

## Article

# Stress Level Detection and Evaluation from Phonation and PPG Signals Recorded in an Open-Air MRI Device <sup>†</sup>

Jiří Přibil \* , Anna Přibilová and Ivan Frollo

Institute of Measurement Science, Slovak Academy of Sciences, 841 04 Bratislava, Slovakia;  
Anna.Pribilova@savba.sk (A.P.); ivan.frollo@savba.sk (I.F.)

\* Correspondence: Jiri.Pribil@savba.sk; Tel.: +421-2-5910-4543

<sup>†</sup> This paper is an extended version of our paper published in 2021 44th International Conference on Telecommunications and Signal Processing (TSP), 26–28 July 2021 (Virtual Conference).

**Abstract:** This paper deals with two modalities for stress detection and evaluation—vowel phonation speech signal and photo-plethysmography (PPG) signal. The main measurement is carried out in four phases representing different stress conditions for the tested person. The first and last phases are realized in laboratory conditions. The PPG and phonation signals are recorded inside the magnetic resonance imaging scanner working with a weak magnetic field up to 0.2 T in a silent state and/or with a running scan sequence during the middle two phases. From the recorded phonation signal, different speech features are determined for statistical analysis and evaluation by the Gaussian mixture models (GMM) classifier. A database of affective sounds and two databases of emotional speech were used for GMM creation and training. The second part of the developed method gives comparison of results obtained from the statistical description of the sensed PPG wave together with the determined heart rate and Oliva–Roztocil index values. The fusion of results obtained from both modalities gives the final stress level. The performed experiments confirm our working assumption that a fusion of both types of analysis is usable for this task—the final stress level values give better results than the speech or PPG signals alone.

**Keywords:** stress detection and evaluation; GMM-based classification; photo-plethysmographic wave analysis



**Citation:** Přibil, J.; Přibilová, A.; Frollo, I. Stress Level Detection and Evaluation from Phonation and PPG Signals Recorded in an Open-Air MRI Device. *Appl. Sci.* **2021**, *11*, 11748.  
<https://doi.org/10.3390/app112411748>

Academic Editors: Norbert Herencsar, Francesco Benedetto and Jorge Crichigno

Received: 12 November 2021  
Accepted: 8 December 2021  
Published: 10 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Magnetic resonance imaging (MRI) is used to visualize anatomical structures in various medical applications. Apart from whole-body MRI, open-air and extremity MRI also have wide usage. Every MRI scanner contains a gradient coil system generating three orthogonal magnetic fields to scan the object in three spatial dimensions. All these devices produce significant mechanical pulses during the execution of a scan sequence resulting from rapid switching of electrical currents that accompany rapid change in the of direction of the Lorentz force. This mechanical vibration is the source of the acoustic noise radiating from the whole system with possible negative effect on the patients as well as the health personnel [1] manifesting as a stress during or after MRI scanning.

MRI is also used to obtain vocal tract shapes during the articulation of speech sounds for the articulatory synthesis [2]. An open-air MRI scanner can be used for this purpose where the examined articulating person lies directly on the plastic cover of the bottom gradient coil while a chosen MR sequence is run. Here the stress-evoking vocal cord tension has an influence on the recorded speech signal [3] by modifying its suprasegmental and spectral features, so it can bring about errors and inaccuracy in the calculation of 3D models of the human vocal tract [4]. This physiological and mental stress can effectively be identified by the parameters derived from the photo-plethysmography (PPG) signal, as heart rate (HR), Oliva–Roztocil index (ORI) [5] pulse transit time [6], pulse wave velocity [7], blood oxygen saturation, cardiac output [8], and others. The amplitude of the picked-up

PPG signal is usually not constant, and it can often be partially disturbed or degraded [9]. The stress is associated with the autonomic nervous system and it can be expressed by higher variability in interbeat intervals (IBI) assessed from the PPG wave as pulse rate variability (PRV) and from the electrocardiogram (ECG) as HR variability (HRV). The variety of frequency spectra determined from PPG and ECG signals can be used for more precise determination of changes in the PRV and HRV values. They are in principle not equivalent because they are caused by different physiological mechanisms. In addition, the level of agreement between the PRV and HRV statistical results depends on several technical factors, e.g., the used sampling frequency or the method of IBI determination [10].

In many people, exposure to acoustic noise and/or vibration causes negative psychological reaction that can be identified with negative emotional states of anger, fear, or panic. Recognition of these negative affective states in the speech signal of the noise-exposed speaker may be used as another stress indicator. All discrete emotions including the six basic ones (anger, disgust, fear, sadness, surprise, joy) can be quantified by two parameters representing dimensions of valence (pleasure) and arousal [11]. The valence dimension reflects changes of the affect from positive (e.g., surprise, joy) to negative (e.g., anger, fear); the arousal dimension ranges from passive (e.g., sadness) to active (e.g., joy, anger) [12]. For emotion detection in the speech signal, various approaches have been used so far. Hidden Markov models were used for performance evaluation of different features: log frequency power coefficients, linear prediction cepstral coefficients, and standard mel-frequency cepstral coefficients (MFCC) [13]. The support vector machines (SVM) [14] employed features extracted from cross-correlograms of emotional speech signals [15]. Another group of speech emotion recognition methods uses artificial neural networks [16]. Recently, machine learning and deep learning approaches have been utilized in this context [17,18]. However, the technique using Gaussian mixture models (GMM) [19] remains the method of choice when dealing with speech emotion recognition [20,21]. Much better scores are achieved by a fusion of different recognition methods, e.g., GMM and SVM in speaker age and gender identification [22] or in speaker verification [23], or SVM and K-nearest neighbour in speech emotion recognition [24]. Another improvement may be achieved by multimodal approach to emotion recognition using a fusion of features extracted from audio signals, text transcriptions, and visual signals of face expressions [25]. In this sense, we use two modalities for stress detection in this paper: the recorded speech signal and the sensed PPG signal.

Our research aim is to detect and quantify the effect of vibration and acoustic noise during the MR scan examination on vocal cords of an examined person. In the performed experiments, the tested person articulated while lying in the scanning area of the open-air low field MRI tomograph [26]. The levels of the vibration and noise in the MRI depend on several factors [27,28]. At first, they comprise a class of a scan sequence based on a physical principle of generation of the free induction decay (FID) signal by the non-equilibrium nuclear spin magnetization precession (gradient or spin echo classes). Next, they depend on the used methodology of MR image construction from received FID signals (standard, turbo, hi-resolution, 3D, etc.). Finally, the basic parameters of MR scan sequences (repetition time TR, echo time TE, slice orientation, etc.) and additional settings (number of accumulations, number of slices, their thickness, etc.) are chosen depending on the required final quality of MR images. All these parameters together with an actual volume depending on a tested person's weight have influence on the intensity of the produced vibration and noise, on the time duration of the MR scan process, and finally on the stimulated physiological and psychological stress in the examined persons. In previous research [29,30] the measured PPG signals together with the derived HR have already been used to monitor the physiological impact of vibration and acoustic noise on a person examined inside the MRI scanning device.

This paper describes the current experimental work focused on stress detection and evaluation from speech records of vowel phonation picked up together with PPG signals. The whole experiment consists of four measurement phases representing different stress

conditions for the tested person. The PPG and phonation signal measurement of the first and the fourth phases is realized in the laboratory conditions; in the second and third phases the tested person lies inside the MRI equipment; the third measurement phase is realized after exposure to vibration and noise during scanning in the MRI device. The first part of the proposed method for stress detection and evaluation uses the recorded phonation signal. From this signal, different speech features are determined for statistical analysis and evaluation with the help of a GMM classifier. For GMM creation and training, one database of affective sounds and two databases containing emotional speech are used. The second part of the stress evaluation method gives comparison of the results obtained from the statistical processing of HR and ORI values determined from the PPG signal. This is supplemented by comparison of energetic, time, and statistical parameters describing the sensed PPG waves. The fusion of the results obtained from both types of stress analysis methods gives the final stress level.

## 2. Description of the Proposed Method

### 2.1. Detection and Evaluation of the Stress in the Phonation Signal Based on the GMM Classifier

The GMM-based classification works in the following way: the input data investigated are approximated by a linear combination of Gaussian probability density functions. They are used to calculate the covariance matrix as well as the vectors of means and weights. Next, the clustering operation organizes objects into groups whose members are similar in some way. The  $k$ -means algorithm determining the centers is used for GMM parameters initialization. This procedure is repeated several times until a minimum deviation of the input data sorted in  $k$  clusters  $S = \{S_1, S_2, \dots, S_k\}$  is found. Subsequently, the iteration algorithm of expectation-maximization determines the maximum likelihood of the GMM [19]. The number of mixtures ( $N_{\text{MIX}}$ ) and the number of iterations ( $N_{\text{ITER}}$ ) have an influence on the execution of the training algorithm—mainly on the time duration of this process and on the GMM accuracy. The GMM classifier returns the probability/score ( $T, n$ )—for the model  $SM_n(n)$  corresponding to each of  $N$  output classes using the feature vector  $T$  from the processed signal. The normalized scores (in the range from 0 to 1) obtained in this way are further processed in the classification/detection/evaluation procedures.

The proposed method uses partially normalized GMM scores obtained during the classification process for three output classes:

- $C_{1N}$  for the normal speech represented by a neutral state and emotions with positive valence and low arousal,
- $C_{2S}$  for the stressed speech modeled by emotions with negative pleasure and high arousal,
- $C_{3O}$  comprising the remaining two of six primary emotions (sadness having negative pleasure with low arousal and joy as a positive emotion with high arousal).

The developed stress evaluation system analyzes the input phonation signal of five basic vowels (“a”, “e”, “i”, “o”, and “u”) obtained from voice records together with the PPG signal sensed in  $M$  measuring phases  $MF_1, MF_2, \dots, MF_M$ . During the GMM classification we obtain  $M$  output matrices of normalized scores with dimension  $P \times N$ , i.e., for  $P$  processed input frames of the analyzed phonation signal and for each of  $N$  output classes—see the block diagram in Figure 1. Then, the relative occurrence parameters  $RO_{C_{1N}, C_{2S}, C_{3O}}$  [%] are calculated as partial winners of  $C_{1N}, C_{2S}, C_{3O}$  classes (with maximum probability scores) separately for each of the analyzed vowels recorded in the  $MF_1$  to  $MF_M$  measuring phases. Then, summary mean values of the  $C_{1N}$  and  $C_{2S}$  class occurrence percentage ( $\overline{RO_{C_{1N}}}, \overline{RO_{C_{2S}}}$ ) quantify differences between measuring phases. The stress factor in [%] is defined as

$$L_{\text{STRESS}}(n) = \overline{RO_{C_{2S}}}(n) - \overline{RO_{C_{2S}}}(1) \quad \text{for } 1 \leq n \leq M \quad (1)$$

This practically corresponds to the mean percentage occurrence for the  $C_{2S}$  class relative to the first recording phase as the baseline—which means  $L_{STRESS}(1) = 0$ . The same methodology is used for  $L_{NORMAL} [\%]$  calculation

$$L_{NORMAL}(n) = \overline{RO_{C_{1N}}}(n) - \overline{RO_{C_{1N}}}(1) \text{ for } 1 \leq n \leq M, \quad (2)$$

which expresses changes corresponding to the normal speech type. While the sum of occurrences of  $RO_{C_{1N}}, C_{2S}, C_{3O}$  parameters is always 100%, actual values of  $L_{STRESS}/L_{NORMAL}$  depend not only on  $C_{2S}/C_{1N}$  classes but also on the current distribution of the class  $C_{3O}$ —compare graph examples in Figure 2.

