



SPEAKtoCOPD: a flashmob study to collect COPD speech

Loes van Bommel¹, Lauren G Reinders¹, Folkert Brijker², Bas Holverda³, Frits M.E. Franssen⁴,
Hanneke van Helvoort⁵, Visara Urovi⁶, Marieke Spreuwenberg⁷, Sami O. Simons⁴

¹Institute for Nutrition and Translational Research in Metabolism, Maastricht University, the Netherlands ²Spaarne Gasthuis, The Netherlands ³Longfonds, The Netherlands ⁴Maastricht University Medical Center, the Netherlands ⁵Radboud University Nijmegen, The Netherlands

⁶Institute of Data Science, Maastricht University, The Netherlands ⁷Care and Public Health Research Institute, Maastricht University, the Netherlands

{loes.vanbommel, lauren.reinders, v.urovi, m.spreuwenberg}@maastrichtuniversity.nl,
{s.simons, f.franssen}@mumc.nl, fbrijker@spaarne-gasthuis.nl, basholverda@longfonds.nl,
hanneke.vanhelvoort@radboudumc.nl

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is the most common chronic respiratory disease in the world. Fluctuations in symptoms frequently occur in COPD and might require additional treatments. Automatic monitoring or diagnosis of these fluctuations may aid the management of people with COPD. Speech analysis is a possible remote non-invasive method to achieve this. However, this analysis is disease and possibly language specific and the development of speech models requires high-quality and high-quantity data, which are often not available. In this study we introduce SPEAKtoCOPD: a dataset of Dutch speech collected with a flash mob study targeting people with a respiratory disease (specifically COPD). In this paper, we describe the flash mob methodology, our collected data, and perform an automatic quality check to assess whether the speech tasks have been performed correctly. We aim to publish our gathered data for researchers to improve speech analysis for respiratory health.

Index Terms: COPD, data acquisition, flash mob research, crowdsourced data

1. Introduction

Speech analysis is a novel non-invasive method to monitor or detect respiratory diseases remotely [1]. However, to develop models that can achieve these tasks with sufficient accuracy, high-quality and high-quantity data are needed. Pathological speech data is known to be low-resource and often not publicly available due to privacy concerns with medical data [2]. Especially for specific pathologies (i.e. respiratory diseases), in specific languages, data are scarce. We aim to address this lack of data by collecting speech from people living with and without chronic respiratory diseases through a website (www.SPEAKtoCOPD.com). In a short time frame, we have collected speech and self-reported medical and demographic data from as many participants as possible by campaigning on different (social) media channels and appointing so-called “ambassadors” to collect data in healthcare settings. This study design follows the flash mob research methodology.

1.1. Flash mob research

Flash mob research, which gets its name from the once-popular “flash mob” dances, consists of a single time point (or limited period) in which as many data are collected as possible. Typically, data are collected from multiple data acquisition sites. This methodology invites researchers to be quick in both acquisition and analysis of the data, in order to accelerate the entire

research process [3]. Previous examples of studies that utilized the flash mob data collection principle included 2005 participants in one day [4], 1850 participants in 50 hours [5], and 258 participants in 2 weeks [6], showing the different ranges of studies that fall under flash mob research.

Flash mob studies typically contain three main teams [3]: (1) the core research team who are responsible for setting up the study, the campaign strategy, and the subsequent data analysis; (2) the ambassadors, who coordinate the study in their local data acquisition site by spreading promotion materials and providing instructions for data collectors; (3) the “mob”, which are the participants themselves in our case. Moons [3] notes that a flash mob study needs to be simple, short, and easily executable by untrained participants. Finally, a certain “fun factor” will motivate people to join the study.

1.2. Novelty

In this paper, we introduce the SPEAKtoCOPD dataset which was gathered using the flash mob principle of data collection. This novel dataset contains four speech tasks from predominantly Dutch participants living with or without a respiratory disease. It contains self-reported information from the participants on demographic group and health status. Such a dataset is to our best knowledge not yet available and paves the way for more research and improved speech analysis for this population.

