PAPER • OPEN ACCESS

Voice signal analysis techniques for cognitive decline (stress) assessment

View the article online for updates and enhancements.

Voice signal analysis techniques for cognitive decline (stress) assessment

C P Loizou[1]* and P Christodoulides[2]

[1] Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, Limassol, Cyprus
[2] Faculty of Engineering and Technology, Cyprus University of Technology, Limassol, Cyprus

christosl.loizou@cut.ac.cy

Abstract. Speech signals may provide important information for measuring and modelling human behaviour, especially for assessing mental health as well as estimating the emotional state of a person. Speaker's physiological and/or physical state may be thus identified by detecting the cognitive decline (CD) or stress levels using signal analysis of voice. This preliminary study presents a survey on methods introduced for CD and stress voice detection in humans. It is shown that increase in signal's fundamental frequency (f_0) as well as its frequency formants, are the most common effects of CD and stress. Additional voice parameters could be used to identify normal from CD and normal from stressed voice as well as the cognitive state of the elderly. The present study poses the initiation of a project for the development of a mobile application for the automated voice detection and analysis on the fly, which will aid in the detection of early signs of CD and stress. Further investigation with application on a large number of voice samples is required for validating the methods.

Keywords: Cognitive decline and stress voice analysis; Fundamental frequency; Voice analysis; Formant frequencies index.

1. Introduction

The transitory step between physiological aging and dementia is known as mild cognitive impairment (MCI). An early diagnosis of the decline of cognitive function is crucial to allow patients to get anticipatory access to pharmacological and cognitive stimulation therapies, which has been shown to increase the duration of independent living 0. Psychological stress is specified as a psychological state that is a reaction to a perceived thread or task demand and is accompanied by specific emotions such as (fear, anger, disgust, sadness, etc.) 0. The importance of detecting physiological stress, lies in its long-term occurrence, which may have serious health consequences **Ошибка! Источник ссылки не найден.**, as well as mental degradation 0. The characteristics of the voice in stress and cognitive decline (CD) are well associated 0-0. The importance of human voice in CD and stress assessment has been shown by many studies 0, 0 as well as that voice is affected by speaker's emotions which influences the vocal tract 0, 0. Characteristics of the voice signal such as glottal spectral slope, its fundamental frequency (f_0, the lowest frequency of the voice signal or pitch) may offer information of the voice stress levels 0, which is affected by a speaker's emotions. A detailed definition and discussion on human voice CD and/or stress is given in 0. It was also documented in other studies 0-0 that the f_0 and the signal jitter, increases with stress, due to a stress-associated tensing of the musculature, specifically the cricothyroid muscle. Additionally, subglottal pressure, jitter, shimmer, maximum airflow declination rate, voice onset, vocal intensity and speaking rate could contribute to CD or stress increase levels. The f_0, of the voice is reported to increase by 3-6 Hz per unit measure (cmH_2O) in subglottal pressure 0.

In the present work, we performed a brief survey on various methods presented in the literature for stress voice detection and/or CD and identify quantitative parameters from normal and stressed voice

