WHY VOICE MATTERS IN MEDICINE

CONVERSATION WITH PROF. KRZYSZTOF IZDEBSKI

Kraków, Poland; La Quinta, California; 29.07.2020

Krzysztof Izdebski

Bartosz Ziółko

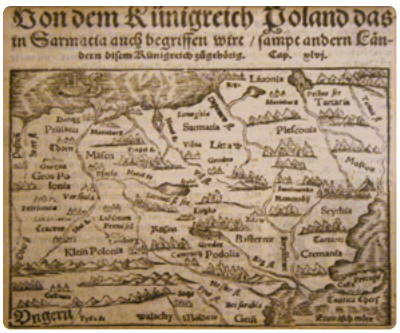
Bartosz Ziółko and Krzysztof Izdebski are two Poles who – especially for you – decided to carry this discussion in English. They spoke during the growing COVID-19 pandemic, which has caused more fatalities this year than Malaria. This pandemic also caused the closure of traditional medical practices based on physical contact due to a risk of COVID-19 infection. Because of communicating via smart devices, the topic of voice interfaces and technologies is getting a lot more attention. The main question of this chat is: What can we say about someone's medical condition based solely on voice analysis of the speaker?

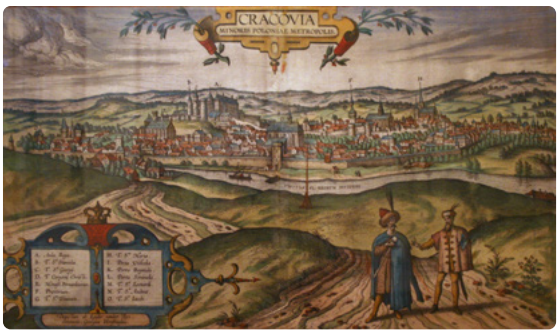
Krzysztof Izdebski [K.I.]: Dear Bartosz, I'm pleased and honored by your invitation to this chat. I trust it will be an informative and mutually productive conversation. Can you tell us first a bit about your VAMP project?

Bartosz Ziółko [B.Z.]: Thank you, Krzysztof. As you know, we're running the VAMP project – Voice Analysis for Medical Professionals. At the moment, we have software prototypes which detect arteriosclerosis with 83 % F-measure, and 71.5 % for prediabetes on our test patient database. This is why it has inspired me to ask you several questions. You are an authority in regarding voice and medicine, so could you say a bit more about yourself for those of us who may not know about who you are?

K.I.: Since you asked me, Bartosz, to introduce myself, here we go. Since my birth, I go by my first and last name, Krzysztof Izdebski, regardless of where I live, giving some linguistic headaches to more than a few people. My profession, Voice-Speech Patho-Physiologist, is a branch of Speech-Language Pathology (SLP). I am licensed by the State of California Board of Medical Quality Assurance to provide clinical services: diagnosis, treatment, and rehabilitation of voice and speech disorders. After graduation from the University of California at San Francisco (UCSF) as the 1st ever doctoral student in this field within the U.C. system. After graduation, I established at UCSF clinical and research Voice Lab, also one of the first-ever in the USA. I was hired as an Professor in the Dept. of Otolaryngology, Head and Neck Surgery at UCSF. Later, I held academic positions (either concurrently or sequentially) in similar departments at UC Davis, UCLA, and Stanford. I also serve on the advisory boards of Santa Clara University's Bioengineering Dept and at the Brain Plasticity Institute In San Francisco, CA. For nearly three decades, I am affiliated with the San Francisco Conservatory of Music, now as a Professor of Voice Science. For the last 30 years, I also conduct my private practice in S.F. For nearly three decades, I have been the Chairman of the Pacific Voice & Speech Foundation (PVSF) and the Vice President of the World Voice Consortium, with headquarters in Oporto, Portugal, EU. Between my academic-clinical practice and private practice, I had an opportunity to study, examine and treat nearly 30 000 voice/speech cases, ranging from transgender voice problems, all types of vocalists to all kinds of systemic, neurological, oncological, and functional dysphonia. My work leads to the world-wide establishment of diagnostic criteria for

dysphonia evaluation protocols. During the last 27 years, I organize yearly PVSF Conferences, held in the USA and abroad (three times in Poland).
My personal life gave me two children, a daughter, and a son. To stick with the idea of being Polish outside Poland, I collect Polonica, especially antique maps. I was active for nearly four years as a D.J. at the University of S.F. Radio station, promoting Polish issues via Polish music, musicians, and music on Polish themes by non-Polish composers. As a professional hobby, I produce vocal concerts called „One Larynx-Many Voices®," In the free time, I try to write other stuff than scientific papers.





