

# Psychological Stress Detection in Speech Using Return-to-opening Phase Ratios in Glottis

Miroslav Stanek<sup>1</sup>, Milan Sigmund<sup>1</sup>

<sup>1</sup>*Brno University of Technology, Department of Radio Electronics,  
Technicka 12, 61600 Brno, Czech Republic  
mirek.stanek@phd.feec.vutbr.cz*

**Abstract**—This paper is focused on investigation of psychological stress in speech signal using shapes of normalised glottal pulses. The pulses were estimated by two algorithms: the Direct Inverse Filtering and Iterative and Adaptive Inverse Filtering. Normalised glottal pulses are divided into opening and return phase, and a feature vector characterizing each glottal pulse is calculated for a series of  $n$ -percentage interval in time domain. Each feature vector is created by parameters describing its return-to-opening phase ratio, namely chosen intervals, kurtosis, skewness, and area. Further, psychological stress is detected by feature vector and four different classifiers. Experimental results show, that the best accuracy approaching 95 % is reached with Gaussian Mixture Models classifier. All the best results were obtained regarding only the interval of 5 % from both phase durations, i.e. for and after pulse peak, where the most significant differences between normal and stressed speech in feature vector are occurred. Presented experiments were performed on our own speech database containing both real stressed speech and normal speech.

**Index Terms**—Emotion recognition, glottal pulses, psychological stress, speech processing.

## I. INTRODUCTION

The first application of glottal pulses can be found in speech synthesis where precise understanding of glottal pulses and its estimation lead to high-quality synthetic speech. For instance, the novel method called Glottal Spectral Separation (GSS) is published recently by Cabral *et al.* [1]. By suitable combination of mixed excitation model and noise component, the high-quality speech can be produced by the GSS method using suitable combination of mixed excitation model and noise components. Another method of speech synthesis was introduced by Raitio *et al.* [2], where synthetic voice is utilized by Hidden Markov Models (HMM) and Iterative and Adaptive Inverse Filtering (IAIF) leading to subjectively highly natural synthetic speech. Similar HMM-based speech synthesizer based on the Liljencrants-Fant (LF) model of the glottal flow is published by Cabral *et al.* [3]. Glottal pulses can be also used in music, for instance the speech (sing) resynthesis [4].

The next application field of glottal pulses is so-called expressive speech processing used for expressing emotions, dynamic and varying voice quality and articulation during

phonation. In 1980, the dynamic changes varying on phonation type, exactly on glottal source signal, was published by Laver [5]. Differences between prosodic and glottal feature were statistically processed and published in [6], where glottal feature shows significant differences for all 30 emotion pairs contrary to prosodic features. Using the suitable combination of prosodic and glottal features for emotion recognition is also described in [7], where Support Vector Machine (SVM), Artificial Neural Network and Gaussian Mixture Models (GMM) classifiers were applied on Berlin emotional speech database. The symmetry of glottal pulse shape has been used for recognizing between six spoken emotional states [8] reaching average efficiency 66.5 % for well recorded speech signal and 47 % for noisy speech respectively. A number of observed parameters including glottal features are described in [9] varying on the type of psychological stress influence. A set of chosen speech features was tested and observed by different classifiers for gender and emotion recognition [10].

Glottal flow analysis can be also applied in speaker recognition [11]. The efficiency of glottal source component derived from Linear Prediction (LP) residual was preliminary experimentally tested for speaker recognition using Auto Associative Neural Networks models on total 20 speakers [12]. Other studies using, for instance, Glottal Flow Cepstrum Coefficients [13] and vocal source model [14] were experimentally tested in the case of speaker recognition.

Glottal pulse analysis can be applied also in biomedical field. Recently, the detection of Parkinson's disease from dysphonia measurements is described as a promising intermediate phase to non-diagnostic diagnostic method [15]. Glottal pulses can be also utilized for analysis vocal disorders [16], alcohol intoxication [17] as well as for Alzheimer's disease detection [18]. Other possible disease detection using voice analysis can be found in the review published by Saloni *et al.* [19]. In general, a survey oriented on glottal source processing and its applications was written by Drugman *et al.* [20].

