An Overview on Audio, Signal, Speech, & Language Processing for COVID-19

Gauri Deshpande¹,², Björn W. Schuller²,³

¹TCS Research India, India  
²Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany  
³GLAM – Group on Language, Audio, & Music, Imperial College London, UK

gauri1.d@tcs.com, schuller@ieee.org

Abstract

Recently, there has been an increased attention towards innovating, enhancing, building, and deploying applications of speech signal processing for providing assistance and relief to human mankind from the Coronavirus (COVID-19) pandemic. Many ‘AI with speech’ initiatives are taken to combat with the present situation and also to create a safe and secure environment for the future. This paper summarises all these efforts taken by the research community towards helping the individuals and the society in the fight against COVID-19 over the past 3-4 months using speech signal processing. We also summarise the deep techniques used in this direction to come up with capable solutions in a short span of time. This paper further gives an overview of the contributions from non-speech modalities that may complement or serve as inspiration for audio and speech analysis. In addition, we discuss our observations with respect to solution usability, challenges, and the significant technology achievements.

Index Terms: COVID-19, digital health, audio processing, computational paralinguistics, affective computing

1. Introduction

As of today, there are more than 4 million confirmed cases of coronavirus-induced COVID-19 infection cases in more than 200 countries across the world. This led to increased need for screening, diagnostics, awareness, and post-care equipment and methods for a huge population. This not only demands for reaching out to a larger population in shorter time, but also to come up with effective solutions at each stage of the pathological, psychological, and behavioural problems. For the diagnosis purpose, the world health organisation (WHO)¹ has described the COVID-19 key symptoms as high temperature, coughing, and breathing difficulties, and an expanded list of symptoms that includes chills, repeated shaking with chills, muscle pain, headache, sore throat, and a new loss of taste or smell. These expanded list of symptoms are found to appear after 2-14 days of the onset of virus. However, the Coronavirus is so new to the mankind, that the unique symptoms of the virus and human immunity towards it are still found to vary from person to person. There are some cases found where a non-infected individual with no symptoms can still be a carrier of the disease.

Due to such uncertainties, it is advisable not to completely depend on digital sensing to diagnose the virus, where many physicians are also concerned about the false alarms that AI based diagnosing would generate. However, looking at the huge number of infected individuals across the globe, detecting early indications or screening will help in reducing this count. Continuous self-monitoring by the suspicious individuals can also help in decelerating and eventually stopping the spread of the virus. Not only does identification and monitoring need digital assistance, but also the post trauma phase would need it.

As per the WHO department of mental health and substance abuse, the current scenario of COVID-19 is susceptible to elevate the rates of stress or anxiety among individuals. Especially, lock-down and social distancing, quarantine and its after effects in the society might have adverse effects on the levels of loneliness, depression, self-harm, or suicidal behaviour. A special attention is needed for the health care providers, having to face the trauma directly and spending long working hours in such scenarios. Both physical and mental health needs have increased and require AI to provide faster and easy to access solutions.

The authors of [1] have already discussed about the opportunities, possible use cases, and the challenges that remain in the space of speech and sound analysis by artificial intelligence to help in this scenario of COVID-19. After this analysis was done, much more work has progressed in this direction which we are summarising in this paper. As depicted in Figure 1, we are giving an overview of the speech based applications for COVID-19 in this paper. Table 1 explains past speech related work done to provide solutions for COVID-19 related health problems. This can be used as a ready reference by those who want to provide immediate solutions in this space. This table also describes the machine learning and deep learning technologies used on the given data-sets to provide mentioned accuracy. For each speech or audio application, there is a vast space in the literature. We have selected only a handful of them, considering their relevance in the COVID-19 situation. Not everything that can be developed, can be used in this scenario considering other factors such as social distancing and personal and environmental hygiene. Hence, the developments are required to be driven by guidance from the clinicians and health care providers.

Figure 1: Speech and Audio Applications for COVID-19

¹www.who.int

Table 1: Past speech related work done to provide solutions for COVID-19 related health problems.
2. Screening and Diagnosing for COVID-19

There exists a fine line between screening and diagnosing, where screening gives an early indication of the presence of a disease and diagnosing confirms the presence/absence of disease. Screening is probabilistic, whereas diagnosis is binary in nature. We will talk about different algorithms/applications using audio processing for screening and diagnosis of COVID-19.

2.1. Cough Detection

Cough detection is not only identifying the cough sound and differentiate it from other sounds such as speech and laughter but also, to identify COVID-19 specific cough. This requires to collect the COVID and non-COVID cough sound samples so as to develop an AI model that can differentiate between them on its own. For development of such a model, relevant data needs to be collected. Cough audio samples can be collected using a simple smartphone mic, hence, major efforts are seen in collecting it using smartphone mics such as done by the Carnegie Mellon University and the Cambridge University.

