A Phonatory System Simulator for testing purposes of voice-monitoring contact sensors

Original
A Phonatory System Simulator for testing purposes of voice-monitoring contact sensors / Casassa, Federico; Troia, Adriano; Astolfi, Arianna; Carullo, Alessio; Vallan, Alberto; Schiavi, Alessandro; Corona, Davide. - ELETTRONICO. - (2017). (Intervento presentato al convegno I2MTC 2017 tenutosi a Torino nel 22-25 maggio 2017).

Availability:
This version is available at: 11583/2670127 since: 2017-09-29T11:35:37Z

Publisher:
IEEE

Published
DOI:

A Phonatory System Simulator for testing purposes of voice-monitoring contact sensors

F.Casassa, A.Carullo, A.Vallan
DET – Electronic and Telecomunications Dept.
A.Astolfi
DENERG – Energy Dept.
Politecnico di Torino
Torino, Italy
Email: federico.casassa@polito.it

A.Troia
Metrology for Quality of Life
A.Schiavi , D.Corona
S.T.A.L.T.
I.N.Ri.M
Torino, Italy

***Abstract*—Monitoring the health of the phonatory system plays a key role for the voice professionals, which are people who use the voice as a work tool. The voice quality can be assessed by means of parameters extracted from the vocal signal sensed by a contact sensor placed at the base of the neck, that is sensitive to the skin vibration related to the phonation. The use of a contact sensor is preferable to a microphone in air because of its lower sensitivity to the acoustical background noise. This paper focus on the development of a system for the characterization of the contact sensor that can be used for this purpose: an apparatus that can mimic the phonatory system and can act like a standard for testing the sensors. Such a system has been created and preliminary tested and characterized, in order to point out the similarity with the real phonatory system and to bring out any critical issue.**

Keywords—biomedical transducers, vibrations, vibration measurements, human voice, acoustic testing, wearable sensors, phantoms

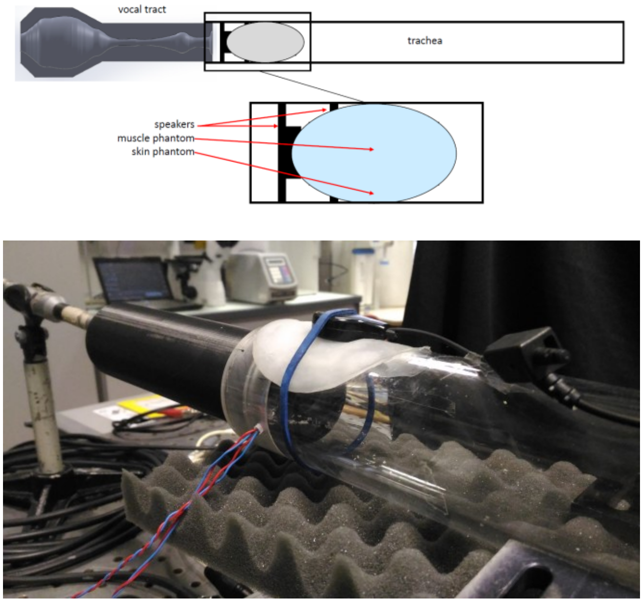

Fig. 1. Phonatory system simulator scheme, horizontal section [top], a detail [middle] and a picture of the preliminary tests [bottom].

I. INTRODUCTION

Vocal monitoring consists in recording the voice activity of a person in order to obtain parameters and descriptors of the voice quality. The main purpose is the early detection of disorder that are correlated to the prolonged use of voice in the so-called voice professionals [1-6], such as teachers and singers. Many devices and methods for voice monitoring has been developed [7-12] that are commonly based on a contact sensor, which senses the skin vibration at the base of the neck (at the jugular notch) produced by and related to the phonation: the vocal folds vibrate and create a pressure signal that is the source of the phonation, but they also produce a vibration in the tissues next to the glottis (the neck region around the adam's apple).

