Transactions of the VŠB – Technical University of Ostrava, Mechanical Series
No. 1, 2019, vol. LXV
article No. 2051

Milan JIČÍNSKÝ[*], Jan MAREŠ[**]

SOFTWARE TOOL FOR VOICE DISORDER DIAGNOSTICS

APLIKACE PRO DIAGNOSTIKU PORUCH HLASU

Abstract

Causes of voice disorders may vary the same way as treatment techniques. Surgical intervention is the most used treatment method. As for not so serious cases, vocal exercises can be efficiently used instead. Still, there are only few methods to find right diagnosis or classify the scale of impact on patient's voice. Nowadays, it is possible to extract key features (fundamental frequency, sound pressure level and less common such as cepstral coefficients, zero crossing rate or spectral energy) from patient's speech using state-of-the-art voice processing methods. So, the software called Voice disorder diagnostician was designed. The tool that can store patient's data and immediately provide the results of analysis.

Abstrakt

Příčiny vzniku poruch hlasu mohou být různé stejně tak jako jejich léčba. Nejčastěji se poruchy hlasu léčí chirurgicky tj. operací nebo v méně závažnějších případech hlasovými cvičeními. Přesto však v současné době existuje jen velmi málo metod, které mohou spolehlivě určit diagnózu pacienta a míru poškození hlasu. S využitím aktuálních metod pro zpracování řeči je možné určit klíčové parametry jako je základní tón hlasu, hladina akustického tlaku, ale i méně běžné jako například kepstrální koeficienty, počet průchodů nulou, či spektrální energie. Pro tyto účely byla vytvořena aplikace s názvem Voice disorder diagnostician. Nástroj, který umožňuje ukládat data od pacientů a provádět jejich okamžitou analýzu.

Keywords

Voice disorder, diagnostic application, diagnostics, voice processing, microphone calibration, sound pressure level, pitch, feature extraction.

1 INTRODUCTION

After certain types of head surgery, patients are often affected by changes in voice. Because rehabilitation takes several months, doctors and patients are able to monitor progress quantitatively and objectively. At present, such quantification is subjective and highly dependent on the doctor's opinion and experience. Thus, we here introduce the use of new computational-intelligence based tool for objective measurement and analysis of voice disorders.

The work is a part of complex project devoted to development of computational and cybernetics systems for modern diagnostics. This approach profits from a combination of both digital signal processing methods based on advanced mathematical algorithms and experience of specialists in medicine. Many research groups all over the world started to deal with research in this

[*] Ing., Process Control, Faculty of Electrical Engineering, University of Pardubice, Studentská 95, Pardubice, tel. (+420) 46 603 7208, e-mail milan.jicinsky@student.upce.cz

[**] doc. Ing., Ph.D., Process Control, Faculty of Electrical Engineering, University of Pardubice, Studentská 95, Pardubice, jan.mares@upce.cz

multidisciplinary area where knowledge of information engineering and biomedicine are necessary. For all we name only:

- neural networks in the brain cancer diagnosis [1],
- advanced statistical analysis in the detection of nasopharyngeal cancer [2],
- computational intelligence in a diagnosis of neck and head cancer [3],
- analysis and diagnosis of Parkinson's disease [4], [5].

2 VOICE DISORDER DIAGNOSTICIAN

Diagnostics of voice disorders can be very complex task from user's point of view - collecting and storing data, calibration according to the standards, keeping patient database and displaying results of voice processing algorithm using feature extraction. All these features are implemented in innovative user-friendly software called Voice disorder diagnostician. The software is completely designed in Matlab. The main idea is to simplify the whole process as much as possible for endpoint user. Earlier version with limited functionality and mathematical background of feature extraction algorithm was described in [6].

2.1 Main menu

There are 6 options in the main menu which will be described in the following chapters in detail. Language can be set by pressing language button located in lower right corner. Only English and Czech language is supported. Patient information section below the buttons is reserved for displaying information about current patient either entered as a new patient or imported from the database. The help button is added in this version. When clicked, digital PDF manual will show up. The manual contains all necessary information about component connection, calibration description, requirements for fulfilling standards during capture of voice, the whole procedure described in few steps. Even an unexperienced user can learn how to handle the software correctly.