## II. MINING THE GLOTTAL PULSES

Despite the years to be the research of obtaining the real glottal course from speech signal worked on, recently best results are reached only for the base of glottal flow estimation. Glottal flow can be characterized by a set of glottal pulses repeated by fundamental period  $T$ .

Manuscript received December 22, 2014; accepted April 13, 2015.

Research described in this paper was financed by Czech Ministry of Education in frame of National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used.

An example of glottal flow is illustrated in Fig. 1. Briefly, the whole glottal pulse is composed by two instances—primary opening  $T_o$  and return  $T_r$  phase. The space between particular pulses is called closed phase  $T_c$  during which the glottis is closed and the air does not flow through the gap. Detail description of each individual glottal flow part including physical changes and processes can be found in [21].

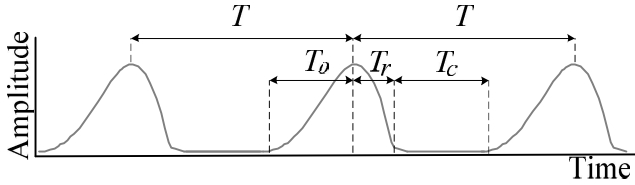


Fig. 1. Illustration of glottal pulses series and its description.

The mostly used methods for the estimation of glottal flow are DIF (Direct Inverse Filtering) and IAIF. Both methods are based on all-pole modelling of speech signal using LP analysis for considering the transfer function of vocal tract with impulsive or periodic source substituting glottis. The topic of inverse filtering and its impact on voice research and therapy was published by Lofqvist [22] and Nwachuku [23]. Other methods of glottal pulses estimation are described in [24].

Basically, the DIF method can be classified as a traditional autoregressive modelling-based inverse filtering method [25]. The IAIF estimating method can be simply described as a suitably connected serial-parallel combination of DIF pair. This method is described in detail in [26], characterized by better results of glottal flow estimation and it is more computationally challenging.

estimation is based on four parameters LF glottal model [27]. Thus, beginning, maximum position and the end of analysed glottal pulse (its absolute height and width) are defined by mined LF model and used in further processing.

Under psychological stress, people tend to make syllables, exactly entire words, shorter, therefore the stability of glottal pulse uniqueness has been also observed in dependency on the word duration. The first row of Fig. 2 illustrates the signal form (light grey) of the Czech syllable “u” containing mainly the vowel /u/ and estimated glottal pulses. The glottal flow of shortly spoken word is black (left column), the longer version of the same word is dark grey (right column). In the next subfigure (second row), both waveforms of estimated glottal flows are illustrated in the same time scale for showing the shape differences and fundamental periods  $T_{short}$  and  $T_{long}$  of the short and long versions of the same word, where fundamental frequency of shortly spoken word is little bit higher (82 Hz versus 78 Hz). Obviously, the glottal pulses do not vary upon the duration of spoken word nor fundamental period which leads to finding verification by two-dimensional normalisation of mined glottal pulses showed in the last row, where for example five normalised pulses are set over themselves for each speech tempo (see Fig. 2). Due to this fact, only individual glottal pulses are used in further experiments and not the whole glottal flow periods where the time interval of closed glottis is seemed to be not so representative in the case of characterizing the actual state of speaker.

### III. GLOTTAL FEATURE EXTRACTION

This section describes the method used for extracting chosen parameters, exactly their ratios. Basically, used method exploits only glottal pulses, composed by return and primary opening phase  $T_r$  and  $T_o$  (see Fig. 1). Each of mined glottal pulses is normalised to value 1 in time and amplitude domain leading to dimensionally uniform glottal pulses keeping original shape. In these two-dimensionally normalised pulses, the primary opening and return phase are processed separately. Both phases are transferred into relative time scale reaching the zero level at the position of current pulse's peak and the maximum (100 %) in both directions, i.e. at the end and at the start of current phases.