The use of a contact sensor is preferable to a microphone in air because it provides useful information on the vocal activity of the monitored subject ensuring a very low influence of the noise present in the room where the monitoring takes place. Various type of vibration transducers are used: piezoelectric transducers, accelerometers and electret contact microphones. The quantification of the voice quality is carried out on the base of frequency and intensity –based parameters like *jitter, shimmer* and *CPPS*, which are extracted from the vibration signal obtained from the subject under monitoring [13-15]. For this reason, the vibration sensor characterization is of primary importance to distinguish between the actual vibration signal related to the phonation and the possible artifacts due to the sensor characteristics (load effect, sensitivity to air-pressure signal, frequency response, weight, dimension). Unfortunately, the specifications provided by the manufacturer are often incomplete and unusable for voice monitoring purposes for two main reasons: the low cost devices engineered to be used like laringophones are not conceived for research or diagnostic/medical purposes, and the common vibration transducers are often designed only for mechanical and industrial use. In the existing literature, the vocal monitoring systems are tested and characterized on the base of tests on human subjects [7-12], but there is a lack of a standard that can be used to analyze and compare the response of the various transducers for vocal-monitoring use. The tests performed to obtain this kind of characterization usually involve various subjects that perform simple voice tasks (reading a passage of a text, vowel vocalization at different pitch and intensity, free speech), and the performance of the sensors are compared in terms of the estimated vocal parameters. These tests carried out on different sensors [16] have shown difficulties in obtaining an effective sensor characterization, because of the poor reproducibility of the system under test, i.e. human subjects. For this reasons, the availability of a system that is able to mimic the biological system on which the sensors have to work is advisable. Such a device has to act as a source-filter system that gives two well-known outputs to a source signal: an acoustical signal, which can be detected by a microphone in air, and a vibration signal that can be detected on a skin-like material.

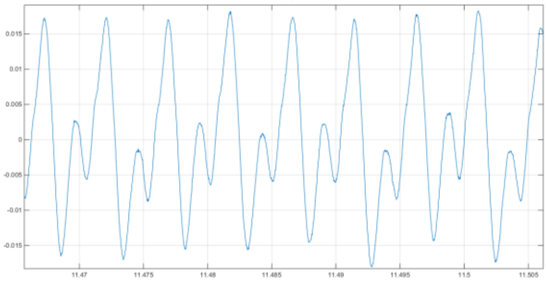

Fig. 2: vibration signal acquired with an electrect contact microphone on the simulator with EGG source signal.

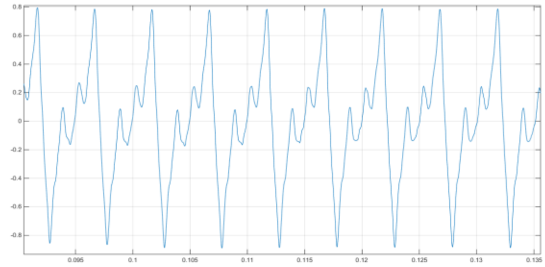

Fig. 3: vibration signal acquired with an electrect contact microphone on a human subject while emitting a /a/ vowel.

In a biological framework, a well-known relationship between the vibration at the base of the neck and the acoustical vocal signal is hard to be accurately obtained. Both the signals are related to the source of phonation, i.e. the air that flows through the glottis and drives the vocal folds into a self-oscillating motion, but the responses of the two filters (skin and vocal tract) are not well reproducible due to various factors: tissue thickness, body fat, sensor position, attachment methods, sensitivity to acoustical stimuli. Furthermore, it is hard to discriminate between the artifact due to the sensor characteristics and the features of the vibration signal originated by a particular subject. In order to obviate to this problems a system has been already developed [17, 18] and used for a preliminary sensor characterization, but its architecture was engineered to mimic the respiratory system. This work instead deals with an apparatus that has been conceived to be similar to the phonatory system, which acts as a stable generator that makes contact-sensor characterization free from reproducibility problems. Moreover, the proposed system allows other characteristics of the sensors to be estimated that cannot be obtained in vivo, such as the load effect of the sensor due to its dimension, shape and weight. The sensitivity to acoustical noise can also be easily investigated by using an external sound source, avoiding the problems that have been faced during in-vivo measurements [16].

