ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2011, Vol. 40, No. 2

ANALYSIS OF VOICED SPEECH EXCITATION DUE TO ALCOHOL INTOXICATION

Milan Sigmund, Petr Zelinka

Brno University of Technology, Faculty of Electrical Engineering and Communication, Dept. of Radio Electronics, Purkynova 118, 612 00 Brno, Czech Republic
e-mail: sigmund@feec.vutbr.cz

crossref http://dx.doi.org/10.5755/j01.itc.40.2.429

Abstract. A significant part of information carried in speech signal refers to the speaker. This paper deals with investigating alcohol intoxication based on analyzing recorded speech signal. Speech changes resulting from alcohol intoxication were investigated in the waveform of glottal pulses estimated from speech by applying the Iterative Adaptive Inverse Filtering (IAIF). Experimental results show that analysis of glottal excitation appears to be a useful approach to provide evidence of alcohol intoxication of over 1‰. At this alcohol level, the associated negative events influence professional performance and may involve fatal accidents in some cases. Via analyzing the speech signal, the speaker could be automatically monitored without their active co-operation. For use in our experiments, a new collection of Czech alcoholized speech consisting of phonetically identical speech data spoken in both sober and intoxicated state was created.

Keywords: speech processing, glottal pulses, alcohol intoxication.

1. Introduction

For many years it has been known that alcohol affects speech in a variety of ways. These changes include both the content of speech and its acoustic form expressed physically by parameters of speech signal. However, intensive scientific research into alcohol recognition from speech signals was started by the accident of the U.S. oil tanker Exxon Valdez, which ran aground in Alaska in March 1989. A suspicion arose that the captain had been influenced by alcohol during the accident but it was impossible to prove it, because tests for alcohol in the blood were performed too late. A tape with the recording of a dialogue between the captain and a terrestrial radio communication station was the only material that could clarify the situation. Two years later, it was confirmed that the captain's speech immediately before and after the accident exhibited significant changes of the sort associated with alcohol consumption [4].

A number of studies on alcohol and speech have been reported in the scientific literature during the past decades. A good general review of early research may be found in the monograph [5]. Some works presented the effect of alcohol from specific points of view. The main physiological effects of alcohol on the articulators are discussed in [24]. A global ability of four speech features (LPC coefficients, cepstral coefficients, PARCOR coefficients, log area ratio coefficients) to recognize alcohol intoxication was compared in [13]. Various prosodic features with alcohol recognition rates between 62% and 85% are reported in [12]. Fundamental frequency, signal-to-noise ratio, and formant frequencies were used for the recognition of low-level alcohol intoxication in [10]. A study analyzing various features derived from speech rhythm and formant frequencies F1-F4 presents some significant differences in alcoholized speech independent of gender and speaking style [18]. Problems of forensic analysis of alcohol intoxication in speakers appear in [11]. An interesting project focused on the accuracy of human listeners (professional vs. lay listeners) to judge the presence and level of intoxication by assessing appropriate speech samples was reported in the study [8]. The purpose of that research was to discover if actors could simulate intoxication when actually sober and simulate sobriety although seriously intoxicated. Currently, no software for automatic detection of alcohol intoxication is available. To our knowledge, no algorithms leading to a practical investigation of speech signal for accurate estimation of alcohol intoxication have yet been presented.

Our paper is organized as follows: Section 2 gives brief information about alcohol intoxication, Section 3 deals with speech data used in experiments while Section 4 describes the algorithms for estimation and processing of glottal pulses. Section 5 presents experimental results achieved with speech data from own speech collection. Final Section 6 gives a short

conclusion of the paper and suggests some topics for future work.

