Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Contemporary Clinical Trials 107 (2021) 106500

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

An innovative protocol for the artificial speech-directed, contactless administration of laboratory-based comprehensive cognitive assessments: PAAD-2 trial management during the COVID-19 pandemic

K. Shin Park *, Jennifer L. Etnier

Department of Kinesiology, University of North Carolina at Greensboro, United States of America

ARTICLE INFO

Keywords:
Alzheimer's disease
Artificial intelligence
Executive function
Neuropsychological tests
Speech synthesis
Text-to-speech
Physical activity
Working memory

ABSTRACT

The COVID-19 pandemic resulted in suspending in-person human subject research across most institutions in the US. Our extensive cognitive assessment for a phase-2 clinical trial, Physical Activity and Alzheimer's Disease–2 (PAAD-2), was also paused in March 2020. It was important to identify strategies to mitigate the risk of COVID-19 transmission during our testing, which initially required substantial human speech and close person-to-person contact for test directions and instant feedback on paper/pencil tests. Given current understanding of the COVID-19 transmission, we dramatically adjusted the testing protocol to minimize the production of speech droplets and allow social distancing while maintaining the integrity of testing. We adopted state-of-the-art speech synthesis and computerization techniques to create an avatar to speak on behalf of the experimenter for all verbal instructions/feedback, used a document camera to observe the paper/pencil tests from the required distances, and automated the testing sequence and timing. This paper aims 1) to describe an innovative laboratory-based cognitive testing protocol for a completely contact-free, computer-speaking, and semi-automated administration; and 2) to evaluate the integrity of the modified protocol ($n = 37$) compared with the original protocol ($n = 32$). We have successfully operated the modified protocol since July 2020 with no evidence of COVID-19 transmission during testing, and data support that the modified protocol is robust and captures data identical to the original protocol. This transition of data collection methods has been critical during the pandemic and will be useful in future studies to mitigate the risk of contagious disease transmission and standardize laboratory-based psychological tests.

Trial registration: ClinicalTrials.gov NCT03876314. Registered March 15, 2019

1. Introduction

The COVID-19 pandemic resulted in the suspension of in-person human research activities across most institutions in the United States. In response to the serious health threat at global and national levels, universities acted quickly to vacate their campuses by converting teaching to online, sending students, faculty, and staff home to perform their duties remotely, and putting all non-essential research on pause [1–3]. Wigginton and colleagues [4] estimated that 80% of on-site research activities at the authors' universities were halted by limiting building access and only permitting studies on animals, patient safety, or COVID-19. Our extensive cognitive testing for a phase-2 clinical trial, Physical Activity and Alzheimer's Disease–2 (PAAD-2) [5], was also paused in March 2020.

It is known that COVID-19 is transmitted from human to human mainly through respiratory droplets that are spread when an infected person coughs, sneeze, or talks particularly when in close contact (within 6 ft) [6]. It was important to identify strategies to mitigate the risk of disease transmission during our four-hour laboratory testing session. Our original protocol required substantial verbal instructions for informed consent and test directions. In many instances, the experimenter was in close contact with a participant to observe cognitive performance on a computer monitor, mobile device, or paper form in order to instantly evaluate and provide necessary feedback. Therefore, our primary goal with protocol modifications was to minimize the possibility of directly and indirectly transmitting aerosolized droplets in our interactions with participants in the laboratory testing.

Mitigation of risk was partially attained by following the university's

* Corresponding author.
E-mail address: k_park4@uncg.edu (K.S. Park).

https://doi.org/10.1016/j.cct.2021.106500
Received 14 December 2020; Received in revised form 28 June 2021; Accepted 28 June 2021
Available online 2 July 2021

requirement of wearing face coverings and maintaining social distance for every person involved in testing. We further attempted to minimize the production of small speech droplets, which can cause airborne transmission of COVID-19 in confined environments [7–10]. Moreover, as speaking under face coverings imposes vocal fatigue and discomfort, difficulties in coordinating speech and breathing, and can make speech more difficult to understand [11], our goal was to reduce the need for speaking by the experimenter. As such, we employed state-of-the-art speech synthesis and computer programming techniques to have an avatar speak on behalf of the experimenter for all verbal instructions. We also used a document camera to allow the experimenter to observe participants' performance from the required or even farther distances and employed computer program to automate the testing sequences and timing control.