Both the universities have provided platforms for the general population to upload their cough sounds along with some additional information such as their age, gender, location, and if they had been tested positive. They would need these audio samples to build their machine learning algorithms for identifying COVID-19 cough. The Cambridge University’s web based platform asks the participants to also read a sentence so that their speech can also be recorded. Hence, their app considers the sounds of cough, breathing and voice. Coughvid is another such app from EPFL (Ecole Polytechnique Fdrale de Lausanne) to detect COVID-19 cough from other cough categories such as normal cold and seasonal allergies. ‘Breath for Science’ – a team of scientists from NYU –, have developed a web based portal to register the participants where they can enter similar details along with phone number. On pressing a ‘call me’ button, the participants receive a callback where they have to cough three times after the beep. With this, they plan to create a dataset of cough sounds for the research purpose. Another web interface ‘CoughAgainstCovid’ for COVID-19 cough sample collection is an initiative by Wadhwani AI group in collaboration with Stanford University.

There are others who are ready with COVID-19 cough identifiers such as a smartphone app described in [13], which detects the COVID-19 cough.

Table 1: Past speech and audio analysis related to COVID-19 health problems

| Application | Ref. | Technology | Dataset | Accuracy |
|-------------|------|------------|---------|----------|
| Tuberculosis cough detection | [2] | STFT, MFCC, MFB with CNN | Google audio set extracts from 1.8 million Youtube videos and Freesound audio database | .946 AUC |
| Asthmatic detection | [4] | INTERSPEECH 2013 Computational Paralinguistics Challenge baseline acoustic features | Speech from 47 asthmatic and 48 healthy controls | 78 % Accuracy |
| Avoiding speech recording while sensing cough and hence providing privacy attribute | [5] | PCA on audio spectrograms, FFT coefficients and RandomForest Classifier | Acted cough from 17 patients having cough due to common cold (n=8), asthma (n=3), allergies (n=1), and chronic cough (n=5) | True positive rate of 92 % and false positive rate of 0.5 % |
| Obstructive sleep apnea (OSA) detection from breathing sounds during speech | [6] | MFCCs with single layer neural network for breathing detection and MFCCs, energy, pitch, kurtosis and ZCR with SVM for OSA classification | 90 Male subjects’ speech and sleep quality measures using WatchPAT | Cohen’s kappa coefficient of 0.5 for breathing detection and 0.54 for OSA detection |
| Automatic speech breathing rate measurement | [7] | Cepstrogram and Support Vector Machine with radial basis function | Speech recordings of 16 participants of age group 21 years mean | 89 % F1 score and RMSE of 4.5 breaths/min for the speech-breathing rate |
| Breathing signal detection for conversational speech | [8] | Spectrogram with CNN and RNN | 20 healthy subjects’ speech recorded using microphone and breathing signal using two respiratory transducer belts | 91.2 % sensitivity for breath event detection and mean absolute error of 1.01 breaths per minute for breathing rate estimation |
| Stress detection from speech in situations such as car accident, domestic violence, situations close to death, and so on | [9] | LLD and functional features extracted using openSMILE with k-NN, SVM, and CNN classifiers | 31 emergency call recordings of the Integrated Rescue System of 112 emergency line from Czech Republic | Accuracy of 87.9 % with SVM and 87.5 % with CNN in classifying stress from neutral speech |

[1]https://www.covid-19-sounds.org/en/
[2]https://www.stanford.edu/
[3]https://www.coughagainstcovid.org/
[4]https://www.breatheforscience.com/
[5]https://coughvid.epfl.ch/
Although these algorithms are attaining high accuracy levels of above 90%, however, from the hygiene perspective it is not advised to cough on open surface. As explained in [18], the infection is primarily transmitted through large droplets generated during coughing and sneezing by symptomatic and also by asymptomatic people before onset of symptoms. These infected droplets can spread 1-2 m, deposit on surfaces and can remain viable for days in favourable atmospheric conditions but can be destroyed in a minute by common disinfectants. Hence, for the collection procedures, it is important to remind the participants to cover the mouth and then only provide the cough sound samples. Otherwise, this may result in further spreading of the disease. Also, after giving the samples, the smartphone surface should be applied with a disinfectant.

