Computer Audition for Healthcare: Opportunities and Challenges

Kun Qian [1]*, Xiao Li [2], Haifeng Li [3], Shengchen Li [4], Wei Li [5], Zuoliang Ning [6], Shuai Yu [5], Limin Hou [7], Gang Tang [8], Jing Lu [9], Feng Li [10], Shufei Duan [11], Chengcheng Du [12], Yao Cheng [13], Yujun Wang [14], Lin Gan [15], Yoshiharu Yamamoto [1] and Björn W. Schuller [16,17,18]

[1] Educational Physiology Laboratory, The University of Tokyo, Tokyo, Japan, [2] Department of Neurology, Children's Hospital of Chongqing Medical University, Chongqing, China, [3] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, [4] Institute of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing, China, [5] School of Computer Science and Technology, Fudan University, Shanghai, China, [6] Shanghai Computer Music Association (SCMA), Shanghai, China, [7] School of Communication and Information Engineering, Shanghai University, Shanghai, China, [8] School of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing, China, [9] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China, [10] Department of Computer Science and Technology, Anhui University of Finance and Economics, Bengbu, China, [11] Department of Information and Computer, Taiyuan University of Technology, Taiyuan, China, [12] Ennova Health, Langfang, China, [13] Department of Music Technology, Shenyang Conservatory of Music, Shenyang, China, [14] Speech Group, AI Lab, AI Department, Xiaomi, Beijing, China, [15] School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin, China, [16] GLAM – Group on Language, Audio & Music, Imperial College London, London, United Kingdom, [17] Chair of Embedded Intelligence for Health Care and Wellbeing, Augsburg University, Augsburg, Germany, [18] audEERING GmbH, Gilching, Germany

1. INTRODUCTION

In the past decades, computer audition (CA), as an emerging interdisciplinary subject that includes acoustics, signal processing, machine learning, and deep learning technologies to provide computers with audio processing abilities similar to or even beyond human beings, has been increasingly studied for its applications in healthcare. Benefiting from its non-invasive characteristics, CA can facilitate both the clinical practice and home monitoring in almost every aspect of advanced intelligent medical systems like machine-listening-based diagnosis (1), mental disease screening (2), music therapy (3), and many others. On the one hand, fast development of the internet of things (IoT) and machine learning (ML) makes it easy to collect and analyze the health-related audio data using the most prevalent devices. On the other hand, even though the market/demand is great, CA for healthcare applications is still a young field compared to automatic speech recognition (ASR) (4) and music information retrieval (MIR) (5). To provide an overview of CA for healthcare concerns, **Figure 1** shows a word cloud generated by key topics related to CA for healthcare in the past two decades on Google Scholar.

A forum on future audio technologies for healthcare was organized at the Harbin Institute of Technology, which was held on 28 December 2019 during the 7th Conference on Sound and Music Technology (CSMT) in Harbin, P. R. China[1]. This forum and its summary report in this paper present the current consensus and opinions from a broad range of leading scientists from the expertise of audio technologies, mobile Health (mHealth), IoT, AI, smart wearables, cognitive sciences, neuro sciences, biomedical engineering, and clinical practice. The authors hope this discussion can be a good start for not only attracting more attentions for this promising interdisciplinary field but also for providing a venue for colleagues from multiple fields to understand where we are and the future trends in the development of CA for healthcare.

---

[1]http://www.csmcw-csmt.cn/

FIGURE 1 | Word cloud generated by the number of references (patents and citations excluded) related to the key topics in CA for healthcare (searched by Google Scholar, years 2000 to 2019).

2. CLINICAL DEMANDS AND BIG DATA

In clinical practice, demand is increasing for personalized and human-centered medical service. The cutting-edge technologies in ML and its subset, *deep learning* (DL) (6), are increasing the capabilities of CA to play an important role in medical applications. Moreover, it is even easier to capture big audio data from the increasingly prevalent sensor devices now used in daily life. The question of how to leverage the power of AI and Big Data to better the healthcare field is now attracting attention and global efforts.

