

Detailed analysis of voice period fluctuations in speakers under psychological stress

Pavel Horský¹, David Haisman², Jan Cícha³

Abstract—Stress is one of the phenomena characterizing today's modern human society. High stress is also reflected in voice and speech in many ways. This paper presents a detailed analysis of voice period fluctuations when a speaker is exposed to psychological stress. The initial features determined are the length of the fundamental period and local jitter extracted from the voiced regions of the speech signal. Statistical parameters of their distribution and other derived features were taken into account in the experimental analysis. Multiple parameters were compared pairwise for normal speech and speech under stress, and parameters with significant differences caused by stress were searched for. The best reliability of 100% was achieved for two parameters. The test speakers were a group of nine university students under real stress, which was induced by oral exams in front of an examination board.

Keywords: speech under stress, speech signal, fundamental period, jitter-derived parameters

1 Introduction

Speech is the most natural way of communication between people. In addition to the linguistic content of speech, the human voice reflects various personal information. Important characteristics such as the age, gender and much of the emotional state of the speaker can be recognized simply by listening. On the other hand, some specific information related to psychological and health status, which is also embedded in the voice, can only be obtained through computer signal processing. This poses a challenge for research into nuanced analysis of speech.

Nowadays, stress can be identified more reliably using speech signal processing methods than by listening from trained experts. Standard speech signal features used for stress detection are fundamental frequency F_0 , harmonic-to-noise ratio (HNR), formants, mel-frequency cepstral coefficients (MFCC), and derived variants of these features. The fundamental frequency in stressed speech has been widely investigated in several studies, for example [1, 2]. Stress is generally associated with increased mean F_0 and greater variability in F_0 . Changes in F_0 due to stress compared to changes due to other negative emotions are discussed in [3]. Typical increase in HNR in stressed speech is described in [4]. In the spectral domain, vowel formants are of primary interest. Authors [5] observed individual formants F_1 to F_4 and found that formant frequencies F_1 and F_2 decrease under stress. The results in [6] show that formant frequency F_3 always

decreases, but the changes in F_1 and F_2 depend on the speech material. Combinations of multiple formants structured into vowel polygons were investigated in [7] in terms of area and centroid. A common feature widely used in many areas of speech processing is the MFCC [8]. The potential of MFCC also for stress detection is shown, for example, in [9]. Effectiveness of seven different sets of cepstral coefficients in classifying speech into three stress levels is compared in an experimental study [10]. Some researchers attempted to extract stress-indicating information from glottal source signal derived from speech signal. Several specific glottal pulse parameters were analysed, for example, the normalized amplitude quotient [11], the ratio of the return to opening phase [12] or the spectral tilt [13]. A physical model of glottal flow influenced by stress is studied in [14]. The experimental results reported in the publications are based on various speech databases as well as on own datasets collected by the researchers. The most widely used database, especially for aviation and military safety, is the Speech under Simulated and Actual Stress (SUSAS) database [15], which contains English speech. The findings published so far suggest that there is no single strong feature that could serve as the sole criterion for detecting stress. The search for the optimal combination of stress features in the voice is still a subject of research. In our work, we focused on changes in the length of individual cycles in speech signal generated under real stress. The presented experiments with uncontrolled period variability in-

^{1, 2, 3} Department of Radio Electronics, FEEC, Brno University of Technology, Brno, Czech Republic
¹ 256484@vut.cz, ² 256476@vut.cz, ³ 256457@vut.cz

clude both the average (standard jitter) and detailed analysis of other jitter-derived parameters.

The remaining parts of this study are organized as follows: Section 2 provides information on the role of the fundamental frequency in speech and its basic properties. Section 3 describes the feature extraction methods used in our work. Experimental results obtained using the methods defined in Section 3 are summarized in Section 4. Finally, Section 5 presents the conclusions and an outlook for future research to extend this study.

2 Role of fundamental frequency

The fundamental frequency in a speech signal (usually referred to as F_0) is audible and perceived as the pitch (subjective attribute). The listener can sense both the average pitch and the changes in pitch during a sentence. In normal speech, speakers intentionally generate speech with a small F_0 deviation from the personal baseline, which contributes to the natural sound of speech. The changes in F_0 are consciously controlled by the speakers and depend on the linguistic structure of the spoken sentence [16]. The contour of F_0 expresses the intonation of the sentence [17].

Voiced regions of speech, especially vowels, are characterized by cycles in the waveform. Figure 1 illustrates a short word consisting of one voiced phoneme and two unvoiced phonemes. The instantaneous fundamental frequency is directly given by the length of cycles T_i as the inverse of the fundamental period $1/T_i$. The range of the fundamental period for a representative normal voice was empirically established by several authors. However, there is no clear, generally accepted normative for T range. Typical T ranges for male, female, and child voices are compared in Tab. 1 [18]. These ranges correspond to F_0 for male voices between 80 Hz and 200 Hz and for female voices between 150 Hz and 350 Hz. However, the ranges differ not only by gender, but they also vary slightly by the age of the adult speakers. A detailed overview of typical F_0 values tabulated by age groups separately for male and female speakers can be found in [19]. At first glance, in Tab. 1, the female fundamental periods are significantly shorter than the male periods. This fact is widely used for gender classification. In medicine, observing F_0 helps detect some mental and neurodegenerative diseases, see for example [20, 21].

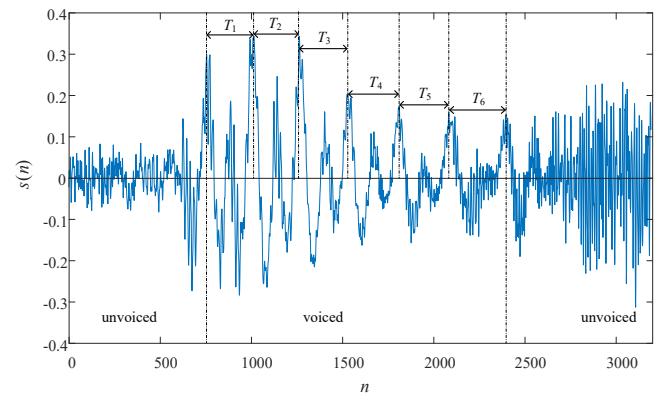


Fig. 1. Waveform of the word “fish” with marked fundamental periods in the voiced region

Table 1. Ranges of voice fundamental period

Speakers	Fundamental period (ms)	
	Minimum	Maximum
Male	5.0	12.5
Female	2.9	6.7
Children	2.0	5.0

3 Method used

The structure of our approach applied for speech analysis is shown in Fig. 2.