signals, as well as from normal and CD voice which may be used to detect CD and/or stress in voice. Based on this initial work, we aimed at developing a mobile application, able to automatically compute acoustic features on-the-fly during phone conversations. Based on the extracted acoustic features and on a learning-based classifier, the mobile app will discriminate between people with a normal and a mildly impaired cognitive function and/or physiological stress, becoming a tool for the daily cognitive monitoring of elderly people in a transparent and non-intrusive manner. Our first aim was to identify potential literature on the subject of study which is in this paper here below presented.
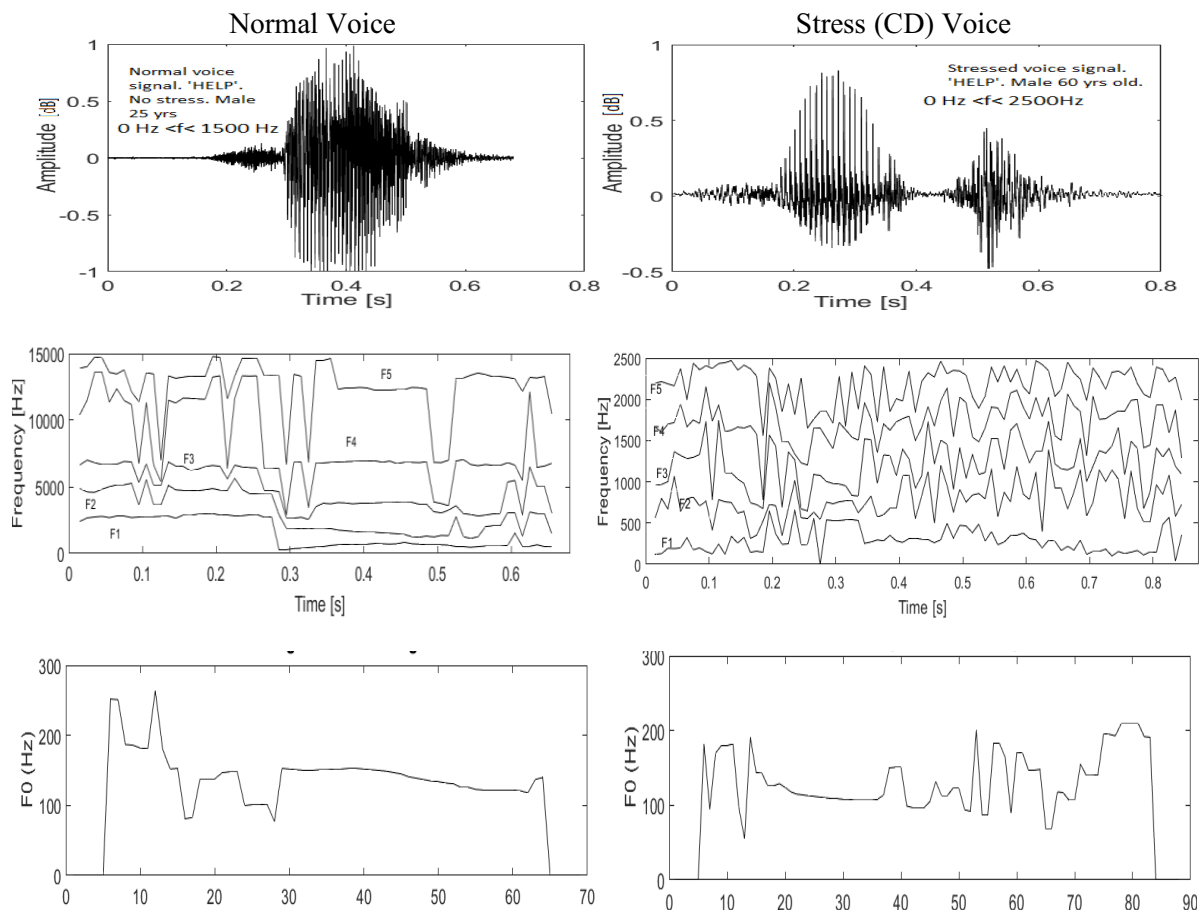
Figure 1. Normal voice signal for the word 'Help', (left column)), CD or Stress voice signal for the word 'Help' (right column). Extraction of the first five frequency formants (f_1-f_5) for the normal and the stressed words (second row, left and right). The fundamental (f_0) frequency (third row) for the normal and the stressed voice in the left and right columns respectively, for the word 'HELP'. The voice signals duration of the signals has a duration of 0.6 secs and a voice amplitude ranging from –1 to 1 dB. The f_0 pitch extraction for the normal and the stressed voice signals (last row left and right) was done using the Summation of Residual Harmonics (SRH) technique **Ошибка! Источник ссылки не найден.**.

