Protocol for a proof-of-concept study of SpeechMate as an experimental speech support system

Birtan Demirel[1, 2], Timothy Denison[1, 2]

[1]Institute of Biomedical Engineering, Department of Engineering Sciences, University of Oxford, Oxford, United Kingdom

[2]Brain Network Dynamics Unit, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[*]Correspondence: birtan.demirel@eng.ox.ac.uk

Abstract

Background: Speaking in unison with an external voice can reduce disfluent syllables in people who stutter during structured speaking tasks, while preserving speech naturalness. In a recent exploratory study, participants spoke in unison with a conventional text to speech guiding voice while reading from paper during a simulated presentation. This reduced disfluency from baseline mean of 10% to below 3% in most participants. SpeechMate Presentation Mode builds on these findings as an experimental mobile speech support system that combines auditory guidance through bone conduction audio, augmented reality text display, and user-controlled pacing during public speaking. SpeechMate Conversational AI Mode extends the same synchronised speech support principle to semi-structured conversation by generating personally relevant replies that appear on augmented reality glasses and are played through bone conduction audio as a guiding voice. This provides an experimental platform for studying supported speech during naturalistic conversation.

Methods: This randomised, active controlled, blinded outcome rated, within participant crossover proof of concept study will include adults who stutter, adults with apraxia of speech, and typically fluent adults. In Experiment 1, adults who stutter will complete a controlled public speaking task in two conditions: baseline Teleprompter Mode (augmented reality text display without guiding voice) and SpeechMate Presentation Mode. They will then be asked to use SpeechMate during up to three self-selected speaking situations over one month and complete brief use logs. In Experiment 2, participants will complete semi-structured conversational scenarios with and without Conversational AI Mode. Typically fluent speakers will complete the same SpeechMate tasks as a reference group. Adults with apraxia of speech will complete an exploratory Teleprompter Mode task comparing original and artificial intelligence (AI) simplified text. The primary outcome in both experiments will be percentage of disfluent syllables. Some of the secondary outcomes will assess speech naturalness, willingness to speak in public, and logged use outside the laboratory.

Funding: This study is funded by the University of Oxford Medical and Life Sciences Translational Fund (MLSTF), supported by MRC IAA and EPSRC IAA funding streams. Development of the SpeechMate research programme is also supported by the Dominic Barker Trust.

Introduction

Stuttering is a complex speech fluency disorder with interacting neural and psychological components [1]. People who stutter (PWS) are at increased risks of developing social anxiety disorder[2–5]. The majority of PWS pursue speech therapy, with 90% in one survey and 87% having done so more than once[6]. In another study, 95% of adults entering therapy requested greater control over their stuttering, focusing on improving fluency, well-being, and understanding of their stuttering[7]. In our recent exploratory study, speaking in unison with a computer-generated voice using headphones during simulated presentations reduced stuttering from a baseline mean of 10% disfluent syllables to below the 3%[8]. In one participant, disfluency changed from 49 percent to 2 percent, speech naturalness ratings improved, and presentation time decreased from approximately 19 to 4 minutes. 83% of the 14 participants expressed interest in using this method in real-life presentations. Some, however, reported discomfort with speaking in unison with a robotic voice generated by a conventional text-to-speech (TTS) decoder, which lacked natural prosody, and flexible speech rate. These findings suggest that synchronised external speech may provide immediate speech support during structured public speaking, especially if delivered through a more personalised and usable guiding voice.

SpeechMate is an AI-assisted mobile system being evaluated as an experimental speech support tool for public speaking and semi-structured conversations. In Presentation Mode (Fig. 1), SpeechMate integrates four components into a portable system that can be adapted to individual communicative needs. First, it delivers personalised auditory guidance through bone conduction, using an adjustable, naturalistic AI-generated voice to facilitate synchronised speech during the study tasks. Second, it provides real time pace control, allowing the user to slow the cue when they anticipate difficulty or when they wish to regain synchrony with the guiding voice. Third, it uses augmented reality (AR) glasses as a teleprompter, projecting the spoken text into the user's visual field so that they can maintain a more natural posture and eye contact with listeners. Fourth, the system includes an exploratory text adaptation function that simplifies syntactic structure and replaces longer or more complex words with shorter alternatives where appropriate, to examine whether text complexity affects speech performance[9,10].

The second component, Conversational AI Mode (Fig. 2), addresses a gap that has limited prior research in stuttering. Stuttering is highly heterogeneous within and between individuals, and many PWS can speak fluently in private speech when alone[11], yet show disfluency when speaking to another agent with communicative intent. This suggests that the context in which fluency is practised can shape what transfers to daily life. In previous fluency training research, participants read the same script in synchrony with a researcher[12], which does not include communicative intent and turn-taking. This may

train fluency under a narrow situation and help explain why fluency improvements were significant for reading than for spontaneous conversation in the long term[12].