2. Methods

2.1. Flash mob research

Our flash mob research was centered around COPD awareness month in November 2024 and had two main data acquisition moments, the first moment being the release of the patient magazine ‘Longkracht’ on the 4th of November and the second moment being World COPD Day 2024 on the 20th of November. The data acquisition time frame of our flash mob study was a month, from the 4th of November to the 4th of December 2024.

2.1.1. Research teams

The core research team consisted of the authors and colleagues of the Department of Respiratory Medicine of Maastricht University. The local ambassadors were mostly identified through respiratory and professional networks and consisted of nurse specialists and pulmonologists across the Netherlands. The “mob” were primarily people with COPD or another respiratory disease, but any adult interested in the study was able to participate. Most campaign efforts were made in the Nether-

lands, resulting in predominantly Dutch participants. A limited sample of participants across the globe were able to be included through efforts by international respiratory patient associations.

2.1.2. Data acquisition

The campaign strategy started well before November 2024 and was mainly handled by the core research team and chosen ambassadors. Presentations at NVALT (Dutch Association of Physicians in Chest Medicine and Tuberculosis) and CAHAG (Dutch General Practitioners Expert group on COPD and Asthma) paved the way for more small-scale collaboration with local healthcare providers. Other ambassadors were recruited through professional networks, social media and mouth-to-mouth promotion. The study sponsors, AstraZeneca and GSK, had internal and external promotion campaigns to further identify local ambassadors in other hospital settings. Patient associations, such as the Lung Foundation Netherlands, prompted the first data acquisition moment with the release of the ‘Longkracht’ article targeted at people living with respiratory diseases and their families¹. Media such as LinkedIn, WhatsApp, Facebook, the local radio station, and the hospitals’ intranets were also used to promote the study and recruit participants. This resulted in a second recruitment wave on World COPD Day itself, where both the core research team and local ambassadors went out in the field (hospitals) to recruit participants and spread awareness about the day.

2.2. Study website

To conduct the study and collect the data, a website was made for both mobile and desktop devices in six languages (English, Dutch, German, Spanish, French, and Polish). Accessibility and simplicity were the main goals of the site’s design so that participants could independently conclude the study. The site contains an informed consent form (ICF), around 20 demographic, health-related and symptom-related questions and four speech tasks. A full overview of the questions can be found on GitHub². Screenshots of the site can be seen in Figure 1. The site was largely inspired by the COVID-19 Sounds study³ [7].

To improve the usability of the website for our target audience, the design of the site and questions were tested with multiple “speaking out loud” interview sessions with people with COPD and respiratory health care professionals. During these sessions, they were asked to navigate the site while providing commentary. Following these interviews, the wording of questions was adapted and the font size was enlarged.

Instructions were given to the participants through the information folder available on the site and through an instruction video. In the video, the speech tasks were explained and a recommendation was given to use a mobile phone, keeping it 10 to 15 cm from the face, similar to Awan et al. [8]. Before the speech tasks, participants were given a dummy recording to assess the loudness of their voice and background noise. Human listeners are known to be the golden standard in speech quality assessment [9] and this method reduced data traffic and kept data processing to a minimum.

At the end of the study, the participant was shown a graph (see right screenshot of Figure 1) depicting their jitter values derived from speech task 1 (green) compared to jitter values from

¹<https://www.longfond.nl/nieuws/zeg-eens-aaa>
²<https://github.com/Loes5307/SPEAKtoCOPD/blob/main/questions.md>
³<https://www.covid-19-sounds.org/>

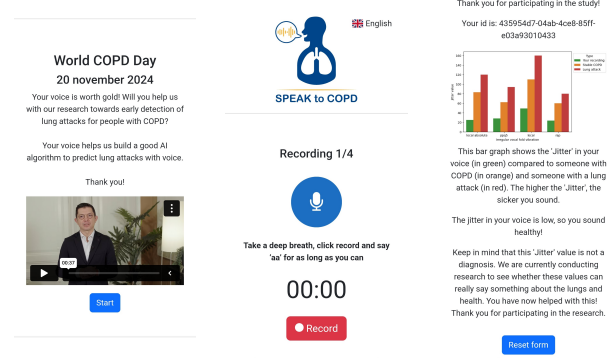


Figure 1: Screenshots of the SPEAKtoCOPD site showing the main page (left) and the first speech task (middle) and the resulting graph (right).

a speaker with COPD in stable state (orange) and exacerbated state (red). The site did not provide any medical advice to users and the graph was added to increase the “fun factor” and motivate people to participate.

