acoustics.ac.uk
THE UK ACOUSTICS NETWORK

UK Research and Innovation

INDUSTRIAL STRATEGY

PROJECTS FINAL REPORT

Call round: - 2

Project Title: An investigation of temporal variability in non-pathological speech: a pilot study towards a robust protocol for remote speech collection for psychological assessment

PI: Dr Nicholas Cummins	**Research Organisation:** Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London	
Department: Biostatistics & Health Informatics	**Start Date:** 1st April 2023	**Duration** 6 months
Cost of award (80%): £44,323.39	**Value of co-investment:** **In kind:** Dr Cummins, Dr Dineley and Dr Carr have also been part funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. Work in the project was supported by two interns from the King's Undergraduate Research Fellow and one intern from the Wellcome Trust's Biomedical Vacation Scholarship scheme. The project team were able to secure the use of test room for the duration of the project. The use of this room had to be secured from a separate department at KCL. We also utilised equipment bought in a previous grant. **Cash:** N/A	
Co-I and associated RO: **Co-I:** Dr Judith Dineley (ECR), Dr Ewan Carr, Dr Johnny Downs, Dr Faith Matcham, Professor Richard Dobson	**Acoustic Research themes:** speech, health, mobile health	

RO: Ms Samiksha Gurung , Research Grants & Contracts Associate (Pre Award), samiksha.gurung@kcl.ac.uk	

Collaborations & Partnerships involved in project: Tell us about bi-lateral or multi-lateral partnerships/participation by the PI or research team in a network, consortium, multi-centre study.

Throughout the duration of the grant, Dr Cummins and Dr Dineley have been actively involved in a Speech and Language Biomarker group organised by the Senseable Intelligence Group at the McGovern Institute for Brain Research, MIT & Harvard Medical School. Dr Dineley directly benefited from this networking by securing the group's lead, Assistant Professor Satrajit Ghosh, as a mentor on a recent fellowship application (see Outcomes/Impact). If the application is successful, Dr Dineley will undertake a placement with the group.

Dr Cummins and Dr Dineley have also been collaborating with Dr. Thomas Quatieri from the Human Health and Performance Systems Group at Lincoln Laboratory MIT. Dr Quatieri has advised the data processing, feature extraction and interpretation of results. Dr Quatieri will be a co-author on planned publications.

During the project, Dr Cummins has founded an international *Speech in Health Feature Standardisation Working Group* of academic and industry researchers working towards standardisation of speech features — both in collection and processing — for adoption in health research. Dr Dineley is a member of the group, who are currently drafting a position paper highlighting considerations and challenges in the area. Data collected in the grant will form a vital resource for work with this group. Through interactions and discussions with colleagues, Dr Cummins also identified the need for a Speech Processing for Health Journal Club which now has 30+ members from the UK, Europe and USA, including Dr Dineley and Dr Carr, and project's research assistant.

Project Partners N/A

Value and details of in-kind co-investment: -	**Value and details of cash co-investment: -**

Summary: A summary that can be published on our website (please do consider providing photo/images)

As acoustic signals, our voices are a valuable window into our mental health. They reflect not only our mood but also the functioning of our brain and its ability to coordinate the 100+ muscles needed to produce speech. Using recordings made on mobile devices, speech analysis could meet the large unmet need for convenient, objective tools that monitor mental health. However, speech analysis is not yet ready for use as a reliable clinical tool in the general population. In this context, how speech is recorded and which metrics to extract for processing are currently neglected steps in analysis pipelines.

Improving our understanding of the natural variability of an individual's voice, that is not caused by illness, and its impact on AI analyses of speech is one aspect that can facilitate translation. For example, could a croaky 'morning' voice hide an improvement in our mood? To begin to tackle this, we recorded 28 healthy participants speaking in the morning, afternoon and early evening and 26 participants speaking on a Monday, Wednesday and Friday at the same time.

In an initial analysis, we assessed variability of several characteristics of speech known to change with health state, using recordings of participants reading aloud a set text, *The Rainbow Passage*, on a studio microphone that does not apply any processing to the captured signal. This gives us an understanding of variability in an individual's speech over a day and a week, without the complicating effect of audio processing commonly found on mobile devices. This gives

us an understanding of variability in an individual's speech over a day and a week, combined with the inherent measurement uncertainty, without the complicating effect of audio processing commonly found on mobile devices.