First, the speech signal was divided by a rectangular window into 20 ms *frames* to perform short-time analysis. The voicing of the frames was tested and the average fundamental frequency (necessary for subsequent calculations) in each voiced frame was calculated using center-clipping approach [22]. Both operations were performed in the *Center clipping* block. To improve robustness of the approach, the obtained frequency values were validated from the perspective of vocal tract physiology in the following *Frequency alternation condition* block. Firstly, the valid values were restricted to the range for male speakers, and secondly, the occurrence of large frequency changes in sequences of three consecutive frames was checked. If the difference between the first frame in the sequence and the average of all three frames exceeds 30% of the average, it is considered an error because it is unnatural. In that case, the first frame was ignored, and processing moved one frame further into the next three-frame sequence.

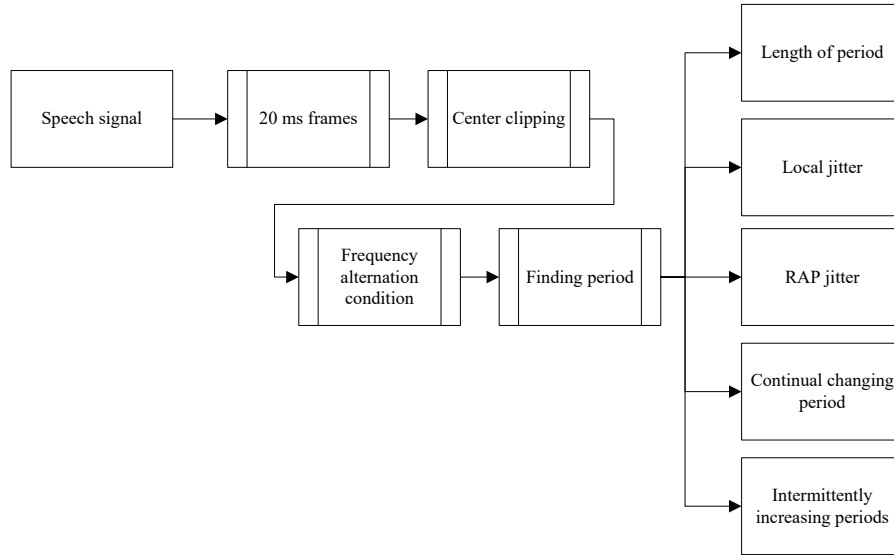


Fig. 2. Block diagram of speech processing

The obtained and validated fundamental frequencies correspond to the frames. However, each 20 ms frame covers more than one fundamental period. Exact determination of the length of individual periods was performed in the *Finding period* block and is depicted in Fig. 3. The algorithm is based on finding the period length by zero crossing instead of considering peaks, which fluctuate in extreme values for some voices. The algorithm starts by finding the maximum of the signal within one average signal period, which was estimated by the center-clipping approach. The first maximum found is marked with a yellow asterisk. The first zero crossing preceding the signal maximum is marked with a purple asterisk. Afterwards, a point shifted from the purple asterisk by the average period (again taken from the center-clipping approach) is found (green asterisk) and then corrected to the nearest zero crossing (red asterisk). Subsequently, the first red asterisk is considered a purple asterisk, and the search process is repeated to determine the length of the next period.

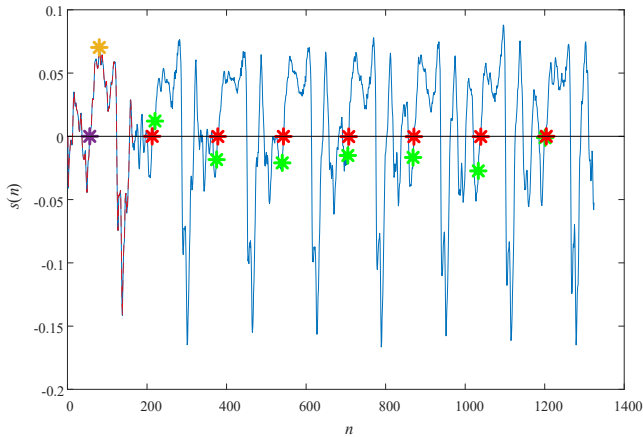


Fig. 3. Determining individual periods in speech

To analyze and assess the variability of voiced speech periods, we calculated several period-based features using the following methods:

3.1 Length of period

First, individual fundamental periods T_i were detected in speech signals, and their length was directly statistically processed in terms of mean and standard deviation. In addition, the period length, tied to the position of the period in the voiced sequence, was used to estimate other features.

3.2 Local jitter

Local jitter J was calculated as the time difference in seconds between two consecutive periods detected in the voiced sequence. Then, four statistical parameters were calculated from the jitter distribution using built-in functions in MATLAB, namely mean μ , standard deviation σ , kurtosis, and skewness. The calculations of kurtosis k_1 and skewness s_1 were carried out according to the formulas

$$k_1 = \frac{\frac{1}{I} \sum_{i=1}^I (x_i - \mu)^4}{\left(\frac{1}{I} \sum_{i=1}^I (x_i - \mu)^2 \right)^2} \quad (1)$$

$$s_1 = \frac{\frac{1}{I} \sum_{i=1}^I (x_i - \mu)^3}{\left(\sqrt{\frac{1}{I} \sum_{i=1}^I (x_i - \mu)^2} \right)^3} \quad (2)$$

In both formulas, x_i are the local jitter values and I represents the total number of jitter values included in the jitter distribution.