2.2.1. Privacy by design

As speech data is identifiable, thus personal data, privacy was considered from the start of the study. No unique identifiable data was collected from the participants other than speech and a randomly generated ID we provided them. This unique ID allows the research team to delete, modify, or show each participant’s data post-study, in compliance with the GDPR, similar to the COVID-19 Sounds study³ [10]. Any data collected on the site was sent to the server for processing only after all tasks had been completed, allowing the participant to change their mind and quit the study without any of their data being stored or processed. A fully online ICF was designed with support of university GDPR experts and requires signing before participation in the study, making the legal basis for data processing explicit informed consent from the participants.

The Ethics Review Board evaluated the study as not applying to the Medical Device Regulation or the Medical Research Involving Human Subjects ACT by the Maastricht Medical Ethical Committee, trial number 2024-0348. The acquired data were stored and processed at Maastricht University and not shared with sponsors.

2.2.2. Speech tasks

The speech tasks on the website were fourfold.

1. Sustained vowel /a/. Instructions: “Take a deep breath, click record and say ‘aa’ for as long as you can”
2. Sustained vowel /u/. Instructions: “Take a deep breath, click record and say ‘uu’ like in the word “wood” for as long as you can”
3. Repeat a sentence three times. Instructions: “Take a breath, click record and say “With my voice I help other people with a lung disease around the world. With my voice I help other people with a lung disease around the world. With my voice I help other people with a lung disease around the world.””
4. Repeat “helicopter” for 20 seconds. Instructions: “Take a breath, click record and repeat ‘helicopter’ for 20 seconds until the clock is on 0.”

Sustained vowels are a well-known task for speech analysis including respiratory disease diagnosis [1, 11]. The sentence

reading task, normally performed with a phonetically balanced sentence such as the Rainbow Passage [12], was adapted for the site to a more “fun” sentence to increase awareness of respiratory diseases. The helicopter repetition task has been used by Zeng et al. [13], suggesting the task’s relevance to respiratory health indicators for people with COPD. The speech task prompts have been translated into six languages, including suitable context words in each languages if needed (i.e. “douze” instead of “wood” for task 2 in French).

2.3. Post-hoc quality control

Since data collection was “in the wild”, quality control is necessary. Ad-hoc quality control as described by Ramani et al. [14] was impossible due to the minimal data processing (see section 2.2.1) and was performed post-hoc instead.

2.3.1. Preprocessing

First, the audio was downsampled to 16kHz. Then, we performed a trim to cut the silent start and end of each recording. For tasks 3 and 4 (the sentence task and repetition task), the audio was trimmed with a Voice Activity Detector (VAD) called SileroVAD [15], taking the ‘start’ of the first spoken segment and the ‘end’ of the final spoken segment. The final repetition of “helicopter” in task 3 was discarded. The sustained vowels (tasks 1 and 2) were trimmed with Librosa [16] and the standard parameters since manual inspection showed that the VAD did not perform well.

2.3.2. Assessing correctly performed tasks

For each speech task, different approaches were used to automatically assess whether the task was performed correctly.

First, any recording with a duration of <1 second was assessed as ‘incorrect’, where we assume people mislicked and ended the recording prematurely. Any recording with a Signal-to-Noise ratio (SNR, specifically WADA-SNR [17]) of <0 is also taken as ‘incorrect’, assuming there is too much background noise. These checks were made for all four tasks.