3. NON-INVASIVE MHEALTH APPLICATIONS

Audio-based methods are marked by being cheap, convenient, and—most importantly—non-invasive. Whether analyzing the audio signals generated by the human body [e.g., snore sounds (1)], or using music for treatment of mental diseases (3), subjects have no need to be equipped with multiple sensors or even be burdened by invasive devices (e.g., endoscopy). Additionally, CA can make it feasible to collect data from subjects via mobile devices (e.g., a smartphone), which can provide the subjects 24×7 monitoring service.

4. DATA COLLECTION, ANNOTATION, AND PARTITION

Open access databases are crucial for a sustainable and reproducible research. However, CA for healthcare is lacking

in standard available public databases. Numerous works were presented by using private databases, which limits the comparability and objectivity of studies on algorithms and methods. In the future, collecting and releasing more publicly accessible databases needs to be a top priority. For a specific topic, the data acquiring system (equipment, environment, place) should be consistent, aiming to minimize the effects caused by humans. Unlike normal applications in other CA fields of application (e.g., speech recognition), healthcare-related projects need specific domain knowledge (e.g., medicine). Annotation of databases is another tough mission. On the one hand, there is a large amount of unlabeled data that can be easily collected by ubiquitous microphone devices. On the other hand, accurately labeled, clean, and high-quality data are rare. To address this (labeled) *data scarcity* issue, unsupervised learning (7), semi-supervised learning (8), active learning (9), and synthesis, such as by generative adversarial networks (GANs) (10), can be explored more in the healthcare area.

5. EVALUATION METRICS

Using suitable and reasonable evaluation metrics is necessary to guarantee a high-quality control progress in developing algorithms and methods. In healthcare applications, screening (such as binary classification of *normal* or *abnormal*) is the prerequisite for almost all cases. Thus, the widely used evaluation metrics in existing works are *accuracy*, *sensitivity*, *specificity*, and *precision*. However, data imbalance is a prevalent phenomenon in numerous healthcare applications. Moreover, multi-class classification and regression can be

more accurate than only screening in clinical practice. Hence, *unweighted average recall* (UAR) (11) is advised rather than the frequently used accuracy due to the latter potentially leading to over-optimistic conclusions. In addition, confusion matrices, the receiver operating characteristic curve, and derived measures such as the area under curve or equal error rate, can provide better insight into a model's performance.

6. FUNDAMENTAL RESEARCH

Fundamental research is always crucial and beneficial for CA applications in healthcare. In a classic ML paradigm, human hand-crafted features can be good indicators for researchers to know the relationship between the acoustical properties (in time and frequency domain) and the pathological symptoms. For instance, the popular large-scale acoustic feature extraction toolkit OPENSMILE (12) can provide thousands of well-designed features that can be used both for the statistical analysis and the ML model building. DL, as a hot sub-discipline of ML, is currently dominating most of the works in AI applications due to its powerful capacity to learn higher representations directly from big data. Particularly, deep end-to-end system can learn features themselves from raw audio data without any human domain knowledge (13). Nevertheless, it is difficult to build an explainable and responsible AI system by DL. In particular, finding the underlying mechanism of the pathological symptoms can never be neglected in any medicine-related subject. We believe that future work should be done by combing both the classic ML and DL methods. Understanding the fundamental knowledge is equally important for building strong enough models.

7. EFFICIENT COLLABORATION ACROSS MULTIPLE FIELDS

As indicated in (14), collaborations across fields of expertise can benefit both the computational scientists and the experimentalists for ML applied to life sciences and medicine. However, breaking the walls between each subject (e.g., medicine and engineering) is still something that needs doing. Experts from an engineering background may look more into the state-of-the-art technologies that can be used but pay less attention to the real clinical practice or subjects' requirement. Medical scientists usually have a stronger interest to uncover the pathology via the help of AI but show less passion for the mechanisms of ML methodologies. However, in order to achieve a major breakthrough, an efficient and thorough collaboration between all involved experts is a prerequisite. Specifically, CA for

healthcare needs even more fields involved, e.g., arts, education, and ethics.