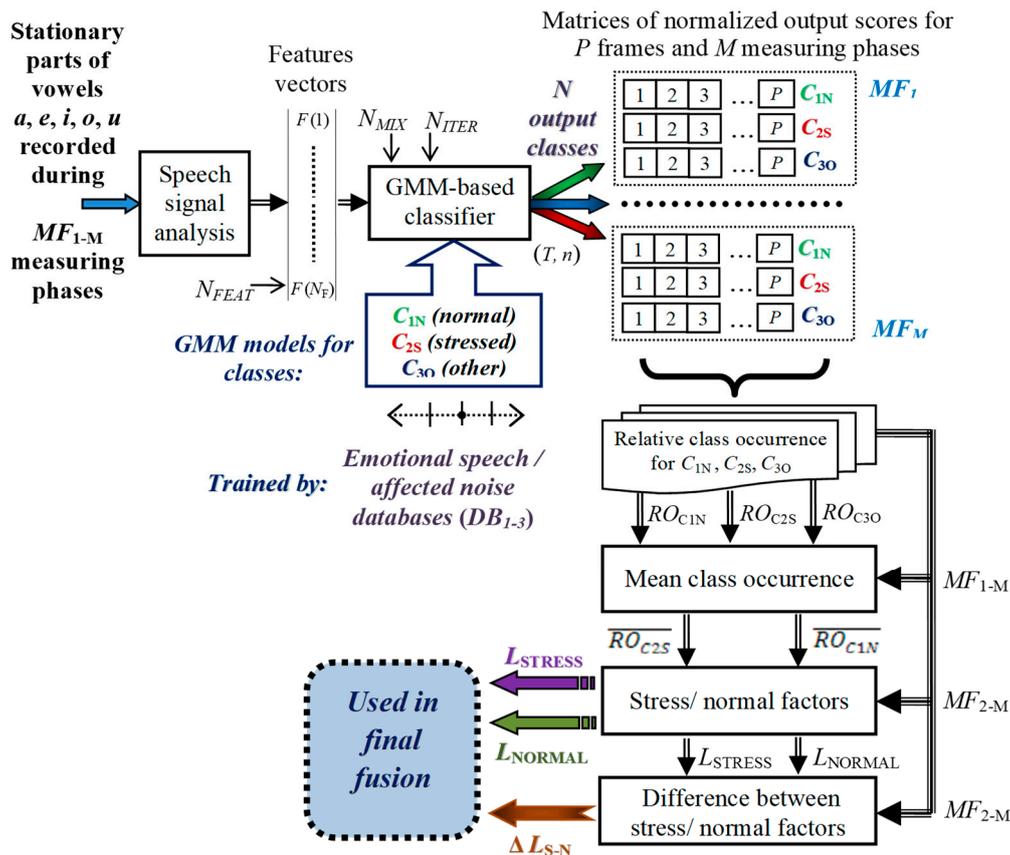


Figure 1. Block diagram of the GMM-based system for stress detection and evaluation in a phonation speech signal.

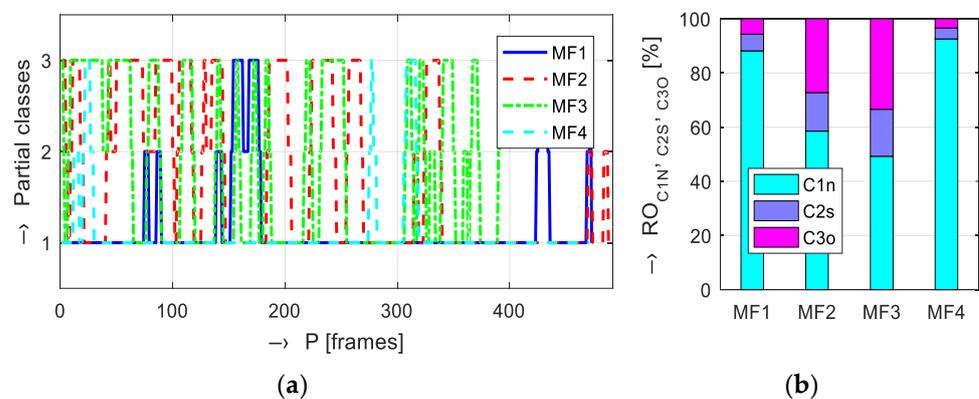


Figure 2. Example of the GMM classification and stress evaluation: (a) sequences of obtained partial winner classes  $C_{1N}$  ("1"),  $C_{2S}$  ("2"), and  $C_{3O}$  ("3") of a vowel "e", (b) bar-graph of relative class occurrences  $RO_{C_{1N}}, C_{2S}, C_{3O}$ .

The desired functionality of the proposed evaluation method expects that the phonation signal produced in the stressed conditions is marked by higher values of  $\overline{RO_{CS2}}$  parameter together with lower  $\overline{RO_{CN1}}$  values. For more significant comparison, the difference  $\Delta L_{S-N}$  between the stress ( $L_{STRESS}$ ) and normal ( $L_{NORMAL}$ ) factors is calculated for  $MF_2$  to  $MF_M$  phases. The negative value of  $\Delta L_{S-N}$  difference corresponds to the  $L_{NORMAL}$  value higher than the  $L_{STRESS}$  value. Sufficiently great differences of  $\Delta L_{S-N}$  between the stressed and normal phonation signals are necessary for proper evaluation processes. While the  $\Delta L_{S-N}$  in the first phase is principally equal to zero, the  $\Delta L_{S-N}$  for the last measuring phase is typically non-zero with lower absolute value and possible opposite polarity compared with previous phases. The  $L_{STRESS}$ ,  $L_{NORMAL}$ , and  $\Delta L_{S-N}$  are used as the GMM classification parameters ( $SP_{GMM}$ ) and they are used together with the PPG signal analysis parameters ( $SP_{PPG}$ ) to form the input vectors for further fusion operation (see the block diagram in Figure 3). The final stress evaluation rate  $R_{SFE}$  is given as

$$R_{SFE}(n) = \sum_{i=1}^Q (w_{GMM}(n, i) \cdot SP_{GMM}(n, i)) + \sum_{j=1}^S (w_{PPG}(n, j) \cdot SP_{PPG}(n, j)), \quad 2 \leq n \leq M, \quad (3)$$

where  $Q$  is the number of GMM parameters,  $S$  is the number of PPG parameters, and  $w_{GMM}/w_{PPG}$  are their importance weights.

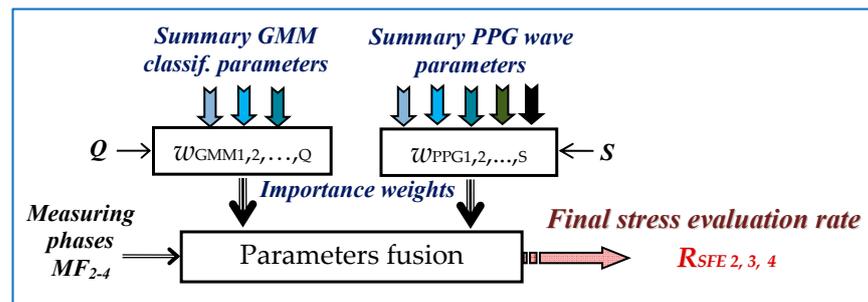


Figure 3. Block diagram of the fusion procedure to obtain the final stress evaluation rate.

### 2.2. Determination of Phonation Features for Stress Detection

For stress recognition in the speech, spectral properties such as MFCC together with prosodic parameters (jitter and shimmer) and energetic features such as Teager energy operators (TEO) are mostly used [31,32]. In the frame of the current experiments, we use four types of parameters for analysis of the phonation signal:

1. Prosodic features containing micro-intonation components of the speech melody  $F_0$  given by a differential contour of a fundamental frequency  $F_{0DIFF}$ , absolute jitter  $J_{abs}$  as an average absolute difference between consecutive pitch periods  $L$  measured in samples, shimmer as a relative amplitude perturbation  $AP_{rel}$  from peak amplitudes detected inside the  $n$ th signal frame, and signal energy  $En_{TK}$  for  $P$  processed frames calculated as

$$En_{TK} = abs \left( \frac{1}{P-2} \sum_{n=1}^{P-2} TEO(n) \right), \quad (4)$$

where the Teager energy operator is defined as  $TEO = x(n)^2 - x(n-1) \cdot x(n+1)$ .

2. Basic spectral features comprising the first two formants ( $F_1, F_2$ ), their ratio ( $F_1/F_2$ ) and 3-dB bandwidth ( $B_{31}, B_{32}$ ) calculated with the help of the Newton–Raphson formula or the Bairstow algorithm [33], and  $H1-H2$  spectral tilt measure as a difference between  $F_1$  and  $F_2$  magnitudes.
3. Supplementary spectral properties consisting of the center of spectral gravity, i.e., an average frequency weighted by the values of the normalized energy of each frequency component in the spectrum in [Hz], spectral flatness measure (SFM) determined as a ratio of the geometric and the arithmetic means of the power spectrum, and spectral

entropy (SE) as a measure of spectral distribution quantifying a degree of randomness of spectral probability density represented by normalized frequency components of the spectrum.

4. Statistical parameters that describe the spectrum: spectral spread parameter representing dispersion of the power spectrum around its mean value ( $S_{\text{SPREAD}} = \Sigma^2$ ), spectral skewness as a 3rd order moment representing a measure of the asymmetry of the data around the sample mean ( $S_{\text{SKEW}} = E(x - \mu)^3 / \sigma^3$ ), and spectral kurtosis being a 4th order moment as a measure of peakiness or flatness of the shape of the spectrum relative to the normal distribution ( $S_{\text{KURT}} = E(x - \mu)^4 / \sigma^4 - 3$ ); in all cases  $\mu$  is the first central moment and  $\sigma$  is the standard deviation of spectrum values  $x$ , and  $E(t)$  represents the expected value of the quantity  $t$ .