2. Alcohol Intoxication

There are two main ways of detecting alcohol concentration in the body, blood-alcohol concentration (BAC) and breath-alcohol concentration (BrAC). Of the two, BAC enjoys some preference, and in fact, BrAC is very often converted to an expression of equivalent BAC. However, the conversion is not utterly reliable and the BrAC value is therefore not admissible as evidence in court in many countries. From a short-term point of view, alcoholic intoxication causes changes both in emotional state and in psychomotorics. Psychomotoric changes are noticeable on levels of over 0.5‰ BAC. Exceeding the level of 1.5‰ BAC, changes in psychomotorics are so distinct that speech defects are usually recognizable by the human ear. Generally, the dose-related effect of alcohol depends on a variety of factors and is highly individual in humans. A comprehensive review and reference source covering a wide range of material, from medico-legal aspects of alcohol metabolism to practical involvement in alcohol metabolism in humans is provided, for example, in the three-volume work [6]. At low doses, alcohol may in some cases actually improve psychomotor performance with a mild euphoria. Increased doses, however, result in the well- known negative effects of alcohol on reaction time, cognitive functions or short-term memory.

Our research is focused on the speech signal produced by normal subjects (non-alcoholics) during a period of acute intoxication of over 1.0‰ BAC. At this level of alcohol concentration, the associated negative states may highly influence professional performance and/or cause dangerous incidents with fatal security consequences in some cases.

3. Speech Data Used

Previous experiments show that vowels and nasals are the most effective phonemes for speaker analysis in general [20]. In terms of speaker-recognition power, the following is the ranking of phoneme classes (in descending order):

vowels, nasals > liquids > fricatives, plosives

Hence, the speaker's alcohol intoxication was investigated from short segments of vowels only by analyzing their glottal excitation. All vowels represent well the voiced excitation of speech and are relatively easy to identify in speech signal [7].

For our experiments, we used our own collection of speech utterances by twelve male speakers (voluntary students) ranging in age from 20 to 28 years. All speakers were native speakers of Czech and self-reported as non-alcoholics drinking alcohol occasionally only. The speakers were asked to give information about some factors which can correlate with alcohol in influencing the voice such as stress, fatigue, negative psychological states, as well as the use of drugs. Each speaker participated in two recording sessions, once without and once with alcohol at two levels of intoxication (from intervals of 0.5-1.0‰ and 1.0-1.5‰ each) reached during increased drinking of liquor. The values of BAC were measured using a breathalyzer before each recording. Recordings in both states, sober and alcoholized, contain the following subsets of read speech: 1) five individual vowels /a/, /e/, /i/, /o/, /u/ pronounced very long; 2) ten individual words comprising five vowels and selected consonants /m/, /n/, /l/, /r/; 3) fluent text read from a book. Subsets 1 and 2 were repeated three times. All recordings took place in a quiet office and the speech signal was stored in the PCM format (16 bit, 22 kHz). The boundary between the two intoxication intervals of interest was set to 1.0‰ with respect to future application of voice analysis in road traffic. By Czech law, driving under alcohol intoxication below 1.0‰ is regarded as an offence, but above 1.0‰ it is regarded as a criminal act.

4. Processing of Speech Signal

4.1. Estimation of Glottal Pulses from Speech Signal

According to a widely used model for speech production [17], the speech signal is modelled by convolution of glottal excitation and vocal tract response. The vocal fold movement and corresponding glottal flow can be measured applying clinical methods, e.g. electroglottography, photo-glottography or pneumotachography [2]. The most frequently used clinical method, the electroglottography, is a non-invasive method of measuring vocal fold contact during voicing without affecting speech production. All these instrumental methods are objective and relatively accurate but it is necessary for the speaker to cooperate with the clinician during the measurement. Thus, the clinical methods can be applied on a person "online" using only some special equipment during speaking.

There are several effective techniques for extracting glottal pulses indirectly from speech. Considering the promising results in our previous experiments with the excitation of speech under stress [21], glottal pulses were estimated from speech signal by applying the IAIF (Iterative Adaptive Inverse Filtering) method. This approach is mainly used in the research of the voice source. The principle of IAIF is to cancel the effect of the vocal tract from a recorded speech signal to acquire the air flow through the glottis. The block diagram of IAIF is shown in Figure 1.

The IAIF method is based on the well-known linear predictive coding (LPC) techniques with multiple use of LPC predictors and inverse filters $H^{-1}(z)$ whereas filters $H(z)$ describe the actual vocal tract by LPC coefficients. The IAIF algorithm operates in two

repetitions, hence the word *iterative* in the name of the method. The first phase (blocks LPC 1st order, filter $H_1^{-1}(z)$, LPC 12th order, and filter $H_2^{-1}(z)$) generates an estimate of glottal excitation, which is subsequently used as input of the second phase (blocks LPC 4th order, filter $H_3^{-1}(z)$, LPC 12th order, and filter $H_4^{-1}(z)$) to achieve a more accurate estimate. A more detailed description of this method may be found in [1]. Other techniques for obtaining glottal pulses have been shown, for example, in [3].

