Impact of ASR on Alzheimer's Disease Detection: All Errors are Equal, but Deletions are More Equal than Others

Aparna Balagopalan[1,2], Ksenia Shkaruta[3], Jekaterina Novikova[1]
1. Winterlight Labs, 2. University of Toronto, 3. Georgia Tech

WINTERLIGHT

1. Introduction

Automatic Speech Recognition (ASR) is a critical component of any fully-automated speech-based dementia detection model. However, despite years of speech recognition research, little is known about the impact of ASR accuracy on dementia detection. In this paper, we experiment with controlled amounts of artificially generated ASR errors and investigate their influence on dementia detection.

We find that :

1. Deletion errors affect detection performance the most, due to their impact on the features of syntactic complexity and discourse representation in speech.
2. We show the trend to be generalisable across two different datasets for cognitive impairment detection.
3. We propose optimising the ASR to reflect a higher penalty for deletion errors in order to improve dementia detection.

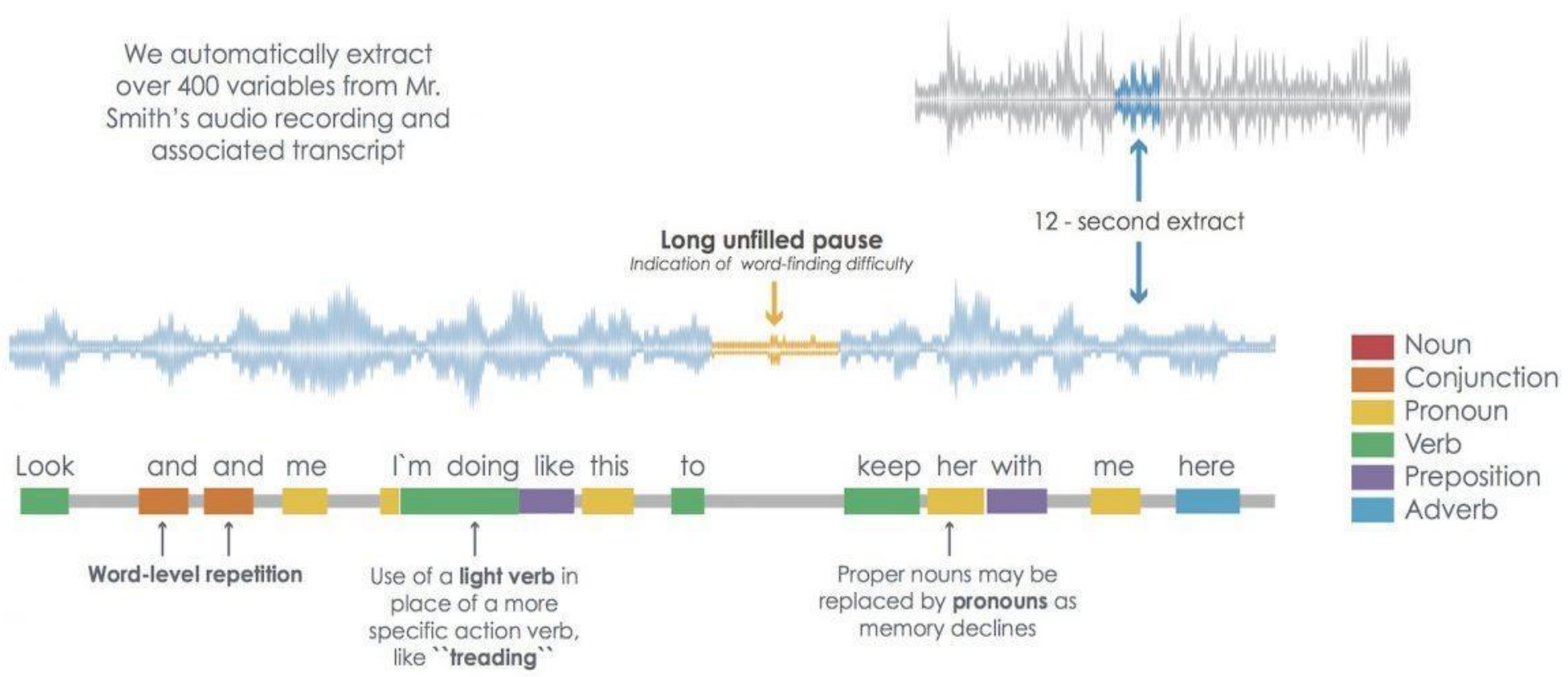
2. Datasets

We use two datasets for all experiments:

- DementiaBank (DB; Becker et al., 1994), contains 409 narrative picture descriptions from healthy and AD participants
- Healthy Aging (HA; Balagopalan et al., 2018), contains 97 picture descriptions. We derive, "healthy" vs "not healthy" labels for all samples based on scores on the Montreal Cognitive Assessment.

3. Feature Extraction

We extract 507 acoustic and lexico-syntactic features of type: Syntactic Complexity, Lexical Complexity and Richness, and Discourse mapping.