### 2.3. PPG Signal Description, Analysis, and Processing

The PPG signal together with its derived parameters (particularly HR and ORI) describe the current state of the human vascular system and, in this way, they can be used for detection and quantification of the stress level [7]. Generally, in a PPG cycle, two maxima (systolic and diastolic) provide valuable information about the pumping action of the heart. For description of signal properties of the sensed PPG waves the energetic, time, and statistical parameters are determined.

The sensed PPG signal representation is typically in the absolute numerical range  $A_{\text{NR}}$  given by the used type of an analog-to-digital (A/D) converter, e.g., output values of the 14-bit A/D converter have a relative unipolar representation in the range from 0 to 16,192 ( $=2^{14} = A_{\text{NR}}$ ). First, from this absolute PPG signal, the local maximum  $Lp_{\text{MAX}}$  and local minimum  $Lp_{\text{MIN}}$  levels of the peaks corresponding to the heart systolic pulses are determined to obtain the mean peak level  $Lp_{\text{MEAN}}$ . Then, the mean signal range  $PPG_{\text{RANGE}}$  is calculated from the global minimum (offset level  $L_{\text{OFS}}$ ) and  $A_{\text{NR}}$  by the equation

$$PPG_{\text{RANGE}} = (Lp_{\text{MEAN}} - L_{\text{OFS}}) / A_{\text{NR}} \cdot 100 [\%]. \quad (5)$$

Finally, we calculate the actual modulation (ripple) of heart pulses in percentage ( $HP_{\text{RIPP}}$ ) as

$$HP_{\text{RIPP}} = (Lp_{\text{MAX}} - Lp_{\text{MIN}}) / Lp_{\text{MAX}} \cdot 100 [\%]. \quad (6)$$

The determined  $Lp_{\text{MIN}}$ ,  $Lp_{\text{MAX}}$ ,  $L_{\text{OFS}}$  together with calculated  $PPG_{\text{RANGE}}$  and  $HP_{\text{RIPP}}$  values are visualized in Figure 4.

The used methodology of heart rate values determined via PPG wave has been described in more detail in [30]. In principle, the procedure works in three basic steps: (1) systolic peaks are localized in the PPG signal, (2) heart pulse periods  $T_{\text{HP}}$  in samples are determined, (3) HR values are calculated using the sampling frequency  $f_s$  by a basic formula

$$HR = 60 \cdot f_s / T_{\text{HP}} [\text{min}^{-1}]. \quad (7)$$

The obtained sequence of HR values is next smoothed by a 3-point median filter and the linear trend (LT) is calculated by the mean square method. For  $LT < 0$  the HR has a descending trend, for  $LT > 0$  the HR values have an ascending trend. The resulting angle  $\varphi$  of LT in degrees is defined as  $HR\varphi_{\text{LT}} = (\text{Arctg}(LT) / \pi) \cdot 180$ . For the final stress evaluation rate determination in the fusion process, the relative parameter  $HR\varphi_{\text{REL}} [\%]$  for the  $q^{\text{th}}$  measurement phase is calculated in relation to the  $HR\varphi_{\text{LT}}$  of the 1st phase

$$HR\varphi_{\text{REL}}(q) = ((HR\varphi_{\text{LT}}(q) - HR\varphi_{\text{LT}}(1)) / HR\varphi_{\text{LT}}(1)) \cdot 100 [\%] \text{ for } 2 \leq q \leq M. \quad (8)$$

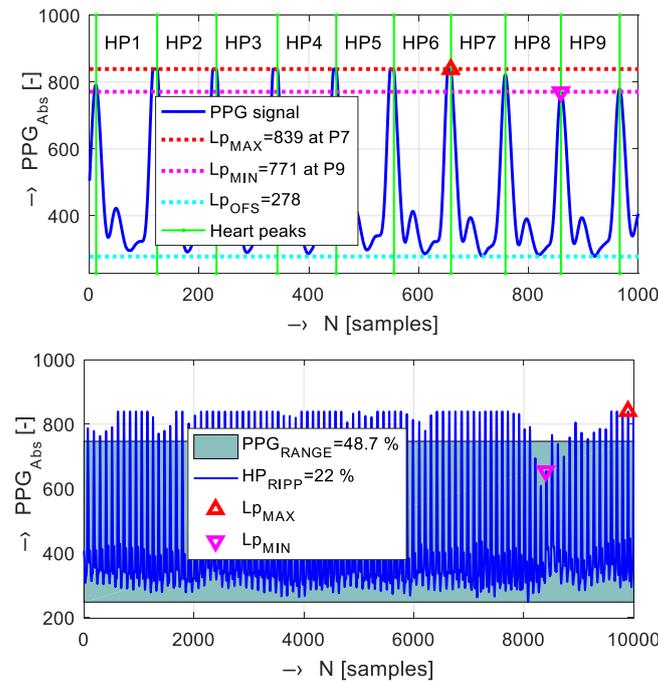
After the mean value  $HR_{\text{MEAN}}$  and LT removal of the smoothed HR sequence a relative variability  $HR_{\text{VAR}}$  based on the standard deviation  $HR_{\text{STD}}$  is calculated as

$$HR_{\text{VAR}} = (HR_{\text{STD}} / HR_{\text{MEAN}}) \cdot 100 [\%]. \quad (9)$$

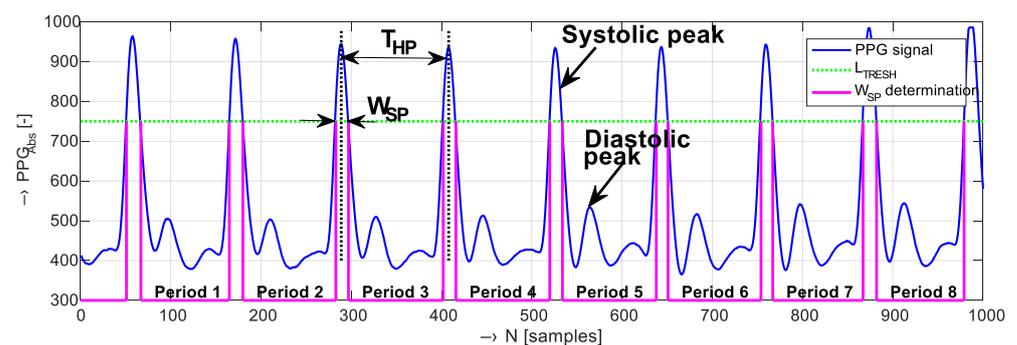
For the purpose of this study, we use the ORI parameter which can also quantify the pain and/or stress in the human cardio-vascular system [6,34]. The typical ORI range lies in the interval of  $\langle 0.1, 0.3 \rangle$  for healthy people in a normal physiological state [10]. This parameter normalizes the width of the systolic pulse  $W_{SP}$  to the heart pulse period  $T_{HP}$  [35]

$$ORI = W_{SP} / T_{HP}, \tag{10}$$

where  $W_{SP}$  is determined typically at the height of two-thirds from the basis (one-third from the top—see Figure 5).



**Figure 4.** Visualization of the PPG signal analysis: detailed 1k-sample example of a PPG wave with localized systolic peaks and partial  $L_{p_{MAX}}/L_{p_{MIN}}, L_{p_{OFs}}$  values (upper graph), 10 k-sample PPG wave used for calculation of  $PPG_{RANGE}$  and  $HP_{RIPP}$  values (lower graph).



**Figure 5.** An example of the PPG signal with localized systolic heart peaks, determined heart pulse periods  $T_{HP}$ , and widths  $W_{SP}$  of systolic peaks at the threshold level  $L_{TRESH}$ .

For the final fusion process, the relative parameter  $ORI_{REL}$  [%] is calculated in a similar manner as  $HR\phi_{REL}$  in (8)—using the mean value  $ORI_{MEAN}$  determined for the phase  $MF_1$

$$ORI_{REL}(q) = ((ORI_{MEAN}(q) - ORI_{MEAN}(1)) / ORI_{MEAN}(1)) \cdot 100 \text{ [%]} \quad \text{for } 2 \leq q \leq M. \tag{11}$$

For the current research, we analyze changes (increase/decrease/stationary state and/or polarity  $\pm$ ) of the mentioned parameters determined from the processed PPG signal. We expect raised PPG ripple and range parameters, higher  $HR\phi_{LT}$  values, higher

HR variability, and smaller ORI (due to narrowed systolic peaks) as indicators of the stress state (equivalent to the  $C_{25}$  class detected during the GMM classification of the phonation signal). In the normal non-stressed state of the tested person, opposite changes are reflected—see a detailed description in Table 1. All these five parameters are used to obtain the final stress evaluation rate. The  $SP_{PPG}$  values become inputs to the fusion procedure in a similar way as the  $SP_{GMM}$  evaluation parameters. Practically, only  $SP_{PPG}$  ( $MF_{2-4}$ ) are applied because the baseline  $SP_{PPG}$  ( $MF_1$ ) is of a zero value.

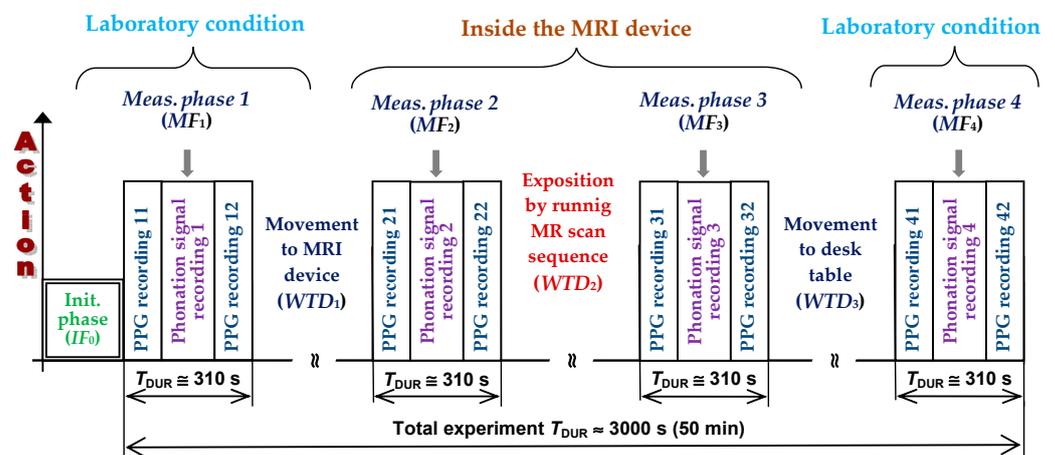
**Table 1.** Corresponding changes of PPG signal properties for stressed and normal states.