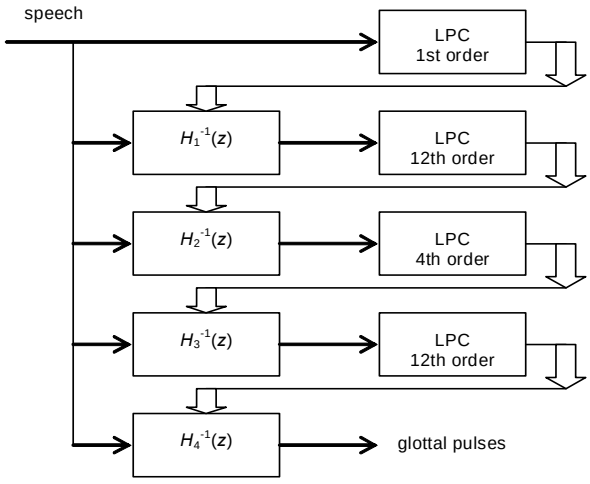

Figure 1. Block diagram of the IAIF algorithm for estimation of glottal pulses

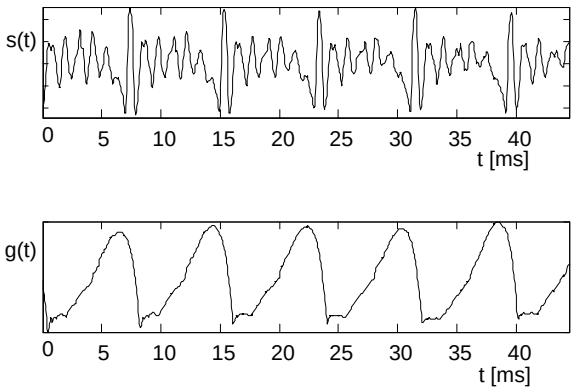

Figure 2. Speech signal of the phoneme /a/ (upper graph) and the corresponding glottal pulses (lower graph) obtained by IAIF

Figure 2 illustrates the acoustic steady-state wave form $s(t)$ of the vowel /a/ spoken separately (without coarticulation) and its corresponding glottal pulses $g(t)$ estimated using the IAIF algorithm. In the clinical examination of the voice, the glottal inverse filtering conveys information about irregularities caused by vocal nodules or polyps or changes in the voicing caused by speaker fatigue [23].

4.2. Glottal Pulse Features

For an analysis of glottal pulses obtained by IAIF, the Liljencrant-Fant (LF) approximation [9] was applied. To identify alcohol intoxication in speakers, the first derivative of the LF approximation was used. The first derivative $v(n)$ of the approximation function $g(n)$ consists of two consecutive temporal segments, $v_1(n)$ and $v_2(n)$, described by Eq. (1) and Eq. (2), respectively:

$$v_1(n) = -E_e \frac{\sin\left[\omega\left(n - T_{op}\right)\right]}{\sin\left[\omega\left(T_e - T_{op}\right)\right]} e^{\alpha\left(n - T_e\right)} \qquad (1)$$

for $T_{op} \le n \le T_e$ and

$$v_2(n) = \frac{-E_e}{\varepsilon T_a}\left[e^{\varepsilon\left(T_e - n\right)} - e^{\varepsilon\left(T_e - T_c\right)}\right] \qquad (2)$$

for $T_e < n \le T_c$.

The four timing parameters, T_{op}, T_e, T_c, and T_a, have a direct physical correspondence with human voicing events [16] as can be seen in Figure 3.