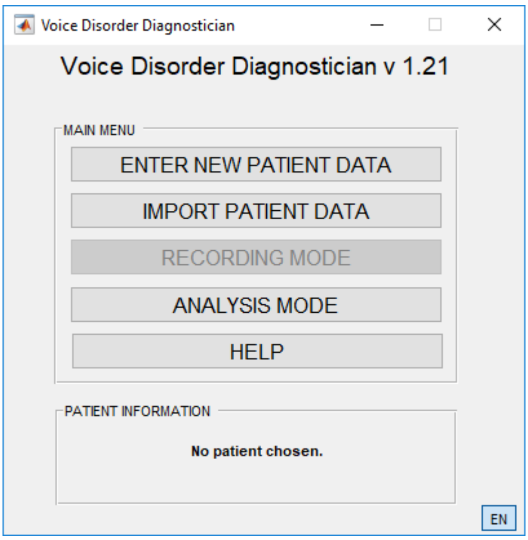

Fig. 1 Voice disorder diagnostician - main menu.

2.2 Entering and importing data

The most important difference between the current version and the old one is the way of handling patient data. In case of adding a new patient to the database user chooses first option (enter new patient data). This is show at Fig. 3 (left window). Physician fills in the name, surname and date of birth of patient. Then physician must select if the patient is either male or female and either native Czech speaker or not. Filling additional information is not compulsory. Patient's unique identifier is generated according to all this compulsory information. Identifier example is shown on Fig. 2, where bold characters are used to generate ID.

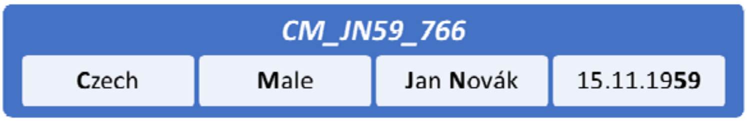

Fig. 2 An example of patient's ID.

This format applies for patient's ID: The first character can be either C – Czech or X – Other nationality. The second character corresponds to genders (M – male, Z – female). Then there is underline followed by the patient's initials and the last 2 digits of the year the patient was born. Finally, it is followed by another underline and specific code related to patient. The code consists of 3-4 digits and its computed by converting name and surname into numbers according to the ASCII standard. The code value is determined as follows.

$$code = \sum_i n(i) + \sum_j s(j) - y + d - 10m \qquad (1)$$

where:

n – name containing i characters converted into the vector of numbers [-],

s – surname containing j characters converted into the vector of numbers [-],

y – last 2 digits of the year the patient was born [-],

m – a month the patient was born [-],

d – a day the patient was born [-].

This approach (when month is multiplied by 10) was chosen because some combinations of name and surname are rather frequent and there was possibility to generate the same IDs for patients whose names and surnames are the same together with the year of birth. But still there is a very small chance of having identical information for 2 patients. Therefore, (only in that case) the identifier would be automatically accompanied by the date of patient registration. Finally, information about patient registration is saved to the file named by patient's ID. These files have *.vdd extension.