We observed that the voices of our healthy volunteers do indeed vary over the day and week. In some characteristics, we could see clear systematic changes in participants' speech. For example, participants spoke faster with each session. This is in part likely due to practice effects; we plan to test the extent of practice effects by comparing speaking rate in recordings of other readings that participants did not have the opportunity to practice. We also observed that participants' pitch increased over the day. Other speech characteristics showed comparatively little variation over a day or a week, such as spectral gravity, which describes the relative distribution of frequencies in someone's speech, where a higher spectral gravity is perceived as a 'brighter' voice. Speech features with less variation over a day or week are beneficial as they are less dependent on the time of day or the day on which recording takes place.

An analysis of the performance of AI models assessing depression severity using speech features selected according to their sensitivity to within- and between-day variations in people's voices will follow.

This project was led by Dr Nicholas Cummins and Dr Judith Dineley, Department of Biostatistics and Health Informatics at the Institute of Psychiatry, Psychology & Neuroscience, King's College London.

Objectives:

1. Record healthy volunteers speaking in the morning, afternoon, and evening of a single day; and on three further days at the same time.

2. Assess the sensitivity of different speech features to variations within individuals' speech.

3. Assess the performance of AI analyses using speech features selected according to their sensitivity to within and between-day variations in speech.

Outcomes/Impact*: Please refer to stated objectives. What impact has this had on the Acoustics Sector? How are the results being applied? Please provide specific examples/evidence to support the provided statements.

Objective 1:

We completed this objective by recording two cohorts of 28 Day and 26 Week participants (see Appendix 1 for cohort details). To capture reliable data from these cohorts, we designed a unique data collection protocol (Appendix 2), which enabled us to collect a dataset with (i) acoustically rich and varied content; (ii) a core amount of fixed phonetic content to enable comparable analyses; (iii) minimal potentially confounding practice effects; and (iv) minimise (human-related) measurement errors.

Our recordings took place in a reverberation-insulated basement test room; a location chosen to minimise ambient noise. Recordings were conducted in specific time ranges for each session with a minimum number of hours between sessions in the 'Day' study and fixed days of Monday, Wednesday and Friday in the 'Week' study. We used six recording devices: an Audio Technica AT2020USB+ condenser microphone, an Apple iPhone 11, a Motorola G5 and G6, a Samsung Galaxy S20 FE 5G, and a Plantronics Office Headset. Our speech elicitation prompts included a reading of *The North Wind and the Sun* and *The Rainbow Passage* in each recording session, as well as an additional reading tasks in session 2 and 3 to minimise potential practice effects; a picture description task with the picture replaced each session, a held vowel vocal exercise in which participants were asked to sustain the vowels /a/, /o/ and /i/ three times in each session for a minimum of 5 seconds each.

Our summer interns and research assistant annotated each task recording from each different devices to enable feature extraction and subsequent analysis. The resulting dataset is unique (see *new research datasets*). Components will be made available to other speech researchers, increasing its impact.

Via questionnaires, we also collected basic sex, age, height and information on voice use habits and minor health issues relevant to speech so these could be controlled for in our analysis. Prior to each recording, we additionally collected information on voice use that day, sleep duration the night before and hours awake, minor health issues, hours since eating and drinking, plus responses to an emotion questionnaire. We also collected ethnicity and gender identity data, purely for reporting purposes.

We also distributed an end of participation questionnaire to get feedback on participants' experiences in the study and understand their views on speech-based health assessment using mobile devices. We received responses from 24 participants. This information will guide future study design.

Objective 2:

We have completed a preliminary analysis, generating descriptive statistics of normative feature values, intra-class correlation coefficients, standardised feature differences, and scatter plots comparing the first and subsequent recordings. See Appendix 3 for details of this work.

We conducted this work on participants' readings of *The Rainbow Passage* to examine speech with fixed phonetic content that required minimal training, recorded on our reference microphone (Audio Technica AT2020USB+ condenser microphone). Key observation from this analysis include:

i. Spectral Gravity and Jitter have minimal within-person variance (i.e., between recordings) for both suprasegmental- and vowel-level extraction, both over one day and over one week.
ii. Pitch and Shimmer, in particular, when extracted from /a/ vowel sounds (identified using forced alignment) demonstrated noticeably higher variability, both over one day and over one week.
iii. Timing measures (Speaking Rate and Pause Rate) showed larger standardised differences across a day or a week. We have used the same reading prompt for this analysis so these could be due to practise effects.
iv. Of the acoustic features, speaking Intensity (loudness) appears to be most consistently affected by recording time and day, with an increase with time of day and day of the week.