4. ASR Setup

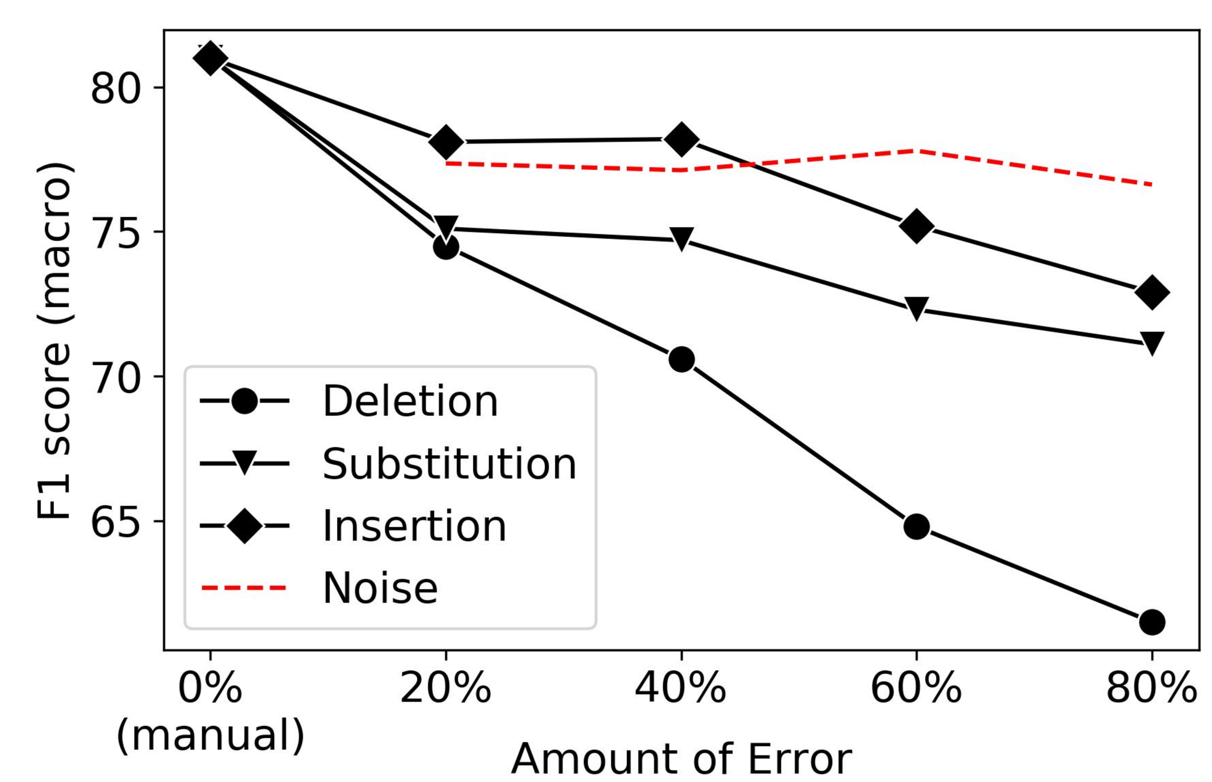
ASR based on the Kaldi toolkit (Povey et al.,2011) is used. ASR uses ASPiRE chain model trained on multi-condition Fisher corpus English as a 3-gram language model.

Dataset		Del (%)	Ins (%)	Sub (%)
DB	HC	54.14	4.27	41.59
	AD	56.98	3.89	39.13
HA	HC	24.37	13.11	62.52
	MCI	21.78	14.81	63.40

Table 1: Rates of ASR errors

5. Artificial Errors and Noise Addition

We follow a method similar to the one used by Fraser et al. (2013) to artificially add deletion, insertion, and substitution errors to manual transcripts at predefined 20%, 40%, 60% and 80% WER rates. We also perturb all lexico-syntactic features, to mimic random sources of errors using Gaussian noise.



6. Change in Classification Performance

We evaluate performance of classifying samples of speech to two classes - AD or healthy. Our findings are:

1. Deletion errors affect performance significantly more than insertion and substitution errors do.
2. F1 score trajectory with varying levels of noise is substantially different from that with varying deletion errors but not that with insertions/substitutions.
3. Model utilizing ASR transcripts retains a level of performance at 74.96%, comparable to effect of deletion errors.

7. Distinctive Effects of Deletion Errors

We identify features maintaining higher correlation with the amount of deletions than that with the amount of insertions and substitutions. Out of the 18 features identified, 83.3% are syntactic, and remaining are discourse-related --- thus showing that syntactic structure of language is much more vulnerable to deletions than to other ASR errors.

We also compute gradient-based importance scores for all samples. Similar results are seen for both datasets.

Feature group	Importance of top-10 features HC	AD	#features	Group rank
Syntactic complexity and Discourse phenomena	0.94	0.95	37	1
Lexical richness	0.91	0.92	18	2

Table 2: Importances of features distinctively correlated with deletions

8. Conclusions and Future Work

We observe that simulated deletion errors have a strong effect on classification performance. A practical suggestion would be to optimise the ASR to reflect a higher penalty for deletion errors. Dealing with deletions in training time is not trivial, so in future work, we will focus on the optimisation of ASR performance with a higher penalty for deletion errors. Careful ASR error management could help enable strong fully-automated speech-based predictive models for dementia detection.

References

1. Kathleen C Fraser, Jed A Meltzer, and Frank Rudzicz.2016. Linguistic features identify Alzheimer's disease in narrative speech.Journal of Alzheimer's Dis-ease, 49(2):407–422.
2. Daniel Povey, Arnab Ghoshal, Gilles Boulianne, LukasBurget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. 2011. The Kaldi speech recognition toolkit. Technical report, IEEE Signal Processing Society
3. Zining Zhu, Jekaterina Novikova, and Frank Rudzicz.2019. Detecting cognitive impairments by agreeing on interpretations of linguistic features. NAACL 2019