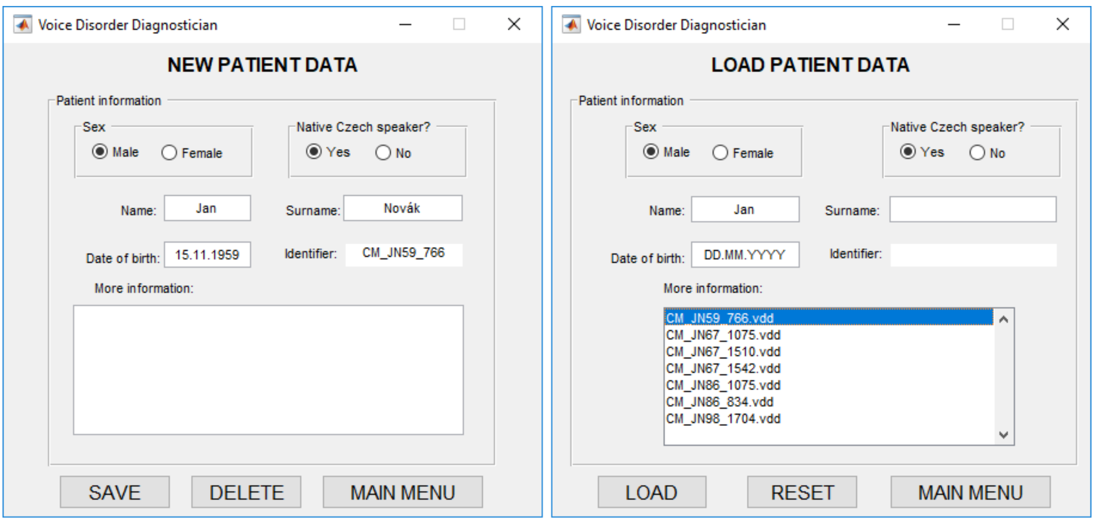

Fig. 3 Entering new patient or importing an existing one.

The second important change in current version is that physician does not have to look for stored patient files manually after opening Windows explorer window. Importing patient data screen is almost the same as for entering a new patient. The comparison is at Fig. 3. The main difference is that there is a list of previously registered patients below. The application loads the list of vdd files stored in predefined folder and displays it. Advanced filtering is implemented. So, when physician enters any information about patient, the results are filtered according to given criteria. When all forms are filled in, patient ID is generated, and physician can be sure of choosing the right file. User is then returned to the main menu where current patient data is displayed.

2.3 Recording mode

The recording mode is reserved for capturing of patient's voice. It shows both sound level meter and microphone signal. Firstly, physician must choose microphone and phase of recording. We distinguish 3 phases: phase 0 – before treatment, phase 1 – after surgery and phase 2 approx. 2 week later. User then moves on to the calibration mode. Recording device and phase can be changed no more when returned to recording mode from the calibration screen. Calibration confirmation is shown in the lower part. Since the device is calibrated both sound level meter and microphone signals should be overlapping. The graph scale corresponds to acoustic pressure. Real time sound pressure level measured in decibels is displayed in lower right corner. The number of recordings is not limited. Files with *.wav extension are saved to predefined folder. The filename consists of patient's ID, phase and the number referring to order of recordings.