2.2. Breathing Analysis
As discussed before, shortness of breath is also one of the symptoms of the virus for which, the smartphone apps are designed to capture breathing patterns by recording their speech signal. Breathing pattern detection has found applications in spectrometry to analyse lung functionalities as well. The authors of [15] have attempted to correlate high quality speech signals captured using an Earthworks microphone M23 at 48 kHz with the breathing signal captured using two NeXus respiratory inductance plethysmography belts over the ribcage and abdomen to measure the changes in the cross-sectional area of the ribcage and abdomen at the sample rate of 2 kHz. They have achieved a correlation of 0.42 to the actual breathing signal and a breathing error rate of 5.6% and sensitivity of 0.88 for breath event detection. In the Breathing Sub-challenge of Interspeech 2020 ComParE [16], the authors have achieved a baseline Pearson’s correlation of 0.507 on a development, and 0.731 on the test dataset, respectively. They have used two piezoelectric respiratory belts for capturing breathing patterns. In addition to the speech signals, blow sound can also help in performing spirometry tests. The authors of [17] have developed a smartphone-based automatic disease classification application built around the spirometry tests.

2.3. Chat-Bots
As the count of positive COVID-19 cases are increasing every day, it brings up the need of automating the conversation a physician would have to understand the presence of symptoms in an individual. Microsoft and CDC have come up with a chatbot named “Clara” for initial screening of COVID-19 patients by asking them questions and capturing the responses. This uses speech recognition and speech synthesis technologies. The risk-assessment test is designed based on advice from the WHO and the Union Ministry of Health and Family Welfare India[8]. “Siri” from Apple is also updated to answer the variations of the general question, “Siri, do I have the coronavirus?” based on WHO guidelines. If a person shows severe symptoms, then it is advised by Siri to call 911. Similarly, Alexa is also updated with answering COVID-19 screening question based on CDC guidelines. Considering the trauma that the health care providers are going through, a web-based chat-bot named Symptoma developed in [18], is a significant step. It can differentiate 20000 diseases with an accuracy of more than 90% and can identify COVID-19 from other diseases having similar symptoms with an accuracy of 96.32%.

2.4. Chest X-Ray
Multiple efforts by several groups are put in the direction of developing a classifier to detect COVID-19 symptoms using chest X-Ray, such as that of in [19]. However, as stated in [21], a multinational consensus is that Imaging is indicated only for patients with worsening respiratory status. Hence, it is advised only for those patients who have moderate-severe clinical features and a high pre-test probability of disease. Hence, unlike Sections 2.1, 2.2 and 2.3 such measures are not suggestive for early identification and are preferred for diagnosis in a clinical setup.

As concluded by [22], the chest X-ray findings in COVID-19 patients were found to be peaked at 10-12 days from symptom onset. Also, it is still required to visit a well-equipped clinical facility for such approaches. On a positive note, in the presence of mobile X-Ray machines, this approach can help in speeding up the diagnosis. The authors of [23] have found from an experimental outcome that the chest X-Ray may be useful as a standard method for the rapid diagnosis of COVID-19 to optimise the management of patients. However, CT is still limited for identifying specific viruses and distinguishing between viruses.

3. Monitoring
The precautions for stopping the spread of the virus include social distancing, wearing a mask, and maintaining respiratory hygiene. Especially at public places, the concerned authorities can monitor the population to confirm that the precautionary measures are adopted by them. This section describes such monitoring tools developed for the COVID-19 scenario.

3.1. Face Mask Detection from Voice
As described in Section 2.1 collecting cough samples without covering mouth can lead to further spreading of the disease, mask wearing has to be mandated for donating a cough sample which requires an app to detect if the donor is wearing a mask or not. This year’s Interspeech 2020 ComParE challenge [16] features a mask detection sub-challenge, where the task is to recognise whether the speaker was recorded while wearing a mask or not. The results from this challenge will be useful to develop a voice monitoring tool which detects the mask or no-mask of a speaker. Also, these algorithms serve as a pre-step for other speech processing algorithms such as speech recognition, emotion recognition and cough detection.

3.2. Cough and Breathing Analysis
While at quarantine, the doctors need to monitor the cough history of patients, which can be done with a continuous cough monitoring device. After we cross the crisis and organisations think of resuming the operations, continuous monitoring of common spaces such as canteens and lobbies can be realised to record the COVID specific coughs. One such monitoring application is developed by the FluSense platform in [24]. It is a surveillance tool to detect influenza like illness from hospital waiting areas using cough sounds. Continuous monitoring and identification of abnormalities from the breathing rate has been done by [25] using image processing.

