COLIVE VOICE :

Identification of vocal biomarkers to monitor the health of people with a chronic disease.

PRINCIPAL INVESTIGATOR :

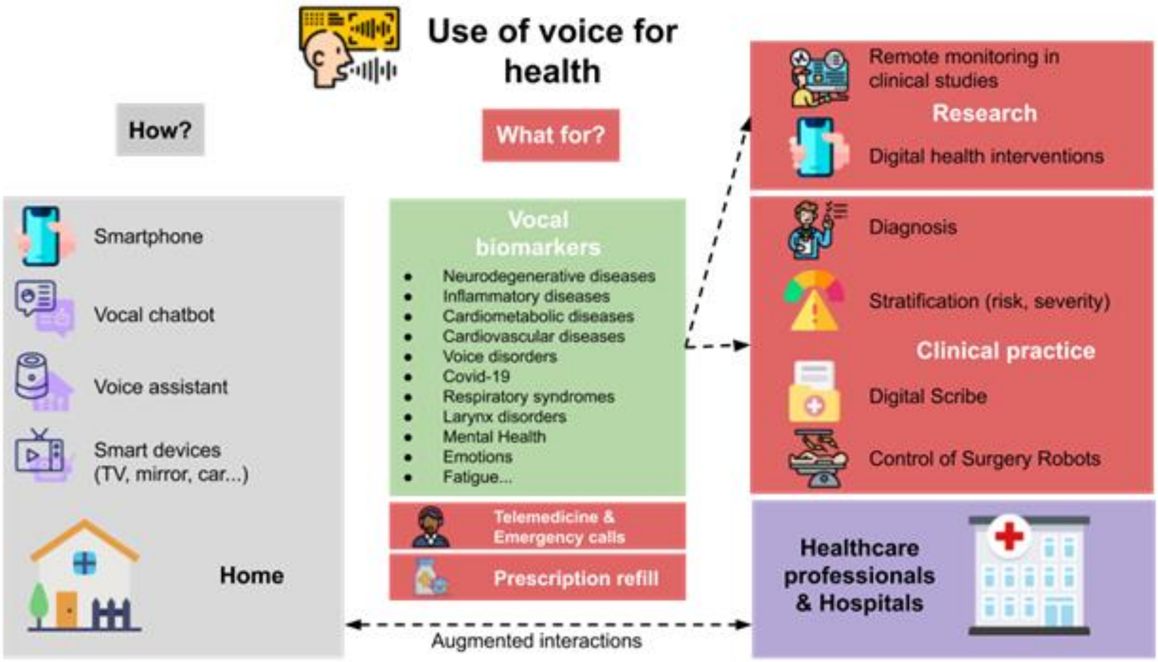
Dr Guy Fagherazzi
Director of Department of Population Health
Group Leader of the Deep Digital Phenotyping Research Unit
Luxembourg Institute of Health

RATIONALE

Human voice is a rich medium which serves as a primary source for communication between individuals. It is one of the most natural, energy-efficient ways of interacting with each other. And it is assumed that the use of voice and voice technologies (such as vocal assistants, smartphone vocal apps, digital scribes…) have the power to give back some humanity to an increasingly digitized healthcare environment. In parallel, technologies, artificial intelligence methods and computer sciences have led the way to new opportunities for the field of digital health, whose ultimate purpose is to ease the lives of people and healthcare professionals through the leverage of technologies. A vocal biomarker can be defined as a signature, a feature or a combination of features from the audio signal of the voice that is associated with a clinical outcome and can be used to monitor patients, diagnose a condition or grade the severity or the stages of a disease or for drug development (Kraus et al, 2018).

With the objective of using vocal biomarkers for diagnosis, risk prediction/stratification and remote monitoring of various clinical outcomes and symptoms, **there is a major need to develop surveys where audio data and clinical, epidemiological and patient-reported outcomes data are collected simultaneously.**

Figure 1. The use of voice for healthcare (adapted from Fagherazzi et al.)