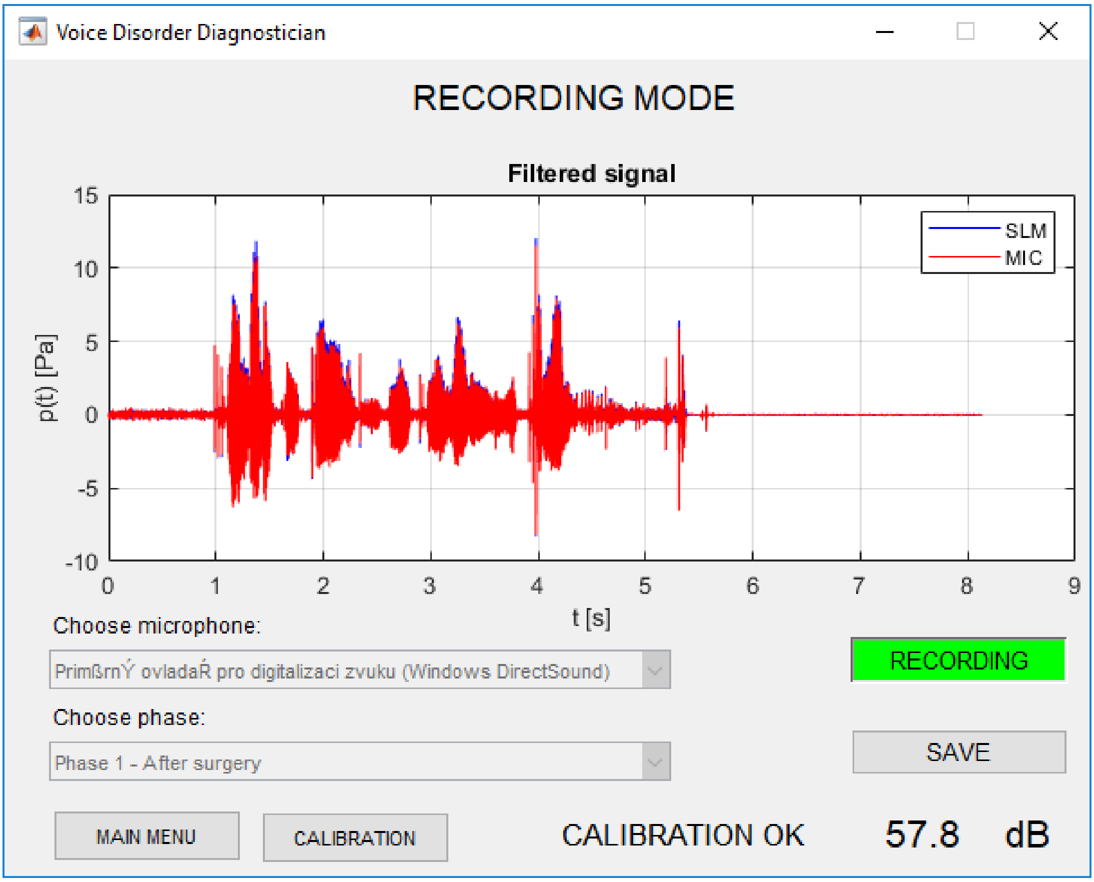

Fig. 4 Voice disorder diagnostician - recording mode.

2.4 Calibration

The most important improvement happened in calibration. Previous versions had no microphone calibration and so the results could not be compared. The current version is using so called 2-step (2-phase) calibration. There are strict requirements for hardware. The microphone needs to have flat frequency response. Dynamic range is also a key parameter. More complex insight about this matter provides [7]. Omnidirectional DPA dfine microphone and Extech sound level meter combined with calibrator are used. Because of bad sound quality of embedded PC sound cards, external Focusrite sound card is used. External sound card has 2 inputs with adjustable gains and phantom voltage (48V). Phantom voltage is necessary to power up the microphone. In general, Voice disorder diagnostician requires a sound card with 2 inputs. Sound level meter must be connected to the left input and microphone to the right input. In other words, single audio track is transferred to the computer, where sound level meter signal must be in the (left) channel 1 and microphone signal must be in the (right) channel 2. Complete connection scheme is shown at Fig. 5.

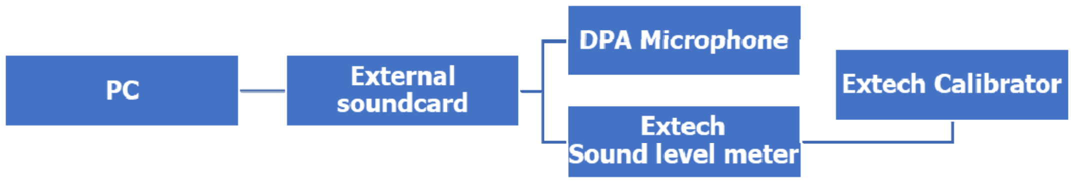

Fig. 5 Connection scheme.

The first step of 2-phase calibration is capturing sound level meter signal from A/C output while the calibrator generates 1kHz sinewave at 94dB. Frequency 1kHz and 94dB are typical values for calibration signal. Sound level meter constant is computed from the calibration signal according to the formula 2. SPL is known (94dB) and rms can be computed using formula 3.

$$SPL = 20 \cdot \log(c \cdot rms) \tag{2}$$

where:

SPL – sound pressure level [dB],

c – device calibration constant [-],

rms – root mean square [-].

$$rms = \sqrt{\frac{\sum_{n=1}^{N} p^2(n)}{N}} \tag{3}$$

where:

p – pressure [Pa],

N – length of pressure vector, total number of values [-].