While the protocol modifications substantially addressed safety concerns relative to COVID-19 transmission, the maintenance of the integrity of testing was critical for the clinical trial. Although the naturalness of synthesized speech has been previously established [12,13], its validity has not been demonstrated for cognitive testing in a laboratory setting. As such, adjustments were made on the protocol in response to pilot testing, and comparisons were made between data collected with the original protocol and the modified protocol. Information regarding the protocol, pilot testing, and these comparisons is intended to assist future investigators to reduce the risk of transmission of contagious diseases and standardize the administration of complex cognitive testing paradigms.

2. Objectives

The purpose of this paper is 1) to describe the detailed methods used to convert a complex cognitive testing protocol that involved close person-to-person contact, substantial human speech, and manual control of the testing sequence and timing into a completely contact-free, computer-speaking, and semi-automated protocol with the goal of significantly minimizing the production of human speech droplets and ensuring the maintenance of social distancing; and 2) to evaluate the integrity of the administration of the modified protocol ($n = 37$) in comparison with the administration of the original protocol ($n = 32$). This transition in terms of the data collection method has been critical during the COVID-19 pandemic and will be useful in future studies to mitigate the risk of transmitting contagious illnesses and to further standardize and automate the administration of laboratory-based cognitive tests. Our detailed description of the modified laboratory testing is intended to be useful for other researchers to partly or entirely replicate similar protocol adjustments during and after the pandemic while also providing a detailed description of the modification of the PAAD-2 protocol [5] as implemented after July 2020.

3. Protocol modification

3.1. Speech synthesis techniques

In 1968, the filmmakers of the epic science movie, “2001: Space Odyssey”, depicted a time 33 years in the future when an artificial intelligence (AI) computer could generate human-like voices to verbally communicate with spaceship crews. This futuristic vision was realized through the use of AI technology about a decade after the imagined year in that machine-generated speech started to become widely available through virtual assistants in smartphones, computers, and other modern devices such as Apple's Siri, Amazon's Alexa, and the Google Assistant [14,15]. Since then, speech synthesis, also known as text-to-speech (TTS) technique, has been rapidly advancing and now the commercial Application Programming Interface (API) platforms[1] allow people to easily create synthetic speech by converting text input into voice output. Recent improvements in one of the TTS synthesis models, called WaveNet [16,17], and subsequent neural network modeling have enabled substantial enhancements in the naturalness of synthesized voices to the extent of rivaling human speech [12,13]. Not just short-form content at the word, sentence, or paragraph level [18], but synthetic voices reading out a long-form article of more than 900 words were found to be comprehensible and pleasant to listen to for several minutes at a comparable level to human voices [19].

As a necessary modification to the PAAD-2 protocol [5] to allow for data collection during the COVID-19 pandemic, we created and operated synthetic voices that directed the entire 4-h testing session for the informed consent and cognitive assessments allowing the experimenter to maintain the required (6 ft) or even farther distances in the laboratory. Specifically, the summary of the consent form and all verbal instructions/feedback of cognitive tests were written as text or Speech Synthesis Markup Language (SSML) input files in the JavaScript Object Notation (JSON) format. We chose the WaveNet voice (en-US-Wavenet-D) and then set the rate of speaking to 0.89–0.93 and the pitch to −2.8. We used the macOS command line interface to convert the text or SSML files into waveform audio file (WAV) formats on the Google Cloud TTS API. More information on the implementation of the Google Cloud TTS API are proprietary but publicly available in a web document [20]. All WAV files of the synthetic voices were then added as sound components in an open-source programming platform, PsychoPy Experiment Builder [21], which are further described in later sections of this paper.