3.3 RAP jitter

Relative Average Perturbation (RAP) represents the average absolute difference of one period and the average of the two nearest neighbouring periods (i.e., the previous and the following period)

$$RAP = \frac{1}{I-2} \sum_{i=2}^{I-1} \left| T_i - \frac{T_{i-1} + T_{i+1}}{2} \right| \quad (3)$$

where T_i is the length of i -th period in seconds and I stands for the number of consecutive periods in the sequence.

3.4 Continuously increasing or decreasing periods

Short sequences of continuously increasing or decreasing consecutive periods without interruption were detected. An illustrative principle is shown in Fig. 4. Each number represents the length of one period, and black arrows above the numbers indicate increasing periods, while black arrows below the numbers indicate decreasing periods. The continuous increase or decrease is terminated by a red number that breaks the given trend. The first sequence contains five increasing periods, and the second sequence contains four decreasing periods.



Fig. 4. Illustration of period sequences containing continuously increasing and then decreasing periods

3.5 Intermittently increasing periods

In addition to detecting continuous increases or decreases in the period length, intermittent increases were also measured. In this case, each pair of consecutive periods without overlapping pairs was considered separately and independently of previous values. The change in period length between consecutive pairs is not registered. An example of five intermittently increasing periods is shown in Fig. 5, where $T_1 < T_2$ (first pair), $T_3 < T_4$ (second pair), and $T_5 < T_6$ (third pair). Changes at the edges of the pairs (indicated by up and down arrows) are ignored but included in the number of periods in short sequences of intermittently increasing periods.

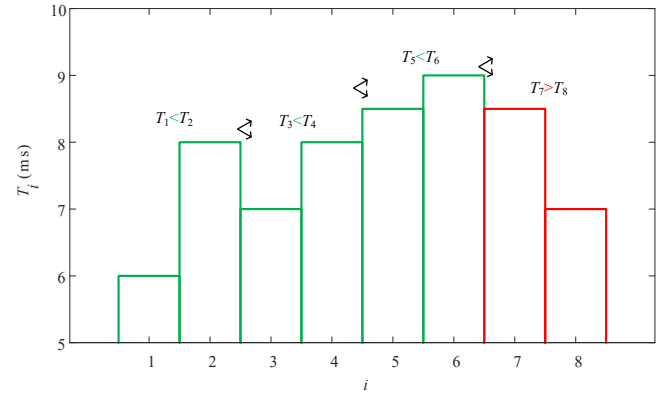


Fig. 5. Intermittent increase of consecutive periods

4 Experimental results

For our research, we used paired speech signals (stressed vs. normal speech) from nine male speakers. The stressed speech was recorded during oral exams at our university in front of an examination board. The normal speech from the same speakers without stress was recorded a few days later. All recordings contain Czech speech in a common indoor environment. The participants in the experiments were volunteer students aged 24 to 26. University exams as an effective psychological stressor have also been used in previous studies by other authors [23, 24]. Speech was digitally recorded in a mono channel with 16-bit fixed point resolution and a sampling rate of 22 050 Hz. For our experiments, all recordings were subsequently converted to 32-bit float precision and scaled in the range from -1 to $+1$. The duration of each recording is approximately 3 minutes with negligible background noise. Speech signals were processed in MATLAB.

The accuracy of determining the fundamental period is limited by the sampling frequency. In our case, for a sampling frequency of 22 050 Hz, the sampling period is $T_{\text{sam}} = 45.35 \mu\text{s}$, i.e., the signal samples are separated by $45.35 \mu\text{s}$. The length of the period can therefore only be determined with an accuracy of $\pm 0.5 T_{\text{sam}} = 22.7 \mu\text{s}$ applied to the beginning and end.

To compare the potential of the used features for stress detection, we evaluated their reliability as the ratio of speakers with a feature change in the same direction to all persons measured in the group of speakers. The reliability corresponds to the recall value. The significance of the results was tested by t-test.

4.1 Length of period

A typical trend in the occurrence of period lengths in speech is illustrated in Fig. 6, where partially overlapping histograms were constructed for speaker #9,

eparately for normal speech and for stressed speech. As can be seen, the periods in the stressed speech (in red) are shorter. The distributions for other speakers show a similar trend. From each distribution, the mean μ and standard deviation σ were calculated. Table 2 contains the average length of the fundamental period for all measured speakers. For 8 out of 9 speakers (except speaker #3) the averages under stress are smaller. This corresponds to an increase in the fundamental frequency under stress, which is also known from other studies as the most common effect of stress, see for example [25, 26]. Based on the results in Tab. 2, the period lengths provide 89% reliability in indicating stress (p-value 0.0014).

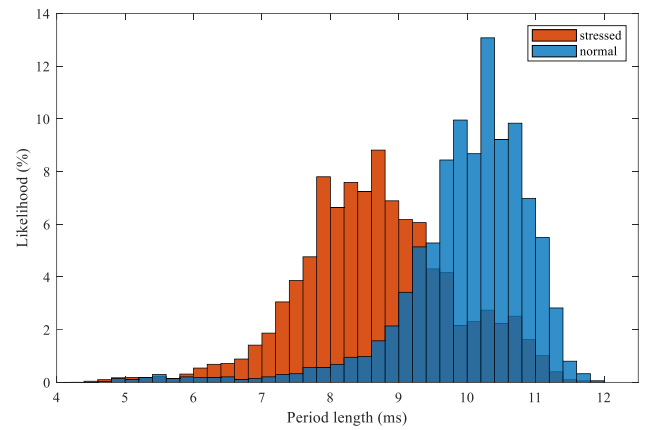


Fig. 6. Chart of normalized distribution of periods

Table 2. Average fundamental periods and standard deviations for individual speakers calculated for periods > 5 ms