The second part of calibration requires stable position. The distance between DPA microphone and the microphone of sound level meter has to be 30 cm and it cannot be changed either during the calibration or during following measurement. The reason is that acoustic pressure or SPL changes with distance. Therefore, the microphone constant can only be evaluated if this condition is met. Equation 4 shows the relationship between sound level meter and microphone constants.

$$c_{slm} \cdot rms_{slm} = c_{mic} \cdot rms_{mic} \tag{4}$$

Microphone signal can be than multiplied by microphone calibration constant which makes it the same scale as sound level meter signal. The main advantage is that microphone provides better sound quality and output corresponds with sound level meter signal. It means that microphone placed beneath patient’s mouth captures sound like if it would be placed in 30 cm distance.

The whole process is very simple from user perspective. Calibration window is divided into two parts. The left half represents sound level meter (left audio channel) and the right side is reserved for the microphone. Calibration starts by clicking the button below the graph and ends by clicking it again. When the first phase (capturing of generated sinewave) is done the sound level meter constant appears in the bottom left. Originally disabled button for phase 2 is now enabled. It allows to calibrate the microphone the same way as user did before in phase 1. At the end of calibration sound level meter signal from phase 1 and both signals for phase 2 are saved and stored in mat files. The filename consists of patient ID, phase and device designation. Finally, user returns to the recording mode and anything subsequently recorded will be calibrated so it can be compared to any other recording.

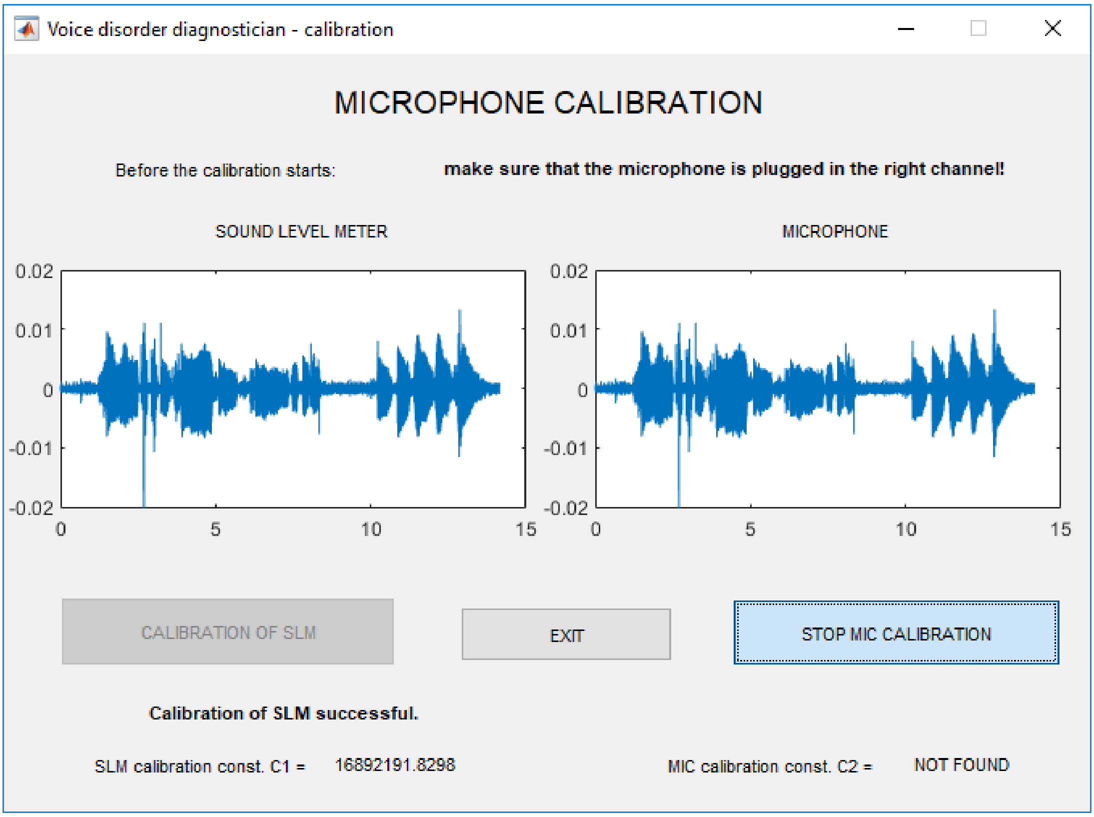

Fig. 6 Voice disorder diagnostician – calibration.