Consequently, our synthetic voice directs the entire testing session. The voice first explains the COVID-19 safety precautions, briefly describes each paragraph of the informed consent form, provides general instructions for the testing session, gives specific instructions and/or feedback for the Montreal Cognitive Assessment (MoCA) [22], the Test of Premorbid Functioning (TOPF) [23], the Rey-Osterrieth Complex Figure Test (ROCFT) [24,25], the Paced Auditory Serial Addition Test (PASAT) [26], the Rey Auditory Verbal Learning Test (RAVLT) [25], the Trail Making Test (TMT) [27], and the Symbol Digits Modalities Test (SDMT) [28]. To maintain consistency with our original protocol, the synthetic voice asks participants to read and follow the text instructions written on the screen for the tests administered with *E*-Prime 3.0 software [29] on a desktop computer and the NIH Toolbox cognition battery on the iPad [30]. The synthetic voice also directs participants to make appropriate transitions between tasks on the computer with a keyboard or mouse, the iPad, or hard copies of documents, and to take a break for certain durations. See Table 1 for the instruments and response formats of each test.

Before the implementation of our modified protocol, we ensured that the naturalness of synthetic speech was acceptable for an extensive cognitive assessment through pilot testing of the entire protocol. We therefore repeatedly tested whether six pilot subjects (a professor and graduate students in psychology) clearly understand the synthesized instructions for test directions and safety precautions and accordingly designed the new protocol to ensure that the artificial speech and automated sequence are easy to follow and identical to the original protocol, which is further addressed in this paper. In the following paragraphs, we further illustrate how the synthetic voices are presented along with an avatar to direct the entire testing protocol based on precise management of timing.

[1] These platforms include Google Cloud (https://cloud.google.com/text-to-speech), IBM Watson (https://www.ibm.com/cloud/watson-text-to-speech), Amazon Polly (https://aws.amazon.com/polly/), and Microsoft Azure (https://azure.microsoft.com/en-us/services/cognitive-services/text-to-speech/) etc.

Table 1
An overview of the PAAD2 cognitive testing protocol modification in comparison to the original protocol.