Used extraction method is based on the observation of both phases only for selected relative division  $n$ . Figure 3 shows the main idea of  $n$ -percentage glottal pulse processing of particular primary opening  $T_{o(n)}$  and return phase  $T_{r(n)}$  leading to following equation

$$RTO_n = \frac{T_{r(n)}}{T_{o(n)}}, [-], \quad (1)$$

where  $RTO$  is Return-To-Opening phase ratio of current  $n$ -percentage interval always symmetric for both used phases in relative scale.

Area, skewness and kurtosis (the third and fourth standardized Pearson's moments) are further calculated for both  $n$ -percentage intervals. Finally for each mined parameter value, the  $RTO$  phase ratio is calculated to sign the domination level of one  $n$ -percentage interval of current parameter.

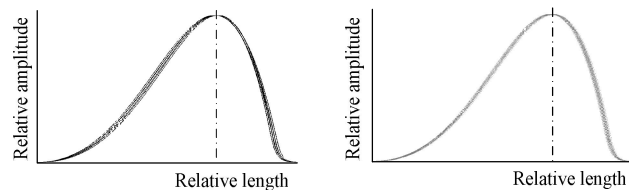
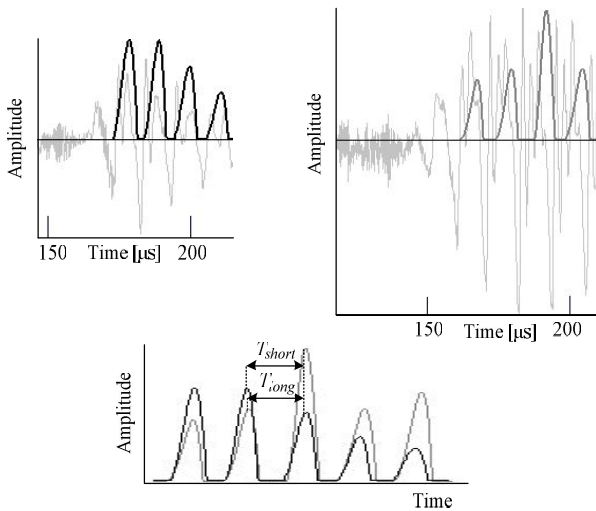


Fig. 2. Differences of glottal pulses depending on the speech tempo. Black signs the fast speech, i.e. shorter version of the same spoken word.

All analysed glottal pulses are mined in speech by modified version of software Aparat [21] where all pulse

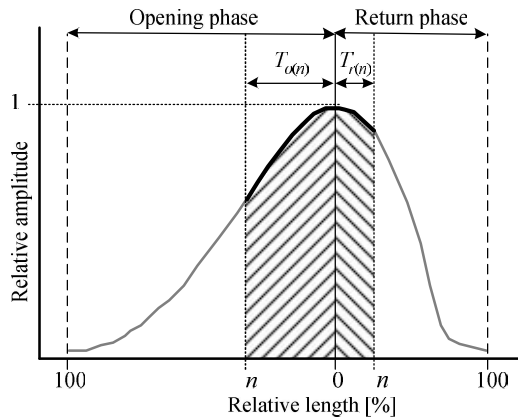


Fig. 3. Division of two-dimensionally normalised glottal pulse into  $n$ -percentage particular intervals of opening and return phase.

Obviously, each part of pulse curve (thick line segment in Fig. 3) corresponding to the  $n$ -percentage division is characterized by three different  $RTO_s$  (kurtosis, skewness and area). These feature values are further used for processing and observing differences between normal and stressed speech. For example, the real values of investigated  $RTO_s$  are listed in Table I, where all values are based on 5 % interval and averaged over all speakers.