8. INTELLECTUAL PROPERTY PROTECTION

The intellectual property (IP) protection is always essential for high-tech research and development. In particular, due to the interdisciplinary characteristic, IP protection cannot be well-implemented by a single field. On the side of experts from a medical background, the data itself should be fully considered as their IP. However, it should be encouraged to publicly release the data for scientific purposes. On the engineering end, the efforts toward developing algorithms, platforms, software, etc., should be valued.

9. DISCUSSION

We fully consider the aspects of clinical demands and big data, non-invasive mHealth applications, data collection, evaluation metrics, fundamental research, efficient interdisciplinary collaboration, and IP protection. We believe that, by reading this brief opinion piece, readers can gain a clear insight on where we are and what we can do in the future by CA for healthcare in this often under catered for area of artificial intelligence (AI). In summary, CA for healthcare is a young and promising field that needs tremendous collaboration across different fields. Future work should aim at the development of non-invasive clinical apparatus, in-home health monitoring system, and personal precision treatment service.

AUTHOR CONTRIBUTIONS

All the co-authors contributed to this work. KQ chaired the forum during CSMT 2019 and organized the writing work of this paper. KQ, YY, and BS conducted this summary and co-wrote this paper. XL, HL, SL, WL, ZN, SY, LH, GT, JL, FL, SD, CD, YC, YW, and LG actively discussed and participated in this work.

FUNDING

REFERENCES

1. Qian K, Janott C, Pandit V, Zhang Z, Heiser C, Hohenhorst W, et al. Classification of the excitation location of snore sounds in the upper airway by acoustic multi-feature analysis. *IEEE*

Trans Biomed Eng. (2017) 64:1731–41. doi: 10.1109/TBME.2016.26 19675

2. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun.* (2015) 71:10–49. doi: 10.1016/j.specom.2015.03.004

3. Hahna ND, Hadley S, Miller VH, Bonaventura M. Music technology usage in music therapy: a survey of practice. *Arts Psychother.* (2012) 39:456–64. doi: 10.1016/j.aip.2012.08.001

4. Benzeghiba M, De Mori R, Deroo O, Dupont S, Erbes T, Jouvet D, et al. Automatic speech recognition and speech variability: a review. *Speech Commun.* (2007) 49:763–86. doi: 10.1016/j.specom.2007.02.006

5. Futrelle J, Downie JS. Interdisciplinary research issues in music information retrieval: ISMIR 2000–2002. *J N Music Res.* (2003) 32:121–31. doi: 10.1076/jnmr.32.2.121.16740

6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539

7. Barlow HB. Unsupervised learning. *Neural Comput.* (1989) 1:295–311.

8. Chapelle O, Schölkopf B, Zien A. *Semi-Supervised Learning.* Cambridge, MA: MIT Press (2006).

9. Settles B. Active Learning. *Synthes Lect Artifi Intell Mach Learn.* (2012) 6:1–114.

10. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Proc. NIPS.* Montreal, QC (2014). p. 2672–80. Available online at: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

11. Schuller B, Steidl S, Batliner A. The Interspeech 2009 emotion challenge. In: *Proc. Interspeech.* Brighton (2009). p. 312–5. Available online at: https://www.isca-speech.org/archive/archive_papers/interspeech_2009/papers/i09_0312.pdf

12. Eyben F, Weninger F, Gross F, Schuller B. Recent developments in openSMILE, the Munich open-source multimedia feature extractor. In: *MM '13: Proceedings of the 21st ACM International Conference on Multimedia.* Barcelona (2013). p. 835–8. Available online at: https://dl.acm.org/doi/pdf/10.1145/2502081.2502224?download=true

13. Schmitt M, Schuller B. End-to-end audio classification with small datasets–making it work. In: *2019 27th European Signal Processing Conference (EUSIPCO).* A Coruña (2019). p. 1–5. Available online at: https://ieeexplore.ieee.org/abstract/document/8902712

14. Littmann M, Selig K, Cohen-Lavi L, Frank Y, Hönigschmid P, Kataka E, et al. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nat Mach Intell.* (2020) 2:18–24. doi: 10.1038/s42256-019-0139-8