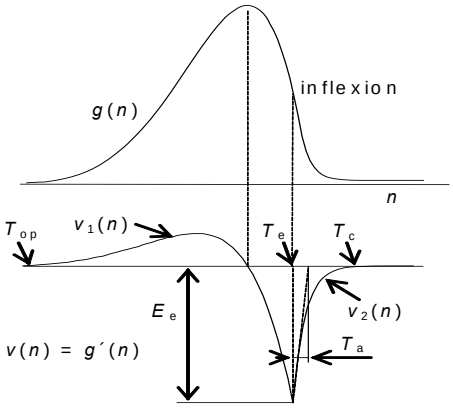

Figure 3. A typical approximation of glottal pulse (upper graph) and its first derivative (lower graph)

Parameter T_a is the second derivative of the volume flow at the minimum of the first derivative. The LF approximation is limited to the interval $T_{op} \le n \le T_c$ representing the open phase of glottal waveform in voiced excitation. Another set of parameters, E_e, α, ω, and ε, was determined from the derivative of $g(n)$ using the iterative method and by the criterion of minimal average quadratic deviation of the curve of glottal pulses obtained by IAIF from its LF approximation $g(n)$.

5. Experimental Results

5.1. Effect of Alcohol on Glottal Pulse Parameters

In the investigation of alcohol intoxication, glottal pulses were estimated from speech segments containing vowels only. The analyzed short-time speech segments span 30 ms in the central part of vowels. The glottal waveform obtained was normalized in maximal amplitude and pitch synchronously before calculating the LF parameters. Figure 4 shows the glottal waveforms of both sober and alcoholized speech obtained

from the long vowel /á/ in the Czech word "garáž" (English "garage"). In this word, the first vowel /a/ is short while the second one is long, which is indicated in text by the acute accent as /á/.

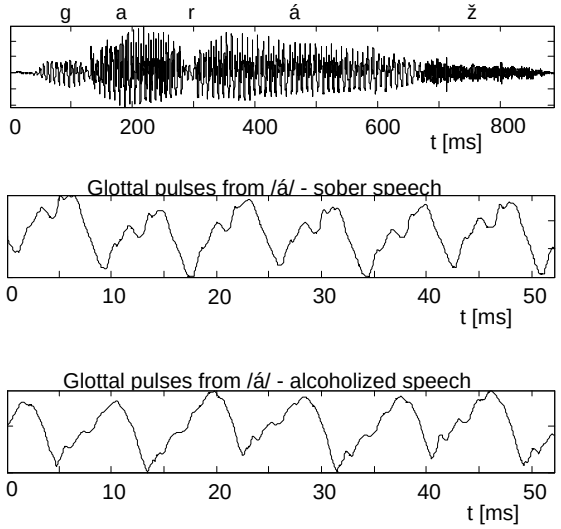

Figure 4. An example of the effect of alcohol intoxication on glottal pulses obtained from the vowel /á/ in the word "garáž"

In our experiments, the parameters E_e, α, ω, T_a, and ε defined implicitly by Eqs. (1) and (2) were measured in both sober and alcoholized speech for each vowel across all speakers and their statistical values obtained were then compared. In the statistical calculations of the mean and standard deviation, 40 segments of each vowel were taken into consideration for each speaker from the individual-vowels and individual words data subsets. Experimental results show that the most sensitive parameters to alcohol intoxication are α, ω, and E_e (in this order).

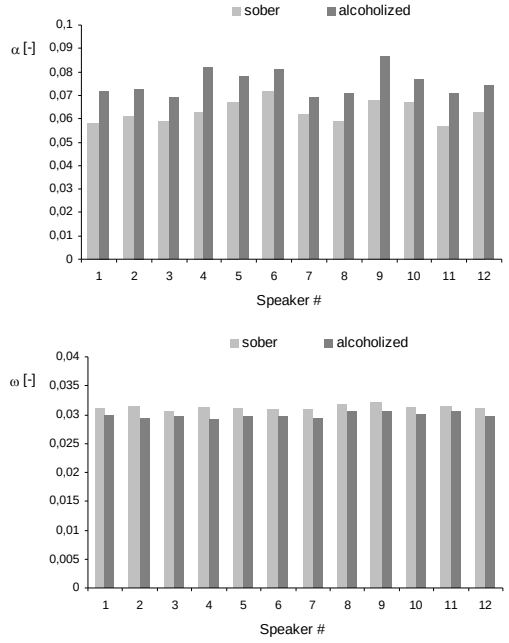

Figure 5. Mean values of the parameters α and ω for the vowel /e/ produced by all speakers while sober and intoxicated at an alcohol level of 1.0-1.5‰