Speaker #	Stressed speech (ms)		Normal speech (ms)		Percentage difference of μ related to normal speech (%)
	μ	σ	μ	σ	
1	6.89	0.878	7.23	0.926	-4.7
2	6.28	0.822	6.77	0.884	-7.2
3	7.11	0.944	7.06	1.22	0.70
4	9.20	0.933	10.2	0.887	-9.6
5	6.23	0.718	6.64	0.740	-6.2
6	6.31	0.487	7.07	0.770	-11
7	7.63	0.843	7.95	0.633	-4.0
8	7.47	0.879	8.62	1.02	-13
9	8.68	1.10	10.0	0.950	-13

4.2 Local jitter

The normalized histogram of local jitter J for speaker #9 is shown in Fig. 7. The histograms for the other speakers have a similar envelope, with no differences in some jitter intervals. The proportion of jitter in the region around zero is higher in stressed speech than in normal speech. Such very small jitter represents stable period lengths (note that zero jitter means constant periods). This effect is consistent with the findings in [22], where short sequences of constant periods were observed.

The following bar graphs compare statistical parameters calculated from individual jitter distributions for all speakers, namely the mean μ , standard deviation σ , kurtosis, and skewness. The best result of 89% reliability was achieved using standard deviation σ (p-value 0.0028). The values of σ are usually smaller for stressed speech (except for speaker #5), as shown in Fig. 8.

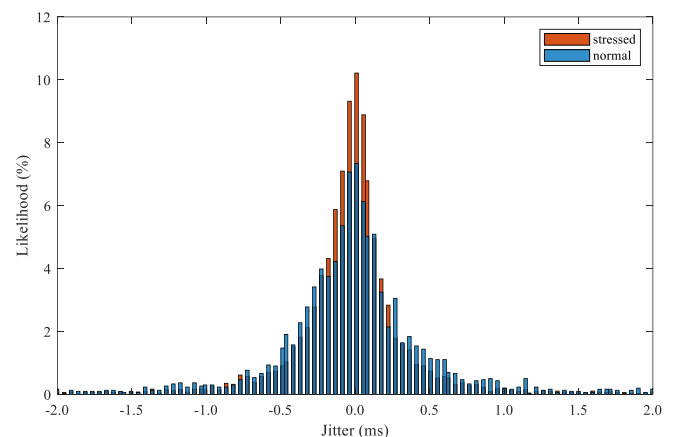


Fig. 7. Normalized jitter distribution

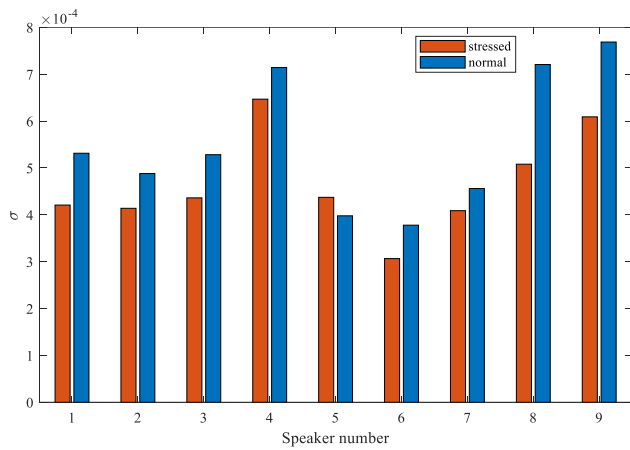


Fig. 8. Comparison of jitter standard deviations σ for all speakers

All jitter averages are negative values very close to zero. As shown in Fig. 9, three speakers have smaller μ in stressed speech, while six speakers have smaller μ in normal speech. This ratio gives a reliability of 66.7% (p-value 0.0898). A similar ratio was found for kurtosis – three speakers have smaller value in stressed speech, and six speakers have smaller value in normal speech, see Fig. 10 (p-value 0.0498). The skewness was smaller for two speakers in stressed speech and for seven speakers in normal speech, giving a reliability of 78%, see Fig. 11 (p-value 0.0962). As can be seen, the statistical parameters of jitter (standard deviation, mean, kurtosis, skewness) differ markedly between speakers. This can be interpreted as a different response to stress in individual speakers, which could be confirmed by measuring the stress hormone cortisol.