3.3. Mental Health – Emotion Detection
The disease spread has equally affected the physical and mental health of the individuals, which is primarily due to plenty
of mandatory precautionary measures such as social-distancing, work from home and the quarantine procedures which usually takes at-least 15 days for the patients to be alone. Also, the health care providers owing to their hectic routines followed by quarantine days are subject to undergo mental health issues. To cater for the growing need of addressing this issue, not only there is a high demand but the physical presence of the available psychologists is also missing. As found in [26], the COVID-19 pandemic has generated unprecedented demand for tele-health based clinical services. Among several initiatives taken against mental health issues such as stress, anxiety, and depression, we are yet to see these emotions being analysed from the speech signal during the COVID-19 period. This demands for the relevant data. Very recently, a study is conducted by the authors of [27] on the speech signal of COVID-19 diagnosed patients. The behavioural parameters detected from speech includes, sleep quality, fatigue, and anxiety considering corresponding self-reported measures as ground truth and have achieved an accuracy of 0.69. This year’s ComParE challenge at Interspeech 2020 [16] has an Elderly Emotion Detection sub-challenge, where the speech captured from elderly people has to be classified into Low, Medium, and High classes of Valence and Arousal. It is found that specific age groups such as that of elderly above 60 years are more prone to the infection, due to which this age group needs to follow the restrictions posed by the pandemic for a larger period in future as well. This shows that it will be crucial to understand the effects of this phase on their psychological parameters such as emotions.

3.4. Face Mask Detection from Image

One of the precaution measures while stepping outside home suggested by the WHO to reduce the chance of getting infected or spreading COVID-19 is to wear a mask. Also, the mathematical analysis presented in [28] suggests that wearing a mask can reduce community transmission and can help reduction in the peak death rate. Hence, similar to what has been outlined above, it is important that the concerned authorities keep a check on people wearing masks or not at public places, especially where the population is quite dense such as airports, railway stations, and hospitals.

The authors of [29] have provided the Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and the Simulated Masked Face Recognition Dataset (SMFRD) for the detection of masked and un-masked face using image processing techniques. They have achieved 95 % accuracy in recognising masked faces. An Apple store app developed by LeewayHertz can be integrated with the existing setup of an Internet Protocol (IP) camera for getting alerts on detecting no-mask on a face. The system provides a facility to add phone numbers for receiving the alerts and also mechanism to see the face not wearing a mask for the admin.

3.5. Text Analysis

The authors of [30] have used machine learning techniques to categorise sentiments from the responses to Press Ganey patient experience surveys. From this analysis, it is found that the patients have expressed negative comments about the cleanliness and logistics and have given positive comments about their interactions with clinicians. In [31], the twitter data with 5GCoronavirus hashtag is analysed to understand the drivers of the 5G COVID-19 conspiracy theory and strategies to help in reducing the spread of mis-information circulating in society.

4. Awareness

Speech recognition and synthesis algorithms have been widely used in the development of chat-bots to provide human like interactions. In this time of crisis, chat-bots are helping in spreading the valuable information about COVID-19 to end users. Once an infected gets cured, they can help researchers with not only their experience but also with donating components such as plasma. CDC has been encouraging recovered people to donate their plasma for development of blood related therapies. For collection of plasma, Microsoft has developed the chat-bot 9 which interacts with individuals to gather the required information such as, the duration for which they are tested negative, their age, weight, and also takes their pin-code to help them know their nearest donation center.

5. Next Steps and Challenges

Since the work culture is moving more towards working-from-home, it will be required to detect some behavioural parameters such as fatigue and sleepiness for the self monitoring of the working professionals. The behavioural parameters such as stress or anxiety need greater attention to be paid in these days. A major challenge given the social distancing norms, is getting the relevant and accurate speech data for developing machine learning models. Speech enabled chat-bots can play a significant role in this. There are other challenges as well from the design perspective of chat-bots. The authors of [32] have expressed both positive effect and drawbacks associated with using chatbots for sharing the latest information, encouraging desired health impacting behaviours, and reducing the psychological damage caused by fear and isolation. This shows that the design of chat-bots should be well thought of for using them, otherwise, they might have negative impact as well. An optimistic approach in these difficult times has been to work towards safe and secure environment for the post pandemic situation so that the society gain the trust and confidence back. This shows the need of accurate and reliable screening and monitoring measures at public places.

6. Conclusion

There is a vast space of physical and mental wellness which needs fast and usable solutions from the digital sensing space. Most of the speech based research works for screening and monitoring of COVID-19 are ready to be customised. Chat-bots, although having the largest possibilities of use-cases, are still in the need of context aware designs. As most of the first level symptoms are related to respiratory system, continuous monitoring of the patients using audio signal processing based techniques can assist the clinicians or health care providers in providing services while following social distancing norms. A multi-modal approach where all three modalities, audio and speech, text, and image together work towards creating a solution framework will be likely the most promising avenue of future efforts.