SpeechMate provides semi-structured conversational scenarios that preserve communicative intent while allowing synchronised speech support to be assessed during a conversation. To keep these conversations personally meaningful, SpeechMate builds a foundational user model from non-sensitive information such as general background information. This profile will then be used to generate responses that reflect the participant's background and self-description style, so synchronised speech support targets content the person is likely to use outside the laboratory.

We plan two experiments to evaluate SpeechMate. Experiment 1 will compare SpeechMate Presentation Mode with Teleprompter Mode in people who stutter. In Teleprompter Mode, participants use augmented reality glasses to view the speech text without auditory guidance. The main outcomes are fluency and speech naturalness. Typically fluent speakers will complete the same task as a reference group. We will also recruit a small exploratory subgroup of adults with apraxia of speech. In this group, we will use Teleprompter Mode to compare speech performance during original and AI-simplified text. The simplified version will reduce syntactic complexity, and word length where appropriate. Experiment 2 will extend the same SpeechMate framework to scenario-based dialogue tailored to each participant's background, comparing conversational blocks with and without Conversational AI Mode

SpeechMate system architecture and control mechanisms

Listing anticipated stuttering events: PWS can predict the words on which they are likely to stutter, from seconds before speech to weeks or months before a planned speaking event. SpeechMate uses this prediction by allowing users to mark words they expect to be difficult, including personally important words such as their own name, or words beginning with sounds that trigger stuttering for them. In Presentation Mode, these marked words will guide local changes in cue timing. The cue may slow briefly or pause before a marked word, giving the speaker more time to begin the word in synchrony with the guiding voice.

Real-time pace control: During SpeechMate guided speaking, participants can dynamically reduce the pace of the guiding voice when they anticipate a stuttering event or detect that they are falling out of synchrony. This can be done either by tapping the phone screen or by tilting the phone to the left. When tilt control is used, the phone accelerometer detects the leftward tilt and reduces playback speed by a user selected amount. When the phone returns to an upright position, the cue resumes its baseline pace. Advance marking of anticipated stuttering events and real-time slowing therefore form a unified control strategy for personalised auditory guidance.

Guiding voice selection: Participants will select a guiding voice before using SpeechMate. They will first choose broad voice preferences, including gender, accent, and voice qualities such as warm, calm, clear, bright, or confident. SpeechMate will then generate three candidate voices that match the selected preferences, among Microsoft Azure's AI voices. Participants will listen to each candidate voice using the same sample text and will adjust the speaking rate with a speed slider. They will then choose the voice that feels most comfortable and natural to follow during a presentation. Where this feature is used, a brief audio recording may also be collected to generate a cloned guiding voice for use during the study. The selected voice profile will be used across Presentation Mode and Conversational AI Mode. Voice selection is included mainly to make the auditory cue comfortable to follow and more predictable at sentence and word onsets.

Bone conduction audio: The guiding voice will be delivered through bone conduction earbuds (Shokz Opendotz One earbuds). This keeps the ears open to environmental sound and allows participants to hear their own speech clearly while following the guiding voice. It may also reduce the tendency to speak louder, which can occur when external audio masks self-monitoring, known as the Lombard effect[13].

AR teleprompter: Text is displayed on AR glasses (Even G1) as a teleprompter overlay that is dynamically aligned to the pacing of the guiding voice. This supports the natural speech and conversations by reducing the need to look down at paper or a screen.