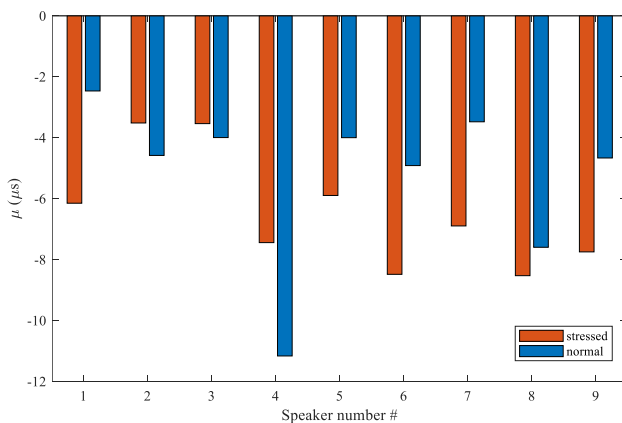


Fig. 9. Comparison of jitter means μ for all speakers

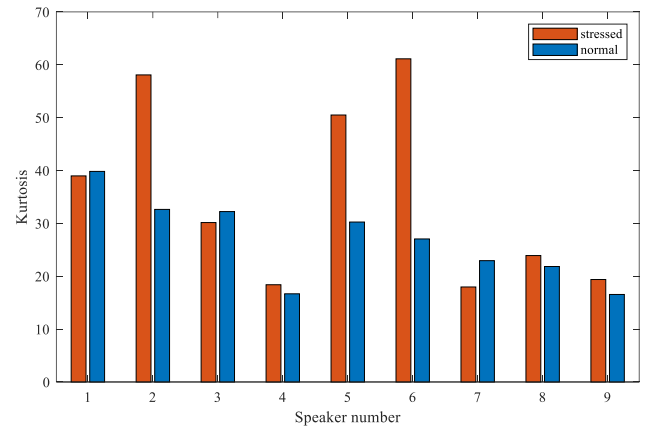


Fig. 10. Comparison of jitter kurtosis for all speakers

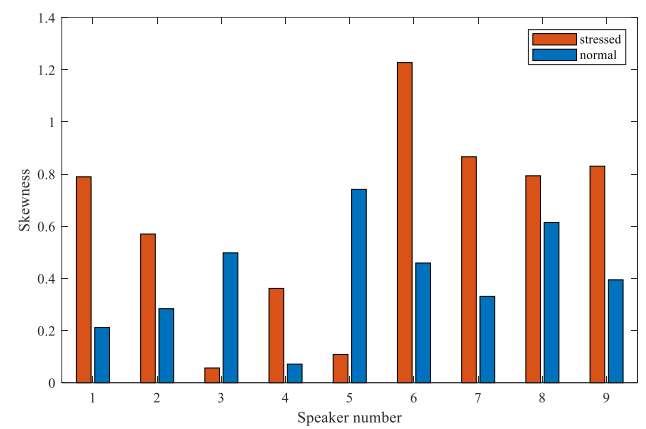


Fig. 11. Comparison of jitter skewness for all speakers

4.3 RAP jitter

The RAP jitter (3) under stress was smaller for eight speakers. However, the differences between stressed and normal speech are very small for three individual speakers (speaker #4, #5, and #7), as shown in Fig. 12 (p-value 0.0042). The largest difference of 72.5 μ s can be seen in speaker #8.

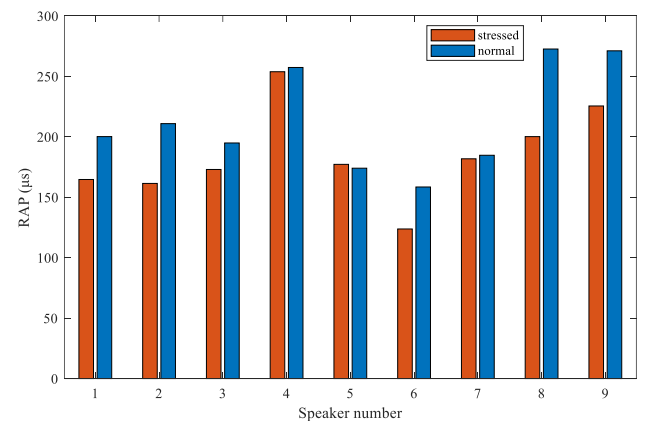


Fig. 12. Comparison of RAP jitter for all speakers

4.4 Continuously increasing or decreasing periods

Sequences of periods with a predefined one direction of change in consecutive periods from 2 to 6 periods were measured separately for continual increase and decrease. The sum of the lengths of the found predefined sequences was related to the total number of all found sequences in the speech and expressed as likelihood. The comparison between stressed and normal speech was made for each speaker using a simple difference:

$$\text{stress likelihood} - \text{normal likelihood}$$

Table 3 summarizes the results achieved by all speakers. Here, “Average change” denotes the average value of the changes in a given sequence calculated from the data of all speakers. Column “Best individual change” is the largest difference achieved by a speaker, which is indicated by the identification number in the “Best speaker #” column. Column “Speakers with $S > N$ ” shows the number of speakers out of all 9 speakers tested whose stress likelihood is greater than

normal likelihood. On the other hand, “Speakers with $S < N$ ” shows the number of speakers with a lower stress likelihood than normal likelihood. Given the average changes and corresponding p-values, the group of changes with increases appears to be more effective than the group of changes with decreases.

The best result of all sequence types, i.e., the largest difference, is for two increases (thus three consecutive periods in the sequence), as shown in the first row of Tab. 3. The result is generated by the likelihoods from individual speakers in Fig. 13. Seven speakers satisfy relation $S < N$, while two speakers satisfy relation $S > N$. This gives a reliability of 77.8%. The speakers with the opposite direction of change in Fig. 13 (speakers #1 and #3) show noticeably larger differences than the others for the sequence of three decreases, as shown in Fig. 14. Combining the corresponding likelihoods in Figs. 13 and 14 in the form of a sum therefore has the potential to achieve 100% reliability.

Table 3. Sequences of continuously increasing periods (rise) and decreasing periods (fall)

Change x times	Average change %	Best individual change %	Speakers with $S > N$	Speakers with $S < N$	Best speaker #	p-value of the t-test
Rise 2	-3.67	-8.05	2	7	2	0.0128
Rise 3	0.60	3.49	5	4	2	0.140
Rise 4	0.74	3.72	5	4	4	0.135
Rise 5	0.87	2.46	7	2	2	0.0162
Rise 6	0.49	1.92	8	1	8	0.0484
Fall 2	-1.61	-14.2	4	5	4	0.221
Fall 3	-0.51	-4.77	5	4	3	0.313
Fall 4	0.63	2.95	5	4	4	0.158
Fall 5	0.62	3.30	6	3	4	0.111
Fall 6	0.22	1.30	5	4	8	0.223