2.5 Analysis

Analysis mode allows physician to see immediate results of feature extraction algorithm. It means that acoustic pressure, log energy, zero crossing rate, spectrogram and pitch (fundamental frequency) are graphically interpreted. If the microphone is calibrated it also displays SPL graph. This is complemented by spectral energy values distributed into 8 frequency bins and MFCC cepstral coefficients. There are no evaluation metrics yet. Analysis mode is currently used as a feedback.

It is also planned to add voice range profile graph. Voice range profile is standardized graphical method used by some physicians during checkup to determine patient's vocal status. It can display patient's voice fundamental frequency and SPL. These two key features can graphically mark an area corresponding to the voice produced by vocal tract. So, the voice range profile (frequency and SPL range) changes according to the patient's condition. Differences between patients suffering from voice disorders and healthy ones are expected to appear affecting voice range profile scale and shape.

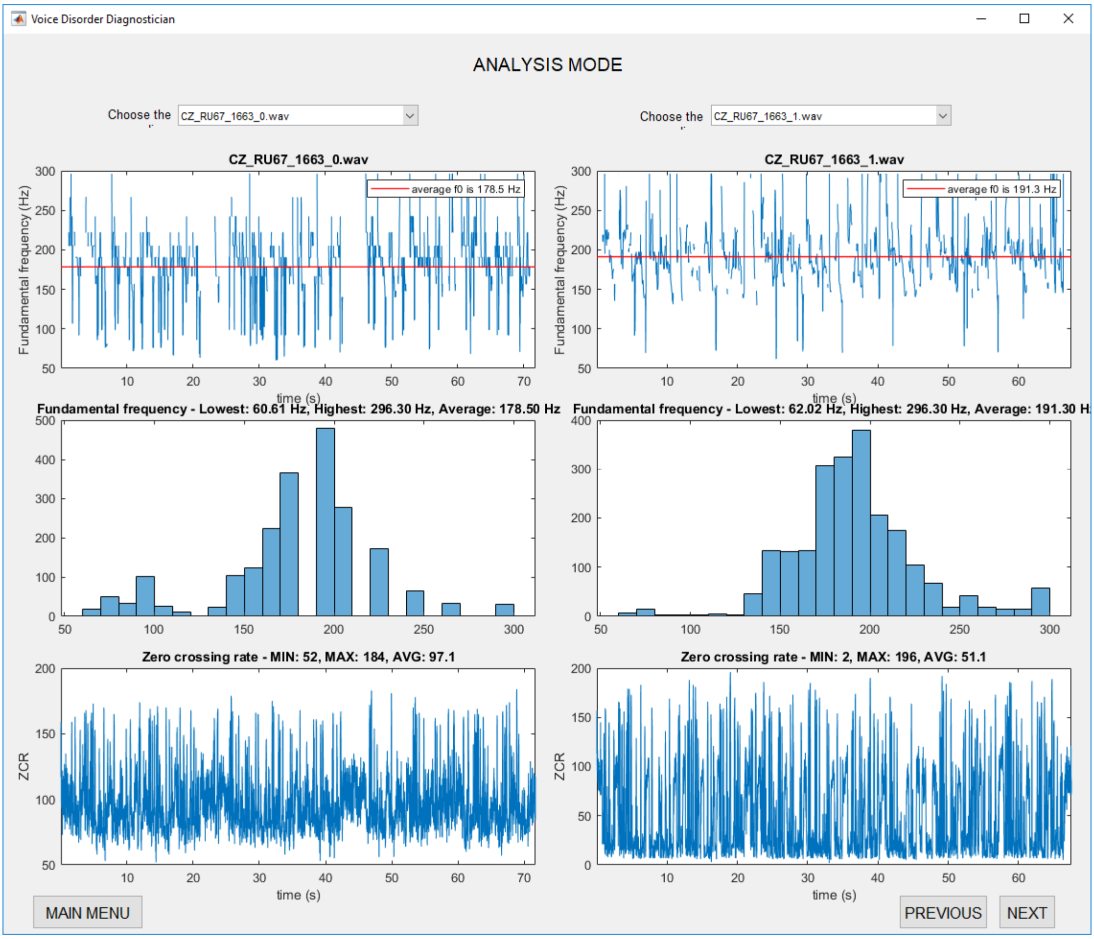

Fig. 7 Feature vector graphical representation within analysis mode.