II. THE APPARATUS

A. *Phonatory system description*

The phonatory system is made up of the trachea, i.e. an air channel that convolve in the larynx the air that comes from the lungs, the larynx itself and the vocal tract. The larynx is the real source of phonation: it contains the vocal folds, two infoldings of thyroarytenoid muscle, and the glottis, the aperture between them. The air column ascends the trachea and drives the vocal folds into a self-oscillation (the glottis opens and closes itself) to create a pressure wave at the exit of the larynx. This also creates the so called subglottal resonances in the back of the glottis (pressure standing waves in the trachea [19]). The pressure wave passes then through the vocal tract, which is formed by the terminal tract of the pharynx and the oral cavity. From an acoustic point of view, the vocal tract acts as a resonator that amplifies the original pressure wave at the so called “formant frequencies” or “formants”. Different vocal tract configurations correspond to different vowels, and each vowel is characterized by their formant frequencies. The vocal folds activity during vocalization can be measured with an electroglottograph (EGG), which is an impedance meter that measure the relative impedance between the vocal folds and give information on the glottis opening-closing cycle. The spectral envelope provided by the EGG is characterized by the greatest amplitude of the fundamental frequency and harmonic components with decreasing amplitude as the frequency increases. In the spectral envelope of the acoustical voice

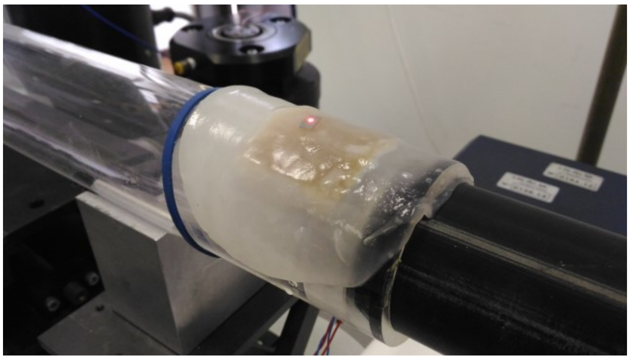

Fig. 4. Simulator characterization: reflective material on the skin phantom to record the vibration with the laser vibrometer.

signal, the higher harmonics (formant frequencies) are instead stronger than the fundamental frequency, due to the effects of the vocal tract.

B. Simulator description

The aim of this project is to develop a system that is able to mimic the vocal apparatus (trachea, glottis/vocal chords, neck tissue and vocal tract) in order to obtain a test system with well-known relations between the vibration of a human-tissue-like material and an acoustic signal. For this reason it is not required to develop an exact mechanical replica of the vocal apparatus like in [20-22], because the goal is the characterization of contact sensors in a stable vocal monitoring framework. The main parts of the vocal apparatus has been recreated as follows, as shown in figure 1:

- trachea: hollow plexiglass tube with a diameter of 5 cm;
- glottis / vocal folds / source signal: two speakers that emit in two different direction, with a $\pi/2$ phase difference;
- vocal tract: a 3D-printed model of Human Vocal Tract;
- neck tissues: tissue-mimicking phantom material (skin and muscles).