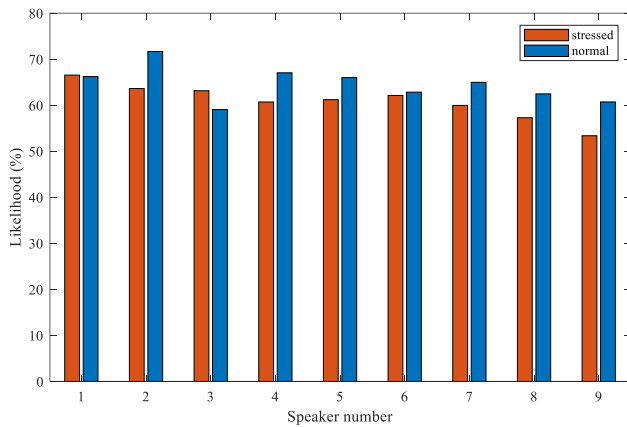


Fig. 13. Comparison of sequences with two consecutive increases

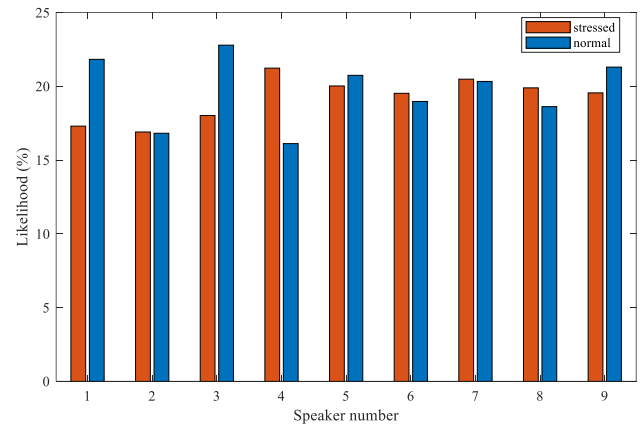


Fig. 14. Comparison of sequences with three consecutive decreases

4.5 Intermittently increasing periods

Groups of fixed-length sequences formed by intermittently increasing periods were observed. The results are summarized in Tab. 4, where the meaning of the columns is similar to that in Tab. 3. In each row, period sequences of exactly one, three, or five intermittent increases are considered. This means that sequences of one increase are not included again in sequences of three and five increases. The results shown in Tab. 4

are better than those in Tab. 3, both in terms of average change and the ratio $(S>N)/(S<N)$. Figures 15, 16 and 17 show the likelihood of sequences with one, three and five intermittent increase for all individual speakers. The best average change corresponds to one increase (see first row of Tab. 4). The best reliability of 100% was achieved by sequences with three and five increases.

Table 4. Sequences of intermittently increasing periods

Intermittent increases	Average change %	Best individual change %	Speakers with $S>N$	Speakers with $S<N$	Best speaker #	p-value of the t-test
1 time	-3.01	-5.41	1	8	9	0.00073
3 times	-1.87	-3.93	0	9	4	0.00082
5 times	-0.75	-1.21	0	9	4	7.6×10^{-6}

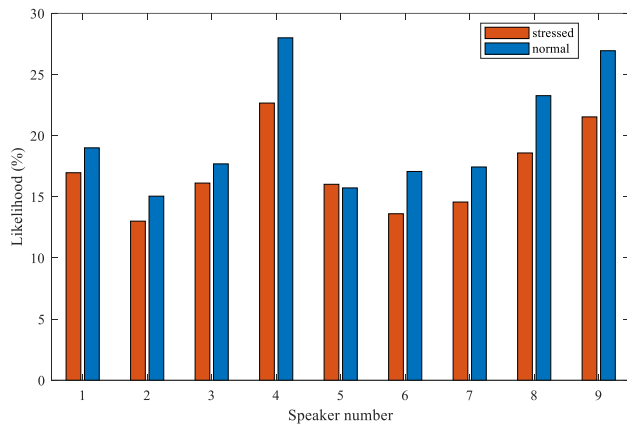


Fig. 15. Comparison of sequences with one intermittent increase

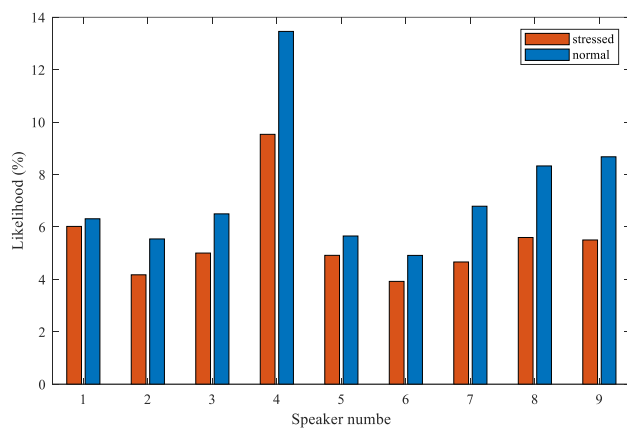


Fig. 16. Comparison of sequences with three intermittent increases

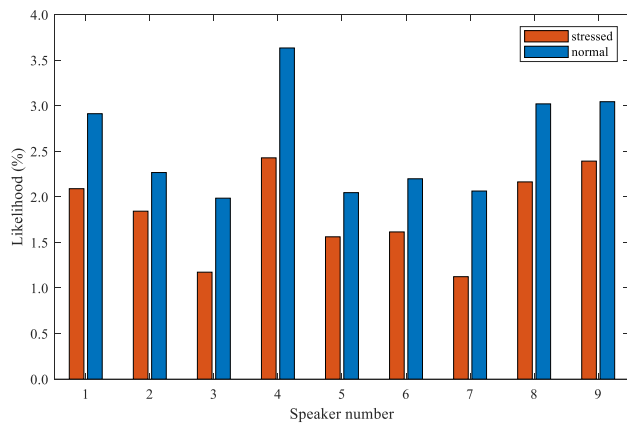


Fig. 17. Comparison of sequences with five intermittent increases

4.6 Comparison of experimental results

In summary, we examined five types of features, some of which were further expanded into multiple variants using several derived parameters. A total of 19 individual parameters were examined as follows:

- Length of period
- Local jitter statistical parameters:
 - mean
 - standard deviation
 - kurtosis
 - skewness
- RAP jitter
- Continuously increasing or decreasing periods
 - Rise times 2, 3, 4, 5, 6
 - Fall times 2, 3, 4, 5, 6
- Intermittently increasing periods
 - Rise times 1, 3, 5

The reliability of individual parameters varies over a relatively wide range. The best reliability of 100% was achieved for two variants of intermittent increase in period length, namely sequences 3 and 5.

