Automatic Speech Analysis for the Detection of Emotional Disturbances in Persons with Dementia

Alexandra König [1,2], Nicklas Linz [3], Johannes Tröger [3], Aurore Rainouard [1], Auriane Gros [1], Jan Alexandersson [3], and Philippe Robert [1]

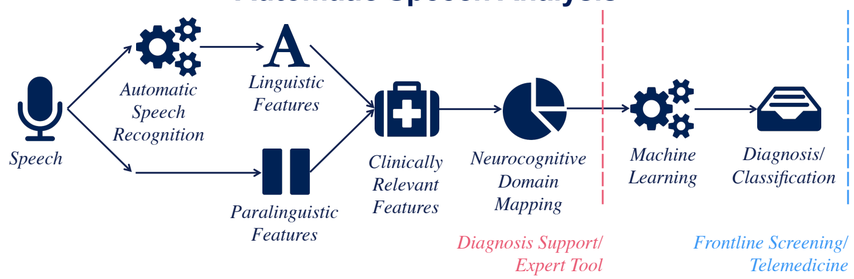
[1] CoBTeK - Cognition Behaviour Technology Research Unit, Université Côte d'Azur, Nice, France, [2] INRIA – French Institute for Research in Computer Science and Automation, Sophia Antipolis, France,

[3] German Research Centre for Artificial Intelligence, Saarbrücken, Germany

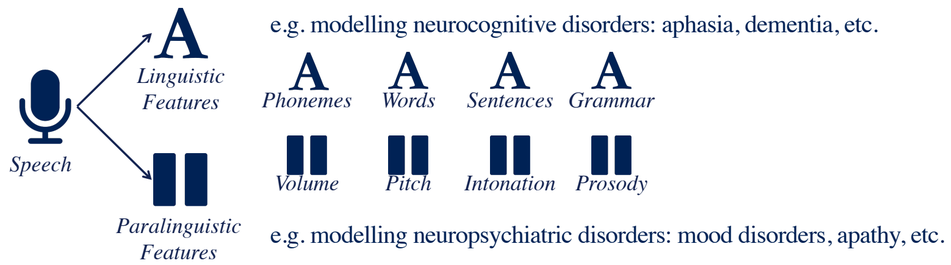
ELEMENT

Emotional disturbances found in dementia patients such as depression, anxiety or apathy have an important impact on quality of life of both patients and their caregivers and represent a strong predictor for illness progression. However, they are often underdiagnosed since current assessment tools rely primarily only on subjective measures.

Automatic Speech Analysis



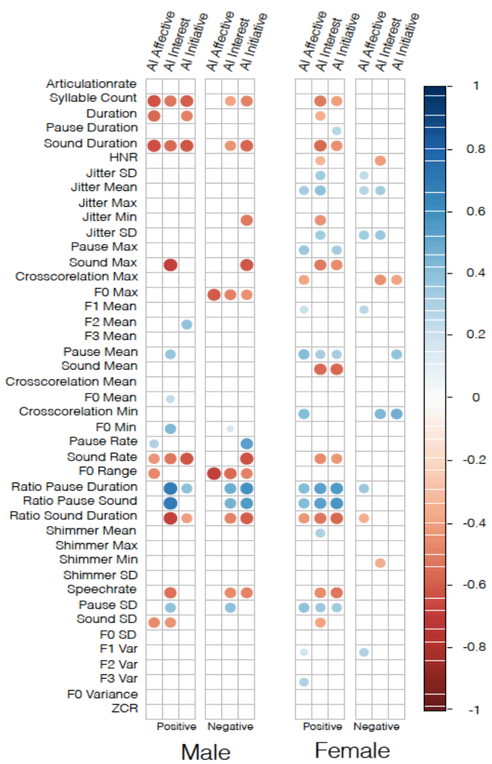
New computational approaches may allow a more objective evaluation of these behaviors that humans struggle to quantify. Therefore the study aims to investigate whether ***automated analysis of linguistic and paralinguistic biomarkers*** derived from audio recordings could be useful to determine emotional disturbances in dementia patients and thus demonstrate potential to improve early diagnosis.