PURPOSE AND OBJECTIVE

The objectives of CoLive Voice are:

1. To launch an international anonymized survey for English speakers where vocal recordings are associated with large validated clinical and epidemiological data, in the context of various chronic diseases or frequent health symptoms in the general population:

    - Multiple Sclerosis
    - Inflammatory Bowel Disease
    - Neurodegenerative diseases
    - Cancer
    - Diabetes
    - Allergy
    - Covid-19
    - Respiratory diseases
    - Stress
    - Anxiety
    - Depression

2. To extract audio features and train supervised machine learning models to identify key candidate vocal biomarkers of the aforementioned chronic conditions or related symptoms.

Those candidate biomarkers could be used in the future to predict disease severity, for diagnosis purposes or for remote symptom monitoring using digital technologies. However, at this stage, the main aim of this study is to identify these candidates and to study the feasibility of using voice to monitor health.

METHODOLOGY

The research field on vocal biomarkers is recent and most projects are currently developed in English. Indeed, vocal features are mainly impacted by the languages, accents, gender and age of the patients and so far validation of vocal biomarkers should be performed in each language. The Colive Voice project is based on an international anonymous survey available through a web-application. For this reason the survey has been designed in English first and will be disseminated worldwide to English Speakers.

We will simultaneously collect voice recordings and clinical, epidemiological and PROs data on the CoLive Voice web platform. Participation is on a voluntary basis, one shot, with no longitudinal follow-up planned, as there will be no way to identify and return to the participants.

Colive voice is a Luxembourgish initiative aiming to collect data from participants of several countries (since access to the survey will be possible through a web platform, potentially people from every country can participate).

People will first answer a detailed questionnaire on their health status (see details in section "Data collected") and then do 5 different voice records:

1) read a 30 sec prespecified text (from the Human Rights Declaration),

2) sustain voicing the vowel /aaaaaa/ as long and as steady as they can at a comfortable loudness

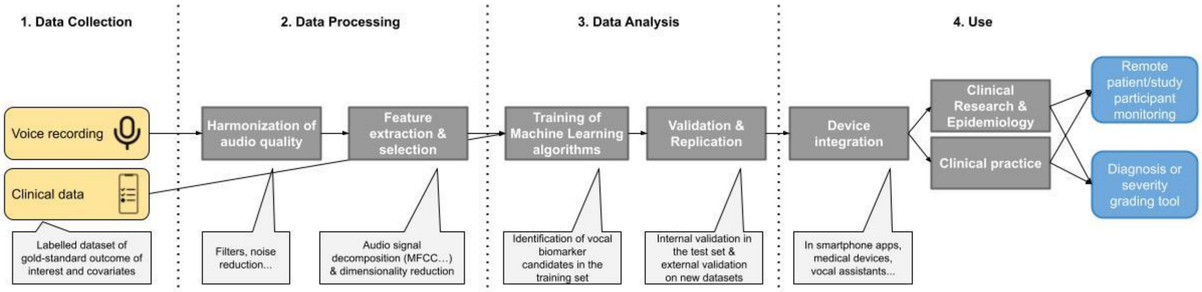
3) cough 3 times

4) breath in and out deeply 3 times

5) Count from 1 to 20 at a normal speed

Preprocessing steps are then necessary before analyzing the data. This includes steps such as resampling, normalization, noise reduction, framing and windowing the data as described in Figure 2 which represents the typical pathway to identify a vocal biomarker.

Figure 2. Pipeline for vocal biomarker identification, from research to practice



Prior to data analysis, there is a need to convert the audio signal into "features", meaning the most dominating and discriminating characteristics of a signal which will later contribute to the training of machine learning algorithms. Various methods are proposed in the literature3 to identify acoustic features from the temporal, frequency, cepstral, wavelet and time-frequency domain. The correct choice of features heavily depends on the voice disorder, disease or type of voice recording. For example, acoustic features extracted from sustained vowel phonations or diadochokinetic recordings are common in detection of Parkinson's disease, whereas linguistic features extracted from spontaneous or semi-spontaneous speech may be a more appropriate choice for mental health disorders.