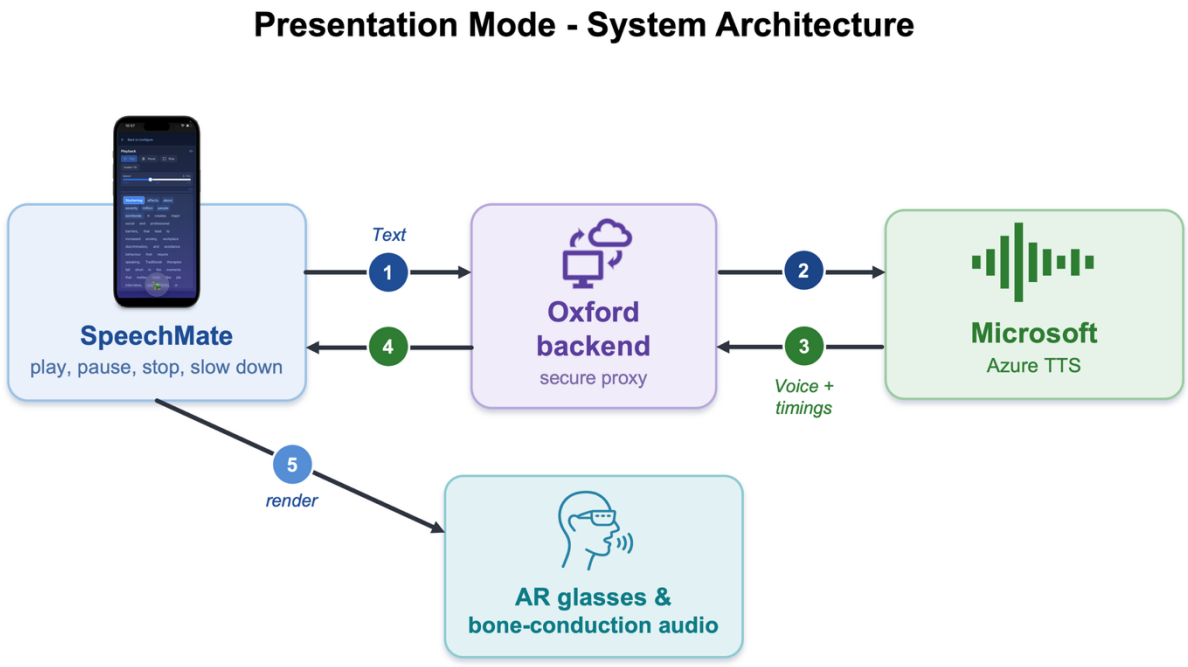

Figure 1 - Presentation Mode. SpeechMate runs as a browser-based app on the participant's smartphone and routes all requests through a secure backend hosted at the University of

Oxford. **(1)** The prepared script text is sent from the app to the Oxford backend. **(2)** The backend forwards the text to Microsoft Azure neural text-to-speech (TTS). **(3)** Azure returns the synthesised guiding voice together with real word-boundary timings. **(4)** The audio and timings are passed back to the app. **(5)** The app renders the script as a teleprompter overlay on the Even G1 AR glasses, time-aligned word-by-word to the guiding voice delivered through bone-conduction earbuds. Synthesis is performed once, ahead of speaking, so there is no in-speech processing lag. During the speech, the participant controls pacing locally with the options to play, pause, stop, and slow down. The real-time slow can work on demand by tapping the screen or tilting the phone, with playback returning to the chosen baseline rate afterwards. The guiding voice can also slow briefly before words marked as anticipated stuttering points.

Creating a foundational model of the participant: Before the conversational session, SpeechMate creates a foundational model of each participant using materials that the user provides. The model (OpenAI, GPT-4o mini), captures background information, interests, and self-descriptive phrasing, and is used to ground the conversational AI in the individual rather than in generic replies. For instance, in structured semi-scenarios such as job interviews, the system generates responses that are consistent with the user's own experience, vocabulary, and qualifications. This allows speech support to be evaluated using personally relevant content and may improve ecological validity.