B.Z.: What are your most significant professional successes?

K.I.: – A: Let me list first some clinical highlights. I am very proud of the close cooperation with a brilliant laryngeal and tracheal surgent, Prof. Herbert H. Dedo. Our work succeeded in decoding neuro-motor causation of laryngeal dystonia affecting voice and speech and establishing for the first time in over the century its effective treatment. This discovery resulted in a world-wide acceptance of our work, having a profound impact on treatment that resulted in improving the quality of life for thousands of patients. [1-2]
At UCSF, I also was a part of the research group working on larynx transplantation [3], and on understanding how pain is expressed in vocalization [4], and there I also studied changes in voice resulting from cochlear implantation. [5]
While working at the Veterans Affairs Hospital, I organized a multidisciplinary team that was very successful in voice-speech restoration in oncological and traumatic cases. [6]
– B: Contributing to detection and successful modeling of coding and decoding of emotions in the human voice. To learn more, look up three volumes entitled „Emotions in the Human Voice" that I edited. [7-9]
– C: Organizing close to the three decades international conferences on voice science, diagnosis, and treatment.
– D: Being awarded by the Voice Foundation (Philadelphia, USA), the highest award known as „STROBOSCARS" for my high-speed recordings of glottic and supraglottic activities in heavy metal singers. [10]

B.Z.: Krzysztof, how many languages do you speak?

K.I.: Few, some fluently, others so-so. But in reality, I can communicate with decreasing fluency in Polish, English, Swedish, Danish, Russian, Norwegian, German, Dutch, Flemish, Italian, Spanish, and few others. I also studied all classical languages, including Sanskrit, Arabic, Latin, Greek, Church Slavonic, Old Norse, etc.

B.Z.: What is more important in speech – words or voice?

K.I.: Both are equally important, but from different angles as both encode crucial information.

[1] Dedo, H.H, and IZDEBSKI, K. Surgical treatment of spastic dysphonia. Contemporary Surgery, 19, 75-90, 1981

[2] IZDEBSKI K, Dedo HH, Boles L. Spastic dysphonia: a patient profile of 200 cases. Am J Otolaryngol. 1984; 5: 7-14

[3] Anthony, JP, Trabulsy, PP, Mahdavian, M, Inoue, Y, Allen, DB, IZDESBKI, K, Mathes, SJ. Canine Larynx Transplantation: Long-Term Prevention of Rejection. in K. Izdebski (ed). Digest. Seventh Annual Pacific Voice Conference, Voice Conservation, Treatment and Restoration After Laryngeal Carcinoma. 36, Singular Publishing Group, San Diego, Ca, 1994.

[4] Levine, J., Feldmesser, M. Tecott, Gordon, N.C., IZDEBSKI, K. Pain-induced vocalization in the rat and its modification by pharmacological agents. Brain Research, 290: 307-315, 1984.

[5] Morgan, DE. Frattali CM, Bosone ZT, Cyr, DG, Hayes D, IZDEBSKI, K., et al. Electrical Stimulation for Cochlear Implant Selection and Rehabilitation. ASHA, Sup.7, vol 34:3, 13-16, 1992