Parameter	Stressed State	Normal Condition
$PPG_{RANGE}$ [%]	Increase	Decrease or constant
$HP_{RIPP}$ [%]	Increase	Decrease
$HR\varphi_{REL}$ [%]	Higher positive (+)	Negative (−) or small
$HR_{VAR}$ [%]	Higher	Smaller
$ORI_{REL}$ [%]	Smaller	Higher

### 3. Experiments

#### 3.1. Basic Concept of the Whole Measurement Experiment

The whole experiment is practically divided into four measurement phases ( $MF_{1,2,3,4}$ ) preceded by the initial phase  $IF_0$ —see the principal measurement schedule in Figure 6. The phase  $IF_0$  serves as preparation and manipulation of the measurement instruments—testing the wireless connection between the PPG sensor and the data-storing device, setting audio levels on the mixer device for phonation recording, etc. Prior to each experiment, the air in the room was disinfected by a UV germicidal lamp for 15 min to minimize risk of COVID-19 infection—the phonation signal recording must be performed without any protective face shield or respirator mask.



**Figure 6.** Principal measurement schedule applied in all measurement experiments.

In the case of the measuring phases  $MF_1$  and  $MF_4$ , the tested person sits at the desk in the MRI equipment control room, while for the measurement in the phases  $MF_2$  and  $MF_3$ , the person lies on the bed inside the shielding metal cage of the MRI device. Each of the measuring phases starts with PPG signal recording—the operation called  $PPG_{x_1}$  (where “ $x$ ” represents the number of the current measuring phase) with duration  $T_{DUR}$  equal to 80 s. Then, the phonation signal is recorded with the pick-up microphone. The signal consists of stationary parts of the vowels a, e, i, o, and u with a mean duration of 8 s interlaced by pauses of 2–3 s. Each vowel phonation was repeated three times, so  $5 \times 3 = 15$  records per person were obtained altogether in every individual measuring phase (total of 55 in the whole experiment). The active measurement is finished by the second PPG signal sensing (operation  $PPG_{x_2}$ —also with  $T_{DUR} = 80$  s, so the summary duration of all the measuring

phases is about 5–7 min. Between each two consecutive measurement phases, a working time delay ( $WTD_{1-3}$ ) with time duration 5–10 min is applied. Therefore, the expected experimental duration is about 50 min in its entirety (without the  $IF_0$  phase). During  $WTD_1$ , the tested person moves from the desk to the MRI device and adapts to the space of the scanning area to stabilize physiological changes in the cardiovascular system after changing body position from sitting to lying. Some people can also have a negative mental feeling inside the MRI tomograph. Both types of changes can evoke the stress that can be detected by the PPG and phonation signals. It holds mainly for  $WTD_2$  when the tested person is exposed by negative stimuli consisting of mechanical vibration and acoustic noise generated by the running MRI device during execution of the MR scan sequence. The last  $WTD_3$  delay part is planned for movement of the tested person to the desk in the control room and short relaxation after changing position from lying to sitting and returning to the “normal” laboratory conditions. Importance weights for input parameters  $SP_{GMM}$  and  $SP_{PPG}$  entered to the fusion process were set experimentally as shown in Table 2.

**Table 2.** Weight settings for parameters entered to the fusion process.

Parameter No.	Phonation Type	Weight [-]	PPG Type	Weight [-]
1	$L_{STRESS}$	$w_{GMM1} = 0.75$	$PPG_{RANGE}$	$w_{PPG1} = 0.25$
2	$L_{NORMAL}$	$w_{GMM2} = -0.5$	$HP_{RIPP}$	$w_{PPG2} = 0.5$
3	$\Delta L_{S-N}$	$w_{GMM3} = 0.25$	$HR\phi_{REL}$	$w_{PPG3} = 0.5$
4	–	–	$HR_{VAR}$	$w_{PPG4} = 0.75$
5	–	–	$ORI_{REL}$	$w_{PPG5} = -1$

In this study, two small databases of the phonation and PPG signals from eight healthy voluntary non-smokers were collected and further processed. The examined persons were the authors themselves and their colleagues: four females (F1, F2, F3, and F4) and four males (M1, M2, M3, and M4). The age and body mass index (BMI) composition of the studied persons is listed in Table 3. During the experiments in the control room as well as inside the MRI device, the room temperature was maintained at 24 °C and the measured humidity was 30%.

**Table 3.** Age and BMI parameters of the persons included in our study.

Parameter/Person	M1	M2	M3	M4	F1	F2	F3	F4
Age (years)	59	53	42	36	59	20	30	58
BMI (kg/m <sup>2</sup> )	24.9	22.2	22.5	23.1	18.3	21.8	19.0	21.2

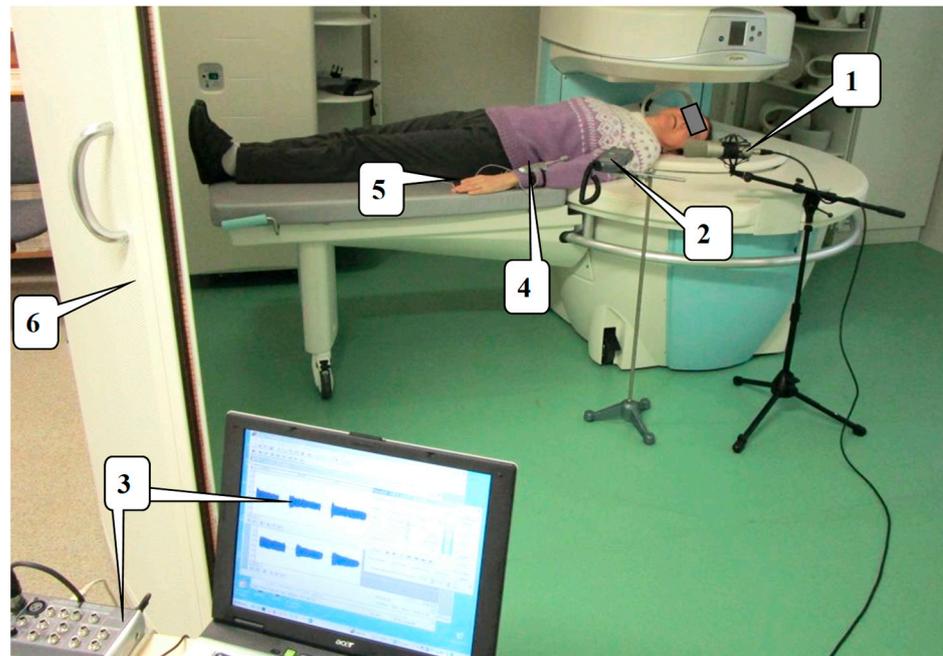
### 3.2. Used Instrumentation and Recording Arrangement

#### 3.2.1. Phonation Signal Recording

In the measurement phases  $MF_2$  and  $MF_3$ , the tested person lay in the scanning area of the open-air, low-field (0.178 T) MRI tomograph Esaote E-scan Opera [36] located at the Institute of Measurement Science, Slovak Academy of Sciences in Bratislava (IMS SAS). In this tomograph, a static magnetic field is formed between two parallel permanent magnets [36]. Parallel to the magnets, there are two internal planar coils of the gradient system used to select slices in three dimensions. In the magnetic field, a tested object is placed together with an external radio frequency receiving/transmitting coil. The whole MRI scanning equipment is placed in a metal cage to suppress high-frequency interference. The cage is made of a 2-mm thick steel plate with 2.5-mm diameter holes spaced periodically in a 5-mm grid to eliminate the propagation of the electromagnetic field to the surrounding space of the control room.

For the phonation signal recording inside the shielding metal cage of this device, the pick-up condenser microphone Mic1 (Soundking EC 010 W) was placed on the stand at the distance  $D_X = 60$  cm from the central point of the scanning area to inhibit any interaction with the MRI’s working magnetic field. Its height was 75 cm from the floor (in the middle

between both gradient coils) and its orientation was 150 degrees from the left corner near the temperature stabilizer. The Behringer XENYX Q802 USB mixer and a laptop used for recording were located outside the MRI shielding metal cage—see an arrangement photo in Figure 7. Another microphone Mic2 (Behringer TM1) was connected to the second channel of the XENYX Q802 mixer for the phonation signal pick-up in the recording phases  $MF_1$  and  $MF_4$  with the tested person sitting at the desk in the MRI equipment control room. Both professional studio microphones are based on the electrostatic transducer with a 1-inch diaphragm and they have very similar cardioid directional patterns as well as frequency responses at 1, 2, 4, 8, and 16 kHz.



**Figure 7.** An arrangement of the phonation and PPG signal recording in the MRI Opera (for measurement phases  $MF_2$  and  $MF_3$ ): (1) pick-up microphone Mic1, (2) noise SPL meter, (3) recording devices outside the shielding cage, (4) electronic part of the wearable PPG sensor, (5) reflection optical pulse sensor on the forefinger of the left hand, (6) door of the shielding cage.

Between the measurement phases  $MF_2$  and  $MF_3$ , the scan sequence 3D-CE (with  $TE = 30$  ms,  $TR = 40$  ms; 3D phases = 8) was run with a total time duration of about 8 min. This type of our most used MR sequence produces a noise with a sound pressure level (SPL) of about 72 dB (C); the background SPL inside the metal shielding cage is produced mainly by the temperature stabilizer and reaches about 55 dB (C) [29]. In this case, the physiological effect of the noise and vibration on the human organism and auditory system is small but still measurable and detectable [30]. During the phonation signal pick-up in the  $MF_1$  and  $MF_4$  measurement phases, the control room background level was up to 45 dB (C). In all cases, the SPL values were measured by the sound level meter Lafayette DT 8820 mounted on the holder at the same height from the floor as the recording microphone (75 cm). For purpose of this study, we are not interested in MR images that are automatically generated by the MRI control system after finishing the currently running scan sequence [36]. To prevent their creation and storage, it is possible to manually interrupt passing of the running scan sequence from the operator console. This approach was practically applied in all our experiments, so no MR images of the tested persons were collected or stored.