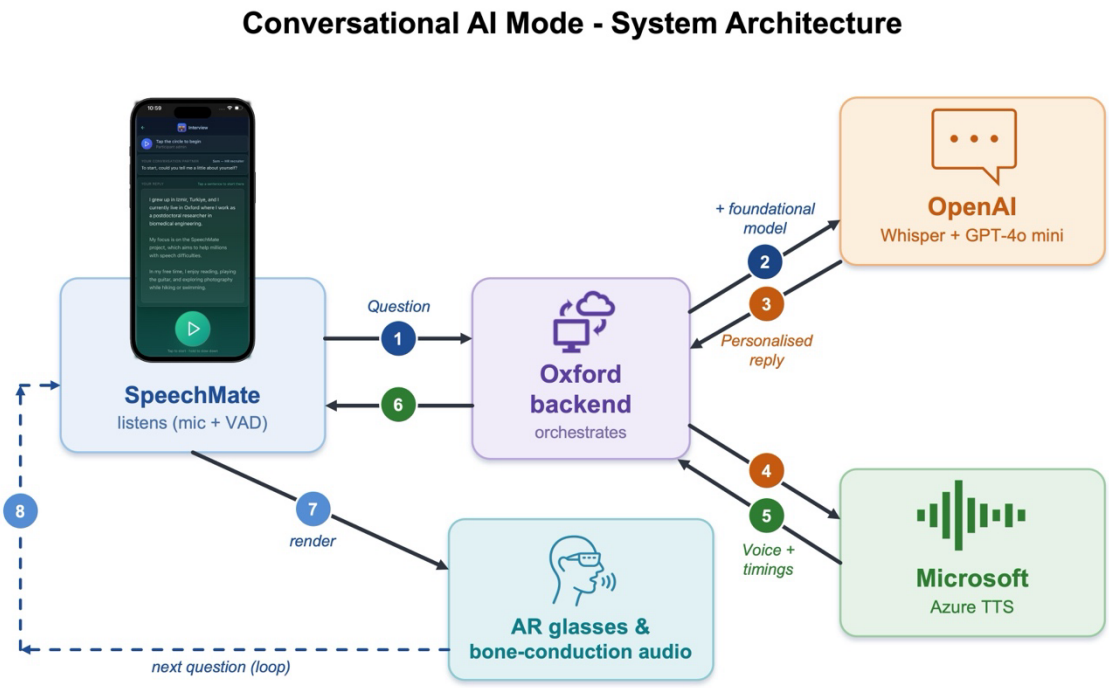

Figure 2 - Conversational AI Mode. (1) The conversation partner's speech is captured by the app, which detects end-of-turn with voice-activity detection (VAD), and the recorded audio is

sent to the Oxford backend. **(2)** The backend forwards it to OpenAI, where Whisper transcribes the speech and GPT-4o mini generates a personalised reply grounded in the participant's foundational model. **(3)** The reply text is returned to the backend. **(4)** The backend sends the reply text to Microsoft Azure neural text-to-speech (TTS). **(5)** Azure returns the synthesised guiding voice together with real word-boundary timings. **(6)** The audio and timings are passed back to the app, which **(7)** renders the turn as a teleprompter display on the Even G1 AR glasses, time-aligned to the guiding voice delivered through bone-conduction earbuds. **(8)** Control returns to the app, which listens for the next question, closing the loop. Replies are produced sentence-by-sentence: the backend streams each sentence from GPT-4o mini and synthesises it on Azure while the previous sentence is still playing, so speech begins before the full reply has been generated. The first audio is expected to begin approximately in 1.7s after the conversation partner stops speaking.

Tap controlled Conversational AI Mode: In Conversational AI Mode, the participant can switch between two forms of speech support from a single control on the smartphone. A single tap starts the guiding voice for a response generated by SpeechMate and displayed on the AR glasses. The audio is delivered through bone conduction earbuds, allowing the user to begin the response at a chosen moment and speak in synchrony with the cue.

If the reply provided by the AI is not preferred by the participant, a tap and hold activate a metronome cue. Metronome-paced speech can also reduce disfluency immediately, with the drawback of compromising speech naturalness[14]. In this option, the user speaks spontaneously and aligns their syllables to the metronome, with the pace set in SpeechMate settings. This option preserves participant control over speech content, while allowing the study to record any effect on speech naturalness.