5.2. Evaluation of LF Parameters of Individual Vowels

To test the extracted glottal pulse parameters for their ability to detect alcohol intoxication, a weighted distance classification was applied based on the distance measures for individual vowels both for the alcoholized speech

$$d_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}},\mathbf{t})=\sum_{n=1}^{N}\frac{1}{\sigma_{\mathrm{A}}^{2}(n)}\left[r_{\mathrm{A}}(n)-t(n)\right]^{2} \tag{3a}$$

and for the sober speech

$$d_{\mathrm{S}}(\mathbf{r}_{\mathrm{S}},\mathbf{t})=\sum_{n=1}^{N}\frac{1}{\sigma_{\mathrm{S}}^{2}(n)}\left[r_{\mathrm{S}}(n)-t(n)\right]^{2} \tag{3b}$$

In Eqs. (3a) and (3b) $t(n)$ stands for the n-th parameter of tested vowels, $r_{\mathrm{A}}(n)$ and $\sigma_{\mathrm{A}}(n)$ are the reference values (mean and standard deviation) for the n-th parameter of alcoholized speech while $r_{\mathrm{S}}(n)$ and $\sigma_{\mathrm{S}}(n)$ are the reference values (mean and standard deviation) for the n-th parameter of sober speech. All reference values used in the test are speaker-specific. Figure 5 shows the mean values of parameters α and ω for /e/ calculated individually for each speaker and used as reference values. Table 1 illustrates the reference values of α and ω (mean and standard deviation) for all vowels obtained from male speaker No. 1 (M1) in both sober and alcoholized speech.

Table 1. Values of LF parameters for speaker M1

	Sober (0 ‰)				Alcoholized (1.0-1.5 ‰)			
	α [-]		ω [-]		α [-]		ω [-]	
	$r_{\mathrm{S}}(\alpha)$	$\sigma_{\mathrm{S}}(\alpha)$	$r_{\mathrm{S}}(\omega)$	$\sigma_{\mathrm{S}}(\omega)$	$r_{\mathrm{A}}(\alpha)$	$\sigma_{\mathrm{A}}(\alpha)$	$r_{\mathrm{A}}(\omega)$	$\sigma_{\mathrm{A}}(\omega)$
/a/	0.051	0.005	0.030	0.001	0.060	0.005	0.029	0.001
/e/	0.058	0.004	0.031	0.001	0.072	0.003	0.030	0.002
/i/	0.053	0.003	0.029	0.002	0.057	0.004	0.028	0.002
/o/	0.061	0.004	0.030	0.001	0.073	0.005	0.028	0.002
/u/	0.059	0.006	0.031	0.002	0.055	0.005	0.031	0.001

In the final test, three most effective glottal pulse parameters, namely α, ω, and E_e, were considered for the estimation of distances d_{A} and d_{S}, i.e., the total number of terms in the sum in Eqs. (3a) and (3b) were N=3. For testing, the vowels from the data subset of read text were used. Hence, the training and testing speech data were from two disjunct classes. The number of speech segments for each tested vowel varied between 50 and 112. The final classification of alcoholized speech was performed using the minimal distance $\min\{d_{\mathrm{A}}, d_{\mathrm{S}}\}$. Table 2 summarizes the detection rate of alcohol intoxication computed in the binomial classification (sober/intoxicated) for a group of twelve male speakers. In the experiment, speech spoken under higher intoxication (interval 1.0-1.5‰) was taken into account.

The most effective individual vowel for the detection of alcohol intoxication seems to be /o/ followed by /e/. However, considering the phoneme occurrence statistics, the vowel /e/ can be the most important individual Czech phoneme for investigation of alcohol

intoxication because it is the most frequently used phoneme in the Czech language, with its relative frequency of 9.2% (in comparison with a frequency of 7.9% for the vowel /o/) [15]. A typical Czech /e/ is pronounced quite broadly. The average duration of short /e/ is 83 ms for spontaneous speech and 77 ms for read speech while the average duration of long /é/ is 125 ms for spontaneous speech and 122 ms for read speech [15]. The duration is in all cases sufficient enough to enable a reliable phoneme analysis. Contrary to English, Czech long and short vowels are the same sound only the duration for which they are pronounced is different. The vowels are never reduced and undergo no assimilations. Vowel modifications such as nasalization do not occur in Czech. The acoustical form of Czech vowels is presented, for example, in the pronunciation guide Local Lingo, which is available online via the website [25].

