextual relationships between words. The choice of these models was motivated by their proven success in several classification tasks related to cognitive impairment assessment. Additionally, we employed acoustic embeddings from Wav2Vec 2.0 and Whisper to capture the speech patterns that may indicate cognitive decline. In addition to this, speech timing features have been considered since speech pauses can be considered as an early marker of cognitive decline.

[1] Amazon web services (AWS) transcribe

3.1. Linguistic Models

This research explores "word-embeddings" based approaches, which seek to understand the semantic and syntactic connections between words along with their contextual meanings. We evaluated two pretrained word-embedding models derived from the "Transformers" architecture's encoder [13]: the Bidirectional Encoder Representations from Transformers (BERT) [14] and A Robustly Optimized BERT Pretraining Approach (RoBERTa) [15]. These methods process all elements simultaneously by forming direct connections between individual elements through a process known as attention.

BERT incorporates multiple parallel attention heads, enabling the capture of a wide array of word relationships through multi-head attention. It bases its training on transfer learning, initially focusing on two unsupervised tasks: Masked Language Modeling (MLM) for predicting a portion of masked words in a sentence, and Next Sentence Prediction (NSP) to determine if one sentence logically follows another.

RoBERTa modifies BERT's model by omitting the NSP component and employing larger batches. It introduces dynamic masking in the MLM process, where the masked tokens vary across training epochs. Our analysis utilized the BERT-Base and RoBERTa-Base models, trained on Wikicorpus data spanning 102 languages and CommonCrawl data filtered from 100 languages [16], respectively. For both models, the final layer, comprising 768 units, serves as the basis for the word-embedding representation utilized in the classification tasks. Additionally, the mean value of these word-embeddings is calculated for classification purposes. The corresponding source

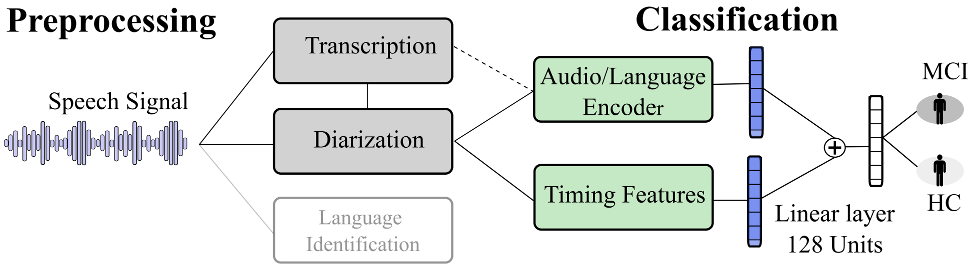

Figure 1: *Fusion methodology considered in this study for merging acoustic/linguistic embeddings with speech timing features*

code has been made publicly accessible online[2].

3.2. Acoustic Embeddings

Wav2Vec 2.0 [17] and *Whisper* [18] represent state-of-the-art models in the field of speech processing, each employing unique learning strategies to interpret and transcribe raw speech signals effectively.

On the one hand, ***Wav2Vec 2.0*** employs self-supervised learning techniques to derive representations directly from raw speech signals. The architecture is composed of three key elements: feature extraction, a context network, and a projection to the final output. It processes 16 kHz raw audio, segmenting it into 30 ms slices with a 10 ms overlap. During feature extraction, temporal convolutions transform the speech data into a latent space representation. the audio segments are then masked and quantized for self-supervised training. It considers the encoder component from the Transformer-based architecture together with contrastive learning-based training to generate contextual representations. We focus on the different pretrained versions: for multilingual, Wav2Vec XLSR-53[3]. We explored the use of different layer, seeking for higher performance that could generalize between language.

Whisper, on the other hand adopts a weakly-supervised learning strategy, focusing on speech-to-text tasks across multiple languages and translating non-English languages into English. The model's architecture is built around an encoder-decoder Transformer setup, processing 16 kHz audio into 30 ms segments. Contrary to Wav2Vec, Whisper employs Mel-frequency representations as audio input, which enhances its transcription accuracy. A similar investigative approach is applied to Whisper, exploring the model's layers to identify those most conducive to superior transcription and translation performance across varied linguistic contexts.