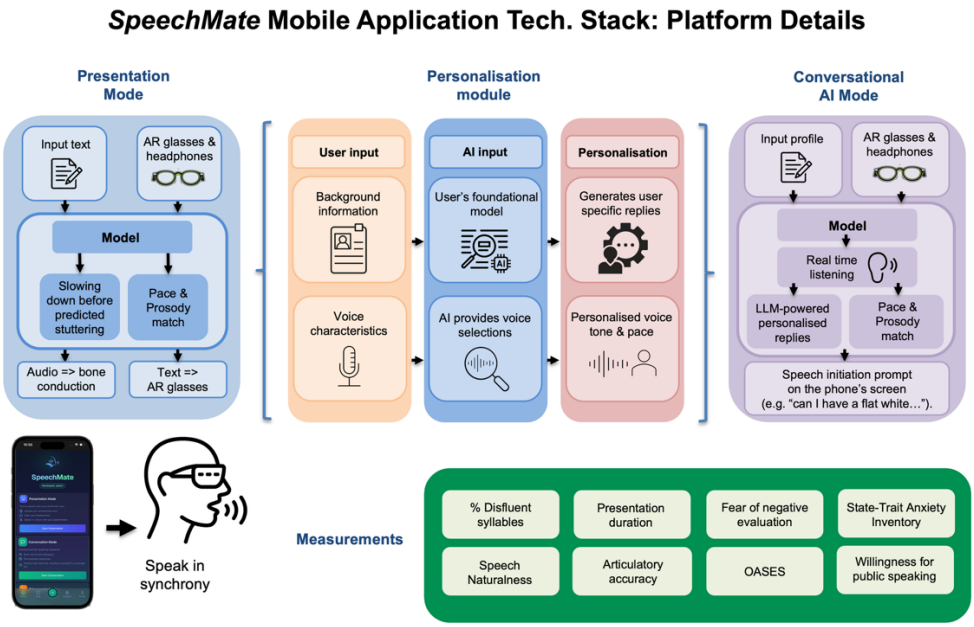

Figure 3 - SpeechMate technical stack and study linked components. Left, the **Presentation Mode** is tested during structured public speaking in Experiment 1. An uploaded script is presented through AR glasses, while a personalised guiding voice is delivered through bone conduction earbuds. The guiding voice is matched to the user's preferred pace and prosody, and can slow before marked anticipated stuttering words or under user's control during speech**.** Centre, the **Personalisation Module** uses background information and guiding voice selection to adapt the guiding voice, pace, and tone to the user. Right, the **Conversational AI Mode** will be evaluated in Experiment 2. The system uses the participant profile, real time listening, AR glasses, and bone conduction audio to support semi-structured dialogue with personalised replies and matched guiding voice. An on-screen button or external controller (e.g., smart ring) can be used to trigger speech initiation. Bottom, the **Measurement Panel** summarises the main outcomes, including percentage of disfluent syllables, speech naturalness, presentation duration, articulatory accuracy, willingness to speak in public, and standardised measures of stuttering severity, anxiety, fear of negative evaluation, and the Overall Assessment of the Speaker's Experience of Stuttering (OASES).

Method

Participants

A previous within subject exploratory study showed a large reduction in percentage of disfluent syllables during choral speech with a conventional TTS decoder in a simulated presentation task (r = 0.89, Cohen's dz approximately 3.91)[8]. We will therefore recruit 20 adults who stutter for the main within subject comparison. Small exploratory groups of adults with apraxia of speech and typically fluent adults will also be recruited. The typically fluent group will provide a reference range for speech measures and delivery ratings, allowing us to examine whether SpeechMate supported speech in adults who stutter moves closer to the range observed in typically fluent speakers. The apraxia subgroup will be included to assess feasibility and speech related outcomes during visual text based SpeechMate tasks.

The study received favourable ethical opinion from the Oxford Central University Research Ethics Committee (CUREC), Medical Sciences Interdivisional Research Ethics Committee (MS IDREC 2487591). All participants will provide written informed consent before taking part. Participants will be informed that participation is voluntary and that they may withdraw from the study in accordance with the approved protocol.

Participants will be recruited through social media advertisements (e.g., British Stammering Association), University mailing lists, posters, word of mouth, and local stuttering group (e.g., Oxford Stammering Meet-up group). Eligibility will be confirmed by a brief video call and screening questionnaire. To minimise ceiling effects, recruitment will prioritise PWS with moderate to very severe

stuttering, based on self-report, brief video call, and screening measures. Participants will be included if they are aged 18 to 65 years and speak English at a native or near native level. They will have normal or corrected to normal hearing and vision. PWS will be eligible either with a prior clinical diagnosis, or by self-report of stuttering. The apraxia subgroup will be eligible with a self-reported clinical diagnosis of apraxia of speech.

Participants will be excluded if they have uncorrected hearing or vision problems, any medical or neurological condition that makes wearing smart glasses or audio devices uncomfortable, or a diagnosed speech, language, or neurological disorder that would confound interpretation of the speaking outcomes. For the stuttering group, additional speech disorders other than stuttering will be exclusionary. For the apraxia subgroup, additional conditions will be recorded and reported, and the analysis will be framed as exploratory.

Procedure

This is a randomised, active controlled, within participant crossover proof of concept study with blinded outcome rating. Participants will experience visible or audible differences between conditions, but they will be blinded to the expected direction of the study hypotheses before completing the laboratory tasks. Speech outcomes will be rated by trained raters blinded to condition and study hypotheses.

Two experiments will be conducted. Experiment 1 will evaluate SpeechMate Presentation Mode during a controlled public speaking task, followed by a one-month real world follow up in which participants use SpeechMate for self-selected presentations. Experiment 2 will evaluate SpeechMate Conversational AI Mode during semi-structured speaking scenarios. Each participant will attend one laboratory visit lasting approximately 60 to 90 minutes.

Upon arrival, participants will complete baseline questionnaires and then provide speech samples for assessment with the Stuttering Severity Instrument, Fourth Edition (SSI-4). SSI-4 scoring will be based on both reading and spontaneous speech samples recorded during the session. Participants will first read a short, standardised passage aloud and then respond to brief open-ended questions designed to elicit spontaneous speech. Audio and video recordings will be used to score frequency of stuttering events, and physical concomitants.

Following the SSI-4 assessment, participants will select a guiding voice in SpeechMate. They will choose the preferred voice gender, accent, voice characteristic, and speaking pace. Participants will then complete a short practice task in Presentation Mode using texts that are not used in the experimental trials. This practice phase will familiarise participants with the guiding voice, AR text display, and pace control before the experimental conditions begin. Participants will be instructed to speak in synchrony

with the guiding voice word by word, consistent with choral speech, or they may follow the voice with a short delay, consistent with shadowing.