Additionally, an Automatic Speech Recognition (ASR) transcription was made for the sentence and repetition tasks (task 3 and 4). Each recording was transcribed with faster-whisper and the large-v3 model⁴, since this model performs well in Dutch and elderly speech according to Fuckner et al. [18] and the Dutch Open Speech Recognition Benchmark⁵. The Word Error Rate (WER) compared to the prompt was calculated compared and any WER>0 was assessed as ‘incorrect’. Task 3 had two ‘correct’ prompt options: the fully correct prompt where the sentence was spoken three times and the incomplete correct prompt where the sentence was spoken only once. The WER for both these prompt options was calculated and the task was assessed as ‘correct’ if either of them were 0.

Finally, we used the ‘spoken’ segments of the audio determined by the SileroVAD to determine the same measures as Ramani et al. [14] did, namely Speech duration, Number of speech segments, Speech rate, Number of pauses, Pause duration and proportion pause duration.

⁴<https://huggingface.co/Systran/faster-whisper-large-v3>

⁵https://opensource-spraakherkenning-nl.github.io/ASR_NL_results/

3. Results

3.1. Flash mob research

The final website was launched on 1-11-2024. While there was a continuous collection throughout November, most participants joined the study around the two days of interest (4 and 20 November), as shown in Figure 2. Note that for the ‘Longkracht’ release, the number of participants was delayed due to magazine delivery time. Most data were gathered on World COPD Day with 213 new participants within one day.

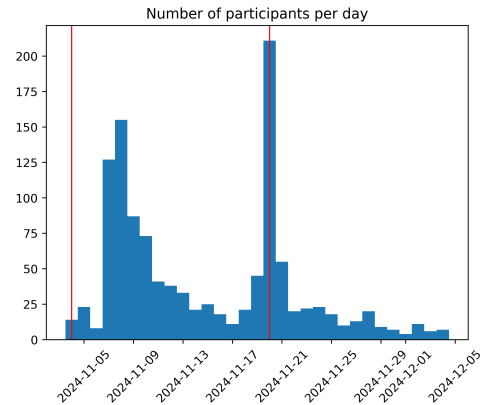


Figure 2: Number of participants per day. The red lines show the release of the ‘Longkracht’ magazine (left) and World COPD Day (right).

3.2. Data

Data from 1176 participants were gathered on their own devices. Smartphones (96%) were used in the majority of recordings, followed by desktop computers or laptops (3.8%) and one single iPad. The speech was captured in .wav files with a sampling rate of 48kHz and a bitrate of 1536 kbps. Table 1 shows an overview of the participants per demographic groups.

| | Female | Male | Total | Incorrect |
|---------------------------|--------|------|-------|-----------|
| Total | 772 | 404 | 1176 | 36% |
| Dutch first language | 738 | 391 | 1129 | 35% |
| Other first language | 34 | 13 | 47 | 57.4% |
| COPD | 283 | 204 | 487 | 45.4% |
| Other respiratory disease | 146 | 44 | 190 | 32.1% |
| No respiratory disease | 343 | 156 | 499 | 28.3% |
| 18-30 | 62 | 19 | 81 | 27.2% |
| 31-40 | 62 | 21 | 83 | 28.9% |
| 41-50 | 95 | 25 | 120 | 33.3% |
| 51-60 | 150 | 54 | 204 | 27% |
| 61-70 | 244 | 121 | 365 | 33.7% |
| 71-80 | 148 | 135 | 283 | 49.5% |
| 80+ | 11 | 29 | 40 | 55% |

Table 1: The number of participants in different demographic groups. % column depicts the percentage of participants in the demographic group who did not perform all four speech tasks fully correctly (>0 mistakes).

3.3. Quality

In total, 753 or 64% of the participants successfully performed all four speech tasks according to our metrics. The distribution of ‘incorrect’ performances can be found in Table 2. For task 3, 68 participants perfectly completed the task and repeated the sentence three times. 744 participants said the sentence perfectly, but only once, resulting in 364 ‘incorrect’ counts.