3.3. Speech Timing

Prior research on AD has also employed these characteristics, yielding satisfactory outcomes as noted in [19]. This set of features comprises duration-oriented descriptors derived through an energy-based Voice Activity Detection (VAD) algorithm, which differentiates between speech and pause intervals. It covers 16 descriptors related to the number of pauses and speech intervals per second, the proportion of speech intervals relative to pauses, and six statistical measures (mean, standard deviation, kurtosis, skewness, minimum, and maximum) that detail the lengths of both pauses and speech intervals.

[2] `https://github.com/PauPerezT/WEBERT`

[3] `https://huggingface.co/facebook/wav2vec2-large-xlsr-53`

3.4. Optimization, Classification, and Regression

A Radial Basis Function-Support Vector Machine (RBF-SVM) and a Multi-Layer Perceptron (MLP) were used as a classifier for the MCI classification task. For the MLP model, we incorporated a single hidden layer that transformed the dimensionality of the input vector into 128 units. Subsequently, this transformed output was passed through the classification layer of two units.

The optimal parameters of the RBF-SVM were found through a grid search where $C \in \{10^{-4}, 10^{-3}, ..., 10^{4}\}$ and $\gamma \in \{10^{-4}, 10^{-3}, ..., 10^{4}\}$. The regression task for the prediction of the MMSE was performed by using a Support Vector Regressor (SVR), and a Linear Regression (LR) model.

The validation of our experiments followed a bootstraping strategy (70% training, 15% validation, and 15% test), where the process was repeated five times, to ensure variability in the different sets. It is important to notice, that we re-trained our model with the whole training set when testing, which indicates that the results may vary. For the classification and regression different fusion strategies were considered by merging sets of features before performing the classification/regression and making the final decision. For the MLP case, when the timing features where applied, we noticed that it improved in performance when adding to the embedding vector, as illustrated in Figure 1. In addition, notice that the preprocessing part applied for all approaches in this study.

When training, balancing samples was to essential getting more stable results, since we did not had any demographic information about the participants in the test set. This was made according to MMSE ranges and languages.

4. Experiments and Results

Table 3 presents the challenge results for the classification task focused on discriminating MCI vs. HC subjects. The study used several feature sets including embeddings extracted from Whisper (WHP), speech timing features (Timing), and Wav2vec embeddings. The tasks were conducted in both English (EN) and Chinese (ZH), with performance metrics including Unweighted Average Recall (UAR), Sensitivity (Sens), and Specificity (Spe). Notice that to maintain speaker independence, the features extracted from each image description were merged using an early fusion approach. However, some of the results only consider training in one single task (image description), since we noticed that in some cases the performance dropped and become unstable when training with the three descriptions together for our approach.

Our criteria for selecting the test models from the initial classification results are as follows:

- (Unilingual) The best-performing model for between languages
- (Crosslingual) The overall best-performing model when evaluating across both languages simultaneously
- (Crosslingual) The most effective model trained on Chinese data and applied to both languages simultaneously
- (Crosslingual) The top model is trained on English data and then applied across both languages
- (Unilingual) A linguistic model, noting that it did not yield strong results

The highest classification result was obtained by training language-independent models using Timing and acoustic (Whisper) features (Validation: UAR= 80%; Test = UAR= 83%), which shows how important the prosodic information is for the detection of MCI from speech. However, the main limitation is that prosodic information is language-dependent; which may explain why it only appears in the top features when a language model is considered for evaluation. The acoustic information provided by Whisper seems to capture patterns related to the clinical condition; however, using this embedding representation is not enough to have strong discriminatory patterns in people with MCI.

Table 3: *Classification results for the automatic detection of MCI patients using acoustic, prosodic, and language-based features. Only the top-5 performing models are reported.*

Feature set	**Classifier**	**Task [EN/ZH]**	**UAR**	**Sens**	**Spe**
		Validation			
Timing-WHP (ZH) & Timing-WHP (EN)	MLP	1/3	80.00	73.35	86.66
WHP	MLP	3/3	77.30	54.55	100.00
Timing-WHP_6$_L$	SVM	1,2,3/1,2,3	65.15	63.64	66.67
WHP_6$_L$	SVM	2,3/2,3	70.50	69.48	72.125
BERT	MLP	1/1	57.59	79.73	35.45
		Test			
Timing-WHP (ZH) & Timing-WHP (EN)	MLP	1/3	83.30	66.67	100.00
WHP	MLP	3/3	65.10	63.5	66.67
Timing-WHP_6$_L$	SVM	1,2,3/1,2,3	56.10	50.8	61.40
WHP_6$_L$	SVM	2,3/2,3	54.10	71.43	36.84
BERT	MLP	1/1	53.60	80.95	26.32
Baseline	MLP	–	59.18	–	–