The number and duration of practice trials will be allowed to vary between participants according to their familiarity with the technology and their ability to operate the guiding voice, augmented reality display, and pace control. Different practice texts will be used before the main experimental trials to reduce adaptation effects from repeated reading of the same material, as repeated utterances can produce larger and more retained fluency gains than novel utterances in PWS[15]. Practice will continue until the participant confirms that they understand the task and can use the relevant controls. The number of practice trials, approximate practice duration, and any technical support required will be recorded for each participant and reported descriptively.

After Experiment 1, participants will complete a brief practice trial in Conversational AI Mode as well. During this practice trial, participants will learn how to initiate an AI-generated reply once it appears on the AR. Practice trials will be completed in a different corner of the room from the main testing setup, to reduce habituation to the experimental speaking tasks.

Experiment 1: Presentation Mode

The experiment will use a within subject design with two conditions.

1. Baseline Teleprompter Mode, no guiding voice but using AR glasses.
2. Presentation Mode, personalised guiding voice and AR glasses.

For each condition, a presentation text provided by the researcher will be read aloud. In the baseline condition, the text will appear on the AR glasses in Teleprompter Mode without auditory guidance. Scrolling will be controlled manually from the phone screen using upwards and downwards arrows. In Presentation Mode, the same text will appear on the AR glasses with auditory guidance from the selected guiding voice. The guiding voice will use the participant's chosen voice profile and speaking pace, and speech will be synchronised with this voice. Participants will be instructed to speak in synchrony with the guiding voice word by word, consistent with choral speech, or they may follow the voice with a short delay, known as shadowing. The text display will advance in time with the audio. During speech, the guiding voice can be slowed in real-time by tapping the phone screen, or tilting the phone to the left.

For the exploratory apraxia subgroup, the presentation task will be adapted to test text complexity rather than guiding voice. Therefore, participants in this group will complete Teleprompter Mode only. They will read two versions of the same text: the original version and an AI-simplified version with reduced

syntactic complexity, shorter word forms where appropriate, and lower articulatory demand. This will be the same text for each participant. This design was informed by pilot testing with an adult with apraxia of speech. In that pilot session, auditory guidance was not preferred during real time speaking, whereas it was observed that the complexity of the text appeared to affect speech production. The exploratory apraxia task will therefore examine speech performance during AR-based teleprompting with original versus AI-simplified text.

Condition order will be randomised and counterbalanced across participants using a pre-generated Latin square schedule. The allocation sequence will be created before recruitment using a computer-generated randomisation list, and participants will be assigned to the next available sequence after enrolment. After each talk, participants will rate comfort, perceived control, and speech naturalness. Following the laboratory visit, participants will receive a study link and participant ID for using SpeechMate on their own smartphone. They will be asked to use the app during up to three self-selected speaking situations over one month. These may include talks the participant would give anyway, or practice speaking situations arranged for the study, such as a work briefing, a practice talk to friends or family, or a voice note sent through a social media platform. After each use, participants will complete a short log describing context, perceived benefit, and any problems. AR glasses will be provided where possible. If glasses are not available, participants will use SpeechMate on the smartphone only.

Experiment 2: Conversational AI Mode

Experiment 2 will use semi-structured conversations modelled on common challenging or everyday speaking situations. The scenario library will include a job interview, booking a doctor appointment, speaking to authority (e.g., passport control at an airport), cafe order, and introducing oneself to a stranger. As in Experiment 1, the order of baseline and assisted blocks within each conversational scenario will be randomised and counterbalanced across participants using a pre-generated schedule

Before the session, each participant will upload non-sensitive interests, hobbies, and general background descriptions into SpeechMate. The system will build a foundational model of each user from these documents. This model will cover education, roles, and interests. During scenarios the suggested responses will be generated from this profile so that the dialogue reflects the participant's own background rather than generic content. This grounding will be used only to run the experiment and will be stored under the participant's study identifier.

Scenario questions will be delivered from a laptop using the SpeechMate Questions Console. The researcher will operate the console, but the questions will be presented in a standardised voice for all participants within each scenario. Different scenarios may use different question voices. If a participant

gives a brief response in a baseline condition, the console will provide standardised follow up prompts, such as “Could you tell me a bit more about that?”

Each scenario will be completed in two blocks.

1. Baseline sessions, participant responds without SpeechMate guidance
2. Assisted sessions, SpeechMate generates suggested responses and the participant speaks in synchrony with them using guiding voice and AR glasses

In assisted sessions, participants will tap the play button on the phone to start the AI-generated guiding response. The response will be displayed on the AR glasses and delivered through bone conduction audio. Participants will speak in synchrony with the guiding voice. A tap and hold will activate the metronome-paced speech option. This option will be used when the AI-generated response is not appropriate, or when the participant wishes to respond spontaneously. For the exploratory apraxia subgroup, Conversational AI Mode will follow the same visual-only approach as the Experiment 1: AI generated replies will be displayed on the AR glasses without auditory guidance.