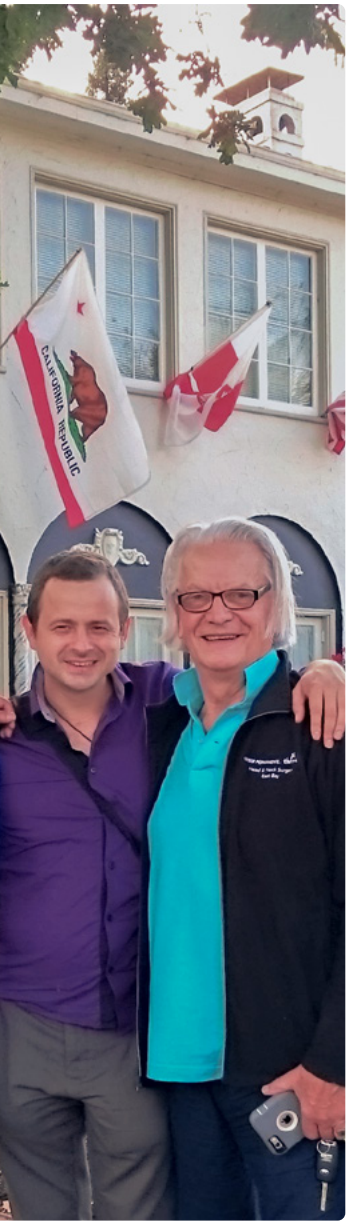
[6] IZDEBSKI K, Reed CG, Ross JC, Hilsinger RL Jr. Problems with tracheoesophageal fistula voice restoration in totally laryngectomized patients: a review of 95 cases. Arch Otolaryngol Head Neck Surg. 1994 Aug; 120(8):840-45

[7] IZDEBSKI, K. (Ed) Emotions in the Human Voice. Foundations, Vol 1. Plural Publishing, San Diego, London, 2008

[8] IZDEBSKI, K. (Ed) Emotions in the Human Voice. Clinical Evidence, Vol 2. Plural Publishing, San Diego, London, 2008

[9] IZDEBSKI, K. (Ed) Emotions in the Human Voice. Culture & Perception, Vol 3. Plural Publishing, San Diego, London, 2008

[10] https://www.insidescience.org/video/heavy-metal-singers-are-big-babies



B.Z. & K.I., Oakland, California, 2015

Speech-words reveal semantic elements. Voice reveals prosody. Both complement our perception and understanding of the entire message, though some research points out that we get bored very quickly if words are produced without prosodic elements. Why? Because prosodic vocal features, i.e., F0, dB, SNR, Jitter, Shimmer, Rate, etc., encode emotions. The process of listening is also accompanied by visual observation of bodily gestures that supplement meaning. Semantics/meaning is represented by speech articulation and features hidden in formant structures. Many politicians use these methods to mesmerize voters.

B.Z.: What information is hidden in the voice? What is it that Krzysztof Izdebski's ear can hear in the sound of a voice?

K.I.: Voice is specific to a single speaker, and it represents specific physiological features of the individual generated by the vibrations of the vocal folds (VF). These vibrations create audible pulses that form the laryngeal sound source and determine the pitch, loudness, and tone of the voice. These features can be modulated to express a variety of emotions. The human voice can also reveal the age and gender of the speaker.

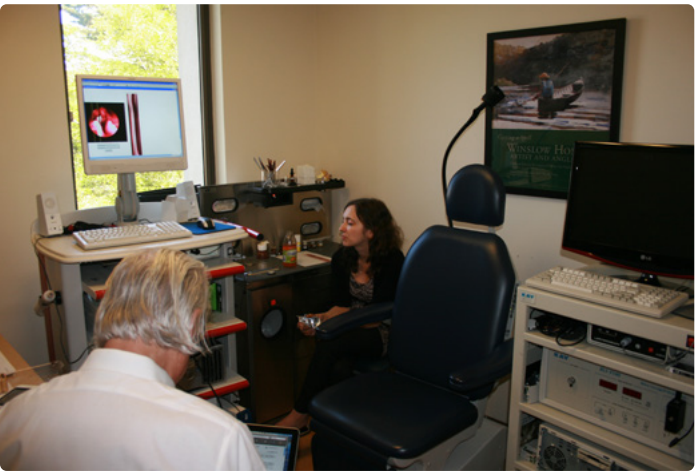
B.Z.: Is it possible that there are some correlations between health and voice quality, which we don't know about yet because the human ear can't hear them?