These results complement observations reported by Pierce et al (2021)[1], whose main conclusion from a study into speech variation across the day in a healthy female population is that speech typically varies over short periods of time, with morning voice having a measurable affect. Our results are consistent this conclusion on a more diverse cohort. Our results also support the main finding in Stegman et al (2020)[2] that, in both a healthy and ALS cohort, many speech features extracted using standard open-source pipelines have low repeatability scores. Both Pierce et al (2021)[1] and Stegman et al (2020)[2] did not consider speech captured on mobile devices. Our future work will, therefore, include expanding our analysis to include signals from our different mobile devices.

1: https://doi.org/10.1044/2021_JSLHR-21-00018
2: https://doi.org/10.1159/000511671

Objective 3:

Data collection, annotation and analysis for Objective 2 have taken longer than expected and we have yet to complete this part of our analysis. The project team are finalising a technique to rank features that is statistically valid and meaningful. Various aspects of the coding are complete for the AI analysis, we plan to submit this work to a future publication, see *Plans to Publish*.

Other Impacts:

Staff Training: Working on this pilot has allowed the project team to improve their project team's data collection protocols and analytics pipeline, and provided Dr Cummins and Dr Dineley in particular, with valuable project management experience. The grant gave invaluable experience four inspiring researchers. Dr Cummins and Dr Dineley hosted three undergraduates (see co-investment) who assisted in data annotation and analytics. They also trained a graduate research assistant, who is now an NIHR pre-doctoral fellow supervised by Dr Cummins and Dr Carr. All four were from groups underrepresented in data science and were Dr Dineley's first formal co-supervision roles.

***What activities have you undertaken to engage with research users, special interest groups and the general public to inform them about the research?**

Dr Cummins gave a UKAN organised presentation "The potential of smartphones voice recordings to monitor depression severity". He has also spoken about the project and presented initial findings at KCL events, most recently to a broader (non-technical) audience and the School of Mental Health & Psychological Sciences School Festival.

Dr Cummins and Dr Dineley both spoke at the inaugural meeting of the KCL and SLAM CAMHS Digital lab. They are also part of a cross-disciplinary working group planning infrastructure for technology-based research in the new Pears Maudsley Centre for Children and Young People. Insights from the pilot will feed directly into these efforts.

Dr Cummins has spoken on the project during an invite talks at mWELL, an international workshop on Affective

***Have any new research tools or methods been created or commissioned, if so, provide details:**

As mentioned previously the project team has been updating our analytics pipelines. This has included creating Parselmouth (a Python library for the Praat software) scripts of the extraction of speech features, and the use of Whisper and the Montreal Forced Alignment toolkit to generate word and phone level segmentation of all speech prompts.

Quality control checking on the extracted features allowed us to identify and mitigate weakness in our code. We will leverage this improved code for speech feature extraction in future projects**.** Plans are being made to release this code as part of the standardisation work (see *Collaborations & Partnerships*)

***Have any new research datasets, databases and models making, or potential to make, significant difference to your research (or that of others), been created, if so, provide details: -**

Recordings of healthy participants completing standard speech elicitation tasks, made on a condenser microphone, 3 smartphones and a consumer headset in a controlled, supervised environment:

(1) 28 participants recorded three times in one day
(2) 26 participants recorded on three days in one week, at the same time of day.

Computing for Mental Wellbeing. He will include results from the work in an upcoming keynote presentation at ASRU 2023.

Dr Dineley recently contributed to the Tomorrow's Engineers social media initiative.

Dr Cummins was involved in the UKAN+ EPSRC AI Hub Scoping Workshops and presented initial finds from the study at the UAKN AGM. He has volunteered to be a reviewer for UKAN + Call 3 and attended the recent *How to Review* seminars.

Both Dr Dineley and Dr Cummins have signed up the *Coach to 500K* course being offered through UKAN.

Using connections made through UKAN, Dr Dineley has named Dr Simone Graetzer as a mentor on a training grant application. See *Plans for follow-on activities/grants*.

34 participants consented to their recordings being made available to other non-commercial researchers subject to the completion and review of a data use agreement.

All 54 participants consented to the sharing of non-personally identifiable sex, age, height, voice use activity and speech features, all labelled only with participant study ID numbers.