Methods: 150 participants including both dementia patients and healthy control subjects were recorded while answering several mood--related open questions (i.e. 'can you tell me something about a pleasant event coming to your mind ?') along with a cognitive assessment Apathy Inventory.
Speech signal processing techniques were applied to extract features to compare to these baseline scores.

Table 2: Feature definition of acoustic markers. Name, definition and intuition of features sorted by category is presented.

Category	Feature	Definition	Intuition
Prosodic	F_0	Mean, Max, Min, Range,, Variance and Standart deviation of F_0	Statistics over the perceived auditory pitch (speech melody)
	Kurtosis	Mean kurtosis of signal amplitudes	Measure of 'tailedness' of a signals amplitude
	Skewness	Mean skewness of signal amplitudes	Representative for a lack of symmetry in the distribution of the amplitude of the speech signal. Perceived as a 'tense' voice
	Periodicity	Mean, Max and Min cross-correlation of speech signal	Measure of the regularity of the speech signal
Formant	$F_1 - F_3$	Mean and Variance of the first three formant frequencies	Indicative of the class of speech sound
Source	*Jitter*	Average absolute difference between consecutive signal periods, divided by the average period length	Indicative for a lack of control for vibration of the vocal cords
	Shimmer	Average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude	Indicative for reduction of glottal resistance
	Harmonics-to-Noise-Ratio (HNR)	Ratio between periodic components and non periodic components comprising voiced speech	Measure of voice quality
	Normalised amplitude quotient (NAQ)	Normalised ratio between the amplitude of the glottal flow pulse and the negative amplitude of the main excitation in the flow derivative	Measure of phonation type / glottal closure
	Quasi-open quotient (QOQ)	Ratio between the duration of a positive wave (pre- and post-peak amplitudes) and the local glottal period	Measure of vocal intensity
Temporal	*Sounding segments*	Mean, Max, Min and Standart Deviation of sounding segment lengths determined based on intensity	Statistics over length of connected speech segments
	Silence segments	Mean, Max, Min and Standart deviation of silence segment lengths determined based on intensity	Statistics over length of continuous pause segments
	Duration	Total length of recording	Total length of recording
	Speech duration	Total length of all sounding segments	Amount of speech
	Silence duration	Total length of all silence segments	Amount of pause
	Speech proportion	Ratio of *Speech duration* and *Duration*	Proportion of recording participant is talking
	Speech Rate	Ratio of number of syllables, detected using [11], and *Duration*	Measure of information density
	Articulation Rate	Ratio of number of syllables, detected using [11], and *Speech duration*	Measure of speech tempo



	Male		Female	
	N	A	N	A
N	12	12	16	16
Age	78.75 (4.97)	79.58 (5.45)	78.88 (5.41)	78.62 (5.48)
MMSE	23.33 (3.55)	19.42 (4.17)	22.50 (4.43)	19.69 (5.79)
AI	1.5 (1.17)	6.0* (1.60)	0.44 (0.96)	5.19* (1.97)
AI-Intr	0.67 (0.89)	2.42* (0.90)	0.12 (0.34)	2.38* (0.39)
AI-Init	0.83 (0.94)	2.67* (0.89)	0.31 (0.70)	2.12* (1.09)
AI-Affect	0.00 (0.00)	0.92*(0.90)	0.00 (0.00)	0.69* (1.14)
NPI-Apathy	1.50 (2.07)	6.50* (3.73)	0.38 (0.72)	4.88* (2.58)
NPI-Depression	0.58 (0.90)	1.50 (2.68)	0.25 (0.58)	1.69 (3.05)
NPI-Anxiety	1.50 (2.07)	2.75 (3.33)	0.94 (1.12)	3.50 (3.65)

Table 1: Demographic data for population by gender and apathy; Mean (standard deviation); Significant difference ($p < 0.01$) from the control population inside each gender in a Wilcoxon-Mann-Whitney test are marked with *.