2. Main features sets extracted

The first row of figure 1 illustrates the original normal and stressed (or CD) voice signals for the word 'Help' (left and right), while the second row of figure 1, presents their first five frequency formants. A larger variability for all formant frequencies is observed for the stressed signal in figure 1, with stronger peaks and fluctuations, when compared with those of the normal voice signal. The third row of figure 1 presents the f_0 of the normal (left) and stressed (right) voice signals respectively. A larger variability may be observed for the stressed voice signal.

The f_0 (see third row of figure 1), and its first five formant frequencies (f_1-f_5) 0 (see second row of figure 1), are usually estimated from the voiced areas of the signal (0.2 to 0.6 secs in figure 1 and figure 2), in order to estimate possible stress voice levels 0-0. Additional voice signal features that could be extracted for estimating stress voice levels include, the signal pitch 0, 0, 0, 0, 0 (see also third row of figure 1), the signal shimmer 0, jitter 0, and the frequency formants (usually the first five, f_1-f_5, as shown in the second row of figure 1) 0, 0. Furthermore, the glottal closure instants (closings and openings of the glottal signal) 0, could provide additional information on the voice stress levels as well as voice parameters as proposed in 0.

3. Voice stress and CD analysis studies

Table 1 illustrates voice stress and CD analysis and detection techniques proposed in the literature, tabulated under chronological order. The method, problem investigated, medical condition assessed, and signal characteristics investigated are also given. Furthermore, we provide details on the number and type of subjects each method was applied. Finally, in the last column of Table 1, we provide few important results derived from each study.

The voice corpus used in this study is the simulated dataset of SUSAS 0. It consists of 9 male speakers, uttering isolated words in a quiet environment. In this work only normal versus stress types were used 0.

A number of voice analysis studies (also presented in Table 1), have been proposed in the literature, for detecting voice under stress or CD in voice. These includes, measuring, the overall intensity, Teager energy operator, Mel-Frequency Cepstral Coefficients, and the extractions of the functionals of f_0 0. More recently, stress was detected using spectral slope measurements 0, based on the probability density function of the glottal source. Other studies focus on depression detection based on facial landmark features detection 0, using voice acoustic features 0, or using facial cues from videos 0. In 0, the Stroop test for stress voice detection was utilized, whereas linear predictive coding was used in 0. However, with the exception of 0, no other study has been reported where stress was investigated in the elderly. Stress may be observed in voice signals, having larger amplitude and more asymmetrical glottal pulse as compared to neutral condition. These changes have an impact in the intensity of the spectrum as stress is concatenated in higher frequencies 0. Additionally, the distribution and/or spectral tilt of the spectrum's energy has to also be considered 0. Important characteristics may be extracted from the voice signal (intensity, pitch and harmonics), detected by a sensor, which may signalize stress levels. Stress identification in voice may be used for monitoring elderly subjects in their environment, which might be presenting a progressive decline in mental abilities and provide feedback to caregivers 0. This maybe estimated as a subtle linguistic impairment by using voice analysis. These impairments may be detected using the f_0 and the f_1-f_5 of the voice signal (see also figure 1), as it was also shown in other studies 0, 0, 0, 0, 0-0.

There is only one study 0, found in the current literature, where CD was investigated in 39 elderly subjects (see also table 1). However, the analysis in 0, was done through a neurophysiological screening for visuospatial abilities (memory, language, executive functions, and attention). A large set of linguistic features was extracted for following up the development of CD using classification analysis. It was found in 0, that these features are very important parameters in the analysis of cognitive alterations. It was furthermore shown, that some linguistic features may be used to discriminate between healthy and impaired elderly subjects (spectral centroid mean, statistics of voice signal, openings and closing voice intervals and f_0-f_5).