This data will enable us to

- report normative speech feature values in our cohort
- describe intra-speaker feature variability
- assess the efficacy of standard feature extraction tools
- observe mobile device performance against a benchmark microphone

Data shared with other non-commercial researchers could be used in

- Speech technology development, such as automatic speech recognition

Conclusion: What is the primary outcome of this research?

Speech analysis has the potential to help meet large unmet needs for convenient objective measurement tools to monitor health remotely in research and clinical settings (e.g., [multilingual paper]). Such use-cases require speech collection and analysis to be much more robust and reliable. However, our preliminary analysis indicates that there are speech characteristics which are affected by simply *when* we choose to record speech, either at different times of day or on different days of the week. Our findings complement observations in the literature concerning the repeatability of commonly used speech features both in healthy[1] and pathological populations[2].Our results highlight the urgent need to develop evidence-based protocols for the collection of speech to help minimise the effects of know sources of variability, which could otherwise be erroneously attributed to a change in health state. Such a protocol must be co-designed with end-users. A lesson learnt from our study is the need to plan how to optimise participants' experience. For example, participants frequently reported feeling uncomfortable performing held vowels and instructing participants in this task was a learning curve for the project team, such that anecdotal evidence indicates participant experience improved over the course of the study.

Developing our analytic pipeline also highlighted vulnerabilities related to using automated tools; even utilising speech collected in highly controlled conditions on normative speech did not prevent output errors. This is due in part to the complexity of leading speech extraction toolkits which require considerable computer science knowledge to operate reliably, and inherent limitations in processes such as pitch estimation. Therefore, there is also a need for an (open source) extraction toolbox specifically designed for clinical research application and clinical use. Again, such as toolbox should be subject to participant input and co-designed with end-users.

We have two immediate research goals utilising the collected data. First, we need to extend our preliminary analysis to include the speech captured on our mobile devices. This is important as the use of consumer devices allows speech can be captured remotely, providing access to health information from people who would not otherwise come to timely clinical attention. However, smartphone specifications vary widely, and the cost of technology is an important intertwined factor here, with implications for both recording performance and equity of access. We also need to complete the work associated with Objective 3. Machine learning models in the speech-health domain typically use

combination of features (as opposed to single features). However, there is very little work assess the performance of these models in regard to the sensitivity of the individual features they comprise.

1: https://doi.org/10.1044/2021_JSLHR-21-00018
2: https://doi.org/10.1159/000511671

Plans for follow-on activities/grants: How are these results being used to further the area of research or its application in an industrial setting?

- **Replication Analysis:** Dr Dineley (PI) and Dr Cummins (Co-I) secured funding (£9,958) through an Institute of Physics and Engineering in Medicine (IPEM) Innovation Award for a replication study of the Day study (see Objectives) of the current UKAN award.
- **Grants submitted:** Insights and results from the pilot have been used in four grant submissions: (i) UKRI Future Leaders Fellowship (Dr Cummins, awaiting outcome); (ii) NIHR Development and Skills Enhancement Award (Dr Dineley, awaiting outcome); (iii) King's Health Partners: Developing translational research capabilities (reject, out of scope); and (iv) MRC - Funding for early-stage development of new healthcare interventions (awaiting outcome).

Weblink: (to the outcome of the project, the Open Access repository for the data[1], or press releases):

We are planning release of data and code to coincide with the publication of our planned publications. We will update UKAN when these links are available.

List of publications: In peer reviewed or non-peer reviewed literature. If no publications are available, what are the plans to publish? Please follow UKRI guidelines for Open Access **https://www.ukri.org/manage-your-award/publishing-your-research-findings/**

[Presented] Poster presentation on interim findings at the UKAN+ annual meeting, September 2023 (Attached, see Appendix 4)

[In progress] Publication on protocol, preliminary results and feasibility lessons learned from the pilot. Target Journal: JMIR Formative Research, submission mid-January 2024.

[Future] Publication combining pilot results with a replication analysis: Target Conference: Interspeech 2024, submission early-March 2024.

[Future] Publication on the impact of feature selection on machine learning model performance. Will utilise RADAR-CNS data as well as insights from previous [ref] and upcoming replication analysis. Target Journal: IEEE/ACM Transactions on Audio, Speech, and Language Processing, submission mid-April 2024.

[1] As a UKRI award holder you must follow their research data policies- https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open