TABLE I. AVERAGED REAL VALUES OF THREE RETURN-TO-OPENING PHASE RATIOS IN 5 % SELECTED INTERVAL, IAIF METHOD, NORMALISED SOUND VOWEL'S BEGINNING.

Vowel	$RTO_s$ [-]					
	Normal mood			Psychological stress		
	kurtosis	skewness	area	kurtosis	skewness	area
/a/	8.33	2.39	0.28	12.60	3.14	0.05
/e/	12.32	3.14	0.05	16.96	3.53	0.05
/i/	7.66	2.43	0.05	20.69	3.99	0.04
/o/	14.75	3.44	0.04	12.67	3.17	0.04
/u/	11.56	3.01	0.05	11.91	3.04	0.04

#### IV. REAL STRESS DATABASE

Research presented in this paper has been performed on created database containing speech under real psychological stress as well as normal speech. The first part of used database is formed by 18 different Czech speaking male speakers from ExamStress database [28] previously used for observation of vowel polygon differences varying on speaker's state [29]. Second part of used database is formed by another 6 Czech male speakers recorded by microphone PCB 378B02 suitable for infrasonic applications and sound interface USB-9234 produced by National Instruments. All Czech speakers in both parts of used database were recorded during the thesis defence in frame of final exam for capturing the real psychological stress influence. Few days later each speaker repeated the same text in more self-comfortable conditions for recording speaker's normal mood.

#### V. EXPERIMENTAL RESULTS

This section describes results achieved by realized experiments. In fluent speech performed by second part of used database (six speakers), Czech vowels were automatically detected and separated for further processing [30]. Then, separated vowels were manually divided to begging and centre vowel parts from which glottal pulses were estimated by DIF and IAIF methods. From the first

part of used database, vowels were separated manually in fluent speech and further were processed similarly to achieve the most pure training data further used in designed classifiers.

For naturally dynamic speech, the efficiency of emotional state (stress and normal mood) recognition was achieved for two types of glottal flow estimation methods (DIF and IAIF) in beginning and centre vowel part for 20 different  $n$ -percentage intervals (5 % to 100 % by step 5 %). Mentioned ways of efficiency testing were also applied on each 10 ms segment of normalised speech leading to the impact observation of glottal pulse uniformity on dynamic range limitation. The efficiency, exactly the uniformity of glottal pulses under normal and stress conditions, is tested by four different classifiers embedded in standard MATLAB version and further appropriately trained, validated and applied.

In following text, the recognition of used various glottal pulses are defined as:

- Method 1 for DIF, vowels' beginning
- Method 2 for DIF, vowels' beginning, normalised sound
- Method 3 for IAIF, vowels' beginning
- Method 4 for IAIF, vowels' beginning, normalised sound
- Method 5 for DIF, vowels' centre
- Method 6 for DIF, vowels' centre, normalised sound
- Method 7 for IAIF, vowels' centre
- Method 8 for IAIF, vowels' centre, normalised sound.

The  $k$ -Nearest Neighbour (kNN) was chosen as the first classifier. The best results are reached for the 5 % observed interval of glottal pulses, where the most significant differences are occurred between normal and stressed speech. Almost the efficiency of 95 % is reached by Method 2 on 5% selected interval. Further, accuracy over 90 % is reached by Method 1 and the Method 4 for 5 % and 10 % selected intervals. This method is the most successful on higher  $n$ -percentage intervals, where its recognition efficiency lies between 70 % and 80 %. The worst efficiency of kNN classifier was reached by Method 8 reaching efficiency values lower than 40 % for higher  $n$ -percentage intervals. Generally for kNN, the average recognition efficiency is approximately 60 % over all used methods and intervals.