Although a number of techniques and evaluation variables have been proposed in the literature for stress detection (see also table 1), it should be noted that the vocal response to stress may be as unique as the voice itself. Findings from the literature review are heterogeneous, but in almost all studies investigated, it was indicated that a consistent universal trend of an increase in f_0 is demonstrated with the voice stress (see also figure 1 and table 1). An explanation of this increase maybe stress-associated tensing of the musculature (cricothyroid muscle) 0. Other possible explanations of this increase may include increase in heart rate, blood pressure and bronchodilation 0. In addition, reduced noise as

reflected in signal jitter 0, and shimmer 0, and an associated increase in vocal jitter is documented, but less frequently.

A limitation in evaluating stress and proposing objective parameters contributing to vocal quality, is the difficulty in reproducing natural stress in the laboratory. Additionally, different environmental parameters, noise, gender, sex, and the response to stress, which varies greatly by individual may contribute to actual stress reproducibility 0-0. Therefore, still the need exists for finding new and more reliable indicators or variables of voice stress. Perhaps complex constructs such as personality, birth and others may also be taken into consideration 0, 0. Further investigation with a large number of samples is required for validating the methods and for estimating reliable indicators of stress and voice evaluation metrics.

4. Conclusions

In this paper we presented a survey on methods introduced so far in the current literature for stress voice detection in humans. However, within this literature research we aimed at developing a monitoring for detecting CD and/or stress levels from voice signals. To reach this aim, we plan to use free speech during phone calls, without the need to record any voice samples. The current work will pose the basis for the development of a mobile application, which could be use daily favouring a longitudinal assessment of the cognitive function of elderly people and therefore supporting the detection of early signs of functional cognitive decline. From the present research it was shown that increase of f_0 as well as alterations on its frequency formants (f_1-f_5), are the most common effects of CD and/or stress. Additional voice parameters could be used to identify normal from CD and/or stressed voice as well as the cognitive state of the elderly. Further investigation with a larger number of samples is required for validating the methods, a task that is currently investigated from our group.

5. Acknowledgments

Table 1. Voice stress detection and analysis techniques proposed in the literature.

Investigator	Year	Problem	Method	Condition	Signal Characteristics	Database (Nr)	Age	Results
Zhou 0	2001	S voice classification	Non-linear TEO	TEO-FM-Var, TEO-Auto-Env, TEO-CB	f_0, TEO, FM component	Simulated vs S voice N=43 0	NA	TEO-CB-Auto-Env better performance. N, A, Lou, Lo
Rothkrantz 0	2005	Pathological voice stress	Stroop test	Loudness, f_0, jitter, zero-crossings 0	Jitter	Nr=108 Native Dutch speakers	NA	Most important: Jitter and f_0
Naylor 0	2007	GCI	Dypsa	Phase-slope function and DP	GCI	APLAWD (10M/5F), SAM (4)	NA	Correct GCI detection 95.7%
Lefter 0	2011	Emotions: Military	Ambulance telephone calls	Emotions parameters (A, Fr, S, N)	Spectral and prosodic features	SUSAS, SUS-0/1, DLP 0, 0(N=3000)	NA	f_0-f_n, spectrum. Training required. LLR better detects stress
Sathe-Pathak 0	2012	3 emotional states (NU, H, S)	CAM, PD, LPC, f_0-f_n	NU: 270< f_1 <730Hz, H: 840< f_2 <2290Hz, S: 1690< f_3 <3010Hz,	Pitch, energy, intensity, F_0-F_n	Nr=10: TU Berlin (5 F+5M)	NA	N/A: f_0 = 130/336Hz H: $f_1/f_2/f_3$ = 338Hz/802Hz/1628Hz
Giddens 0	2012	Voice stress indices	Voice recordings	F_0, Jitter most common	F_0 and Jitter incre. with stress	Electronic databases	NA	Review 50 studies: F_0 and jitter most common
Patil **Ошибка! Источник ссылки не найден.**	2013	Stress: true versus lie	MFCC 0, FFT	Voice stress analysis	Fourier and graph characteristics	5 persons answered questions	NA	True vs lie can be identified
Sigmund 0	2013	Stress, M/F: f_0=80-170Hz / 150-260Hz), Children: f_0 = 300-500 Hz	Auto-correlation	M/F: f_0 = 110Hz / 210Hz, N vs S	f_0, mean (μ), standard deviation (σ), skewness (σ³), kurtosis (σ⁴)	14 M speakers under stress.	NA	N / S: M subjects: μ = 115 / 140, σ = 7 / 12, σ³ = 0.0139 / 0.126, σ⁴ = 4.13 / 0.99, f_0 higher in S
Black 0	2013	Identify gender in married couples	Classification	Prosodic, spectral, voice quality features	voice-derived audio cues	Nr=63 (age 18-65)	42±8.6	74.1% correct classification
Dixit 0	2014	MC	Time, frequency, LPC	MC influences voice parameters	f_0, pulse, pitch, harmonics	None	NA	Review study. MC affects voice parameters
Mongia 0	2014	Stress vs vocal tract function	51 Parameters, LPC, MFCC	f_0-f_n + 51 other statistical measures	Pitch, plots, zero-plots	SAVEE database 0	NA	Vocal tract function for each vowel. 25 out of 51