The most right column in Table 1 shows the percentages of participants who did not perform all four speech tasks correctly.

| | Task 1 | Task 2 | Task 3 | Task 4 |
|-------------------|--------|--------|--------|--------|
| duration ≤ 1 | 20 | 28 | 31 | 31 |
| SNR ≤ 0 | 13 | 25 | 12 | 13 |
| WER > 0 | - | - | 364 | 83 |
| Total 'incorrect' | 23 | 41 | 364 | 88 |

Table 2: The number of 'incorrectly' performed tasks per criteria from $N=1176$. Note that some recordings meet multiple 'incorrect' criteria.

| Task | Demographic group | Speech duration (s) | #Speech segments | Speech rate (segments/s) | #Pauses | Pause duration (s) | Proportion pause duration (%) |
|------|-------------------|---------------------|------------------|--------------------------|-----------|--------------------|-------------------------------|
| 1 | COPD | 12.2 (6.4) | 1.3 (1) | 0.3 (0.2) | 0.3 (1) | 0.8 (2.4) | 5.9 (16) |
| 1 | Other | 12.6 (6.6) | 1.2 (0.7) | 0.2 (0.1) | 0.2 (0.7) | 0.4 (1.6) | 3 (11.5) |
| 1 | Healthy | 16.9 (7.9) | 1.2 (0.8) | 0.2 (0.1) | 0.2 (0.8) | 0.6 (2.7) | 3.4 (13.4) |
| 2 | COPD | 12.7 (6.9) | 1.2 (1) | 0.3 (0.1) | 0.2 (1) | 0.4 (1.5) | 3.8 (13.3) |
| 2 | Other | 12.3 (6.5) | 1.2 (0.6) | 0.2 (0.1) | 0.2 (0.6) | 0.6 (2.4) | 4.1 (14.9) |
| 2 | Healthy | 17.2 (8) | 1.2 (1) | 0.2 (0.2) | 0.2 (1) | 0.5 (2.5) | 2.9 (12.8) |
| 3 | COPD | 15.7 (4.3) | 4.3 (2.1) | 0.3 (0.1) | 3.3 (2.1) | 1.8 (2.3) | 10.2 (6.7) |
| 3 | Other | 15 (2.7) | 3.7 (1.9) | 0.3 (0.1) | 2.7 (1.9) | 1.3 (1) | 8 (6) |
| 3 | Healthy | 14.4 (3) | 3.5 (1.7) | 0.3 (0.1) | 2.5 (1.7) | 1.1 (0.9) | 7.2 (4.9) |
| 4 | COPD | 16.7 (3.2) | 7.7 (4.9) | 0.5 (0.2) | 6.7 (4.9) | 2.9 (2.6) | 15.2 (13.9) |
| 4 | Other | 16.9 (3.2) | 7.5 (5.1) | 0.5 (0.2) | 6.5 (5.1) | 2.3 (2.2) | 11.9 (11.3) |
| 4 | Healthy | 17.6 (2.9) | 7.4 (5.7) | 0.5 (0.2) | 6.4 (5.7) | 2.1 (2.3) | 11.1 (11.9) |

Table 3: Timing metrics for three demographic groups for 'correct' recordings. N : COPD = 266, Other respiratory diseases = 129, Healthy = 358.

A subsequently performed χ^2 test showed a significant association between incorrectly performing the speech tasks and having a respiratory disease or belonging to an older age group ($p < 0.01$). Timing metrics are shown in Table 3, where we observe that speakers with COPD or other respiratory diseases have a shorter sustained vowel duration (task 1 and 2) and proportionally more pauses compared to healthy speakers.

3.4. Publication of data

From the 1176 participants, 424 participants were willing to share their data with researchers within and outside of the EU while an additional 550 participants agreed to share their data within the EU only. The dataset is currently still growing. Data sharing after the signing of a data sharing agreement and assessment by an access committee consisting of experts in medicine, data science, and data privacy, will be facilitated by the DELAD platform⁶. Please note that consent for data sharing by participants was given for research only, and data will not be shared for commercial purposes.

4. Discussion

We have, to our knowledge, gathered the largest dataset available for Dutch COPD speech with a flash mob research during COPD awareness month 2024. Through our multilevel data acquisition approach, the study reached the target population with the vast majority of participants being Dutch (96%) and living with a chronic respiratory disease (677 participants, 57.6%). By discussing the site design with end-users and prioritizing inclusivity we succeeded in designing the speech task to be understandable. The majority of our 1176 participants completed all four speech tasks correctly, with proportionally more older people and people with COPD making mistakes.