Following the selection of features, machine or deep learning algorithms such as Support Vector Machines, Hidden Markov Models, convolutional or recurrent neural networks, just to name a few, can be trained to automatically predict or classify any clinical, medical or epidemiological outcome of interest, from vocal features alone or in combination with other health-related data. Algorithms are usually trained on one dataset (the "training set") and then tested on a separate dataset ("validation set").

The performance metrics of the different Machine Learning algorithms will be compared to find the best model. We will test different methods, namely: Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest and XGBoost. Furthermore, we will also train Deep learning algorithms (Convolutional Neural Networks) after converting the audio records into spectrograms and analyze the audio as images.

The detailed process for pre-processing, feature extraction and analysis is shown in figure 3.

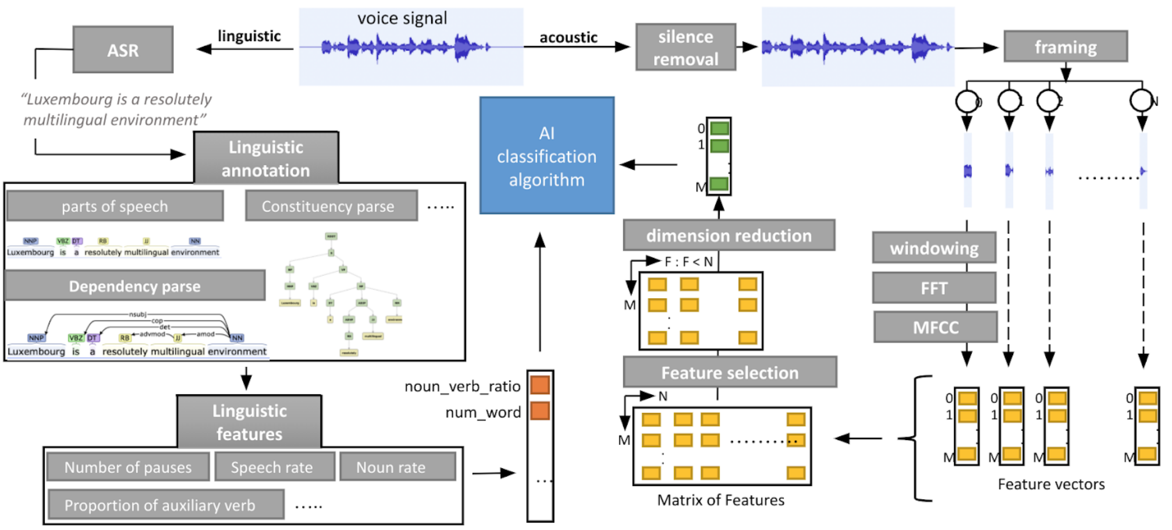


Figure 3. Vocal data pre-processing, feature extraction and analysis. Adapted from Fagherazzi et al. 2021

Representation of a typical voice signal pre-processing and linguistic and acoustic feature extraction. Voice signal represents the sound of the following sentence (e.g "Luxembourg is a resolutely multilingual environment"). ASR refers to Automatic Speech Recognition. Linguistic annotation includes part-of-speech, dependency and constituency parses and sense tagging. In this diagram, linguistic annotation is applied using tools like CoreNLP. Number of pauses, speech rate, noun rate are linguistic features and extracted using BlaBla package which is a clinical linguistic feature extraction tool. Acoustic features are extracted using Mel-Frequency Cepstral Coefficients (MFCC). Framing step refers to a signal segmentation into N samples. Windowing is multiplying of the signal sample by a window function like Hamming to minimize discontinuous signals that can cause noise in the subsequent Fast Fourier Transform (FFT) step. In this diagram, dimension reduction is represented by Principal Component Analysis method (PCA), reducing feature space to one dimensional vector.

STUDY PROCESS

The study is planned to start during the 1st trimester of 2021 and will run during up to 10 years.