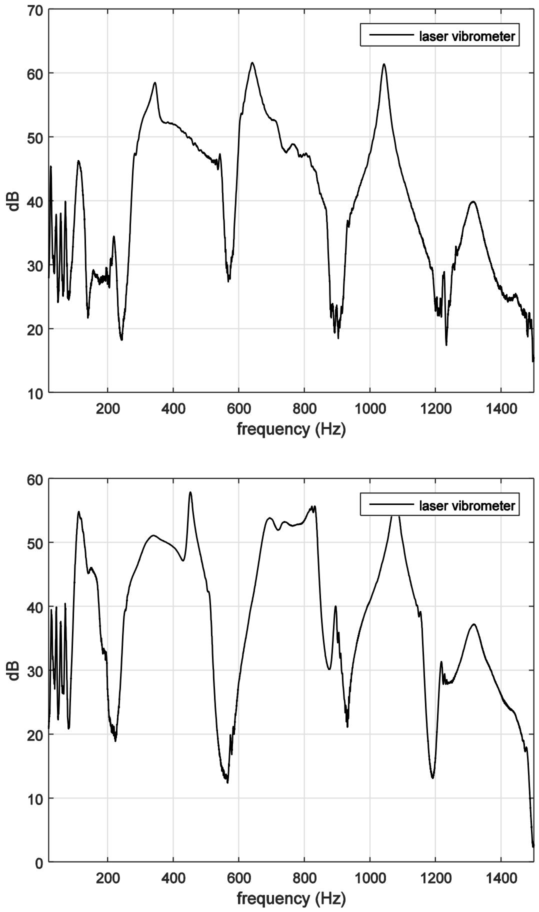

Fig. 5. Vibrometer spectral envelopes obtained with a frequency sine sweep source signal, closed-end [top] and open-end [bottom] configuration.

The two speakers recreate the glottis activity by emitting two pressure signals: one signal propagates in the plexiglass tube (trachea) and the other one propagates in the vocal-tract simulator.

The plexiglass tube used to simulate the trachea is 55 cm long and it is equipped with a moving removable end, in order to recreate the subglottal resonances by means of the closed/open tube modes. These can be tuned by changing the length of the plexiglass tube, to obtain the right frequencies to mimic the subglottal resonances.

The Vocal tract is a 3D-printed hollow resonator, which is based on a model of the vocal tract while emitting an /a/ vowel proposed by Švec et al. [23]. This resonator allows the formant frequencies from the pressure signal originated from the speaker to be amplified.

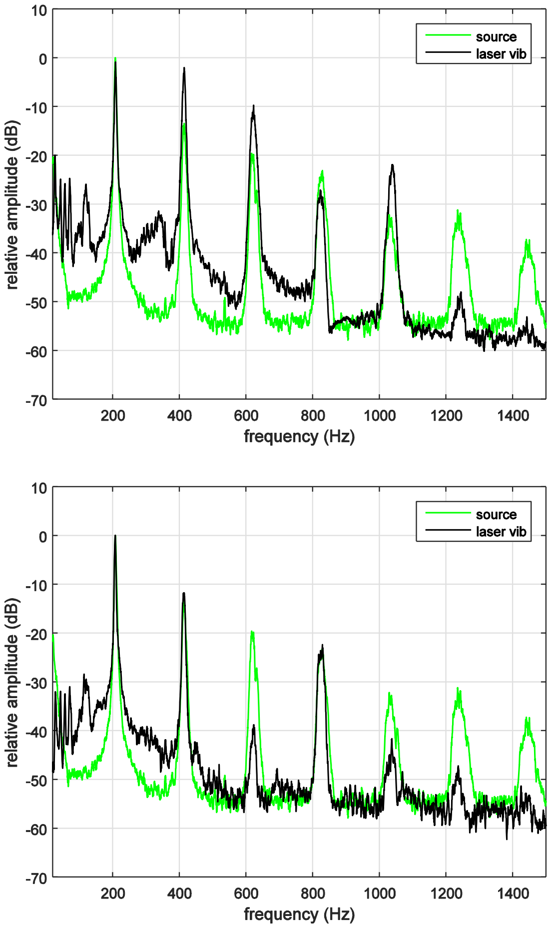

Fig. 6. Source signal and vibrometer spectral envelopes obtained on the simulator with EGG source signal, closed-end [top] and open-end [bottom] configuration.

As a coupling phantom, a multi component Tissue Mimicking Materials (TMM) has been realized. Two types of muscle/throat simulating tissue have been investigated as TMM: a stiff Gellan Gum based hydrogel containing kieselghur and silicon carbide solid particles, whose detail on preparation and characterization of its acoustic and mechanical properties can be found in [24-25]. Alternatively, a simple latex rubber based TMM, which showed similar properties but with a better time stability, is appeared as ideal to perform several trials required for a complete characterization of the contact sensors. The results presented in this paper are obtained with the latex rubber based TMM. A polyvinyl alcohol based phantom have been prepared with freezing/thaw cycle technique and used as skin simulating tissue. Further details on its preparation can be found in [26].