The efficiency of stress detection of chosen classifier and actual  $n$ -percentage interval and method is calculated as follows

$$Efficiency(n) = \frac{N_{cdn} + N_{cds}}{N_n + N_s} \times 100, [\%], \quad (2)$$

where  $N_n$  is the total number of used normal state glottal pulses,  $N_s$  the total number of glottal pulses under psychological stress,  $N_{cdn}$  is the number of correctly detected normal mood glottal pulses and  $N_{cds}$  is the number of correctly classified stressed glottal pulses. The numeric example of calculated efficiency for kNN classifier and Method 3 on 35 % selected interval is computed as

$$Efficiency(35) = \frac{(815 + 941) \times 100}{1049 + 1176} = \frac{1756 \times 100}{2225} \cong 78.9\%. \quad (3)$$

Significant efficiency increase was obtained by the SVM classifier where average efficiency value approaches to 70 % over all methods and intervals. Generally, efficiency reached by SVM can be regarded as more satisfactory with more possible used  $n$ -percentage glottal pulse intervals for correct psychological stress detection. The best results approaching 95 % accuracy are received by Method 4 for 5 % interval as well as in the selected interval range 75 %-95 %. The most significant differences between observed features ratios of normal and stressed speech can be found in 5 % (Method 6 and Method 7) and 65 % (Method 3) selected intervals where also accuracy approaches to 95 %.

As the third classifier, GMM was used. The high efficiency values of psychological stress detection were reached over all possible  $n$ -percentage intervals of glottal pulses. On the other hand, the generally lowest accuracy values are also achieved by GMM, exactly for 10 % (Method 5) and for 35 % (Method 1) selected intervals where the accuracy approaches only to 10 % in stress detection. In some selected intervals, each method reaches the efficiency almost 95 % which signs the highest uniformity of observed features varying on actual state of speaker, and targets on GMM as a suitable classifier for stress detection. The average efficiency value over all used methods and interval approaches 82 %, but the best results were achieved by Method 4.

Figure 4 illustrates reached efficiency for Method 3 and its sound normalised equivalent Method 4 for showing the impact of sound normalisation on stress detection in the form of more stable reached efficiency results. Obviously, the best and the most constant results can be found in the range of  $n$ -percentage intervals 50 %-90 % for Method 4.

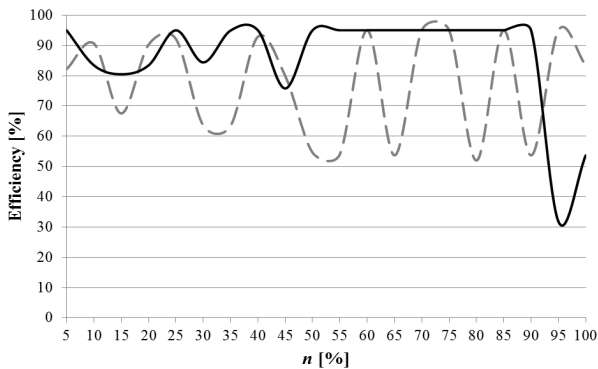


Fig. 4. Efficiency of stress detection for Method 3 (dashed grey line) Method 4 (solid black line) depending on selected  $n$ -percentage interval for using the GMM classifier.

The Probabilistic Neural Network (PNN) was used as the fourth classifier. Comparing to previous results, similar observations were occurred. The highest uniformity of  $RTOs$  varying on speaker's state can be found in the usage of 5 % selected interval (Method 2, Method 4 and Method 7) and higher intervals 75 %-100 % used only by Method 4. Absolutely highest accuracy (almost 94 %) in stress detection is achieved by Method 2. The worst efficiency results were achieved in intervals higher than 70 % by Method 3 and Method 8. These both methods are not suitable for psychological stress detection with the PNN classifier. Generally, the average efficiency of PNN classifier using  $RTOs$  is approximately 62 %.

Obviously, the chosen method and classifier are not important in the case of stress recognition as well the appropriate selected interval, but generally the best results are reached by GMM classifier.