Table 2. Detection rate of alcohol intoxication by individual vowels

Speaker	Detection rate (in %)				
	/a/	/e/	/i/	/o/	/u/
M1	61	72	69	79	71
M2	76	82	77	75	70
M3	54	67	68	62	61
M4	78	81	80	83	76
M5	72	78	73	78	75
M6	75	74	72	76	73
M7	62	73	76	75	76
M8	68	74	77	83	69
M9	72	85	80	81	70
M10	71	82	79	77	65
M11	75	79	76	79	74
M12	68	72	69	76	71
Average	69.3	76.6	74.7	77.0	70.9

To do the signal experiments in our research, the proposed algorithms were implemented in MATLAB. In addition, for inverse filtering of speech signal, we developed our own software tool and implemented it in the MATLAB GUI environment [22].

6. Conclusions and Future Work

Human voice can be taken into account as a possible indicator of alcohol intoxication in speakers. While investigating alcohol in speakers we are only concerned with the speech features which are physically measurable from the speech signal. Besides these features, other verbal factors (e.g. word repetition, discontinuity of speech, excessive or incoherent talking, etc.) can be observed in the speech spoken under the influence of alcohol.

An approach for speaker-dependent detection of alcohol intoxication based on signal analysis of vowel glottal pulses was presented. The detection rate in the binomial classification (sober or intoxicated) varies between 69.3% and 77.0%, analyzing individual vowels only. The most suitable vowels for alcohol detection seem to be /o/ (77.0%) and /e/ (76.6%). The achieved detection rate is comparable with other methods based on phonetic and prosodic features given, for example, in [12] and [18]. An advantage of the glottal pulse parameters is their relative independence from the voluntary changes in speech and therefore usability for extended speaker recognition in the biometric security systems [14]. It was evident that the acoustic correlates of alcohol in the speech signal are subject to individual differences. The problem of individual variability is not limited to alcohol and speech research but to alcohol research in general. The obtained results should be verified on a large-scale database of speakers including also female voices. Furthermore, a comparison with foreign sober and intoxicated native speakers will be useful. However, the only available corpus of alcoholic speech is the German professional database called Alcohol Language Corpus which was recorded at the University of Munich [19]. Future work could be oriented at robust detection of alcohol intoxication with respect to eliminate the detection error of type "false alarm". For that reason, other factors affecting also the glottal pulses such as some emotions, psychological stress, etc. should be investigated and compared with alcohol intoxication. In general, the two main spheres in which alcohol testing play a role at present are vehicular traffic and workplaces.

Acknowledgements

The research leading to these results has received funding from the European Social Fund under grant agreement CZ.1.07/2.3.00/20.0007 (the WICOMT project) and by the research program MSM 0021630513 (ELCOM).

References

[1] **P. Alku, B. Story, M. Airas.** Evaluation of an inverse filtering technique using physical modeling of voice production. *In Proceedings of International Conference on Spoken Language Processing*, *Jeju Island*, 2004, 497-500.

[2] **R.J. Baken, R.F. Orlikoff.** Clinical Measure-ment of Speech and Voice. *San Diego: Singular Publishing*, 2000.

[3] **M. Bostik, M. Sigmund.** Methods for estimation of glottal pulses waveforms exciting voiced speech. *In Proceedings of Eurospeech´*03, *Geneva*, 2003, 2389-2392.

[4] **M. Brenner, J. R. Cash.** Speech analysis as an index of alcohol intoxication – The Exxon Valdez accident. *Aviation, Space, and Environmental Medicine*, 1991, *Vol*. 62, *No*. 9, 893-898.

[5] **S.B. Chin, D.B. Pisoni.** Alcohol and Speech. *San Diego: Academic Press*, 1997.