Table 1. Voice stress detection and analysis techniques proposed in the literature.

Investigator	Year	Problem	Method	Condition	Signal Characteristics	Database (Nr)	Age	Results
Sondhi 0	2015	Emotional stress vs voice signal changes	f_0, f_1-f_4, jitter and shimmer	f_0-f_n, Spectrograms distinct	PRAAT software analysis pitch.	11 subjects, questions vs answers. Radio broadcasts	NA	(N/S voice): f_0: 185 / 216, f_1: 336 / 269Hz, f_2: 2256 / 1941Hz, f_3: 2860 / 3142Hz, f_4: 4267 / 3987Hz, Jitter: 1.36/1.18%, Shimmer: 8.6/6.82%
Meilan 0	2014	Alzheimer's vs normal voice measures	Spectrogram, f_0, jitter, shimmer	Distinguish dementia in Alzheimer's via voice analysis	PRAAT software analysis	Alzheimer's (N=30) vs Normal (N=36)	78.6±9 .4/ 74±9.7	(N/Alzheimer's): f_0: 161±25 / 180±29, Jitter: 2.83±0.61 / 2.8±0.95%, Shimmer: 12.9±4.1/13.3±4.7%
Koenig 0	2015	Predementia & Alzheimer's	Voice analysis	Alzheimers vs MCI vs normal	PRAAT software analysis	Alzheimer's=26, Normal=15, MCI=23	76 (70-82)	Three groups could be classified successfully using vocal markers
Simantiraki 0	2016	Emotional stress	Spectral tilt of glottal source	f_0, spectral slopes	Stress affects higher frequency regions	SUSAS 0, 9 M	NA	92.06% recognition accuracy
Beltrami 0	2016	CD in elderly. Dementia	OPLON project	Linguistic features extracted to identify cognitive frailty	Acoustic, Rhythmic, Lexical features	96 elderly subjects (48 healthy/48 with CD)	50-75	Significant differences found between features for NU vs CD
Kumar 0	2016	A, H, S, E (Stress levels)	LPC	Peak accuracy for S=70%	Pitch, power, f_0-f_n, prosodic features	5 Simulations compared with TU Berlin	NA	Peak accuracy (A/H/S/E) :70%, 55%, 82%, 77%, (TU Berlin vs 0)
Rektorova 0	2016	Cognitive decline	Acoustic voice analysis	Prediction of cognitive decline	f_0, REM questionnaire	44 Parkinson's patients	66±6.1	Variation of the f_0, rhythmicity speech index most significant.
Pambouchidou 0	2016	Depression	Video, text, voice signal analysis	Visual vs voice parameters	All above	68 Facial landmarks, recordings	NA	All signal measures outperformed the visual measures