People will be invited to participate through advertisements on social media, press releases, and a dedicated website will be developed for the information of the general public and potential participants.

The application will be available in English and accessible by following a link to the Colive Voice platform (no application to download through Android or Apple stores).

There will be no need to create an account, no login nor password and participants are expected to complete the survey only once.

An information and consent page will be implemented in the application and interested people will have to click on "Participate". Application will ask to access the device's microphone and participants will have to agree in order to perform the voice records.

Participation to the survey is "one shot" as mentioned in the Methodology part and consists in the completion of a detailed questionnaire on health status (see details in section "Data collected" and in the questionnaire detail in attachment) and in doing 5 different voice records as described in the Methodology and Data collected sections. Figures 3 and 4 show screenshot of the global process in the Colive Voice web application.

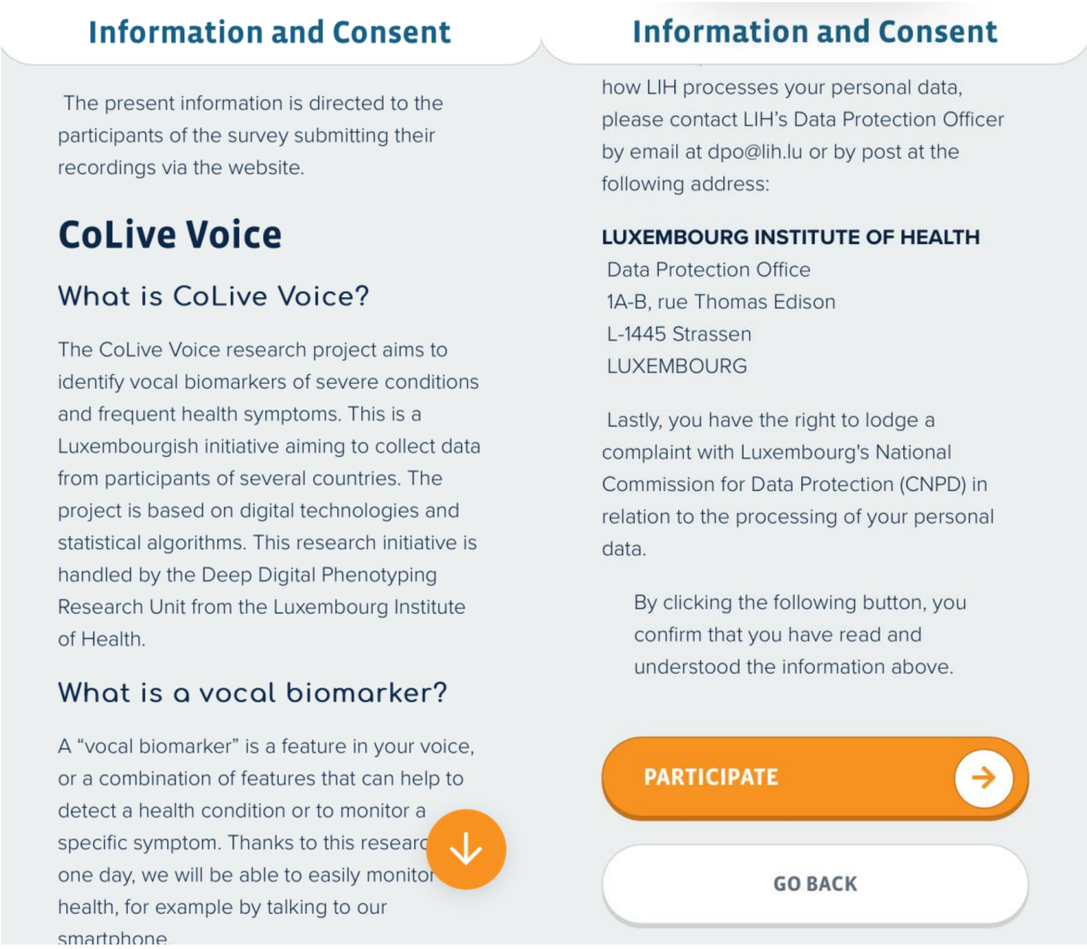


Figure 3: Screenshots of the consent page in the Colive Voice app

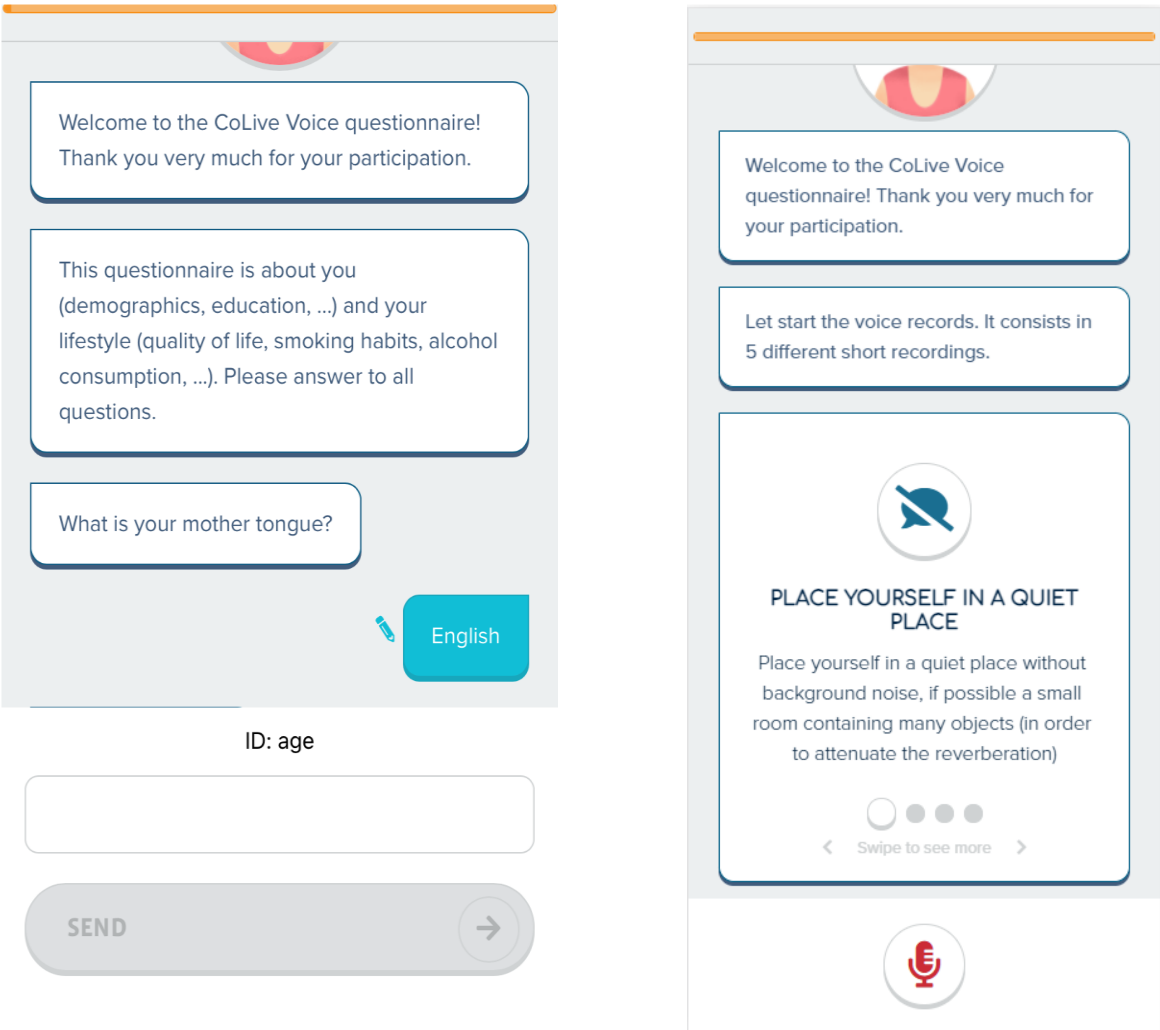


Figure 4: Screenshot of the first questionnaire page and on the onboarding page for voice records on the Colive Voice app

Adults and adolescents above 15 years, regardless of their health status, will be able to participate in the survey. The provided information has been written in a form understandable by people above 15 years.