The phonation/sound signal was analyzed by a pitch-asynchronous method with a frame length of 24 ms and a half-frame overlap. For calculation of spectral properties, the number of fast Fourier transform (FFT) points was  $N_{FFT} = 1024$ ; for estimation of the formant frequencies and their bandwidths, the complex roots of the 18th order LPC

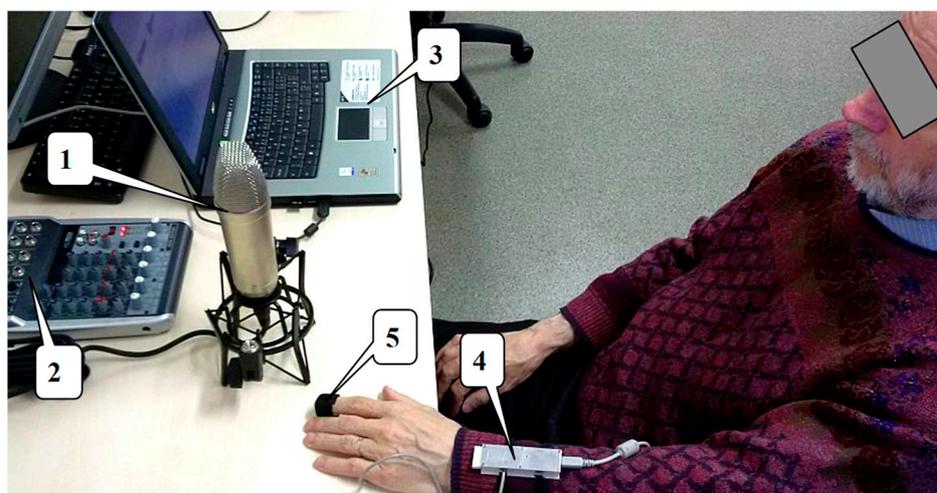
polynomial were used. In contrast with our first-step work [26] and with the aim to obtain results with higher precision, computation of the full covariance matrix [19] and 512 mixtures were finally applied. The length of the input feature vector for GMM creation, training, and classification was set experimentally to  $N_{\text{FEAT}} = 32$ , and  $N_{\text{ITER}} = 1500$  iterations were used. The phonation signal processing as well as implementation of basic functions for the GMM classifier was currently realized in the Matlab environment (ver. 2019a).

### 3.2.2. PPG Signal Recording

Generally, two principles of optical sensors (transmission or reflection) can be utilized in the PPG signal measurement. Both types consist of two basic elements: a transmitter (light source—LS) and a receiver (photo detector—PD). In the transmission mode, the LSs and PDs are placed on the opposite sides of the measured human tissue. In the reflection PPG sensor, the PDs and LSs are placed on the same skin surface. In this research, the optical sensors working on the reflection principle were used and the PPG signals were picked up from fingers [37]. For practical PPG signal recording, a previously developed wearable PPG sensor, PPG-PS1, was used. This also operates in a weak magnetic field with radiofrequency disturbance (in the scanning area of the running MRI device during patient examination) [38]. This PPG sensor realization is fully shielded, assembled only from non-ferromagnetic components, and based on the reflection optical pulse PPG sensor (Pulse Sensor Amped—Adafruit 1093 [39]). For data transmission to the control device, the wireless communication based on Bluetooth standard is utilized. Due to the 10-bit A/D converter implemented in the microcontroller of the whole PPG sensor, the absolute unipolar PPG signal representation lies in the range from 0 to 1024 ( $A_{\text{NR}} = 1024$ ). This wearable sensor enables real-time PPG wave sensing and recording for the sampling frequencies from 100 to 500 Hz.

The typical PPG cycle frequency corresponding to the HR of healthy adults is in the range 1 to 1.7 Hz (from 60 to 106  $\text{min}^{-1}$ ) [37], so the  $f_S$  about 150 Hz is sufficient to fulfil the Shannon sampling theorem. In addition, the commercial wearable PPG sensors use typical sampling frequencies between 50 and 100 Hz. Using different  $f_S$  from the investigated range does not change the subsequently detected pulse period and the finally determined heart rate; only the precision of the systolic and diastolic peaks decreases in the case of lower  $f_S$ . For the purpose of this study the precise shape of peaks is less relevant, only the detected  $T_{\text{HP}}$  and  $W_{\text{SP}}$  parameters are necessary for HR and ORI calculation. As we statistically analyze the obtained HR and ORI values for final comparison in the fusion block, the statistical stability and credibility is most important for us. From the previously performed analysis, it follows that a decrease in the number of detected HR periods as a consequence of higher used  $f_S$  brings an incorrectness to the results of the statistical analysis due to too small a number of the processed values—the PPG signal is sensed in real-time by the data block samples from the internal memory of a wearable PPG sensor with sizes from 1 to 25 k [38]. This is the main reason why we use the  $f_S = 125$  Hz for sensing of the PPG signal in our experiments.

The optical part of the PPG sensor is fixed on a forefinger of the left hand by an elastic ribbon. The PPG signal pick-up is begun just before the start of the human voice phonation and the PPG sensing is finished immediately after the end of the phonation recorded by the microphone Mic2—see an arrangement photo in Figure 8 obtained during the  $MF_1$  measurement phase.



**Figure 8.** An arrangement of the phonation recording and PPG signal measurement in the laboratory conditions (for  $MF_1$  and  $MF_4$  phases): (1) a pick-up microphone Mic2, (2) the analogue mixer XENYX Q802, (3) a control and recording device, (4) body of the wearable PPG sensor with BT data transfer, (5) a reflection optical pulse sensor mounted on the forefinger of the left hand.

### 3.3. Used Databases for GMM-Based Stress Detection and Evaluation in the Phonation Signal

Three different audio corpora were used to create and train the GMM models for the classes of the normal and stressed speech. Our first corpus (further called  $DB_1$ ) was taken from the International Affective Digitized Sounds (IADS-2) [40] comprising 167 sound and noise records with duration of 6 s. The database is standardized and rated using Pleasure and Arousal (P-A) parameters in the range of  $\langle 1\sim 9 \rangle$ . The second created corpus ( $DB_2$ ) was extracted from the emotional speech database Emo-DB [41]. It contains sentences of the same content with six acted emotions and a neutral state by five male and five female German speakers with time durations from 1.5 s to 8.5 s. We used sentences in a neutral state and a surprise for the  $C_{1N}$  class; a fear, an anger, and a disgust for the  $C_{2S}$  stress class, and a sadness with a joy for the  $C_{3O}$  class—separately for both genders (234 + 306 in total). The third audio corpus ( $DB_3$ ) was extracted from the audiovisual database MSP-IMPROV [42] recorded in English. This database has sentences also evaluated in the P-A scale but in the range from 1 to 5. For compatibility with the  $DB_1$ , all the applied speech records were resampled at 16 kHz and the mean P-A values were recalculated to fit the range from 1 to 9 of the  $DB_1$ . We have used only declarative sentences with acted speech in a neutral state by three males and three females, in total  $2 \times 250$  sentences (separately for male and female voices) with duration from 0.5 to 6.5 s.

Applied P-A ranges and mean values for basic emotions are shown in Table 4. For the class  $C_{1N}$ , the records with  $P = \{3.5\sim 5.5\}$ ,  $A = \{4\sim 6\}$  corresponding to the neutral state and joy were finally used. The sound/noise records with  $P \leq 3$ ,  $A \geq 6$  corresponding to the anger, disgust, and fear emotions were used for the stressed class  $C_{2S}$ . The class  $C_{3O}$  represented negative emotions of sadness (with both P and A parameters low) and a positive emotion joy (both P and A parameters high)—compare the 4th and the 7th line in Table 4. These three described audio databases were used because their records are freely accessible without any fee or other restrictions.

**Table 4.** Ranges and mean values of P-A parameters related to discrete basic emotions.

Emotion	Pleasure Range/Mean	Arousal Range/Mean	Emotion Location in the P-A Space
Anger <sup>2</sup>	(1.0~3.0)/2.40	(6.0~8.0)/6.04	
Disgust <sup>2</sup>	(3.0~4.5)/3.50	(4.5~6.5)/5.73	
Fear <sup>2</sup>	(1.5~3.5)/2.97	(4.0~6.5)/5.72	
Sadness <sup>3</sup>	(2.0~3.5)/3.04	(3.0~5.0)/3.88	
Neutral <sup>1</sup>	(4.0~6.0)/5.14	(2.5~4.5)/3.45	
Surprise <sup>1</sup>	(4.5~7.0)/5.67	(4.5~7.0)/4.81	
Joy <sup>3</sup>	(7.0~9.0)/8.44	(4.5~8.0)/5.88	

<sup>1</sup> used for normal speech class C<sub>1N</sub>; <sup>2</sup> used for C<sub>2S</sub> class; <sup>3</sup> used for C<sub>3O</sub> class.

#### 4. Discussion of Obtained Results

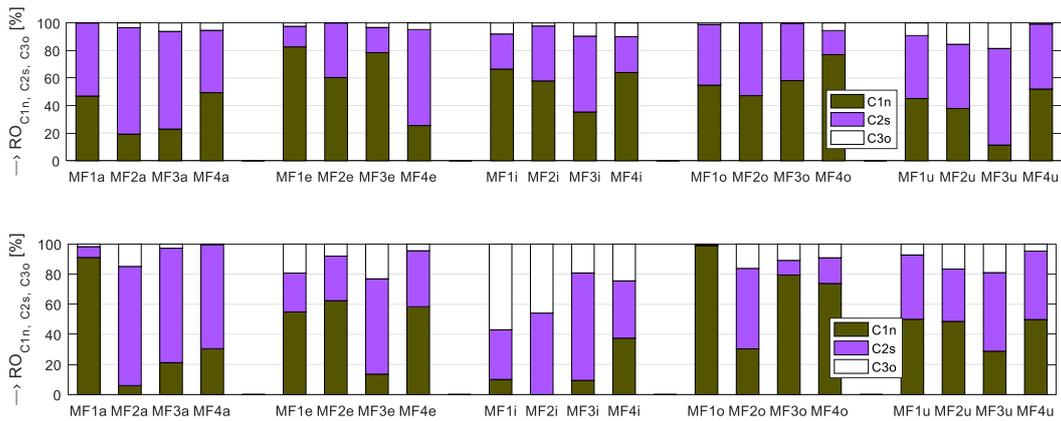
Obtained results are structured by the applied stress evaluation methods: at first, using the GMM-based classification parameters  $SP_{GMM}$  from the phonation signals, next the statistical parameters  $SP_{PPG}$  determined from the PPG signals (both for  $MF_{1-4}$  measuring phases), and finally the stress evaluation rates for  $MF_2$  to  $MF_4$  phases are calculated by the fusion of the  $SP_{GMM}$  and  $SP_{PPG}$  parameters. Summary results are next divided by gender of a tested person—values for groups of males, females, and for all participating persons are subsequently visualized and compared.