4.7 Results achieved in the works of other authors

A direct objective comparison of the achieved results with those of other authors in the field of stress detection should ideally be performed with the same set of test speech signals. However, based on recently published reports, the current state-of-the-art in the success of stress detection in speech can be compared indirectly. From the perspective of separate evaluation of individual feature types, 91% accuracy is reported for gammatone cepstral coefficients in Staš et al. 2025 [6] and 92.8% accuracy for fundamental frequency in Sigmund 2013 [22]. Higher accuracy can be achieved by using more types of features. For example, the fusion of Teager energy operator and MFCC showed an average accuracy of 95.6% for male and female speakers together in Nordin 2022 [9] and 93.9% for female speakers only in Zainal 2025 [27]. This set of features was also used to classify speech into two-level stress classes in Staš 2023 [28] with an accuracy of 64%. By combining mel-spectrogram and MFCC, an accuracy of 97.1% was achieved in Chyan et al. 2022 [29]. The multi-feature approach proposed in Kumar et al. 2024 [30] was tested on patients with an accuracy of 98%. A systematic review of 12 recent studies on stress detection using machine- and deep-learning was presented in Demareva 2026 [31]. Most of the models in this review achieved accuracies in the range from 55 to 85%.

A clear summary of the achieved accuracies is provided in Tab. 5.

Table 5. Overview of previous results by other authors

Features used		Accuracy	Author	Year	Citation
gammatone cepstral coefficients		91%	Staš et al.	2025	[6]
fundamental frequency		93%	Sigmund	2013	[22]
fusion of Teager energy operator and MFCC	male and female speakers together	96%	Nordin et al.	2022	[9]
	female speakers only	94%	Zainal	2025	[27]
speech derived parameters		64%	Staš	2023	[28]
combining mel-spectrogram and MFCC		97%	Chyan et al.	2022	[29]
multi-feature		98%	Kumar et al.	2024	[30]
machine- and deep-learning		55% - 85%	Demareva	2026	[31]

5 Conclusion

This study presents measurable changes in voice based on period fluctuations in the speech signal, which are induced by psychological stress. Intermittent increase in the period length is a new potential feature for stress detection that emerged from our experimental work. The introduction of this feature is the original contribution of our work. The promising success of 100% reliability achieved with our small group of 9 speakers needs to be validated with a larger and more representative number of people, as well as with other real-world stress scenarios. A valuable finding seems to be the fact that each speaker has a “best parameter” among the observed parameters that is most suitable for indicating stress in them. This may simplify the optimization of long-term stress monitoring that is tailored to the speaker. Speaker dependent optimization of stress detection is a relevant motivation for searching for new speech features reflecting stress.

In future work, the general significance of the presented results will be verified with more speakers, including women. It is known that some standard voice features differ between men and women [32, 33]. Furthermore, a study [34] found that women are more sensitive to stress than men due to their biological factors. It will be interesting to see how gender specificities are reflected in the proposed stress features. Future research also should investigate the influence of ambient noise, especially non-stationary noise [35], and varying vocal effort [36] on the robustness of the features used. We further plan to explore the applicability of the proposed features for purposes other than stress indication, for example, for

detection of some neurodegenerative diseases that are reflected in the voice, such as Parkinson’s disease [37], dysarthria [38], and the like.

References

- [1] A. Protopapas and P. Lieberman, “Fundamental frequency of phonation and perceived emotional stress,” *The Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2267–2277, 1997. doi:10.1121/1.418247.
- [2] R. K. Garg and R. K. Sharma, “Stress effects on fundamental frequency of human voice: A review,” *International Journal of Healthcare Technology and Management*, vol. 12, no. 5/6, pp. 390–404, 2011. doi:10.1504/ijhtm.2011.042370.
- [3] L. Schewski, “Measuring negative emotions and stress through acoustic correlates in speech: A systematic review,” *PLOS One*, vol. 20, no. 7, 2025. doi:10.1371/journal.pone.0328833.
- [4] M. Kappen, “Speech as an indicator for psychosocial stress: A network analytic approach,” *Behavior Research Methods*, vol. 54, no. 2, pp. 910–921, 2021. doi:10.3758/s13428-021-01670-x.
- [5] S. Sondhi, “Acoustic analysis of speech under stress,” *International Journal of Bioinformatics Research and Applications*, vol. 11, no. 5, pp. 417–432, 2015. doi:10.1504/ijbra.2015.071942.
- [6] J. Staš, S. Ondáš and J. Juhár, “Performance evaluation of different speech-based emotional stress level detection approaches,” *IEEE Access*, vol. 13, pp. 112880–112904, 2025. doi:10.1109/access.2025.3584534.
- [7] M. Stanek and M. Sigmund, “Finding the most uniform changes in vowel polygon caused by psychological stress,” *Radioengineering*, vol. 24, no. 2, pp. 604–609, 2015. doi:10.13164/re.2015.0604.
- [8] A. C. Kelly and Ch. Gobl, “A comparison of mel-frequency cepstral coefficient (MFCC) calculation techniques,” *Journal of Computing*, vol. 3, pp. 62–66, 2011.
- [9] M. S. Nordin, “Stress detection based on TEO and MFCC speech features using Convolutional Neural Networks (CNN),” *2022 IEEE International Conference on Computing (ICOCO)*, pp. 84–89, 2022. doi:10.1109/icoco56118.2022.10031771.