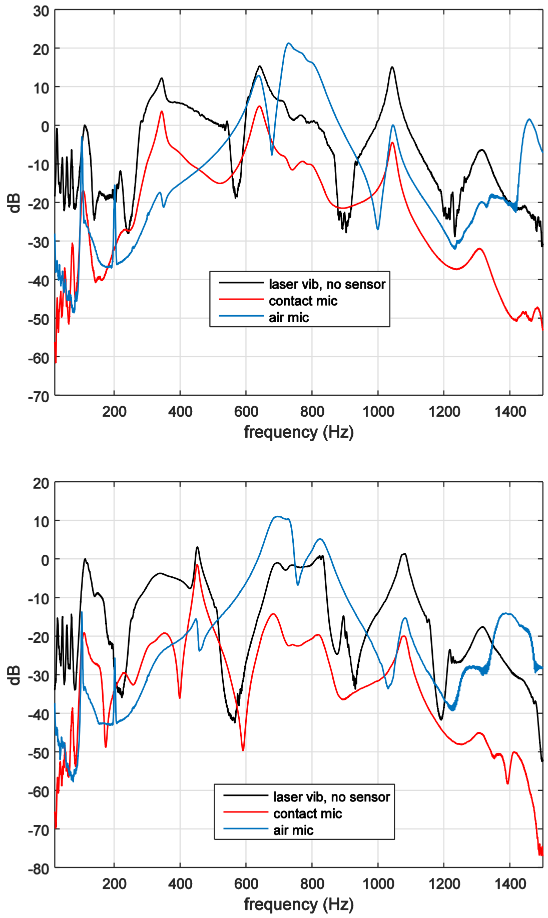

Fig. 7. Vibrometer, contanct sensor and microphone in air spectral envelopes obtained on the simulator with sine sweep frequency source signal, closed-end [top] and open-end [bottom] configuration

III. EXPERIMENTAL RESULTS

With the aim of validating the effectiveness of the phonatory-system simulator, specific tests have been performed in order to characterize the glottis-vibration filter and the glottis-voice filter. The acoustic signals have been sensed with a Behringer ECM2000 microphone, the vibration signals have been sensed with a Midland MIAE38 electrect contact microphone and with a Meas-Spec CM-01B piezo-film contact microphone. The signals has been acquired by means of a National Instruments USB 6211 acquisition board connected to a PC, with a 22 kHz sampling frequency. The tube length was 50 cm in the closed-end configuration and 55 cm in the open-end configuration.

A. *Preliminary tests*

In the first test, the two speakers have been driven with an EGG signal corresponding to a sustained /a/ vowel and the output of the simulator on the skin phantom has been acquired using an electret contact microphone. The obtained signal, which is shown in figure 2, is very similar to the vibration signal acquired on a human subject that emits the same vowel, as showed in figure 3. Even though this result seems to validate the proposed approach, one should note that the compared signals refer to different subjects. In order to better evaluate the simulator performance a further step will concern the simultaneous acquisition of EGG, vibration and acoustic signals from the same subject, and the comparison between those signals and the ones obtained with the simulator driven with the same EGG signal.

B. *Simulator preliminary characterization*

The apparatus has been characterized by means of a Laser vibrometer (Polytec CLV-2534), in order to obtain the system vibration response with no mechanical sensor attached to the phantom. The simulator has been placed on the optical bench and the laser was focused on a piece of reflective material placed on the skin-mimicking phantom (fig. 4).