Abc.: MMSE='Mini Mental State Examination', AI='Apathy Inventory', AI-Intr='AI domain Interest', AI-Init='AI domain Initiative', AI-Affect='AI domain affective', NPI='Neuropsychiatric Inventory', NPI-Apathy='NPI domain apathy', NPI-Depression='NPI domain depression', NPI-Anxiety= 'NPI domain anxiety'

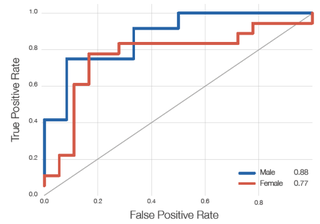


Figure 2: Receiver Operator Curve (ROC) of classifiers trained to detect apathy from speech. The blue and red line represents classifiers trained and evaluated on the male and female populations respectively. Area under the curve (AUC) is reported in the legend.

Results: Vocal features such as speech rate or sound duration correlate significantly with severity levels of certain emotional disturbances like apathy. Classification between pathological and non-pathological groups based on the extracted features obtain 88% accuracy.

Origin	Feature	Significance	Statistic χ^2	Effect size ρ	Direction
	Ratio Pause Duration	**	10.62	0.54	↑
	Ratio Sound Duration	**	10.62	0.54	↓
	Ratio Pause Sound	**	9.61	0.52	↑
	Sound Max	**	6.73	0.43	↓
	Sound Mean	**	8.29	0.48	↓
	Sound Duration	**	8.66	0.49	↓
Positive	Pause Mean	**	6.73	0.43	↑
	Pause Max	*	5.48	0.39	↑
	Pause SD	**	7.06	0.44	↑
	Syllable Count	**	7.23	0.45	↓
	Speechrate	**	8.11	0.47	↓
	Jitter Mean	*	5.33	0.38	↑
	HNR	**	6.73	0.43	↓
	Crosscorrelation Min	**	8.11	0.48	↑
	Crosscorrelation Max	**	7.93	0.47	↓
Negative	Jitter Min	*	5.41	0.39	↓
	Jitter Mean	*	5.05	0.37	↑
	Jitter SD	*	5.33	0.38	↑
	HNR	**	9.42	0.51	↓

(b) Comparison for female population

Origin	Feature	Significance	Statistic χ^2	Effect size ρ	Direction
	Duration	*	5.60	0.39	↓
	Ratio Pause Duration	**	6.75	0.43	↑
	Ratio Sound Duration	**	6.75	0.43	↓
	Ratio Pause Sound	*	5.60	0.39	↑
	Sound Max	*	6.45	0.42	↓
Positive	Sound Mean	*	5.33	0.38	↓
	Sound Duration	***	13.23	0.61	↓
	Pause Mean	*	4.56	0.36	↑
	Syllable Count	***	11.81	0.57	↓
	Speechrate	**	9.36	0.51	↓
	Ratio Pause Duration	*	6.16	0.41	↑
	Ratio Sound Duration	*	6.16	0.41	↓
	Ratio Pause Sound	*	5.60	0.39	↑
	Sound Duration	*	6.45	0.42	↓
	Sound Max	*	4.08	0.34	↓
	Pause SD	*	6.45	0.42	↑
Negative	Pause Mean	*	4.32	0.35	↑
	Pause Max	*	4.08	0.37	↑
	Syllable Count	*	3.85	0.33	↓
	Speechrate	*	5.88	0.40	↓
	F_0 Range	**	9.72	0.52	↓
	F_0 Max	**	9.36	0.51	↓

(a) Comparison for male population

Table 3: Statistical group comparison between non-apathetic and apathetic group using Kruskal-Wallis tests. Features with $p < 0.05$ are reported. Vocal task of origin, p-value, test statistic (χ^2), effect size (ρ) and direction of effect in the apathetic group in comparison to the non-apathetic group are reported.
* : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$

Voice--based emotion detection has many potential applications in healthcare for screening and diagnosis, even remotely, but also in helping identify new behavioral markers of emotional disturbances, as well as to measure intervention response, or test clinical theories about underlying mechanism.

The ***ELEMENT*** project is developing solutions that enable assessment of neurocognitive functions as well as certain neuropsychiatric symptoms. It is based on speech analysis and artificial intelligence, enabling multiple use cases ranging from non-clinical telemedicine screening to expert diagnosis support settings. This results in a faster and earlier diagnosis leading to timely intervention.

Contact: alexandra.konig@inria.fr

German Research Center for Artificial Intelligence

CoBTeK Cognition Behaviour Technology

ki elements

Inria

eit Digital