K.I.: Valid question, specifically for your project. So, let's construct a quick model of how to approach this challenge. You are saying, „we do not know because we may not hear due to ear limitations."
Hearing acuity is affected by a myriad of factors. But, in my opinion, even in an ideal situation, it is not the ear that is a culprit, although you may argue to the contrary. So, let's take this apparat.
The human ear is an acoustic, hydraulic, and neural filter, passing sound information to the auditory cortex for decoding. So, it is not the ear that hears but is the brain that decides what we understand by analyzing the incoming wave.
To distinguish the auditory signal's nuances, our brain must focus on tiny signal perturbations. One possibility is that this cannot be done by all listeners, as the general population has not trained the mind to do so. Detection of just noticeable differences in voice signals is frequent in classical singers and their teachers. By the way, I just learned (7-23-2020) that dogs are successful in differentiating between patients with and without COVID-19 from their odor alone. Can humans or AI match this task by analyzing the sound of the voice?
The second variable in this equation is the idea that both laryngeal and nonlaryngeal medical conditions will augment the voice. Voice changes due to laryngeal conditions are intuitive and logical, but correlations are not always straightforward.
Effects on voice quality by nonlaryngeal conditions is a fundamentally more complex issue.
So, we must ask, is this even theoretically possible. I do not have straightforward explanations, but theoretically, I tend to believe that vocal signal specificity due to systemic conditions is reasonable.
Here are some „pro" facts. Vocal folds (VF) are innervated bilaterally by recurrent laryngeal nerves (RLN) and superior laryngeal nerves (SLN) that determine the character of the mucosal wave. SLN has both efferent and afferent functions. RLN contains not only motor fibers but unmyelinated axons as well. Our combined research between my lab and Sweden has also documented the presence of chemoreceptors in the trunk [11].
These cells seemed to be identical to carotid body chemoreceptor cells and thus to indicate that the cells of the microglomus store catecholamines. High catecholamine levels in the blood are associated with stress, which can be induced from psychological reactions or environmental stressors. The recurrent laryngeal endoneurial microglomus may have a chemosensitive function influencing the laryngeal physiologic reflex mechanism.
Ultrastructurally, these cells seemed to be identical to carotid body chemoreceptor cells and thus

[11] Carlsoo, B., Dahlqvist, A., Domeji, S., Hellstrom, S., Dedo, H.H., and IZDEBSKI, K. Carotid-like body in the recurrent laryngeal nerve. An endoneurial chemosensitive micro-organ? Am J Otolaryngol 4: 334-341, 1983

to indicate that the cells of the microglomus store catecholamines. High catecholamine levels in the blood are associated with stress, which can be induced from psychological reactions or environmental stressors such as loud sounds, bright or intense light, or blood sugar levels. Or a perception of having a medical condition?

Nerve terminals impinging on the cells, classified as both afferent and efferent, were also identified. The recurrent laryngeal endoneurial microglomus may have a chemosensitive function influencing the laryngeal physiologic reflex mechanisms.



K.I. Clinic in San Francisco a short break between seeing patients, 2012

Note that the RLN is a branch of the X Cranial Nerve, the Vagus, the longest nerve in the human body that innervates multiple organs. Vagus innervates not only structures in the head, but also runs through the neck, thorax, and abdomen, supplying most of the visceral organs there. And Vagus is a mixed nerve containing efferent and afferent fibers. Hence this supports the possibility of modulating the vibratory pattern of the VF based on the information from the periphery.

To correlate voice quality with disorders of the VF without their visualization is very difficult, and often unrealistic correlations are produced. Acoustic biopsy, the project I am working on for the last ten years, is still not solved. This challenge has to do with the mechanics of the mucosal wave and the degree of the glottic gap during phonation and how the mucosal wave is affected by the type and location of the morphological nature of the problem within the VF cover. Meaning even small mucosal changes in the strategic location will produce severe dysphonia. At the same time, large lesions away from the anterior commissure and away from the vibratory edge (specifically the lower lip) may cause minimal or even no disruptions of the sound. In other words, glottis seems to behaves in accordance with a real estate law, meaning, its location determines the value of the property. Hence, the value of the sound is determined by the location and type of the vocal fold lesion.