Within the GMM classification part, an auxiliary analysis was also performed to evaluate an influence of the database used for GMMs creation and training. Comparison of  $L_{STRESS}$ ,  $L_{NORMAL}$  and  $\Delta L_{S-N}$  values in Table 5 shows that all three tested databases are usable for this purpose. As shown in the last column, the greatest differences between  $L_{STRESS}$  and  $L_{NORMAL}$  values are obtained when the Emo-DB speech database was used. Therefore, in further analysis, the GMMs were created and trained with the help of the database DB<sub>2</sub>. Next, we analyzed the percentage distribution values of the output classes C<sub>1N</sub>, C<sub>2S</sub>, and C<sub>3O</sub> per each vowel of the phonation signal. The representative results from this analysis performed on the recorded vowels are shown in detail in Figure 9, where a non-uniform class distribution can be seen for vowels recorded in the measuring phases  $MF_{1-4}$ . However, the summary comparison in Figure 10 demonstrates the expected trends of  $L_{STRESS}$  and  $L_{NORMAL}$  values being in correlation with mean  $RO_{C_{1N}, C_{2S}, C_{3O}}$  values calculated for all five vowels together— $RO_{C_{2S}}$  values are increased in  $MF_{2,3}$  phases in comparison to  $MF_{1,4}$  phases. This trend is accompanied with parallel decrease of  $RO_{C_{1N}}$  values in  $MF_{2,3}$  phases and increase in  $MF_{1,4}$  phases.

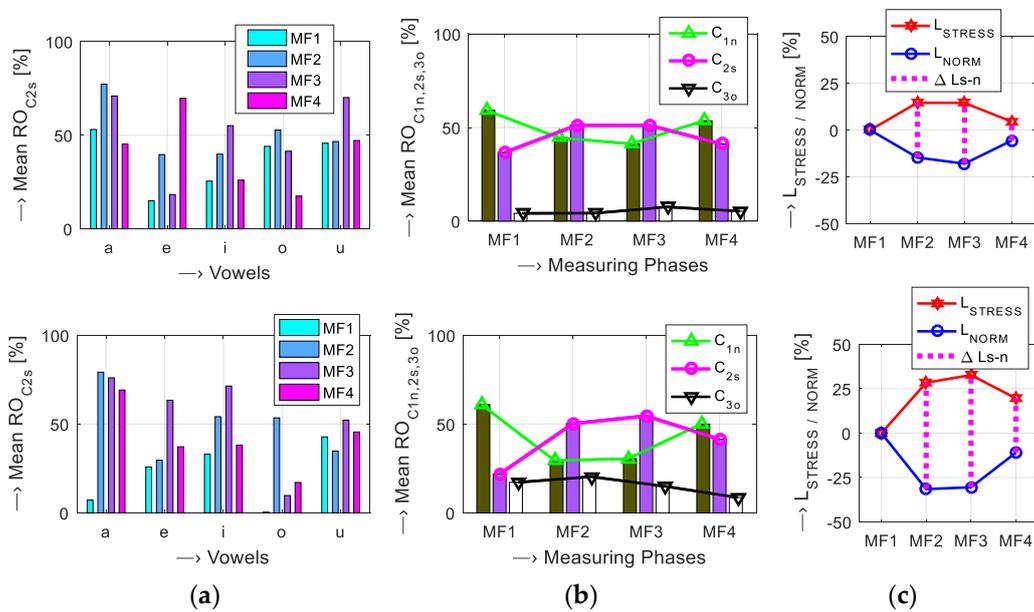
**Table 5.** Influence of different databases used for GMM creation and training on stress evaluation—male speaker M1,  $N_{MIX} = 512$ , full covariance matrix, summarized for all five vowels.

Database Type	$L_{STRESS}$ [%] <sup>1</sup> ( $MF_{2,3,4}$ )	$L_{NORMAL}$ [%] <sup>1</sup> ( $MF_{2,3,4}$ )	$\Delta L_{S-N}$ [%] <sup>1</sup> ( $MF_{2,3,4}$ )
DB <sub>1</sub> (sounds-IADS-2)	8.09, 11.2, -2.09	-29.6, -38.9, 4.41	37.7, 50.1, -6.51
DB <sub>2</sub> (speech-Emo-DB)	56.6, 75.2, -9.76	-2.88, 13.9, 14.60	53.7, 89.1, 4.87
DB <sub>3</sub> (speech-MSP-IMPROV)	15.4, 20.2, 2.23	-30.3, -36.6, 0.08	45.7, 56.7, 2.15

<sup>1</sup> for  $MF_1$  are  $L_{STRESS}/L_{NORM}/\Delta L_{S-N} = 0$  in all cases.



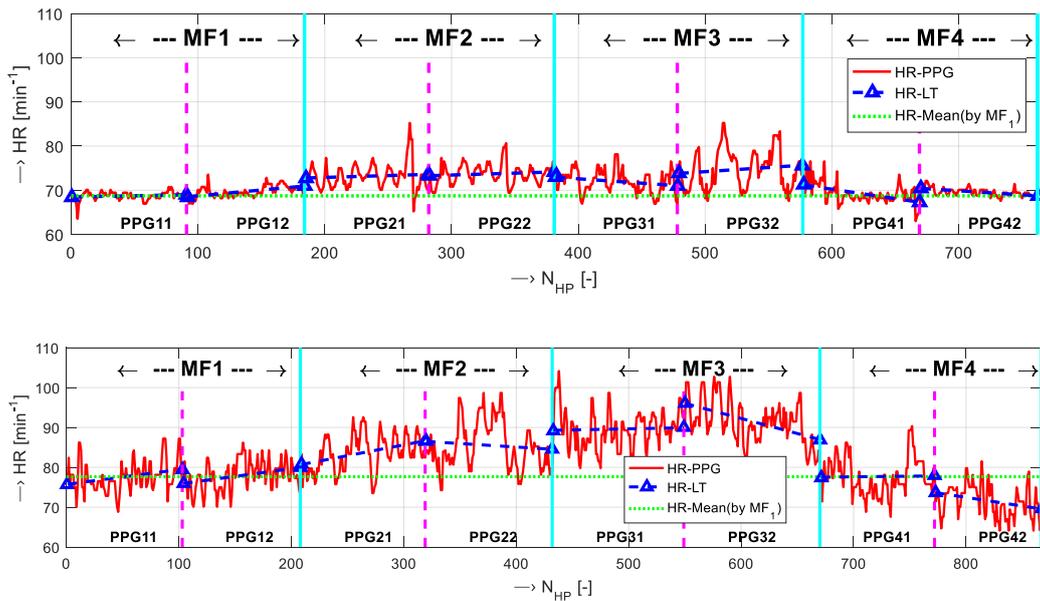
**Figure 9.** Visualization of percentage distribution values of output classes  $C_{1N}$ ,  $C_{2S}$ , and  $C_{3O}$  per each vowel of the phonation signal recording within all four measuring phases ( $MF_{1-4}$ ); from the speech signal by male M2 (**upper** graph) and female F2 (**lower** graph),  $N_{MIX} = 512$ , full covariance matrix.



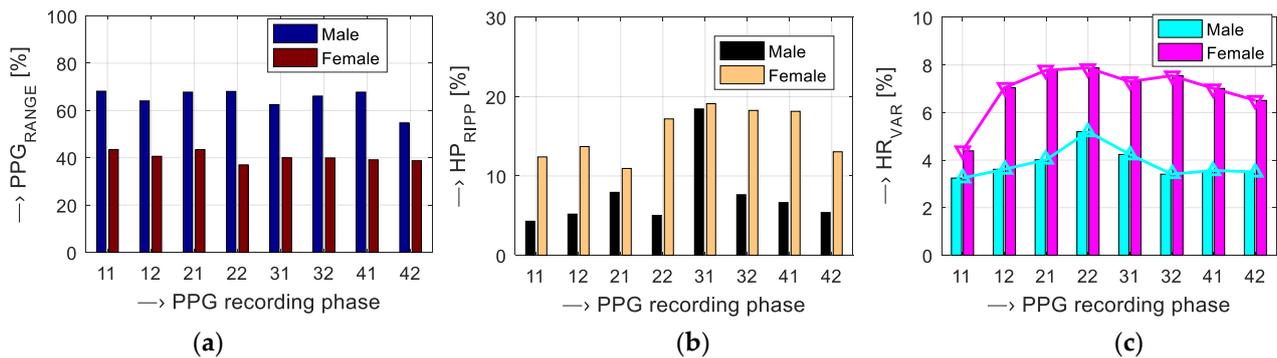
**Figure 10.** Summary GMM-based comparison parameters for male M2 (**upper** graph) and female F2 (**lower** graph): (a) visualization of mean  $RO_{C_{2S}}$  values per each vowel phonated in the measuring phases  $MF_{1-4}$ , (b) bar-graphs of mean  $RO_{C_{1N}}$ ,  $C_{2S}$ ,  $C_{3O}$  values, (c) visualization of  $L_{STRESS}$ ,  $L_{NORMAL}$ , and  $\Delta L_{S-N}$  values calculated relative to the baseline  $MF_1$ ;  $N_{MIX} = 512$ , full covariance matrix.

The results obtained by the second evaluation approach confirm our assumption that the stress level evoked by scanning in the tested MRI device is identifiable and measurable using HR values determined from the PPG signal. From the detailed analysis of filtered HR values concatenated for the recording phases  $PPG_{11-42}$  together with their LT parameter follows that, in the measuring phases  $MF_2$  and  $MF_{23}$ , there is a pronounced increase in the mean HR with a positive LT, while the last phase  $MF_4$  has typically lower mean HR and negative LT. This increase of mean HR values is accompanied by higher variation of discrete HR values. In the first measuring phase  $MF_1$ , lower HR with positive LT is observed. In addition, there are visible differences in HR values determined from the recording phases  $PPG_{11}$  and  $PPG_{12}$ . This was probably due to the load effect of speech (vowels) production by a tested person manifested by a small increase of the mean HR determined from PPG signals recorded after phonation. Figure 11 shows concatenated sequences of HR values for two distinct cases that occurred in a male person M2 (upper graph with minimum changes of HR and LT values) and in a female person F3 (lower

graph with maximum increase of HR and LT values in  $MF_{2,3}$  phases). In summary, the mentioned increase of HR as well as its variance is more pronounced in females. It is also documented by a graphical comparison in Figure 12. During the stress phase  $MF_3$ , the maximum mean HR =  $92 \text{ min}^{-1}$  occurred in the case of the female F1, while during the final phase  $MF_4$  the minimum mean HR =  $61 \text{ min}^{-1}$  was achieved for the male M4, and these mean HR values lie within the HR range for healthy adults [37]. On the other hand, the absolute maxima can be locally higher as documented by HR values in PPG<sub>31,32</sub> phases for the female F3 showed in Figure 11.