The final sorting of used types of stress detection parameters is listed in Table II, where due to the total number 640 of all used types, only the first fifteen (best) and the last five (worst) positions are listed. All types are written only in abbreviations, e.g. GMM\_5\_D\_C\_N represents GMM classifier applied on 5 % selected interval, DIF estimation method, the vowels' centre and normalised sound (Method 6). The values of total analysed normal  $N_n$  and stressed speech  $N_s$  glottal pulses are listed in Table II as well as false detected normal  $N'_{cdn}$  and stressed  $N'_{cds}$  glottal pulses.

TABLE II. FINAL SORTING OF USED TYPES.

Rank	Type	$N'_{cdn}$ [-]	$N_n$ [-]	$N'_{cds}$ [-]	$N_s$ [-]	Eff. [%]
1	SVM_5_I_B_N	37	1049	59	1176	95,7
2	GMM_5_D_C_N	60	1098	55	1454	95,5
3	GMM_5_I_B_N	50	1049	50	1176	95,5
4	GMM_5_I_C	70	1192	52	1452	95,4
5	GMM_65_I_B_N	42	1049	60	1176	95,4
6	GMM_50_I_B_N	45	1049	57	1176	95,4
7	GMM_55_I_B_N	44	1049	58	1176	95,4
8	GMM_65_D_C	78	1098	39	1454	95,4
9	GMM_65_D_B_N	65	1065	62	1626	95,4
10	GMM_60_I_B_N	51	1049	54	1176	95,3
11	GMM_70_I_B_N	51	1049	54	1176	95,3
12	GMM_75_I_B_N	53	1049	52	1176	95,3
13	GMM_80_I_B_N	53	1049	52	1176	95,3
14	GMM_65_I_C	68	1192	56	1452	95,3
15	GMM_65_D_C_N	59	1098	61	1454	95,3
...						
636	KNN_5_I_C	1183	1192	1149	1452	11,8
637	GMM_35_D_B	975	1065	1487	1626	8,5
638	PNN_70_I_C_N	1136	1192	1291	1452	8,2
639	PNN_80_I_C_N	1005	1192	1422	1452	8,2
640	KNN_5_D_C_N	930	1098	1413	1454	8,2

Obviously, the best results are reached by Method 4 and GMM classifier in general. As the best  $n$ -percentage interval can be marked 5 % sector, but the most stress recognition stable range is from 50 % to 80 %.

## VI. CONCLUSIONS

By the comparison of eight different glottal pulse estimation methods and four classifiers, the GMM classifier can be marked as the best for stress recognition with method estimating glottal pulses by IAIF algorithm from normalised sound vowel's beginning. Obviously by presented  $RTOs$ :

- the IAIF estimation is more suitable than DIF algorithm,
- stress influence is better detectable at vowels' beginning,
- sound normalisation leads to more stable efficiency results,
- the biggest differences in  $RTOs$  between normal and stressed speech lie in 5 % interval as well as in 65 %.

Absolutely best results reaching efficiency of 95 % were achieved for GMM classifier, Method 4 on 5 % and 50 %-80 % intervals.

Generally, presented approach corresponds with the similar method detecting stress by means of glottal pulse distribution [31]. However, presented experiments show

higher accuracy (95 %) as the accuracy of 88 % published in [31] or in [32], where Glottal Spectral Slope reached stress detection ratios in the range 18 %–36 %.

Obviously, the combination of automatic vowel detection, e.g. [30], and findings presented in this paper can lead to development new systems recognizing psychological stress in speech which can negatively influence human behaviour. These systems can be practically applied in many fields of usage e.g. machine control, medical applications, etc.

Further, it is necessary to expand real psychological stress database to verify experimentally presented results. In future, described method will be also expanded and adapted to its usage on all estimated glottal pulses in all voiced parts of speech, i.e. not only on found vowels. This modification can lead to higher amount of estimated glottal pulses and to observation if described methods are phoneme-independent in the case of psychological stress detection in speech.