The spectra presented in figure 5 refer to the acceleration signal obtained with the vibrometer; in this case the source signal is a frequency sine sweep (10 seconds duration, 100Hz - 1500 Hz). The peaks represent the acoustical resonances of the speakers and the tube in the two configurations (open-end and closed-end). These spectral envelopes define the filter effect of the simulator, which includes the electrical-to-acoustical response of the speakers, the effect of the tube’s resonances and antiresonance and the attenuation of the phantoms. The resonance frequencies of the tube in the open-end configuration can be calculated as [27, 28]:

$$f = \frac{nv}{2(L+0{,}8d)}, n = (1,2,3,\ldots) \tag{1}$$

For the closed-end configuration, the resonance frequencies are:

$$f = \frac{nv}{4(L+0{,}4d)}, n = (1,3,5,\ldots) \tag{2}$$

where v is the speed of sound in air in (m/s), L is the tube length (m), d is the tube diameter in (m). The peaks correspond to these frequency according to the simulator dimensions.

Moreover, the apparatus has been tested with the EGG signal as source, and the results are presented in figure 6. The spectral envelope of the acceleration signal is in good agreement with the source one, with attenuation in correspondence of the anti-resonant frequencies highlighted in the previous test (e.g. near 600 Hz and 1200 Hz in open-end configuration). This confirms the effectiveness of the approach: the source signal is well transmitted by the speaker to the phantoms despite of the involved filters effect.

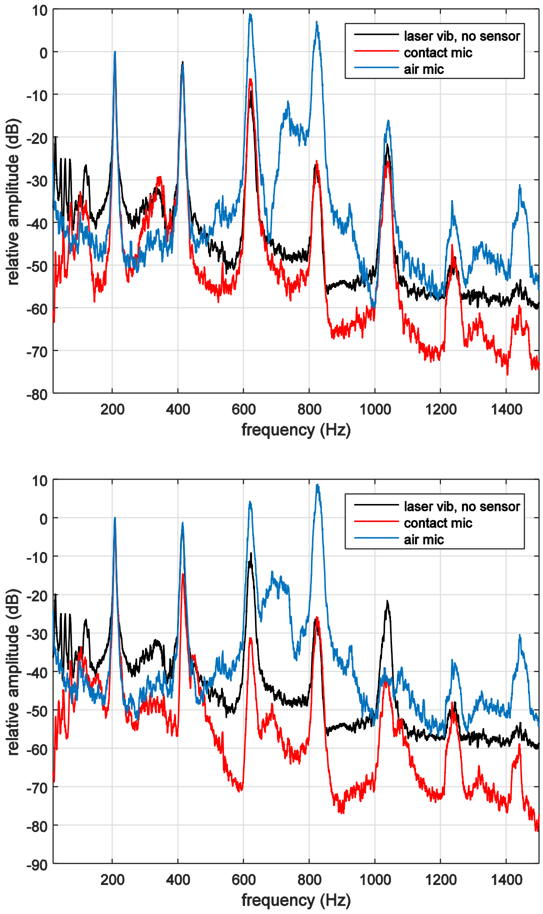

Fig. 8. Vibrometer, contanct sensor and microphone in air spectral envelopes obtained on the simulator with EGG source signal, closed-end [top] and open-end [bottom] configuration

C. *Simulator test*

In order to test the whole system and to exploit the influence of all the parts in the frequency domain, more tests were carried out. The contact sensor (CM-01B piezo-film contact microphone) was attached to the skin phantom's surface by means of a rubber band, and the microphone in air was placed at the resonator's end to acquire the acoustical output of the simulator. In this way it is possible to compare the vibration signal without any sensors attached with the actual vibration signal acquired by the contact sensor, and with the acoustical output of the simulator recorded at the end of the resonator.

The spectra presented in figure 7 are obtained by using the frequency sweep as a source signal. The amplitude is normalized on the first peak of the laser vibrometer spectral envelope. The peaks in the laser vibrometer spectral envelope (black) are present in the contact sensor one too (red), with a different amplitude due to the phantom attenuation and to the load effect of the sensor.

For what concern the acoustical response, it is evident from the microphone in air spectral envelope (blue) that the vocal tract resonator have an important influence on the acoustical output: it enriches the harmonic components by amplifying only some particular frequencies (corresponding to the /a/ formants).