Correlating systemic medical conditions (outside the laryngeal conditions) with specific voice quality is even more enigmatic. But a meta-analysis of data from well-diagnosed patients associated with age, gender, BMI, etc. may be the key. AI's help is needed, as AI works with ones and zeros, while the brain is not a computer. It is a neural network that operates with fractals. Recently, however, there seems to be an upsurge in the literature addressing this possibility, precisely concerning the pathophysiology of many systemic diseases and their effects on the voice. One recent publication states, „difficulty with voice is often the first manifestation of systemic disease." So, it looks that your project is getting more extensive attention. [12]

[12] Hamdan A-L, Sataloff RT, Hawkshaw MJ. Laryngeal Manifestations of Systemic Diseases, Plural Publishing. 2019, ISBN13: 978-1-63550-129-2

B.Z.: So, this is how you make a living. Apart from hospitals and universities for whom were you working for – musicians, courts, police, military?

K.I.: Private Practice, Conservatory of Music, as an expert witness in court cases involving injuries to voice and speech, as a medical examiner of injured workers concerning their compensation claims involving voice, speech, and swallowing losses. I do not work on forensic issues any longer, and recently I have been recruited to assess some aspects of health for candidates applying to officers' schools. [13]

[13] IZDEBSKI, K, Manace ED, Skiljo-Haris, J. The challenge of determining work-related voice/speech disabilities in California. In P.H. Dejonkere (Ed.) Occupational Voice – Care and Cure. Kugler Publ. The Hague, 149-154, 2001.

B.Z.: How has speech and voice analysis changed over the last ten years?

K.I.: Exponentially, and the progress is promising. It is clear that working with voice-speech, we need to cooperate with people representing knowledge that you and your team have.

B.Z.: From your professional point of view, can technologies like VAMP succeed?

K.I.: I am sure that the technology you are using for VAMP to succeed is forward-looking and scientifically elegant. In my opinion, the problem is with the quality of the material you acquire for analysis. To derive a specific profile of the medical condition from voice, a massive amount of data is needed. I listened to the statements of your medical collaborates regarding their „security" in making a final diagnosis based solely on auditory evaluation, and I am concerned about the overconfidence expressed. Other concerns include the detection of deep-fake voices as these can lead to fraudulent underwiring.

B.Z.: Should people be afraid of large IT systems that process the patient's medical data?

K.I.: Some people are afraid of ghosts, other of black cats, others of falling in love. Hence, it is natural that some fear IT interaction because of the idea that robots may be taking over or because of compromising privacy and the consequences that follow. Such fear, in my opinion, is universal but may be more prevalent in countries controlled more stirringly by the authorities.

B.Z.: Will job recruiters and insurance risk analytics start to use voice analysis?

K.I.: At this moment I am aware of such an application in some countries. From a business perspective using IT, AI to analyze voice may result in more accurate and faster decision making and apply to the process of underwriting insurance policies. Specifically, when concerning suspicious applications, fraud analysis, and verification of personal (identity) data.

B.Z.: What would you advise a student of medicine or a freshly graduated physician?

K.I.: Think, be compassionate, and think again before discussing condition and treatment options. Rely on intuition and in-depth knowledge. Use your brain in connection to AI, not the other way around.

B.Z.: Should medical technologies like VAMP be a part of the private or public sector?

K.I.: Here, meaning in the USA, such technologies are predominantly in private hands, but an oversight by federal or state governments is accepted and, in my opinion, relevant. There may be many government driven technological advances, but the authorities tend to keep such progress to themselves.

B.Z.: Do you think the order is one of the forms of chaos?

K.I.: This question requires competence beyond my capacity; however, the way I understand these concepts follows next.
Chaos connotes disambiguation, while the order is more like a binary relation. Order being conscious pattern vs. chaos. As I stated before, the brain is not a computer and does not operate in a binary fashion, instead of in fractals. The mind must make decisions to clear disambiguation (chaos). For example, take the Polish word „GŁOS" (dźwięk-sound) vs. „GŁOS" (wybory-a vote). The pronunciation is the same; the meaning quite different. So, I equate order with repeated patterns, while for me, chaos means more something like „evolution," describing partly random or chaotic phenomena. My time is up: Thank you for this opportunity to discuss your project.

Projekt VAMP
www.voicebiometrics.pl

VAMP Voice Diagnosis

Uniwersytet Medyczny w Białymstoku
www.umb.edu.pl

European Funds Smart Growth | Republic of Poland | The National Centre for Research and Development | European Union European Regional Development Fund