We expect the participation of 50.000 people in the survey.

DATA COLLECTED

Questionnaires (see attached the detailed questionnaire):

- **Basic characteristics (**Mother tongue, age, gender, weight, lifestyle factors, quality of life, alcohol, smoking habits)

- **Symptoms** (stress, anxiety, constipation, pain, sleep disorders, respiratory quality of life, cough, fatigue, fever..)

- **Current treatments** (treatment for pain, cholesterol, diabetes, hypertension, anticoagulants, antidepressants, anti-reflux, hormonal treatments..)

- **Diseases**: communicable diseases (HIV, Covid-19, influenza, tuberculosis, malaria, Zika), chronic diseases (diabetes, CVD..), cancer, endocrine diseases, mental health (depression, stress..), neurological diseases..

- **Technical data (**Date and time of completion of the survey. IP addresses will be temporarily collected and converted by ipstack.com, an external service provider, whose only purpose is to convert the IP address into the country of origin. IP addresses will not be stored by ipstack.com and neither in the Colive Voice study database.

Voice recordings:

5 short voice recordings (read a 30 sec prespecified text (from the Human Rights Declaration), sustain voicing the vowel /aaaaaa/ as long and as steady as they can at a comfortable loudness, cough 3 times, breath in and out deeply 3 times, count from 1 to 20 at a normal speed)

DATA PRIVACY

Participation in the survey is anonymous: there will be NO collection of name, surname, birthdate, email or postal address and IP address will only be transiently used to determine the country of residence as described above. There will be no way to recontact the participants and no way to re-identify them even by crossing the data collected.

Data controller

The Luxembourg Institute of Health (LIH), having its registered address at 1A-B, rue Thomas Edison L-1445 Strassen, LUXEMBOURG, is Controller of the research project “Colive Voice” in the meaning of the European General Data Protection Regulation (GDPR). Indeed, voice records are considered as personal data by the GDPR.

This means that LIH is responsible for the collection, analysis and more generally for processing personal data and ensures their protection, in accordance with the GDPR and any subsequent text replacing or supplementing this text (in particular the law of August 1, 2018 on the organization of the National Commission for Data Protection and GDPR Implementation).

Purpose and legal basis for data processing

The use of personal data is necessary to enable us to achieve the aims of the study, which we are conducting in the public interest as part of our missions and for the purposes of scientific research (art. 6.1e and art. 9.2j of the GDPR).

Data sharing

The data collected in the frame of this project may be shared in the future for research purposes in similar health research areas but only in an anonymous form.

Data hosting

Data will be stored on Microsoft Azure cloud only, in servers located in Europe and we will apply a data separation principle (data from questionnaires and voice records will be stored on separate servers).

Data security and integrity

LIH takes appropriate security measures, depending on the sensitivity of the data concerned, to protect the data from the risk of unauthorised access, loss, fraudulent use, disclosure, modification and destruction. Collected data will be treated as strictly confidential.

Retention period

LIH will retain the data collected via the application for 15 years following the date of collection.

BIBLIOGRAPHY

1. Voice For Health: The Use Of Vocal Biomarkers And Voice Technologies From Research To Clinical Practice: A Review. Guy Fagherazzi, Aurélie Fischer, Muhannad Ismael, Vladimir Despotovic, PhD (UNDER REVIEW)
2. Kraus, Virginia B. 2018. “Biomarkers as Drug Development Tools: Discovery, Validation, Qualification and Use.” Nature Reviews. Rheumatology 14 (6): 354–62.
3. Sharma G, Umapathy K, Krishnan S. Trends in audio signal feature extraction methods. Applied Acoustics. 2020. p. 107020. doi:10.1016/j.apacoust.2019.107020