In order to better underline the vocal tract resonator behavior and in general the relations between the vibration signal and the acoustic signal, in figure 8 the spectra obtained with the EGG source signal, normalized on the magnitude of the first harmonic, are presented. It is noticeable how the contact microphone spectral envelope is quite near to the vibrometer one, with small differences on the 3rd and higher harmonics (the maximum difference is 8 dB). The microphone in air spectral envelope is instead very different from the other, with the 3rd and 4th harmonics amplitude higher than the ones in the other two envelopes. The differences between the microphone in air and the contact sensor in closed-end configuration is 2 dB, 16 dB and 32 dB for the 2nd, 3rd and 4th harmonic respectively, and 12 dB, 35 dB and 34 dB for the same harmonics in open-end configuration.

IV. CONCLUSIONS

The tests carried out shows that the simulator mimics the behaviour of the phonatory system in an efficient way for what concern the initial purpose of the work: testing contact sensors to be used in a voice monitoring framework. The source signal is well transmitted to the phantom's surface, and the vocal tract resonator amplify the high harmonics of the source signal and create the vowel formants. The system has a stable relations between the source signal and the two output (sound and vibration), and it allows to test the contact sensors with a known reference. Hereafter, an artificial EGG-like signal will be created, to be used as a source signal.

To better point out the sensor's load effect it will be necessary to acquire the phantom vibration with the laser vibrometer differently loaded, simulating different sensor shapes and weights.

REFERENCES

[1] D.D. Metha, M. Zañartu, S. W. Feng, H. A. Cheyne, R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform", IEEE Trans. On Biomedical Engineering, vol. 59, n. 11, pp. 3090-3096, Nov. 2012

[2] G. J. Švec, I. R. Titze, P. S. Popolo, " Estimation os sound pressure levels of voiced speech from skin acceleration fo the neck", J. Acoust. Soc. Am, vol. 117, no. 3 pt. 1, pp 1386-1394, March 2005

[3] M. Ghassemi, J. H. Van Stand, D. D. Metha, M. Zañartu, H. A. Cheyne, R. E. Hillman, J. V. Guttag, "Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: initial results for vocal fold nodules", IEEE Trans. on biomedical engineergin, vol. 61, no. 6, June 2014, pp. 1668-1675

[4] P. Bottalico, A. Astolfi, "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms", Journal of the Acoustical Society of America, vol. 131 no. 4, pp. 2817- 2827, 2012.

[5] A. Astolfi, p: Bottalico,A. Accornero, M. Garzaro, J. Nadalin, C. Giordano, "Relationship between vocal doses and voice disorders on primary school teachers", Proc. Euronoise Conference, Prague, Czech Republic, pp. 55-60, 2012.

[6] M. Wirebrand, "Real-time monitoring of voice characteristics using accelerometer and microphone measurements" [Online.] available: http://liu.diva-portal.org/smash/get/diva2:432959/FULLTEXT01.pdf

[7] H.A. Cheyne, H.M. Hanson, R.P. Genereux, K.N. Stevens, R.E. Hillman, “Development and Testing of a Portable Vocal Accumulator”, Journal of Speech Language and Hearing Research, vol. 46, pp. 1457-1467, 2003.

[8] P.S. Popolo, J.C. Švec, I.R. Titze, “Adaptation of a Pocket PC for Use as a Wearable Voice Dosimeter”, Journal of Speech Language and Hearing Research, vol. 48, pp. 780-791, 2005.

[9] A. Carullo, A., A. Vallan, A. Astolfi “Design Issues for a Portable Vocal Analyzer”, IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 5, pp 1084-1093, 2013.

[10] A. Carullo, A. Vallan, A. Astolfi, “A Low-Cost Platform for Voice Monitoring”, Proceedings of I2MTC Conference, Minneapolis, MN (USA), May 6-9, 2013

[11] A. Carullo, A. Vallan, A. Astolfi, L. Pavese, G.E. Puglisi, “Validation of Calibration Procedures and Uncertainty Estimation of Contact-Microphone Based Vocal”, Measurement, vol. 74, pp. 130-142, Oct. 2015.

