

## A Appendices

### A.1 List of Features Extracted from Speech and Transcripts

Full list of lexico-syntactic, discourse mapping and additional language features is in Table A.1, all with brief descriptions and counts of sub-types. Spacy<sup>2</sup> is used for part-of-speech tagging in the linguistic pipeline, where tags belong to Penn Treebank<sup>3</sup>.

### A.2 Noise Addition Method

The strategy followed to add  $k$  level of noise ( $\epsilon$ ) to an input  $X_i$  is as follows:

$$X_{i,j}^{noised^k} = X_{i,j} + \epsilon \quad (4)$$

$$\epsilon \sim \mathcal{N}(0, \sigma_{new}^2) \quad (5)$$

$$\sigma_{new} = k * \sigma \quad (6)$$

where  $i$  is a sample number,  $j$  is a feature number, and standard deviation of noise added is  $k$  times the standard deviation of the original unperturbed feature,  $\sigma$ . Note that the standard deviation ( $\sigma$ ) per feature is calculated over *all* samples.

### A.3 Classification Setup Details

**Hyperparameter Settings:** All 2 hidden layers of our network have 10 units each (Pedregosa et al., 2011). We use the Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 0.001 and  $L2$  regularization parameter for network weights set to 0.0001. The model is trained to a maximum of 200 epochs.

**Evaluation:** For classification experiments on all datasets (DB and HA, varying levels of artificial errors), we perform 10-fold cross-validation stratified at subject level (ensuring that speech samples from the same subject do not in the training and testing sets in each fold). The F1 (macro) - score, averaged across the 10 folds, is chosen as our primary evaluation metric, given the imbalanced nature of our dataset.

---

<sup>2</sup><https://spacy.io/usage/linguistic-features>

<sup>3</sup><http://www.cis.upenn.edu/treebank/>

Feature type	#Features	Brief Description
Syntactic Complexity	36	L2 Syntactic Complexity Analyzer (Lu, 2010) features; max/min utterance length, depth of syntactic parse tree
	104	Number of times a production type occurs divided by total number of productions
	13	Proportion, average length and rate of phrase types
Lexical Complexity and Richness	12	Average norms across all words, across nouns only and across verbs only for imageability, age of acquisition, familiarity and frequency (commonness)
	6	Type-token ratios (including moving window); brunet; Honoré's statistic
		Proportion of demonstratives (e.g., "this"), function words,
	5	light verbs and inflected verbs, and propositions (POS tag verb, adjective, adverb, conjunction, or preposition)
	3	Ratios nouns:(nouns+verbs); nouns:verbs; pronouns:(nouns+pronouns)
	1	Average word length
	18	Proportions of Spacy universal POS tags
Discourse Mapping	53	Proportions of Penn Treebank POS tags
	15	Avg/max/min similarity between word2vec (Mikolov et al., 2013) representations of utterances (with different dimensions)
	5	Fraction of pairs of utterances below a similarity threshold (0.5,0.3,0); avg/min distance
	13	Representing words as nodes in a graph and computing density, number of loops, etc.
	1	Number of switches in verb tense across utterances divided by total number of utterances
Additional Features	15	Avg/min/max cosine distance between word2vec (Mikolov et al., 2013) utterances and picture content units, with varying dimensions of word2vec
	2	Ratios – number of words: duration of audio; number of syllables: duration of speech,
	1	Proportion of words not in the English dictionary
	9	Average sentiment valence, arousal and dominance across all words, noun and verbs
	7	Total and mean duration of pauses; long and short pause counts;
		pause to word ratio; fillers(um,uh); duration of pauses to word durations
	4	Avg/min/max/median fundamental frequency of audio
	2	Duration of audio and spoken segment of audio
	4	Avg/variance/skewness/kurtosis of zero-crossing rate
	168	Avg/variance/skewness/kurtosis of 42 Mel-frequency Cepstral Coefficients (MFCC) coefficients
	10	Proportion of lemmatized words, relating to the Cookie Theft picture content units to total number of content units

Table A.1: Summary of all lexico-syntactic features extracted. The number of features in each subtype is shown in the second column (titled “#Features”).