7. Acknowledgements

We would like to thank all researchers, health supporters, and others helping in this crisis. Our hearts are with those affected and their families and friends. We acknowledge funding from the German BMBF with ZIM grant No. 16KN069402 (Krun).

9https://covi19-plasmaalliance.org/
8. References

[1] B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng, and X. Li, “Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the sars-cov-2 coralisa,” arXiv preprint arXiv:2003.11177, 2020.

[2] I. D. Miranda, A. H. Diacon, and T. R. Niesler, “A comparative study of features for acoustic cough detection using deep architectures,” in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019, pp. 2601–2605.

[3] E. van Miltenburg, B. Timmermans, and L. Aroyo, “The vu sound database,” in Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16), 2016, pp. 2124–2130.

[4] S. Yadav, M. Keerthan, D. Gope, P. K. Ghosh et al., “Analysis of acoustic features for speech sound based classification of asthmatic and healthy subjects,” in ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 6789–6793.

[5] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Rungeval, M. Chetouani, F. Wenninger, F. Eyben, E. Marchi et al., “The interspeech 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism,” in Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyons, France, 2013.

[6] E. C. Larson, T. Lee, S. Liu, M. Rosenfeld, and S. N. Patel, “Accuracy and privacy preserving cough sensing using a low-cost microphone,” in Proceedings of the 11th international conference on Ubiquitous computing, 2011, pp. 375–384.

[7] R. M. Simply, E. Dafna, and Y. Zigal, “Obstructive sleep apnea (osa) classification using analysis of breathing sounds during speech,” in 2020 26th European Signal Processing Conference (EUSIPCO). IEEE, 2018, pp. 1132–1136.

[8] Y. J. Gan, L. Lim, and Y. K. Chong, “Validation study of watchPat 200 for diagnosis of osa in an asian cohort,” European Archives of Oto-Rhino-Laryngology, vol. 274, no. 3, pp. 1741–1745, 2017.

[9] A. Routray et al., “Automatic measurement of speech breathing rate,” in 2019 27th European Signal Processing Conference (EUSIPCO). IEEE, 2019, pp. 1–5.

[10] V. S. Nallanthigal, A. Härma, and H. Strik, “Deep sensing of breathing signal during conversational speech,” in Proceedings of the 20th Annual Conference of the International Speech Communication Association, Interspeech. Graz, Austria:[Sn], 2019, pp. 4110–4114.

[11] P. Partila, J. Tovarek, J. Rozhon, and J. Jalowiczor, “The interspeech 2015 computational paralinguistics challenge: Elderly emotion, breathing & masks,” Proceedings INTERSPEECH. Shanghai, China: ISCA, 2020.

[12] S. Trivedy, M. Goyal, P. R. Mohapatra, and A. Mukherjee, “Design and development of smartphone-enabled spirometer with a disease classification system using convolutional neural network,” IEEE Transactions on Instrumentation and Measurement, 2020.

[13] A. Martin, J. Nateqi, S. Gruarin, N. Munsch, I. Abdarrahmane, and B. Knapp, “An artificial intelligence-based first-line defence against covid-19: digitally screening citizens for risks via a chatbot,” bioRxiv, 2020.

[14] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Pla, F. Al Hossain, A. A. Lover, G. A. Corey, N. G. Reich, and T. Rahman, “Flusense: A contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 1, pp. 1–28, 2020.

[15] Y. Li and L. Xia, “Coronavirus disease 2019 (covid-19): role of chest ct in diagnosis and management,” American Journal of Roentgenology, pp. 1–7, 2020.

[16] F. Al Hossain, A. A. Lover, G. A. Corey, N. G. Reich, and T. Rahman, “Flusense: A contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 1, pp. 1–28, 2020.

[17] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei et al., “Masked face recognition dataset and application,” arXiv preprint arXiv:2003.09093, 2020.

[18] S. Gümüy, C. Daniels, and Z. Childers, “Using ai to understand the patient voice during the covid-19 pandemic,” NEJM Catalyst Innovations in Care Delivery, vol. 1, no. 2, 2020.

[19] M. A.hởmer, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, and B. Knapp, “An artificial intelligence-based first-line defence against covid-19: digitally screening citizens for risks via a chatbot,” bioRxiv, 2020.

[20] A. Martin, J. Nateqi, S. Gruarin, N. Munsch, I. Abdarrahmane, and B. Knapp, “An artificial intelligence-based first-line defence against covid-19: digitally screening citizens for risks via a chatbot,” bioRxiv, 2020.