LY: it refers to the latent layer in models such as Wav2Vec. **UAR**: Unweighted Average Recall. **Sens:** Sensitivity. **Spe:** Specificity. EN: English. ZH: Chinese 6L: the sixth layer from the encoder

In the case of the regression task, Table 4 details the top results for the challenge. The performance was evaluated according to the Pearson's Correlation Coefficient (PCC) and the RMSE.

Our criteria for selecting the test models from the initial regression task are as follows:

- (Crosslingual) The best-performing model considering balance samples by MMSE from both datasets
- (Unilingual) The best-performing/consistent model between languages
- (Crosslingual) The linguistic model with highest results
- (Unilingual) Best performing combination of language and speech considering between languages
- (Unilingual) Highest results for acoustic alone

Similar to the classification task the combination of Whisper embeddins and Timing features achieved the highest results (RMSE = 1.87), while also outperforming the baseline. Notice the inclusion of linguistic information.

Table 4: *Regression results for the prediction of MMSE scores from the subject of the challenge. Only the top-5 performing models are reported.*

Feature set	**Regressor**	**Task [EN/ZH]**	**RMSE**	**PCC**
		Validation		
BERT-Timing-WHP$_{LY}$	SVR	1/1	2.34	0.31
BERT	SVR	1/1	3.39	0.18
BERT-RoBERTa	SVR	1,3/1,3	2.69	0.59
Wav2Vec$_{LY}$-BERT	SVR	1,2,3/1,2,3	3.06	0.15
Timing	LR	3/3	3.64	0.29
		Test		
BERT-Timing-WHP$_{LY}$	SVR	1/1	1.87	0.79
BERT	SVR	1/1	2.23	0.66
BERT-RoBERTa	SVR	1,3/1,3	2.30	0.22
Wav2Vec$_{LY}$-BERT	SVR	1,2,3/1,2,3	2.68	0.44
Timing	LR	3/3	2.86	0.22
Baseline	MLP	–	2.89	–

LY: it refers to the latent layer in models such as Wav2Vec
PCC: Pearson's Correlation Coefficient

5. Conclusions

In this paper, we presented the results obtained in the TAUKADIAL challenge for the detection of MCI and the prediction of cognitive impairment in two different languages. Our proposed approach was first to use a language detection model to label the audio samples and test language-dependent and -independent models. Then, we considered running a speaker diarization algorithm in the challenge data to separate the patient and avoid biases due to the presence of other speakers. This was mainly motivated by our experience working with Alzheimer's data, where we learned that the actions taken by the interviewer during the data collection can also influence the outcome of the classification task. Once we had the samples identified language and patients, we considered different acoustic, prosodic, and language-based features for characterizing the speech signals. For the acoustic and language analysis, we decided to use pre-trained Transformer models, given the success of such architectures not only in modeling language but also in acoustic information. We also considered classical features, such as timing parameters. In our experience, such a set of features has always appeared as one of the top-performing features in our previous experiments.

According to the results, our highest-performing model in the classification task was an MLP trained with the combination of timing and Whisper (acoustic) features to train language-independent models (UAR = 83%). For the regression task, our best-performing model was an SVR trained in one language (English) with BERT, timing, and Whisper features (RMSE=1.87). Again, we see that the timing and Whisper features, and this time were in a single language. These results can be explained considering that the focus of the task is to predict a score that is restricted to a very narrow range for English participants and wider for Chinese. On the one hand, predicting the scores of the English participants lowers the prediction error because most of the participants (roughly) have their MMSE scores between 22 and 30. On the other hand, the base low variation in the MMSE score of the base model seems to act as a reference point making it easier to predict the MMSE scores of the participants with data further away from the trained distributions.

6. Acknowledgements

This work was partially funded by the EVUK programme ("Next-generation Al for Integrated Diagnostics") of the Free

State of Bavaria and by CODI at UdeA grant # PI2023-58010.