Experimental setting

All sessions will take place in a quiet room at the University of Oxford’s Institute of Biomedical Engineering. One researcher will run the protocol and deliver prompts. Speech will be recorded using both video and audio. Video will be captured using a smartphone or a camera on a tripod positioned at approximately eye level, framed to include the face and upper body. Audio will be captured using a collar mounted lavalier microphone connected to a dedicated recorder.

During experiments, participants remain in control of whether and how they use the AI-generated output. In Presentation Mode, participants can start, stop, pause, or slow the guiding voice. During real world use after the laboratory visit, participants will be instructed to use SpeechMate only in situations where use of a phone-based speech support tool is appropriate. In Conversational AI Mode, participants can choose not to use a generated reply if it is unsuitable, inaccurate, uncomfortable, or not what they want to say. In that case, they can speak spontaneously and may use the metronome cue instead. Participants will be instructed not to provide sensitive personal information for the conversational profile.

The system will not be used to diagnose participants, make clinical decisions, or replace speech therapy. It will be tested as an experimental speech support tool. During laboratory testing, a researcher will remain present and can pause or stop the task if the participant becomes uncomfortable or if the system behaves unexpectedly. System errors or usability problems, including unsuitable generated replies,

transcription errors, delays, poor synchrony, or participant rejection of an AI-generated reply, will be recorded descriptively as part of the feasibility assessment.

Measurements

The primary outcome for both experiments will be percentage of disfluent syllables during each laboratory speaking block. Percentage of disfluent syllables will be calculated for each block as the number of stuttered syllables divided by the total number of syllables, multiplied by 100. Stuttering events will include blocks, prolongations, and part-word repetitions. Whole-word repetitions will be included in the primary fluency metric[8,12] but will be excluded from the frequency component when computing SSI-4[16]. Speech naturalness will be rated on a nine-point scale, where 1 indicates highly natural-sounding speech and 9 indicates highly unnatural-sounding speech[16]. Additional measures will capture complementary aspects of speech performance and experience during each laboratory speaking block, including total speech duration, articulatory accuracy, usability and perceived control after each relevant condition.

To characterise participant experience and relevant psychosocial measures, participants will also complete standardised questionnaires. The Overall Assessment of the Speaker's Experience of Stuttering (OASES), which assesses the impact of stuttering on daily communication and quality of life, will be completed by adults who stutter at baseline and after the one month follow up. The Fear of Negative Evaluation scale (FNE), which measures concern about negative social judgment, particularly relevant for conversational and interview contexts, will be completed at baseline and after the one month follow up where applicable. Willingness to give public presentations will be measured at baseline and after the one month follow up to examine change in self-reported speaking motivation. Logged use outside the laboratory will be recorded after each self-selected SpeechMate use over the one month follow up period, including the speaking context, perceived benefit, and any problems.

The apraxia subgroup will be analysed descriptively as an exploratory extension. The same primary and self-reported metrics will be computed during the relevant laboratory speaking blocks, alongside apraxia relevant speech measures including articulatory accuracy during the original and AI-simplified text conditions. Participants will also provide self-reported ratings of speaking effort, control, and perceived fluency after the relevant conditions. Analyses will be framed as feasibility rather than formal hypothesis testing, with individual responses reported alongside group level summaries.

For typically fluent speakers, analyses will emphasise self-reported measures alongside the same objective speech metrics used for the other groups. Participants will provide ratings of comfort, usability, and willingness to speak in front of an audience across conditions.

All speech ratings will be performed by independent raters who are blinded to condition and study aims. Two trained raters will score disfluency, and accuracy measures together with speech naturalness, with inter-rater agreement quantified using an intraclass correlation coefficient. The primary rater will repeat scoring for a predefined subset of recordings after a minimum interval of one week to assess intra-rater reliability.

Analysis and hypotheses

For PWS, we hypothesise that SpeechMate Presentation Mode will be associated with a lower percentage of disfluent syllables than baseline Teleprompter Mode during the structured public speaking task. We will also test whether speech naturalness ratings and participant preference differ between conditions. We further hypothesise that self-reported willingness to give presentations will be higher after one month of SpeechMate use outside the laboratory than at baseline. In exploratory analyses of Conversational AI Mode, we expect assisted conversations to be associated with a lower percentage of disfluent syllables than baseline conversational sessions.