2.6 Current status and future research

University of Pardubice cooperates with University hospital Královské Vinohrady (FNKV) in Prague to develop fully functional standalone application capable of voice disorder diagnostics. FNKV provided a silent chamber so that patients could be recorded in it to fulfill recording requirements and standards. FNKV also created a team of experts who supervise the whole recording process. They also contributed by their knowledge of medical procedures. FNKV will also provide feedback about diagnosis of previously recorded patients. Final diagnosis needs to be known in detail in order to develop diagnostic core of the application. Right now, the application doesn't contain any diagnostic mode. Supervised recording process and data collection started in May 2019. The amount of obtained data is currently insufficient. In order to make progress in diagnostics, machine learning techniques can be used to train models. But that cannot be performed if having lack of data. As soon as the dataset will be enlarged, Voice disorder diagnostician development can continue.

3 CONCLUSIONS

The Voice disorder diagnostician version 1.21 is standalone application suitable for medical purposes as a complement to the regular otorhinolaryngological checkup. It allows physician to add patient to the database or load patient's data. Since the recording device is calibrated patient's voice can be recorded and stored in a file. The recordings can be analyzed anytime. The results of two phases of the same patient or completely different patients can be compared. Application interface is simplified and user friendly. The user doesn't need to know anything about voice processing or any other technique used. The whole procedure can be done following the steps written in application manual. If having appropriate hardware, the Voice disorder diagnostician can be used anywhere (English language supported). Unfortunately, the application is not yet capable of making decisions about patient's diagnose. This feature will be implemented as soon as sufficient number of recordings will be collected.

4 ACKNOWLEDGMENT

This research was supported by students SGS grant at Faculty of Electrical Engineering, University of Pardubice. This support is very gratefully acknowledged.

REFERENCES

[1] JERMYN, M., DESROCHES, J., MERCIER, J., TREMBLAY, M.A., ST-ARNAUD, K., GUIOT, M.C., PETRECCA, K., LEBLOND, F. Neural networks improve brain cancer detection with raman spectroscopy in the presence of operating room light artifacts. *Journal of biomedical optics.* 2016, XXI. Nr. 9.

[2] LIN, D., GE, X., LIN, X., CHEN, G., CHEN, R. Blood surface-enhanced raman spectroscopy based on agand au nanoparticles for nasopharyngeal cancer detection. *Laser Physics.* 2016, XXVI. Nr. 5.

[3] HARRIS, A.T., RENNIE, A., WAQAR-UDDIN, H., WHEATLEY, S.R., GHOSH, S.K., MARTIN-HIRSCH, D.P., FISHER, S.E., HIGH, A.S., KIRKHAM, J., UPILE, T. Raman spectroscopy in head and neck cancer. *Head & neck oncology.* 2010, II. Nr. 1.

[4] MACHIDA, E., CAO, M., MURAO, T., HASHIMOTO, H. Human motion tracking of mobile robot with kinect 3d sensor. In: *2012 Proceedings of SICE Annual Conference (SICE).* Place: IEEE, 2012, pp. 2207–2211. ISBN 978-1-4673-2259-1.

[5] BOYD, J.E., GODBOUT, A., THORNTON, C. In situ motion capture of speed skating: Escaping the treadmill. In: *2012 Ninth Conference on Computer and Robot Vision (CRV).* Place: IEEE, 2012, pp. 460–467. ISBN 978-1-4673-1271-4.

[6] JIČÍNSKÝ M. & MAREŠ J. Speech processing in diagnosis of vocal cords diseases. Unpublished.

[7] ŠVEC, J. G., & GRANQVIST, S. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*. 2010, XIX. Nr. 4, pp. 356–368.