Feature extraction using openSMILE feature

namely “ComParE_2016.conf” for openly available pathological speech dataset-2

Pathological speech dataset II

- Speech databases recorded primarily for medical research but is useful in linguistics as well
- Two databases were recorded:
  - for healthy children's speech (recorded in kindergarten and in the first level of elementary school)
  - for pathological speech of children with a Specific Language Impairment (recorded at a surgery of speech and language therapists and at the hospital).

Development Dysphasia

- Also known as Specific Language Impairment a language disorder that **delays the mastery of language skills** in children who have **no hearing loss or other developmental delays.**
- These children fail to acquire their native language properly/completely, despite having normal non-verbal intelligence, no hearing problems, and no known neurological dysfunctions or behavioral, emotional or social problems
- It is estimated that SLI affects approximately 5–7% of the kindergarten population

The Database

The entire database contains three subgroups of recordings of children's speech from different types of speakers.

- The first subgroup (controls, or H-CH) consists of recordings of children without speech disorders;
- the second subgroup (cases, or SLI-CH I) consists of recordings of children with SLI;
- the third subgroup (cases, or SLI-CH II) consists of recordings of children who have SLI of different degrees of severity (1 –mild, 2 –moderate, or 3 –severe).

Trends in the way that speech changed during the given time period (approximately 3 months) were the determining factors for inclusion in SLI-CH I database, rather than the degree of severity of the children's diagnosis.

Table 2. Description of All Databases. Subgroup H-CH is for controls and subgroups SLI-CH I and SLI-CH II are for cases.

	H-CH		SLI-CH I	SLI-CH II
		WITH DEFECT		
Girls	45	16	13	26
Number of recordings	45	16	22	45
Boys	25	17	33	46
Number of recordings	25	17	64	88
All children	70	33	46	72
All recordings	70	33	86	133
All utterances	4620	2178	5676	8819

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

- The Munich **open toolkit Speech and Music Interpretation by Large Space Extraction** (openSMILE) is a modular and flexible feature extractor for signal processing and machine learning applications.
- It contains a number of default feature sets, some of which are shown:

- Chroma features for key and chord recognition
- MFCC for speech recognition
- PLP for speech recognition
- Prosody (Pitch and loudness)
- The INTERSPEECH 2009 Emotion Challenge feature set
- The INTERSPEECH 2010 Paralinguistic Challenge feature set
- The INTERSPEECH 2011 Speaker State Challenge feature set
- The INTERSPEECH 2012 Speaker Trait Challenge feature set
- The INTERSPEECH 2013 ComParE feature set
- The MediaEval 2012 TUM feature set for violent scenes detection.

OpenSMILE ComParE_2016 feature set

- It has approximately 200 features divided into two groups of Low-Level Descriptors (LLDs).
  - 59 LLDs in group A, 54 functionals are applied
  - 59 delta LLDs of group A, 46 functionals are applied
  - 6 LLDs in group B, 39 functionals are applied
  - 6 delta LLDs of set B, 39 functionals are applied.
- This results in a total of 6,368 features.

LLD and Functionals

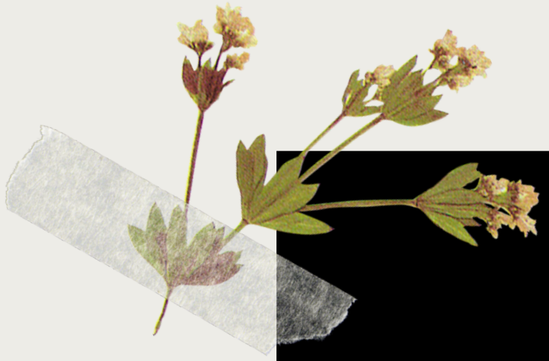
Table 3.5 INTERSPEECH 2013 Computational Paralinguistics ChallengE (Com- ParE) set

LLD	Functionals
Group A: (59)	Arithmetic^{A*,B} or positive arithmetic$^{A\delta,B}$ mean,
Loudness,	Root-quadratic mean, flatness,
Modulation loudness,	Standard deviation, skewness, kurtosis,
RMS energy, ZCR,	Quartiles 1–3,
RASTA auditory bands 1–26,	Inter-quartile ranges 1–2, 2–3, 1–3,
MFCC 1–14,	99th and 1-st percentile, range of these,
Energy 250–650 Hz,	Relative position of max. and min. value,
Energy 1–4 kHz,	Range (maximum to minimum value),
Spectral RoP .25, .50, .75, .90,	Linear regression slope^{A*,B} a, offset^{A*,B} b,
Spectral flux, entropy, variance,	Linear regression quadratic error^{A*,B},
Spectral skewness and kurtosis,	Quadratic regression coeff.^{A*,B} a, b, c,
Spectral slope,	Quadratic regression quadratic error^{A*,B},
Spectral harmonicity,	Temporal centroid^{A*,B},
Spectral sharpness (auditory),	Peak mean value^{A} and dist. to arithm. mean^{A},
Spectral centroid (linear).	Mean^{A} and std. dev.^{A} of peak to peak distances,
Group B: (6)	Peak and valley range^{A} (absolute and relative),
	Peak-valley-peak slopes mean^{A} and std. dev.^{A},
F_0 via SHS, Prob. of voicing,	Segment length mean^{A}, min.^{A}, max.^{A}, std. dev.^{A},
Jitter (local and delta),	Up-level time 25 %, 50 %, 75 %, 90 %,
Shimmer,	Rise time, left curvature time,
logḢNR (time domain).	Linear Prediction gain and coefficients 1–5.

Overview of Low-level Descriptors (LLDs) and functionals applied to these LLDs. Functionals marked with ^{A} and ^{B} are only applied to group A or B LLDs (and deltas), respectively; functionals marked with * or $^{\delta}$ are not or only (respectively) applied to the delta LLDs. Details in Appendix A.1.5

Work Completed So Far

1. Downloaded and understood the dataset and its divisions.
2. Set up openSMILE
3. Extracted the ComParE_2016 feature set for utterances in our dataset.
4. Worked on an automated script for extracting ComParE_2016 feature set for all .wav files in the dataset.
5. [In progress] Writing the code to extract another feature from the dataset.



Demo

...

Thank you!

Presented by
GROUP 8

Harshita Sharma & Aashna Jena