For the apraxia subgroup, no confirmatory hypotheses will be tested. Instead, we will assess feasibility and describe changes in articulatory accuracy, self-reported control, and willingness to speak in individual participants. For typically fluent speakers, we expect SpeechMate to be usable, with stable objective speech measures and favourable ratings of perceived control and willingness to speak in public.

Between group analyses will compare changes in percentage of disfluent syllables, willingness to speak in public, usability, and FNE scores across groups. The primary between group endpoint will be change in percentage of disfluent syllables from baseline to SpeechMate. These analyses will be performed using models that include group as a fixed effect and participant as a random effect. We expect that, during the structured public speaking task, PWS using SpeechMate Presentation Mode will receive delivery quality ratings comparable with those of typically fluent speakers. We further hypothesise that changes in self-reported willingness to speak will be larger in people who stutter than in typically fluent speakers, with corresponding group differences in perceived control, and FNE scores.

Normality of residuals will be assessed using Shapiro Wilk tests and visual inspection. If distributions are clearly non normal, sensitivity analyses will use nonparametric repeated measures tests, including Friedman tests across conditions and Wilcoxon signed rank tests for planned pairwise comparisons with correction for multiple comparisons.

Data sharing and confidentiality

Audio and video recordings contain potentially identifiable speech, facial images, and personal information. These recordings will therefore be handled in accordance with the approved CUREC protocol, participant consent, and University of Oxford data protection procedures. Where participants provide explicit consent for sharing, selected audio and video recordings may be shared for research, education, presentation, publication, or public engagement purposes, according to the consent options approved for the study. Recordings will not be shared beyond the scope of the participant's consent. Identifiable data, including names and contact details, will be stored separately from study outcome data. De-identified summary data, analysis code, and non-identifiable study materials may be published in peer-reviewed journals.

References

1. Yairi, E. & Ambrose, N. Epidemiology of stuttering: 21st century advances. J. Fluency Disord. 38, 66–87 (2013).
2. Ferguson, A. M., Roche, J. M. & Arnold, H. S. Social judgments of digitally manipulated stuttered speech: An evaluation of self-disclosure on cognition. J. Speech Lang. Hear. Res. 62, 3986–4000 (2019).
3. Gerlach, H., Totty, E., Subramanian, A. & Zebrowski, P. Stuttering and labor market outcomes in the United States. J. Speech Lang. Hear. Res. 61, 1649–1663 (2018).
4. Craig-McQuaide, A., Akram, H., Zrinzo, L. & Tripoliti, E. A review of brain circuitries involved in stuttering. Front. Hum. Neurosci. 8, 884 (2014).
5. Tichenor, S. E. & Yaruss, J. S. Variability of stuttering: Behavior and impact. Am. J. Speech Lang. Pathol. 30, 75–88 (2021).
6. National Stuttering Association. The Experience of People Who Stutter: A Survey by the National Stuttering Association. (2009).
7. Sønsterud, H., Halvorsen, M. S., Feragen, K. B., Kirmess, M. & Ward, D. What works for whom? Multidimensional individualized stuttering therapy (MIST). J. Commun. Disord. 88, 106052 (2020).
8. Demirel, B. A novel use of choral speech significantly reduces stuttering in a simulated presentation setting: An exploratory study. J. Fluency Disord. 86, 106168 (2025).
9. Le Dévic, A., Diwersy, S. & Didirková, I. The link between syntactic complexity and stuttering-like disfluencies in French speaking adults. Int. J. Lang. Commun. Disord. 61, e70218 (2026).
10. Strand, E. A. & McNeil, M. R. Effects of length and linguistic complexity on temporal acoustic measures in apraxia of speech. J. Speech Hear. Res. 39, 1018–1033 (1996).
11. Jackson, E. S., Miller, L. R., Warner, H. J. & Yaruss, J. S. Adults who stutter do not stutter during private speech. J. Fluency Disord. 70, 105878 (2021).

12. Chesters, J., Möttönen, R. & Watkins, K. E. Transcranial direct current stimulation over left inferior frontal cortex improves speech fluency in adults who stutter. Brain 141, 1161–1171 (2018).

13. Lombard, E. Le signe de l'élévation de la voix. Ann. Maladies Oreille Larynx Nez Pharynx 37, 101–119 (1911).

14. Howell, P. & El-Yaniv, N. The effects of presenting a click in syllable-initial position on the speech of stutterers: Comparison with a metronome click. J. Fluency Disord. 12, 249–256 (1987).

15. Briley, P. M., Barnes, M. P. & Kalinowski, J. S. Carry-over fluency induced by extreme prolongations: A new behavioral paradigm. Med. Hypotheses 89, 102–106 (2016).

16. Riley, G. D. SSI-4: Stuttering Severity Instrument, Fourth Edition. (Pro-Ed, 2009).