#### REFERENCES

- [1] J. P. Cabral, K. Richmond, J. Yamagishi, S. Renals, "Glottal spectral separation for speech synthesis", *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 2, pp. 195–208, 2014. [Online]. Available: <http://dx.doi.org/10.1109/JSTSP.2014.2307274>
- [2] T. Raitio, A. Suni, J. Yamagishi, H. Pulakka, J. Nurminen, M. Vainio, P. Alku, "HMM-Based speech synthesis utilizing glottal inverse filtering", *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 153–165, 2011. [Online]. Available: <http://dx.doi.org/10.1109/TASL.2010.2045239>
- [3] J. P. Cabral, S. Renals, J. Yamagishi, K. Richmond, "HMM-Based speech synthesiser using the LF-model glottal source", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Prague, 2011, pp. 4704–4707. [Online]. Available: <http://dx.doi.org/10.1109/icassp.2011.5947405>
- [4] F. Villavicencio, "A strategy for LF-based glottal-source & vocal-tract estimation on stationary modal singing", in *Proc. 22<sup>nd</sup> European Signal Processing Conf. (EUSIPCO)*, Lisbon, 2014, pp. 1457–1461.
- [5] J. Laver, *The Phonetic Description of Voice Quality*. Cambridge University Press, UK: Cambridge, 1980.
- [6] R. Sun, E. Moore, J. Torres, "Investigating glottal parameters for differentiating emotional categories with similar prosodies", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, 2009, pp. 4509–4512. [Online]. Available: <http://dx.doi.org/10.1109/icassp.2009.4960632>
- [7] G. Shashidhar, K. Koolagudi, K. S. Rao, "Emotion recognition from speech using source, system, and prosodic features", *IEEE Int. Journal of Speech Technology*, vol. 15, no. 2, pp. 265–289, 2011.
- [8] A. I. Iliev, M. S. Scordilis, "Spoken emotion recognition using glottal symmetry", *EURASIP Journal on Advances in Signal Processing*, vol. 2011, 2011.
- [9] P. K. Mongia, R. K. Sharma, "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", *Journal of Computer Networks and Communications*, vol. 2014, 2014.
- [10] I. Bisio, A. Delfino, F. Lavagetto, M. Marchese, A. Sciarone, "Gender-driven emotion recognition through speech signals for ambient intelligence applications", *IEEE Trans. Emerging Topics in Computing*, vol. 1, no. 2, pp. 244–257, 2014. [Online]. Available: <http://dx.doi.org/10.1109/TETC.2013.2274797>
- [11] J. Kametani, "Speaker recognition with glottal pulse-shapes", U.S. Patent 5091948A, 1992.
- [12] B. Yegnanarayana, K. Sharat Reddy, S. P. Kishore, "Source and system features for speaker recognition using AANN models", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, 2001, pp. 409–412. [Online]. Available: <http://dx.doi.org/10.1109/icassp.2001.940854>
- [13] J. Wang, M. T. Johnson, "Psychologically-motivated feature extraction for speaker identification", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 1690–1694.
- [14] V. N. Sorokin, A. A. Tananykin, V. G. Trunov, "Speaker recognition using vocal source model", *Pattern Recognition and Image Analysis*, vol. 24, no. 1, pp. 156–173, 2014. [Online]. Available: <http://dx.doi.org/10.1134/S1054661814010179>
- [15] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, L. O. Raming, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease", *IEEE Trans. Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012. [Online]. Available: <http://dx.doi.org/10.1109/TBME.2012.2183367>
- [16] H. Herzel, D. Berry, I. R. Titze, M. Saleh, "Analysis of vocal disorders with methods from nonlinear dynamics", *Journal of Speech, Language, and Hearing Research*, vol. 37, 1994.