S: Stressed voice, GCI: Glottal closure instants, TEO: Teager Energy Operator, TEO-FM-Var: TEO-decomposed frequency modulation variation, TEO-Auto-Env: Normalized TEO autocorrelation envelope area, DP: Dynamic programming, TEO-CB-Auto-Env: Critical band based TEO autocorrelation envelope area, CAM: Cepstral analysis method **Ошибка! Источник ссылки не найден.**, PD: Pitch detection, f_0: Fundamental frequency of the signal, f_0-f_n: Formant frequencies, LPC: Linear predictive coding **Ошибка! Источник ссылки не найден.**, LLR: Linear Logistic Regression, Neutral: NU, Happy: H, Sad: S, Excitement: E, Angry: A, Lo: Lombard, Lou: Loud, Frustration: Fr, MFCC: Mel frequency Cepstral coefficient, M: Male subject, F: Female subject, FFT: Fast Fourier transform, μ: mean, σ: standard deviation, σ^3: skewedness, σ^4: kurtosis, MC: Asthma, Aphasia, Parkinson, Depression, Schizophrenia, Autism, Cancer, CD: Cognitive decline, MCI: Mild cognitive impairment.

6. References

[1]   Hansen J H L and Patil S A 2007 Speech under stress: analysis, modeling and recognition *Speaker classification I: Fundamentals, features, and methods*, ed C Mueller (Berlin, Germany: Springer) pp 108–137

[2]   Selye H 1956 *The stress of life* (McGraw-Hill)

[3]   Sigmund M 2012 Influence of psychological stress on formant structure of vowels *Elektronika ir Elektrotechnika* **18(10)** pp 45–48

[4]   Lefter I, Rothkrantz L J M, van Leeuwen D A and Wiggers P 2011 Automatic stress detection in emergency (telephone) calls *Int. J. Intellig. Defence Support Systems* **4(2)** pp 148–168

[5]   Giddens C L, Barron K W, Byrd-Crave J, Klark K F, *et al* 2012 Vocal indices of stress: A review *J. of Voice* **27(3)** pp 21-29

[6]   Zhou G J, Hansen J H L and Kaiser J F 2001 Nonlinear feature based classification of speech under stress *IEEE Trans. Speech & Audio Proces.* **9(3)** pp 201–216

[7]   Godin K W, Hasan T and Hansen J H, 2012 Glottal waveform analysis of physical task stress speech *Interspeech* pp 1648–1651

[8]   Simantiraki O, Giannakakis G, Pampuchidou A and Tsikankis M 2016 Stress detection from speech using spectral slope measurements *MINDCARE, 6th Int. Symp. on pervsive computing paradigms for mental health* pp 1–10, Nov. 28-29, Barcelona, Spain

[9]   Nasir Md, Jati A, Shivakumar P G, Chakravathula S N and Georghiou P 2016 Multimodal and multiresolution depression detection from speech and facial landmark features *Proc. 6th Int. Workshop Audio/Visual Emotion Challenge (AVEC)* pp 43–50

[10]   Sluijter A M and van Heuven V J 1996 Spectral balance as an acoustic correlate of linguistic stress *The J. Acoustical Society of America* **100(4)** pp 2471–2485

[11]   Hansen J H, Bou-Ghazale S E, Sarikaya R and Pellom B 1997 Getting started with SUSAS: a speech under simulated and actual stress database *Eurospeech* **97(4)** pp 1743–1746

[12]   Rothkrantz L J M, Wiggers P, van Wees J-W A and van Vark R J 2005 Voice stress analysis *Int. Conf. on text, speech and dialogue* pp 449–456

[13]   Kumar D, Sharma U and Solanki S 2016 Voice stress analysis using linear predictive coding in Matalb *Int. J. Emerging Technol. Eng. Research (IJETER)* **4**(7) pp 210–213