[12] M. Zañartu, C.C. Ho, S. S. Kraman., H. Pasterkamp, J. E. Huber, G.R. Wodicka, “Air-Borne and Tissue-Borne Sensitivities of Bioacoustic Sensors Used on the Skin Surface”, IEEE Tr. on Biomedical engineergin, vol. 56, no. 2, pp.443- 451, 2009.

[13] Y.D. Heman-Ackah, D.D. Michael, G.S. Goding, “The relationship between cepstral peak prominence and selected parameters of dysphonia”, J Voice,vol.16, pp. 20–27, 2002.

[14] Y. Maryn, M. De Bodt, N. Roy, “The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders,” J. Commun. Disord.,vol. 43, pp.161–174, 2010.

[15] S.N. Awan, N. Roy, M.E. Jette, G.S. Meltzner, R.E. Hillman, “Quantifying dysphonia severity using a sepctral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V,” Clin. Linguist. Phon., vol.24, pp.742–758, 2010.

[16] A. Carullo, F. Casassa, A. Castellana, A. Astolfi, L. Pavese, G. E. Puglisi, “Performance comparison for different contact-microphones used for voice mointoring”, 22nd Interntion congress on sound and vibration, Florence (ITA), June 12-16, 2015.

[17] S.S. Kraman, G. A. Pessler, H. Pasterkamp, G. R. Wodicka, “Desing, construction and evaluation of bioacoustic transducer testing (BATT) system for respiratory sound”, IEEE Trans. on biomedical engineering, vol. 53, no. 8, pp. 1711-1715, Aug. 2015.

[18] S.S. Kraman, G. R. Wodicka, G. A. Pessler, H. Pasterkamp, “Comparson of lung sound transucers using a bioacoustic transducers testing system”, J. Appl. Physiol., vol. 101, pp. 469-476, Aug. 2006.

[19] S. F. Austin, I. R. Titze, “The effect of subglottal resonances upon vocal fold vibration”, Journal of voice, vol. 11, iss. 4, pp. 391-402, December 1997.

[20] J. Horacek, V. Uruba, V. Radolf, J. Vesely, V. Bula, “Airflow visualization in a model of human glottis near the self-oscillating vocal folds model”, J. Apllied and computational mechanics, vol. 5, pp 21-28, 2011.

[21] L. P. Fulcher, R. C. Sherer, T. Powell, “Pressure distribuition in a static physical model of the uniform glottis: entrance and exit coefficients”, J. Acoust. Soc. Am., vol. 129, no. 3, March 2001.

[22] G. C. J. Hofmans, G. Groot, M. Ranucci, G. Graziani, A. Hirshberg, “unsteady flow through in-vitro model of the glotts”, J. coust. Soc. Am, vol. 113, no. 3, march 2003.

[23] T. Vampola, J. Horacek, J. G. Švec, “FE modeling of human vocal tract acoustics. Part 1: production of czech vowels”, Acta acustica united with Acustica, vol. 94, pp 433-447, 2008.

[24] A. Schiavi, R. Cuccaro, A. Troia, “Strain-rate and temperature dependent material properties of Agar and Gellan Gum used in biomedical applications”, Journal of the mechanical behavior of biomedical materials, vol. 53, p. 119-130, 2015.

[25] R. Cuccaro, C. Musacchio, PA. Giuliano Albo, A. Troia, S. Lago, “Acoustical characterization of polysaccharide polymers tissue-mimicking materials”, Ultrasonics, vol. 56, p. 210-219, 2014.

[26] K.J.M. Surry, H.J.B. Austin, A. Fenster, T.M. Peters, “Poly(vinyl-alcohol) cryogel phantoms for use in ultrasound and MR imaging”, Physics in Medicine and Biology, vol. 49, no. 24, pp. 5529-5546, 2004.

[27] “Open VS closed pipes (flutes and clarinets)”, University of New South Wales, [online] http://newt.phys.unsw.edu.au/jw/flutes.v.clarinets.html .

[28] Johan Lijencrants, “End correction at a flute pipe mouth”, [online] http://www.fonema.se/mouthcorr/mouthcorr.htm .