- [10] J. Staš, K. Škotková and L. Hudák, "Comparison of feature extraction algorithms for speech-based emotional stress level detection," *35th International Conference Radioelektronika*, pp. 1–6, 2025. doi:10.1109/radioelektronika65656.2025.11008416.
- [11] X. Yao, "Glottal source related to stressed speech under workload in human-robot interface," *2nd International Conference on Robotics and Automation Engineering (ICRAE)*, pp. 141–145, 2017. doi:10.1109/icrae.2017.8291369.
- [12] M. Stanek and M. Sigmund, "Psychological stress detection in speech using return-to-opening phase ratios in glottis," *Elektronika ir Elektrotechnika*, vol. 21, no. 5, pp. 59–63, 2015. doi:10.5755/j01.eie.21.5.13336.
- [13] O. Simantiraki, G. Giannakakis, A. Pampouchidou and M. Tsiknakis, "Stress detection from speech using spectral slope measurements," *International Symposium on Pervasive Computing Paradigms for Mental Health*, pp. 41–50, 2015.
- [14] X. Yao, "Research on speech under stress based on glottal source using a physical speech production model," *IEEE Access*, vol. 6, pp. 44473–44482, 2018. doi:10.1109/access.2018.2860130.
- [15] J. H. Hansen and S. E. Bou-Ghazale, "Getting started with SUSAS: A speech under simulated and actual stress database," *5th European Conference on Speech Communication and Technology (Eurospeech)*, pp. 1743–1746, 1997. doi:10.21437/eurospeech.1997-494.
- [16] H. Strik and L. Boves, "Control of fundamental frequency, intensity and voice quality in speech," *Journal of Phonetics*, vol. 20, no. 1, pp. 15–25, 1992. doi:10.1016/s0095-4470(19)30250-5.
- [17] W. E. Cooper and J. M. Sorensen, *Fundamental frequency in sentence production*. New York, USA: Springer-Verlag, 2012.
- [18] L. R. Rabiner and R. W. Schafer, *Theory and Applications of Digital Speech Processing*. London: Pearson Prentice Hall, 2011.
- [19] R. J. Baken and R. F. Orlikoff, *Clinical measurement of speech and voice*. San Diego: Singular Thomson Learning, 2000.
- [20] I. Vigo, L. Coelho and S. Reis, "Speech- and language-based classification of Alzheimer's disease: A systematic review," *Bioengineering*, vol. 9, no. 1, p. 27, 2022. doi:10.3390/bioengineering9010027.
- [21] C. Oh, R. J. Morris and X. Wang, "A systematic review of expressive and receptive prosody in people with dementia," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 10, pp. 3803–3825, 2021. doi:10.1044/2021_jslhr-21-00013.
- [22] M. Sigmund, "Statistical analysis of fundamental frequency based features in speech under stress," *Information Technology and Control*, vol. 42, no. 3, pp. 286–291, 2013. doi:10.5755/j01.itc.42.3.3895.
- [23] T. Roy, "The effects of exam-induced stress on EEG profiles and memory scores," *Behavioral Sciences*, vol. 13, no. 5, p. 373, 2023. doi:10.3390/bs13050373.
- [24] Y. H. Gulhane, S. V. Rode and S. A. Ladhake, "Application of fuzzy logic in stress analysis," *International Conference & Workshop on Emerging Trends in Technology*, pp. 679–685, 2011. doi:10.1145/1980022.1980168.
- [25] D. Veiga, "The fundamental frequency of voice as a potential stress biomarker: A systematic review and meta-analysis," *Stress and Health*, vol. 41, no. 5, e70112, 2025. doi:10.1002/smi.70112.
- [26] K. Pisanski, J. Nowak and P. Sorokowski, "Individual differences in cortisol stress response predict increases in voice pitch during exam stress," *Physiology & Behavior*, vol. 163, pp. 234–238, 2016. doi:10.1016/j.physbeh.2016.05.018.
- [27] N. A. Zainal, "Utilizing MFCCs and TEO-MFCCs to classify stress in females using SSNNA," *IJUM Engineering Journal*, vol. 26, no. 1, pp. 324–335, 2025. doi:10.31436/ijumej.v26i1.3411.
- [28] J. Staš, "Analysis and detection of speech under emotional stress," *21st International Conference on Emerging eLearning Technologies and Applications*, pp. 493–498, 2023. doi:10.1109/iceta61311.2023.10343755.
- [29] P. Chyan, A. Achmad, I. Nurtanio and I. S. Areni, "A deep learning approach for stress detection through speech with audio feature analysis," *6th International Conference on Information Technology, Information Systems and Electrical Engineering*, pp. 269–273, 2022. doi:10.1109/ICITISEE57756.2022.10057845.
- [30] A. Kumar, M. A. Shaun and B. K. Chaurasia, "Identification of psychological stress from speech signal using deep learning algorithm," *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 9, 100707, 2024. doi:10.1016/j.prime.2024.100707.
- [31] V. Demareva, "Voice-based stress detection: A systematic review of machine- and deep-learning approaches (2021–2025)," *Chaos, Solitons & Fractals*, vol. 208, 118372, 2026. doi:10.1016/j.chaos.2026.118372.
- [32] M. Sigmund, "Gender distinction using short segments of speech signal," *International Journal of Computer Science and Network Security*, vol. 8, no. 10, pp. 159–162, 2008.
- [33] J. Pribil, A. Pribilova and J. Matousek, "Comparison of formant features of male and female emotional speech in Czech and Slovak," *Elektronika ir Elektrotechnika*, vol. 19, no. 8, pp. 83–88, 2013. doi:10.5755/j01.eee.19.8.1739.
- [34] Y. S. Balhara, R. Verma and C. Gupta, "Gender differences in stress response: Role of developmental and biological determinants," *Industrial Psychiatry Journal*, vol. 20, no. 1, pp. 4–10, 2011. doi:10.4103/0972-6748.98407.
- [35] P. Zelinka and M. Sigmund, "Hierarchical classification tree modeling of nonstationary noise for robust speech recognition," *Information Technology and Control*, vol. 39, no. 3, pp. 202–210, 2010.
- [36] Y. A. S. Lien, "Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception," *Journal of Speech, Language, and Hearing Research*, vol. 58, no. 3, pp. 566–575, 2015. doi:10.1044/2015_jslhr-s-14-0194.
- [37] H. Wright and V. Aharonson, "Vocal feature changes for monitoring Parkinson's disease progression – A systematic review," *Brain Sciences*, vol. 15, no. 3, p. 320, 2025. doi:10.3390/brainsci15030320.
- [38] N. Calaf, "Dysarthria assessment across Spain: A survey study of tools, practices, and needs," *International Journal of Language & Communication Disorders*, vol. 60, no. 5, e70122, 2025. doi:10.1111/1460-6984.70122.

Received 1 February 2026