Healthy speakers had proportionally less pauses and a longer sustained vowel, as seen in Table 3. This trend is also observed by Ramani et al. [14], even with a different population and speech task. The observed difference in our data is most likely caused by a lower lung capacity translating to more effort (and less duration) when speaking aloud for the speakers with COPD and other respiratory diseases.

⁶<http://delad.net/>

4.1. Limitations

Due to the nature of the study being "in the wild" on participants's own devices without any headphones, the data is inherently noisy. The recording environment was not checked other than the participant's own assessment in the initial dummy recording. Previous literature like Awan et al. [8] suggests most modern smartphones are well-suited for data collection, and relatively little recordings had an SNR < 0 , as seen in Table 2, indeed suggesting the background noise is minimal.

The trade-off between the opportunity for a larger number of participants through a simple design and the accuracy of the data by verifying medical records was made. Thus, demographic and health status, including diagnosis, were self-reported. Respiratory symptom questionnaires were added to assess pulmonary health in all participants, but diagnoses like pre-COPD and others might be missing in the data.

The automatic quality assessment for the speech tasks is not infallible. Since the transcription did not work for the sustained vowel tasks, there might be speech other than the sustained vowel present in the recording. Furthermore, transcription might not be perfect and unfortunately it is known that ASR models perform worse for pathological and elderly speech. This could have added to the low number of 'correctly' performed recordings in task 3, as even 1 mispronounced or mistranscribed word will result in an 'incorrect' count.

4.2. Recommendations for future flash mob studies

While the flash mob study was successful, there are some recommendations for future flash mob studies. First, more elaborate interviewing and user testing before launching the site could improve user-friendliness and motivate more people to participate. While all participants were able to complete the study, elderly people and people with COPD were less likely to correctly complete all tasks, especially showing mistakes with the WER metric. A recommendation to mitigate this would be the option for participants to re-record their speech if they mis-spoke. Second, while it was not the main goal of the study, many participants enjoyed seeing the jitter graph on the site (see Figure 1 on the right) and we would recommend such visual feedback in future studies. However, this visual should be considered carefully since some participants were confused by this cluttered graph and at times even concerned when their own jitter values approached those of someone with a COPD exacerbation. Third, we noticed that the video instructions were a great help for many people in completing the tasks correctly. We would thus advise future studies to make an instruction video in either multiple languages or add subtitles. This improves accessibility for participants with different first languages, hearing abilities, and literacy levels. Notwithstanding these recommendations, we would like to echo the previous recommendations made by Moons [3], highlighting the simplicity of the study setup, the importance of ambassadors that can reach your target population and the "fun factor" to increase uptake.

In conclusion, this study shows that a large dataset of voice recordings can be quickly obtained through a flash mob study. Co-creation with end-users facilitates the uptake and the success of completing the speech tasks correctly, even with limited online instruction. To our knowledge, we collected the largest dataset of voice recordings from Dutch people living with COPD and other chronic respiratory conditions. This dataset contains enormously valuable information regarding voice and its clinical correlates, which will be made accessible to other researchers.

5. Acknowledgements

This work was supported by Lung Foundation Netherlands (grant number 7.3.22.013) and the Dutch Research Council (grant number KICHLZ03.21.023). We also thank MEMIC for their work on developing and hosting the website. The development of the website was supported by an unrestricted grant from AstraZeneca and GSK. Additionally, we would like to thank everyone who participated in the study as a participant, ambassador or otherwise.