7. References

[1] H. L. Park, J. E. O'Connell, and R. G. Thomson, "A systematic review of cognitive decline in the general elderly population," *International journal of geriatric psychiatry*, vol. 18, no. 12, pp. 1121–1134, 2003.

[2] S. Luz, S. d. l. F. Garcia, F. Haider, D. Fromm, B. MacWhinney, A. Lanzi, Y.-N. Chang, C.-J. Chou, and Y.-C. Liu, "Connected speech-based cognitive assessment in chinese and english." arXiv, 2024, dOI: 10.48550/ARXIV.2404.nnnnn (TBA).

[3] Q. Yang, X. Li, X. Ding, F. Xu, and Z. Ling, "Deep learning-based speech analysis for Alzheimer's disease detection: A literature review," *Alzheimer's Research & Therapy*, vol. 14, no. 1, p. 186, 2022.

[4] C. Oh, R. J. Morris, and X. Wang, "Prosody analysis as a tool for differential diagnosis of cognitive impairment," in *Proceedings of Meetings on Acoustics*, vol. 50, no. 1. AIP Publishing, 2022.

[5] V. Taler and N. A. Phillips, "Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review," *Journal of clinical and experimental neuropsychology*, vol. 30, no. 5, pp. 501–556, 2008.

[6] H. Lindsay, J. Tröger, and A. König, "Language impairment in alzheimer's disease—Robust and explainable evidence for ad-related deterioration of spontaneous speech through multilingual machine learning," *Frontiers in aging neuroscience*, vol. 13, p. 642033, 2021.

[7] G. Gosztolya, V. Vincze, L. Tóth, M. Pákáski, J. Kálmán, and I. Hoffmann, "Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features," *Computer Speech & Language*, vol. 53, pp. 181–197, 2019.

[8] F. Braun, S. P. Bayerl, P. A. Pérez-Toro, F. Hönig, H. Lehfeld, T. Hillemacher, E. Nöth, T. Bocklet, and K. Riedhammer, "Classifying Dementia in the Presence of Depression: A Cross-Corpus Study," in *Proc. INTERSPEECH 2023*, 2023, pp. 2308–2312.

[9] S. Luz, F. Haider, D. Fromm, I. Lazarou, I. Kompatsiaris, and B. MacWhinney, "Multilingual alzheimer's dementia recognition through spontaneous speech: A signal processing grand challenge," *arXiv preprint arXiv:2301.05562*, 2023.

[10] A. Ablimit, C. Botelho, A. Abad, T. Schultz, and I. Trancoso, "Exploring dementia detection from speech: Cross corpus analysis," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 6472–6476.

[11] K. C. Fraser, K. L. Fors, and D. Kokkinakis, "Multilingual word embeddings for the assessment of narrative speech in mild cognitive impairment," *Computer Speech & Language*, vol. 53, pp. 121–139, 2019.

[12] H. Lindsay, P. Müller, I. Kröger, J. Tröger, N. Linz, A. König, R. Zeghari, F. R. Verhey, and I. H. Ramakers, "Multilingual Learning for Mild Cognitive Impairment Screening from a Clinical Speech Task," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, 2021, pp. 830–838.

[13] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[14] J. Devlin *et al.*, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://www.aclweb.org/anthology/N19-1423

[15] L. Zhuang, L. Wayne, S. Ya *et al.*, "A Robustly Optimized BERT Pre-training Approach with Post-training," in *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, 2021, pp. 1218–1227.

[16] R. Schäfer, "CommonCOW: massively huge web corpora from CommonCrawl data and a method to distribute them freely under restrictive EU copyright laws," in *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, 2016, pp. 4500–4504.

[17] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 449–12 460, 2020.

[18] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[19] P. A. Pérez-Toro, T. Arias-Vergara, F. Braun, F. Hönig, C. A. Tobón-Quintero, D. Aguillón, F. Lopera, L. Hincapié-Henao, M. Schuster, K. Riedhammer, A. Maier, E. Nöth, and J. R. Orozco-Arroyave, "Automatic Assessment of Alzheimer's across Three Languages Using Speech and Language Features," in *Proc. INTERSPEECH 2023*, 2023, pp. 1748–1752.