Speech SIgnal Processing
Project Mid-Eval Report
GROUP 8

Project name : Feature Extraction using OpenSMILE on pathological dataset-2
Members : Aashna Jena (20171095) and Harshita Sharma (20171099)

About the Dataset

This dataset is created by the LANNA research group in the Faculty of Electrical Engineering at Czech Technical University in Prague. This dataset contains two sub-databases :

1. Healthy children's speech (recorded in kindergarten and in the first level of elementary school)
2. Pathological speech of children with a Specific Language Impairment or Developmental Dysphasia (recorded at a surgery of speech and language therapists and at the hospital)

This dataset uses utterances from children aged from 4 to 11 years. The main criteria used in creating the corpus is the selection of a suitable text and the selection of participants for speech recording. The words and phrases were directly selected for research among children with this disorder (specific language impairments), and they were selected while accounting for the physiological and mental development of the children. The specific utterances are divided into 13 parts.

Table 1. Speech database–structure and types of utterances used in our research.

Task code	Type of part	# Patterns	Descripton	
[T 1]	Vowels	5	Czech	"a", "o", "u", "e", "i"
			English	"a", "o", "u", "e", "i"
[T 2]	Consonants	10	Czech	"m", "b", "t", "d", "r", "l", "k", "g", "h", "ch"
			English	"m", "b", "t", "d", "r", "l", "k", "g", "h", "ch"
[T 3]	Syllables	9	Czech	"pe", "la", "vla", "pro", "bě", "nos", "ber", "krk", "prst"
			English	"pe", "la", "vla", "for", "bě", "nose", "take", "neck", "finger"
[T 4]	Two-syllable words	5	Czech	"kolo", "pivo", "sokol", "papír", "trdlo"
			English	"wheel", "beer", "falcon", "paper", "boob"
[T 5]	Three-syllable words	4	Czech	"dědeček", "pohádka", "pokémon", "květina"
			English	"grandfather", "fairy tale", "Pokemon", "flower"
[T 6]	Four-syllable words	3	Czech	"motovidlo", "televize", "popelnice"
			English	"niddy noddy", "television", "dustbin"
[T 7]	Difficult words	2	Czech	"r znobarevný", "mateřídouška"
			English	"varicoloured", "thyme"
[T 8]	Geminate words	3	Czech	"pohádková víla", "kouzelný měšec", "čarotvorný hrnec"
			English	"fairy", "magic pouch", "magic pot"
[T 9]	Accretion of range of words	4	Czech	"voda", "živá voda", "živá a mrtvá voda", "pramen s živou a mrtvou vodou"
			English	"water", "live water", "live and dead water", "source of live and dead water"
[T 10]	Sentence	1	Czech	"Když šla červená Karkulka k babičce, potkala zlého vlka."
			English	"When Little Red Riding Hood went to her grandmother, she met bad wolf."
[T 11]	Auditory differentiation	10	Czech	"pes—nes", "ten—den", "k l—v l", "hrát—brát", "ječí—ježí", "ble—ple", "kloč—kloč", "kvěš—kveš", "šný—šní", "vošl—vočl"
			English	Change in one phoneme in the word. For example: "pes—nes", . . .
[T 13]	Describe the picture	1	English	"Look at the laughable clown."—A spontaneous description of the girl's picture.

OpenSMILE ComParE_2016 feature set

The ComParE feature set was introduced for the INTERSPEECH 2016 Computational Paralinguistics Challenge : Deception, Sincerity & Native Language. It has approximately 200 features divided into two groups of Low Level Descriptors (LLDs).

1. 59 LLDs in group A, 54 functionals are applied
2. 59 delta LLDs of group A, 46 functionals are applied
3. 6 LLDs in group B, 39 functionals are applied
4. 6 delta LLDs of set B, 39 functionals are applied.

This results in a total of 6,368 features. Group A has energy, spectral, and cepstral LLDs like Loudness, ZCR, Spectral Flux, Entropy, Sharpness etc. Group B has F0, Jitter, Prob of voicing, Shimmer, log HNR and δ Jitter. The functionals applied are mean, flatness, moments, quartiles, percentiles, peak mean, peak distance, peak range etc. The table is available in Appendix 1.5 of Springer's thesis.

Few important low level descriptors are described below :

Name	Number of coefficients	Description
pcm_loudness	1	Loudness as the normalized intensity raised to a power of 0.3.
Mfcc	15	Mel-Frequency cepstral coefficients 0–14.
logMelFreqBand	8	Logarithmic power of Mel-frequency bands 0–7 (distributed over a range from 0 to 8 kHz)
lspFreq	8	The 8 line spectral pair frequencies computed from 8 LPC coefficients.
F0finEnv	1	The envelope of the smoothed fundamental frequency contour.
voicingFinalUnclipped	1	The voicing probability of the final fundamental frequency candidate. Unclipped means that it was not set to zero when it falls below the voicing threshold.

Some of the functionals used on these LLDs are described below:

Name	Description
maxPos	The absolute position of the maximum value (in frames).
minPos	The absolute position of the minimum value (in frames).
amean	The arithmetic mean.
linregc1	The slope (m) of a linear approximation of the contour.
linregc2	The offset (t) of a linear approximation of the contour.
linregerrA	The linear error computed as the difference of the linear approximation and the actual contour.
linregerrQ	The quadratic error computed as the difference of the linear approximation and the actual contour.
stddev	The standard deviation of the values in the contour.
skewness	The skewness (3rd order moment).
kurtosis	The kurtosis (4th order moment).
quartile1	The first quartile (25% percentile).
quartile2	The second quartile (50% percentile).
quartile3	The third quartile (75% percentile)
iqr1-2	The inter-quartile range: quartile2-quartile1.
iqr2-3	The inter-quartile range: quartile3-quartile2.
iqr1-3	The inter-quartile range: quartile3-quartile1.
percentile1.0	The outlier-robust minimum value of the contour, represented by the 1% percentile.
percentile99.0	The outlier-robust maximum value of the contour, represented by the 99% percentile.
pctlrange0-1	The outlier robust signal range 'max-min' represented by the range of the 1% and the 99% percentile.

Work Completed So Far

1. We downloaded and understood the dataset and its divisions
2. We set up openSMILE
3. We extracted the ComParE_2016 feature set for utterances in our dataset
4. We wrote an automated script for extracting ComParE_2016 feature set for all .wav files in the dataset
5. We are working on writing the code to extract another feature from scratch