[14]   Black M P, Katsamanis A, Baucom B R, Lee C-C, *et al* 2013 Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features *Speech Commun.* **55** pp 1–21

[15]   Giannakakis G, Pediaditis M, Manousos D, Kazantzaki E, *et al* 2017 Stress and anxiety detection uisng facial cues form videos *Biomned. Sign. Proces. Control* **31** pp 89–101

[16]   Khanagha V, Daoudi K, Yahia H, 2014 Detection of glottal closure instants based on the microcanonical multiscale formalism *IEEE Trans. Audio, Speech & Lang. Proc.* **22(12)** pp 1941–1950

[17]   Sathe-Pathak B V and Panat A R 2012 Extraction of pitch and formants and its analysis to identify 3 different emotional states of a person *IJCSI Int. J. Computer Science Issues* **9(4)** pp 296–299

[18]   Patil V P, Nayak K K and Saxena M 2013 Voice stress detection *Int. J. Electrical Electronics* **2(2)** pp 148–154

[19]   Sigmund M 2013 Statistical analysis of fundamental frequency based features in speech under stress *Inf. Techn. & Control* **42(3)** pp 286–291

[20]   Dixit M V and Sharma Y 2014 Voice parameter analysis for the disease detection *IOSR J. Electr. & Comm. Engin.* (IOSR-JECE) **9(3)** pp 48–55

[21]   Mongia P K and Sharma R K 2014 Estimation and statistical analysis of human voice parameters to investigate the influence of psychological stress and to determine the vocal tract transfer function of an individual *J. Comput. Networks & Communic.* **2014** (290147) pp 1–17

[22]   Sondhi S, Khan M, Vijay R and Salhan A K 2015 Acoustic analysis of speech under *stress Int. J. Bioinformatics Research & Applications* **11(1)** pp 417–432

[23] Beltrami D, Calza L, Gagliardi G, Ghidoni E, *et al* 2016 Automatic identification of mild cognitive impairment through the analysis of Italian spontaneous speech productions *Proc. 10th Int. Lang. Resourc. & Evaluat. (LREC 2016)* Slovenia pp. 2086-2093 May 23–28

[24] Pambouchidou A, Pediatis M, Giannakis G, Marias K, *et al* 2016 Depression assessment by fusing high and low level features from audio, video and text *Proc. 6th Int. Work. Audio/Visual Emotion Challenge* The Netherlands pp 27–34 Oct. 16-18,

[25] Naylor P A, Kounoudes A, Gudnason J and Brokes M 2007 Estimation of glottal closure instants in voiced speech using the Dypsa algorithm *IEEE Trans. Audio, Speech, & Sign. Process.* **15(1)** pp 34–43

[26] SAVEE Database Website, http://personal.ee.surrey.ac.uk/Personal/P.Jackson/SAVEE/

[27] Meilam J J, Martinez-Sanchez F, Carro J, Lopez D E, *et al* 2014 Speech in Alzheimer's disease: can temporal and acoustic parameters discriminate dementia? *Dement & Geriatr. Cognit. Disord.* **37(5-6)** pp 327–334

[28] Koenig A Satt A Sorin A *et al* 2015 Automatic speech analysis for the assemsnetn of patinets with predentia and Alzheimers disease *Alzheimers & Dement: Diagn. & Diseas. Monitor* **1(1)** pp 112–124

[29] Rektorova I, Mekyska J, Janousova E, Kostalova M, *et al* 2016 Speech prosody impairment predicts cognitive decline in Parkisons disease *Parkisonism & Relat. Disord* **29** pp 90–95

Authors' background

Name	*Prefix*	*Research Field*	*Email*
Christos P Loizou	Special Scientist	Signal, Image and video processing and analysis	christos.loizou@cut.ac.cy
Paul Christodoulides	Assistant Professor	Applied Mathematics	paul.christodoulides@cut.ac.cy