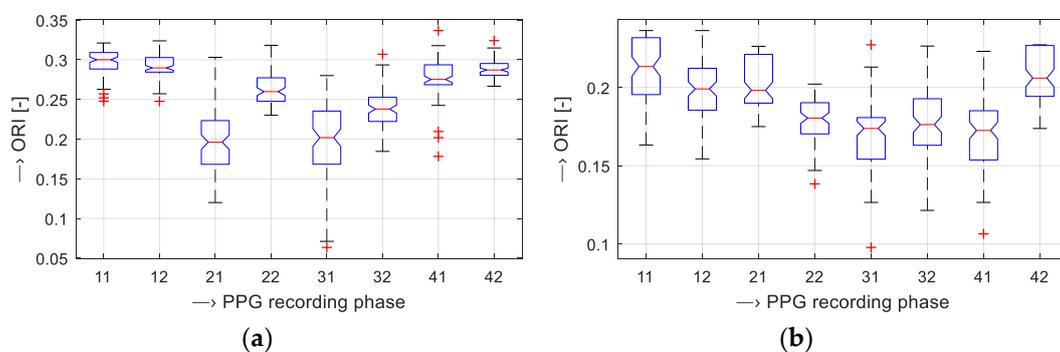
**Figure 11.** Filtered HR values determined from the recorded PPG signals (HR-PPG) concatenated for all measuring phases together with their linear trend (HR-LT) and the mean HR level in the  $MF_1$  phase: for the male M2 (upper graph), for the female F3 (lower graph).



**Figure 12.** Partial results of PPG wave parameters in phases PPG<sub>11-42</sub> for the male M2 and female F3 persons: bar-graphs of (a)  $PPG_{RANGE}$ , (b)  $HP_{RIPP}$  parameters, (c) comparison of  $HR_{VAR}$  values.

Contrary to our expectations, the observed changes in  $PPG_{RANGE}$  and  $HP_{RIPP}$  parameters do not follow the trends presented in Table 1, and they do not seem to be useful for detection of the stress level. The LT (or  $HR\phi_{REL}$ ) and  $HR_{VAR}$  parameters partially exhibit the expected increase in the  $MF_{2,3}$  phases, but these changes are not significant and stable. This effect is similar for male as well as female tested persons, as demonstrated by the graphs in Figure 12. In the case of the  $ORI$  parameter, its changes are not consistent, probably as they are more individual, or because the chosen time duration of the measuring phases as well as the length of working time delays were not set properly. As follows

from the definition of ORI in (10) the resulting value depends on the width of the systolic pulse and the heart pulse period. These two parameters can be affected in synergy or in antagonism. In consequence of this state, we cannot obtain any credible statistical result for precise comparison—see box-plot graphs of basic statistical parameters of ORI values for one male and one female person in Figure 13. Therefore, in this stage of our research, we can only state that in one case of a male person the ORI values start to decrease in the  $MF_3$  phase, and this trend continues also in the final  $MF_4$  phase, while the changes of HR values fulfill our experimental premise—in  $MF_3$  they are higher, in  $MF_4$  they substantially decrease. Next, for one female person during measurements inside the MRI device, the HR and ORI changed in the opposite manner—this was probably caused by her adaptation to the changed position (from standing to lying) and, at the same time, by being rather nervous in a foreign environment inside the shielding cage of the MRI scanning area perceived as somewhat unfriendly. In other cases, some effect of stress on the ORI parameter could also be observed but it was not concentrated in the monitored phases  $MF_{2,3}$ .



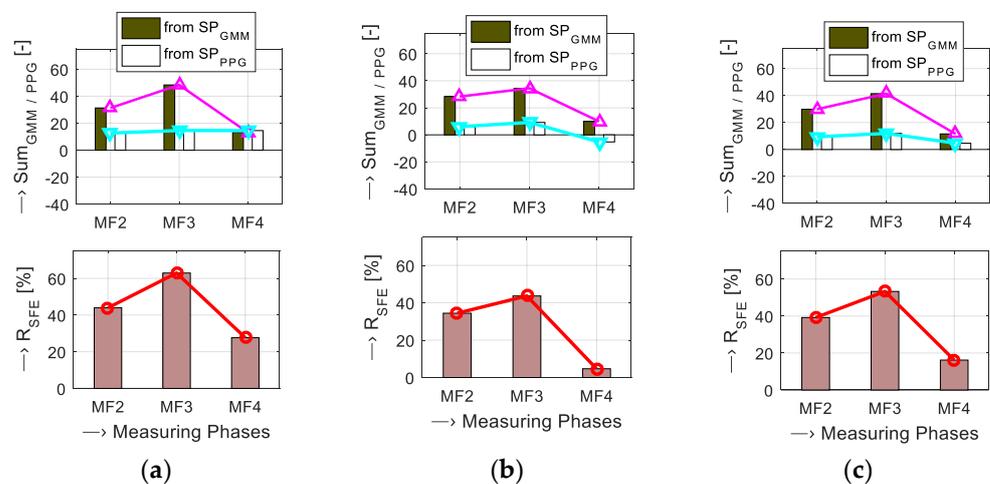
**Figure 13.** Comparison of boxplots of basic statistical parameters of ORI values in the recording phases  $PPG_{11-42}$  for: (a) the male M2, (b) the female F3 tested person.

The process of fusion—calculation of the final stress evaluation rate—is described by a numerical example in Table 6. This shows the entered input parameters from the GMM and PPG stress evaluation parts together with the applied importance weights. In the right part of this table, there are the corresponding partial sums for  $MF_{2,3,4}$  phases together with the final  $R_{SFE}$  values. Application of the  $SP_{PPG}$  parameters brings greater difference in the final  $R_{SFE}$  values between  $MF_{2-3-4}$  phases by 26% (for  $\Delta MF_{2-3}$ ) and 45% (for  $\Delta MF_{3-4}$ ) in comparison with using  $SP_{GMM}$  alone ( $\Delta MF_{2-3} = 10\%$ ,  $\Delta MF_{3-4} = 43\%$ ). Visualization of partial and summary results obtained during the fusion process depending on gender (male, female, and all persons) is presented in Figure 14. These graphical results correspond to numerical values shown in Table 6, i.e., the partial sums calculated from  $SP_{PPG}$  parameters are smaller in comparison to the sums from  $SP_{GMM}$  ones. This trend can be seen especially for female tested persons in a graph in Figure 14b. The bar-graph of the final  $R_{SFE}$  values obtained for all tested persons in Figure 14c practically confirms our working hypothesis about the negative stress effect after examination by the running scan sequence of the MRI device—the  $R_{SFE}$  value for the  $MF_3$  phase is the highest. However, merely lying in the non-scanning MRI device can evoke a non-negligible stress as documented by about 40% increase of the  $R_{SFE}$  value in the  $MF_2$  phase in comparison with the zero-normalized  $R_{SFE}$  in the starting phase  $MF_1$ . Our working presupposition about the human physiological parameters returning to the baseline in the last measuring phase  $MF_4$  was not completely confirmed. In most cases, the  $R_{SFE}$  value was greater than zero in this phase ( $SP_{GMM}$  and  $SP_{PPG}$  stress parameters determined in  $MF_4$  were higher than those in  $MF_1$ ), but there was also a situation with stress parameters lower than in the initial phase, yielding a negative value of  $R_{SFE}$  in  $MF_4$ . The return to the person's initial state could be facilitated by the increase of the working time delay  $WTD_3$ —a longer pause before the last measuring phase. Nevertheless, it was practically unacceptable to the experimenter as well as to the

examined testing persons with respect to a relative long duration of about 50 min for the whole measurement experiment.

**Table 6.** Example of calculation process of the final stress evaluation rate for the male M1.

Parameter Type	$SP_{GMM/PPG} (MF_{2,3,4})$	Partial Sum ( $MF_{2,3,4}$ )	Final $R_{SFE} (MF_{2,3,4})$
$SP_{GMM1}$	8.1, 11.2, -2.1	30.3, 40.4, -3.4	} 27.1, 54.2, 8.8
$SP_{GMM2}$	-29.6, -38.9, 1.4		
$SP_{GMM3}$	37.7, 50.1 -4.5		
$SP_{PPG1}$	7.3, -1.8, -4.8		
$SP_{PPG2}$	-1.9, 1.8, -5.7		
$SP_{PPG3}$	4.4, -14.7, -5.5	-3.2, 13.8, 12.2	
$SP_{PPG4}$	1.2, 0.4, 0.1		
$SP_{PPG5}$	6.8, -20.5, -18.9		



**Figure 14.** Visualization of final  $R_{SFE}$  results obtained during the fusion process depending on the gender of testing persons: (a) partial results for male persons, (b) partial results for female persons, (c) summary results for all joined persons.

**5. Conclusions**

The current article is an extension of our previous work [26], where experiments with sensing and analyzing of a PPG signal have been described. The main limitation of this study lies in the fact that only a small group of tested persons participated in the measurement of phonation and PPG signals. This was caused mainly by a bad COVID-19 situation in our country at the time of the recording experiments. Since the tested persons could not put on any mask during the phonation signal recording, only healthy vaccinated people participated (authors themselves and their colleagues from IMS SAS) for collecting the speech and PPG signal databases. The second limitation lies in the fact that the testing open-air MRI device is the standard equipment for use in medical practice, but our institute is not certificated for work with real patients, so it can be used for non-clinical and non-medical research only.

Nevertheless, the obtained experimental results confirm our hypothesis about the negative influence of the vibration and noise during MRI execution expressed by increased an stress level in the recorded phonation signal as well as increased heart rate and its variation determined from the PPG signal. In addition, the performed experiments confirm our working assumption that both types of analysis are usable for this task—the final stress level values obtained by a fusion of bimodal results are more differentiable. On the other hand, the results obtained in this way cannot be fully generalized, only special and typical cases that occurred during our experiments are described and discussed. Due to processing of a relatively small number of phonation and PPG signal records, it was very difficult to

obtain results with good statistical credibility—so only basic statistical parameters were calculated and compared.

In future, we plan to perform a detailed analysis of speech features applied for GMM-based classification to obtain greater differences in the detected normal and stress classes. We would also like to test this stress detection approach with the help of well-known databases consisting of stressed speech either simulated or recorded under real conditions, the speech under simulated and actual stress (SUSAS) database in English [43], the experimental speech corpus ExamStress in Czech [44], etc. which are not free or have a limited access. In the PPG signal sensing, processing, and analysis we will try to find other parameters for better description of changes in a human cardiovascular system caused by a stress factor. We also plan to test another type of PPG sensor working on the transmission principle (as an oximeter device) enabling measurement and recording of blood oxygen saturation, heart rate, and perfusion index values to the control device via BT connection. In this case, the realization requirement to operate in a low magnetic field must be fulfilled—the PPG sensor must consist of non-ferromagnetic components and all parts must be shielded due to strong RF disturbance in the scanning area of the MRI device.

**Author Contributions:** Conceptualization and methodology, J.P. and A.P.; data collection and processing, J.P.; writing—original draft preparation, J.P. and A.P.; writing—review and editing, A.P.; project administration, I.F.; funding acquisition, I.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Slovak Scientific Grant Agency project VEGA2/0003/20 and the Slovak Research and Development Agency project APVV-19-0531.