6. References

- [1] P. Kapetanidis, F. Kalioras, C. Tsakonas, P. Tzamalidis, G. Kontogiannis, T. Karamanidou, T. G. Stavropoulos, and S. Nikolettas, "Respiratory diseases diagnosis using audio analysis and artificial intelligence: A systematic review," *Sensors*, vol. 24, no. 4, p. 1173, 2024.
- [2] S. A. Sheikh, M. Sahidullah, and I. Kodrasi, "Deep learning for pathological speech: A survey," *arXiv preprint arXiv:2501.03536*, 2025.
- [3] P. Moons, "Flash mob studies: a novel method to accelerate the research process," *European journal of cardiovascular nursing*, vol. 20, no. 2, pp. 175–178, 2021.
- [4] H. M. Wesselijs, E. S. Van Den Ende, J. Alsma, J. C. Ter Maaten, S. C. Schuit, P. M. Stassen, O. J. de Vries, K. H. Kaasjager, H. R. Haak, F. F. Van Doormaal *et al.*, "Quality and quantity of sleep and factors associated with sleep disturbance in hospitalized patients," *JAMA internal medicine*, vol. 178, no. 9, pp. 1201–1208, 2018.
- [5] E. S. van den Ende, B. Schouten, M. N. Kremers, T. Cooksley, C. P. Subbe, I. Weichert, L. S. van Galen, H. R. Haak, J. Kellett, J. Alsma *et al.*, "Understanding what matters most to patients in acute care in seven countries, using the flash mob study design," *BMC health services research*, vol. 21, no. 1, p. 474, 2021.
- [6] A. M. Schols, R. T. Willemsen, T. N. Bonten, M. H. Rutten, P. M. Stassen, B. L. Kietselaer, G.-J. Dinant, and J. W. Cals, "A nationwide flash-mob study for suspected acute coronary syndrome," *The Annals of Family Medicine*, vol. 17, no. 4, pp. 296–303, 2019.
- [7] T. Xia, D. Spathis, J. Ch, A. Grammenos, J. Han, A. Hasthanasombat, E. Bondareva, T. Dang, A. Floto, P. Cicuta *et al.*, "Covid-19 sounds: a large-scale audio dataset for digital respiratory screening," in *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 2)*, 2021.
- [8] S. N. Awan, M. A. Shaikh, J. A. Awan, I. Abdalla, K. O. Lim, and S. Misono, "Smartphone recordings are comparable to "gold standard" recordings for acoustic measurements of voice," *Journal of voice*, 2023.
- [9] P. C. Loizou, "Speech quality assessment," in *Multimedia analysis, processing and communications*. Springer, 2011, pp. 623–654.
- [10] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, and C. Mascolo, "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 3474–3484.
- [11] W. Mayr, A. Triantafyllopoulos, A. Batliner, B. W. Schuller, and T. M. Berghaus, "Assessing the clinical and functional status of copd patients using speech analysis during and after exacerbation," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 20, pp. 137–147, 2025.
- [12] A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Proceedings of the second joint 24th annual conference and the annual fall meeting of the biomedical engineering society [engineering in medicine and biology]*, vol. 1. IEEE, 2002, pp. 182–183.
- [13] B. Zeng, E. M. Williams, C. Owen, C. Zhang, S. K. Davies, K. Evans, and S.-R. Preudhomme, "Exploring the acoustic and prosodic features of a lung-function-sensitive repeated-word speech articulation test," *Frontiers in Psychology*, vol. 14, p. 1167902, 2023.
- [14] J. Ramani Haulcy, "Clac: a speech corpus of healthy english speakers," in *Proceedings of Interspeech*, vol. 2021, 2021, pp. 2966–2970.
- [15] S. Team, "Silero vad: pre-trained enterprise-grade voice activity detector (vad), number detector and language classifier," <https://github.com/snakers4/silero-vad>, 2024.
- [16] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *SciPy*, 2015, pp. 18–24.
- [17] C. Kim and R. M. Stern, "Robust signal-to-noise ratio estimation based on waveform amplitude distribution analysis," in *Proceedings of Interspeech*, 2008, pp. 2598–2601.
- [18] M. Fuckner, S. Horsman, P. Wiggers, and I. Janssen, "Uncovering bias in asr systems: Evaluating wav2vec2 and whisper for dutch speakers," in *2023 International Conference on Speech Technology and Human-Computer Dialogue (SpED)*. IEEE, 2023, pp. 146–151.