[6] **K.E. Crow, R.D. Batt.** Human Metabolism of Alcohol. *Boca Raton: CRC Press*, 1989.

[7] **K. Driaunys, V. Rudžionis, P. Žvinys.** Implementation of hierarchical phoneme classification approach on LTDIGITS corpora. *Information Technology and Control*, 2009, *Vol*. 38, *No*. 4, 303-310.

[8] **H. Hollien, J.D. Harnsberger, C.A. Martin, R. Hill, G.A. Alderman.** Perceiving the effects of ethanol intoxication on voice. *Journal of Voice*, 2009, *Vol*. 23, *No*. 5, 552-559.

[9] **M.R. Iseli, A. Alwan.** Inter- and intra-speaker variability of glottal flow derivative using the LF model. *In Proceedings of International Conference on Spoken Language Processing*, *Beijing*, 2000, 477-480.

[10] **F. Klingholz, R. Penning, E. Liebhart.** Recognition of low-level alcohol intoxication from speech signal. *Journal of the Acoustical Society of America*, 1988, *Vol*. 84, *No*. 3, 929-935.

[11] **H.J. Künzel, A. Braun.** The effect of alcohol on speech prosody. *In Proceedings International Congress of Phonetic Science*, *Barcelona*, 2003, 2645-2648.

[12] **M. Levit, R. Huber, A. Batliner, E. Noeth.** Use of prosodic speech characteristics for automated detection of alcohol intoxication. *In Proceedings of Prosody in Speech Recognition and Understanding*, *Red Bank*, 2001, 22-25.

[13] **R. Menšik.** Recognition of alcohol influence on speech. *In Proceedings of Workshop on Text, Speech and Dialogue*, *Plzen*, 1999, 384-387.

[14] **F. Orsag.** Speaker recognition in the biometric security systems. *Computing and Informatics*, 2006, *Vol*. 25, *No*. 5, 369-391.

[15] **J. Psutka, L. Müller, J. Matoušek, V. Radová.** Mluvíme s počítačem česky. *Praha: Academia*, 2006.

[16] **Y. Qi, N. Bi.** A simplified approximation of the four-parameter LF model of voice source. *Journal of the Acoustical Society of America*, 1994, *Vol*. 96, *No*. 2, 1182-1185.

[17] **T.F. Quatieri.** Discrete-Time Speech Signal Processing. *Englewood Cliffs: Prentice Hall*, 2002.

[18] **F. Schiel, Ch. Heinrich, V. Neumeyer.** Rhythm and formant features for automatic alcohol detection. *In Proceedings of INTERSPEECH*, *Chiba*, 2010, 458-461.

[19] **F. Schiel, Ch. Heinrich, S. Barfüßer, T. Gilg**. ALC-Alcohol language corpus, *In Proceedings of Language Resources and Evaluation Conference*, *Marrakesh*, 2008, 1-5.

[20] **M. Sigmund.** Voice Recognition by Computer. *Marburg: Tectum-Verlag*, 2003.

[21] **M. Sigmund, A. Prokes, Z. Brabec.** Statistical analysis of glottal pulses in speech under psychological stress. *In Proceedings of EUSIPCO*, *Lausanne*, 2008, *paper* 1569105092.

[22] **M. Sigmund, A. Prokes, P. Zelinka.** Detection of alcohol in speech signal using LF model. *In Proceedings of International Conference on Artificial Intelligence and Applications*, *Innsbruck*, 2010, 193-196.

[23] **E. Vilkman, L. Lehto, T. Bäckström, and P. Alku.** Vocal loading of call centre personnel. *In Proceedings of the 6th International Workshop, Advances in Quantitative Laryngoscopy, Voice and Speech Research*, *Hamburg*, 2003, 1-7.

[24] **H. Watanabe, T. Shin, H. Matsuo, F. Okuno, T. Tsuji, M. Matsuoka, J. Fukaura, H. Matsunaga.** Studies on vocal fold injection and changes in pitch associated with alcohol intake. *Journal of Voice*, 1994, *Vol*. 8, *No*. 4, 340-346.

[25] *http://www.locallingo.com/czech/pronunciation/index.html* (*cited* 2011-04-15).

Received November 2010.