- [17] M. Sigmund, P. Zelinka, "Analysis of voiced speech excitation due to alcohol intoxication", *Information Technology and Control*, vol. 40, no. 2, pp. 145–150, 2011. [Online]. Available: <http://dx.doi.org/10.5755/j01.itc.40.2.429>
- [18] P. Gomez-Vilda, O. Perez-Brocano, R. Martinez-Olalla, V. Rodellar-Biarge, K. Lopez de Ipinna Pena, M. Eca, P. Martinez-Lage, "Biomechanical characterization of phonation in Alzheimer's disease", in *Proc. Int. Work Conf. Bio-inspired Intelligence (IWobi)*, Liberia, 2014, pp. 14–20. [Online]. Available: <http://dx.doi.org/10.1109/iwobi.2014.6913931>
- [19] Saloni, R. K. Sharma, A. K. Gupta, "Disease detection using voice analysis: a review", *Int. Journal of Medical Engineering and Informatics*, vol. 6, no. 3, pp. 189–209, 2014. [Online]. Available: <http://dx.doi.org/10.1504/IJMEI.2014.063173>
- [20] T. Drugman, P. Alku, A. Alwan, B. Yegnanarayana, "Glottal source processing: From analysis to applications", *Computer Speech & Language*, vol. 28, no. 5, pp. 1117–1138, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.csl.2014.03.003>
- [21] M. Airas, "TKK Aparat: an environment for voice inverse filtering and parameterization", *Logopedics, phoniatrics, vocology*, vol. 33, no. 1, pp. 49–68, 2008. [Online]. Available: <http://dx.doi.org/10.1080/14015430701855333>
- [22] A. Lofqvist, "Inverse filtering as a tool in voice research and therapy", *Logopedics, phoniatrics, vocology*, vol. 16, no. 1–2, pp. 8–16, 1991. [Online]. Available: <http://dx.doi.org/10.3109/14015439109099168>
- [23] M. A. Nwachuku, "Inverse filtering techniques in speech analysis", *Nigerian Journal of Technology*, vol. 1, no. 1, pp. 38–42, 1991.
- [24] S. Dias, A. Ferreira, "Glottal pulse estimation – a frequency domain approach", (Unpublished).
- [25] H. Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech", *IEEE Trans. Audio and Electroacoustics*, vol. 21, no. 5, pp. 417–427, 1973. [Online]. Available: <http://dx.doi.org/10.1109/TAU.1973.1162506>
- [26] P. Alku, "Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering", *Speech Communication*, vol. 11, no. 2–3, pp. 109–118, 1992.
- [27] G. Fant, J. Liljencrants, Q. Lin, "A four-parameter model of glottal flow", *STL-QPSR*, vol. 26, no. 4, pp. 1–13, 1985.
- [28] M. Sigmund, "Introducing the database ExamStress for speech under stress", in *Proc. 7<sup>th</sup> Nordic Signal Processing Symposium*, Rejkjavik, 2006, pp. 290–293. [Online]. Available: <http://dx.doi.org/10.1109/norsig.2006.275258>
- [29] M. Stanek, M. Sigmund, "Finding the most uniform changes in vowel polygon caused by psychological stress", *Radioengineering*, vol. 24, no. 2, pp. 604–609, 2015. [Online]. Available: <http://dx.doi.org/10.13164/re.2015.0604>
- [30] M. Stanek, L. Polak, "Algorithms for vowel recognition in fluent speech based on formant positions", in *Proc. 36<sup>th</sup> Int. Conf. Telecommunications and Signal Processing*, Rome, 2013, pp. 521–525. [Online]. Available: <http://dx.doi.org/10.1109/tsp.2013.6613987>
- [31] M. Sigmund, A. Prokes, Z. Brabec, "Statistical analysis of glottal pulses in speech under psychological stress", in *Proc. of EUSIPCO*, Lausanne, 2008, pp. 1–5.
- [32] J. H. L. Hansen, S. Patil, "Speech under stress: analysis, modeling and recognition", *Lecture Notes in Computer Science*, vol. 4343, pp. 108–137, 2007. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-74200-5\\_6](http://dx.doi.org/10.1007/978-3-540-74200-5_6)