**Informed Consent Statement:** Ethical review and approval were waived for this study, due to testing authors themselves and colleagues from IMS SAS. No personal data were saved, only PPG waves and phonation signals used in this research.

**Acknowledgments:** We would like to thank all our colleagues and other volunteers who participated in the phonation and PPG signal recording experiments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Steckner, M.C. A review of MRI acoustic noise and its potential impact on patient and worker health. *eMagRes* **2020**, *9*, 21–38.
2. Mainka, A.; Platzek, I.; Mattheus, W.; Fleischer, M.; Müller, A.-S.; Mürbe, D. Three-dimensional vocal tract morphology based on multiple magnetic resonance images is highly reproducible during sustained phonation. *J. Voice* **2017**, *31*, 504.e11–504.e20. [[CrossRef](#)]
3. Hansen, J.H.L.; Patil, S. Speech under stress: Analysis, modeling and recognition. In *Speaker Classification I, Lecture Notes in Artificial Intelligence*; Müller, C., Ed.; Springer: Berlin, Germany, 2007; Volume 4343, pp. 108–137.
4. Schickhofer, L.; Malinen, J.; Mihaescu, M. Compressible flow simulations of voiced speech using rigid vocal tract geometries acquired by MRI. *J. Acoust. Soc. Am.* **2019**, *145*, 2049–2061. [[CrossRef](#)] [[PubMed](#)]
5. Pitha, J.; Pithova, P.; Roztocil, K.; Urbaniec, K. Oliva-Roztocil Index, Specific Parameter of Vascular Damage in Women Suffering from Diabetes Mellitus. *Atherosclerosis* **2017**, *263*, e275. [[CrossRef](#)]
6. Celka, P.; Charlton, P.H.; Farukh, B.; Chowienczyk, P.; Alastruey, J. Influence of mental stress on the pulse wave features of photoplethysmograms. *Healthc. Technol. Lett.* **2020**, *7*, 7–12. [[CrossRef](#)]
7. Rundo, F.; Conoci, S.; Ortis, A.; Battiato, S. An advanced bio-inspired photoplethysmography (PPG) and ECG pattern recognition system for medical assessment. *Sensors* **2018**, *18*, 405. [[CrossRef](#)] [[PubMed](#)]
8. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1–R39. [[CrossRef](#)] [[PubMed](#)]
9. Blazek, V.; Venema, B.; Leonhardt, S.; Blazek, P. Customized optoelectronic in-ear sensor approaches for unobtrusive continuous monitoring of cardiorespiratory vital signs. *Int. J. Ind. Eng. Manag.* **2018**, *9*, 197–203. [[CrossRef](#)]
10. Charlton, P.H.; Marozas, V. Wearable photoplethysmography devices. In *Photoplethysmography: Technology, Signal Analysis and Applications*, 1st ed.; Kyriacou, P.A., Allen, J., Eds.; Elsevier: London, UK, 2022; pp. 401–438.
11. Harmon-Jones, E.; Harmon-Jones, C.; Summerell, E. On the importance of both dimensional and discrete models of emotion. *Behav. Sci.* **2017**, *7*, 66. [[CrossRef](#)]
12. Nicolaou, M.A.; Gunes, H.; Pantic, M. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Trans. Affect. Comput.* **2011**, *2*, 92–105. [[CrossRef](#)]

13. Nwe, T.L.; Foo, S.W.; De Silva, L.C. Speech emotion recognition using hidden Markov models. *Speech Commun.* **2003**, *41*, 603–623. [CrossRef]
14. Campbell, W.M.; Campbell, J.P.; Reynolds, D.A.; Singer, E.; Torres-Carrasquillo, P.A. Support vector machines for speaker and language recognition. *Comput. Speech Lang.* **2006**, *20*, 210–229. [CrossRef]
15. Chandaka, S.; Chatterjee, A.; Munshi, S. Support vector machines employing cross-correlation for emotional speech recognition. *Measurement* **2009**, *42*, 611–618. [CrossRef]
16. Nicholson, J.; Takahashi, K.; Nakatsu, R. Emotion recognition in speech using neural networks. *Neural Comput. Appl.* **2000**, *9*, 290–296. [CrossRef]
17. Jahangir, R.; Teh, Y.W.; Hanif, F.; Mujtaba, G. Deep learning approaches for speech emotion recognition: State of the art and research challenges. *Multimed. Tools Appl.* **2021**, *80*, 23745–23812. [CrossRef]
18. Andrade, G.; Rodrigues, M.; Novais, P. A Survey on the Semi Supervised Learning Paradigm in the Context of Speech Emotion Recognition. *Lect. Notes Netw. Syst.* **2022**, *295*, 771–792.
19. Reynolds, D.A.; Rose, R.C. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans. Speech Audio Process.* **1995**, *3*, 72–83. [CrossRef]
20. He, L.; Lech, M.; Maddage, N.C.; Allen, N.B. Study of empirical mode decomposition and spectral analysis for stress and emotion classification in natural speech. *Biomed. Signal Process.* **2011**, *6*, 139–146. [CrossRef]
21. Zhang, G. Quality evaluation of English pronunciation based on artificial emotion recognition and Gaussian mixture model. *J. Intell. Fuzzy Syst.* **2021**, *40*, 7085–7095.
22. Yucesoy, E.; Nabyev, V. A new approach with score-level fusion for the classification of the speaker age and gender. *Comput. Electr. Eng.* **2016**, *53*, 29–39. [CrossRef]
23. Asbai, N.; Amrouche, A. A novel scores fusion approach applied on speaker verification under noisy environments. *Int. J. Speech Technol.* **2017**, *20*, 417–429. [CrossRef]
24. Al Dujaili, M.J.; Ebrahimi-Moghadam, A.; Fatlawi, A. Speech emotion recognition based on SVM and KNN classifications fusion. *Int. J. Electr. Comput. Eng.* **2021**, *11*, 1259–1264. [CrossRef]
25. Araño, K.A.; Orsenigo, C.; Soto, M.; Vercellis, C. Multimodal sentiment and emotion recognition in hyperbolic space. *Expert Syst. Appl.* **2021**, *184*, 115507. [CrossRef]
26. Přibil, J.; Přibilová, A.; Frollo, I. Experiment with stress detection in phonation signal recorded in open-air MRI device. In Proceedings of the 44th International Conference on Telecommunications and Signal Processing, TSP 2021, Virtual, 26–28 July 2021; pp. 38–41.
27. Prince, D.L.; De Wilde, J.P.; Papadaki, A.M.; Curran, J.S.; Kitney, R.I. Investigation of acoustic noise on 15 MRI scanners from 0.2 T to 3 T. *J. Magn. Reson. Imaging* **2001**, *13*, 288–293. [CrossRef]
28. Moelker, A.; Wielopolski, P.A.; Pattynama, P.M.T. Relationship between magnetic field strength and magnetic-resonance-related acoustic noise levels. *Magn. Reson. Mater. Phys. Biol. Med.* **2003**, *16*, 52–55. [CrossRef]
29. Přibil, J.; Přibilová, A.; Frollo, I. Analysis of the influence of different settings of scan sequence parameters on vibration and voice generated in the open-air MRI scanning area. *Sensors* **2019**, *19*, 4198. [CrossRef] [PubMed]
30. Přibil, J.; Přibilová, A.; Frollo, I. First-step PPG signal analysis for evaluation of stress induced during scanning in the open-air MRI device. *Sensors* **2020**, *20*, 3532. [CrossRef] [PubMed]
31. Sigmund, M. Influence of psychological stress on formant structure of vowels. *Elektron. Elektrotech* **2012**, *18*, 45–48. [CrossRef]
32. Tomba, K.; Dumoulin, J.; Mugellini, E.; Khaled, O.A.; Hawila, S. Stress detection through speech analysis. In Proceedings of the 15th International Joint Conference on e-Business and Telecommunications, ICETE 2018, Porto, Portugal, 26–28 July 2018; pp. 394–398.
33. Shah, N.H. *Numerical Methods with C++ Programming*; Prentice-Hall of India Learning Private Limited: New Delhi, India, 2009; p. 251.
34. Korpas, D.; Halek, J.; Dolezal, L. Parameters Describing the Pulse Wave. *Physiol. Res.* **2009**, *58*, 473–479. [CrossRef]
35. Oliva, I.; Roztocil, K. Toe Pulse Wave Analysis in Obliterating Atherosclerosis. *Angiology* **1983**, *34*, 610–619. [CrossRef] [PubMed]
36. E-Scan Opera. *Image Quality and Sequences Manual*; 830023522 Rev. A; Esaote S.p.A.: Genoa, Italy, 2008.
37. Jarchi, D.; Salvi, D.; Tarassenko, L.; Clifton, D.A. Validation of instantaneous respiratory rate using reflectance PPG from different body positions. *Sensors* **2018**, *18*, 3705. [CrossRef] [PubMed]
38. Přibil, J.; Přibilová, A.; Frollo, I. Wearable PPG Sensor with Bluetooth Data Transmission for Continual Measurement in Low Magnetic Field Environment. In Proceedings of the 26th International Conference Applied Electronics 2021, Pilsen, Czech Republic, 7–8 September 2021; pp. 137–140.
39. Pulse Sensor Amped Product (Adafruit 1093): World Famous Electronics LLC. Ecommerce Getting Starter Guide. Available online: <https://pulsesensor.com/pages/code-and-guide> (accessed on 16 July 2020).
40. Bradley, M.M.; Lang, P.J. *The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual*; Technical Report B-3; University of Florida: Gainesville, FL, USA, 2007.
41. Burkhardt, F.; Paeschke, A.; Rolfes, M.; Sendlmeier, W.; Weiss, B.A. Database of German emotional speech. In Proceedings of the Interspeech 2005, Lisbon, Portugal, 4–8 September 2005; pp. 1517–1520.
42. Busso, C.; Parthasarathy, S.; Burmania, A.; AbdelWahab, M.; Sadoughi, N.; Provost, E.M. MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception. *IEEE Trans. Affect. Comput.* **2017**, *8*, 67–80. [CrossRef]

43. Hansen, J.H.; Bou-Ghazale, S.E.; Sarikaya, R.; Pellom, B. Getting started with SUSAS: A speech under simulated and actual stress database. In Proceedings of the Eurospeech 1997, Rhodes, Greece, 22–25 September 1997; pp. 1743–1746.
44. Sigmund, M. Introducing the database ExamStress for speech under stress. In Proceedings of the NORSIG 2006, Reykjavik, Iceland, 7–9 June 2006; pp. 290–293.