Batch	Testing components	Operating instrument	Response format	Comparisons of test duration (min, $M \pm SE$)			
				Original ($n = 32$)	Modified ($n = 37$)	ET (p)	t-test (p)
0	Safety Precautions	Python†	P/P	–	5 ± 0.20	–	–
	Informed Consent Pre only	Python†	P/P	≈ 15	16 ± 0.58	< 0.05	0.09
	MoCA Δ Pre & Post only	Python†	P/P^{Cam}, Verbal^{Mic}	≈ 10	10.2 ± 0.30	< 0.001	0.49
1	General Instructions	Python†	–	≈ 3	2.8 ± 0.24	< 0.05	0.48
	Test of Premorbid Functioning Pre only	Python†	Verbal^{Mic}	3.4 ± 0.13	3.4 ± 0.12	< 0.01	0.83
	NIH TB List Sort Working Memory	iPad	Verbal	8.4 ± 0.22	8.1 ± 0.21	< 0.05	0.26
	NIH TB Picture Sequence Memory Δ	iPad	Screen touch	7.7 ± 0.23	7.7 ± 0.25	< 0.01	0.98
	Mnemonic Similarity Task Δ	*E*-Prime	Key press	10.9 ± 0.10	11.4 ± 0.24	0.16	0.06
	Perceptual Discrimination Task Δ	E-Prime	Key press	5.78 ± 0.38	4.64 ± 0.20	0.75	< 0.01
2	ROCFT – Copy	Python†	P/P	3.2 ± 0.23	5.8 ± 0.59	0.89	< 0.001
	Break	Python†	–	3	3 ± 0.00	< 0.001	1
	ROCFT - Immediate Recall	Python†	P/P	2.7 ± 0.20	3.87 ± 0.39	0.39	< 0.01
	Stroop Color-Word Task	E-Prime	Key press	6.1 ± 0.21	5.83 ± 0.15	0.06	0.37
	PASAT - 3′ and 2’	Python† fd	Verbal^{Mic}	≈ 12	12 ± 0.21	< 0.01	0.63
	(Break; for 30 min delay)	Python†	–	(≈ 5; 30)	(4.5 ± 0.42; 29.8 ± 0.16)	< 0.01	0.21
	ROCFT - Recall	Python†	P/P	1.8 ± 0.15	2.5 ± 0.24	0.33	< 0.05
	ROCFT - Recognition	Python†	P/P	≈ 3	2.8 ± 0.25	< 0.01	0.34
	Tower of London - Freiburg version Δ	VTS	Mouse click	10.5 ± 0.43	10.2 ± 0.31	< 0.05	0.58
	Break	Python†	–	≈ 7	7 ± 0.00	< 0.001	1
3	RAVLT - Learning & Recall Δ	Python†	Verbal^{Mic}	≈ 8	8.2 ± 0.17	< 0.01	0.89
	NIH TB Dimensional Change Card Sort	iPad	Screen touch	5.1 ± 0.05	5.1 ± 0.06	< 0.05	0.25
	NIH TB Flanker Inhibitory Control	iPad	Screen touch	3.3 ± 0.03	3.3 ± 0.03	< 0.05	0.41
	Spatial Working Memory	E-Prime	Key press	11.0 ± 0.31	10.7 ± 0.19	< 0.05	0.46
	Trail Making Test - A & B	Python† fd	P/P^{Cam}	≈ 5	5.0 ± 0.09	< 0.001	0.74
	(Break; for 30 min delay)	Python†	–	(≈ 5; 30)	(4.7 ± 0.28; 30 ± 0.00)	< 0.001	1
	RAVLT - Recall Δ	Python†	Verbal^{Mic}	≈ 1.4	1.4 ± 0.02	< 0.05	0.06
	RAVLT - Recognition Δ	Python†	P/P	≈ 3.5	3.4 ± 0.11	< 0.01	0.33
4	Paired Associates - Learning & Recall Δ	Python†	Verbal	≈ 3	2.96 ± 0.08	< 0.01	0.60
	Matrix Reasoning Δ	E-Prime	Mouse click	14.1 ± 0.34	14.1 ± 0.34	< 0.01	0.90
	Digits Span - Forward	E-Prime	Verbal	3.1 ± 0.30	2.8 ± 0.12	0.08	0.23
	(Break; for 20 min delay)	Python†	–	(≈ 2; 20)	(1.8 ± 0.17; 20 ± 0.00)	< 0.001	1
	Paired Associates - Recall Δ	Python†	Verbal	≈ 1.5	1.5 ± 0.08	< 0.01	0.60
	Digits Span - Backward	E-Prime	Verbal	2.3 ± 0.16	2.6 ± 0.25	< 0.05	0.34
	Break	Python†	–	≈ 5	4.7 ± 0.19	< 0.01	0.17
5	Logical Memory - Learning & Recall Δ	Python†	Verbal^{Mic}	4.1 ± 0.08	4.2 ± 0.09	< 0.05	0.37
	Spatial Relations Δ	E-Prime	Mouse click	11.0 ± 0.38	10.6 ± 0.31	< 0.05	0.47
	SDMT Written & Oral Trials	Python† fd	P/P^{Cam}, Verbal^{Mic}	≈ 7	6.7 ± 0.09	0.12	< 0.01
	SDMT Incidental Learning Δ	Python†	P/P	0.82 ± 0.07	1.1 ± 0.11	0.12	0.08
	(Break; for 20 min delay)	Python†	–	(≈ 1; 20)	(1.4 ± 0.32; 20.3 ± 0.42)	< 0.01	0.32
	Logical Memory - Recall Δ	Python†	Verbal Mic	1.9 ± 0.05	1.9 ± 0.05	< 0.05	0.47

Abbreviation: MoCA, Montreal Cognitive Assessment; PASAT, Paced Auditory Serial Addition Test; P/P, Paper/Pencil; RAVLT, Rey Auditory Verbal Learning Test; ROCFT, Rey-Osterrieth Complex Figure Test; SDMT, Symbol Digits Modalities Test; VTS, Vienna Test System.
Note: One- or two-sample *t*-tests or equivalence tests (ET) were conducted to compare each test duration between two protocols. Results indicate that two protocols are significantly equivalent in terms of its duration. Equivalence margin (Cohen's d, δ) was ±0.7. Some test duration of the original protocol was not measured and thus estimated (≈). Total protocol duration is about 3.5 h at pre-test and 3 h at mid- and post-test. Synthetic voice is used for all necessary verbal instructions and feedback for informed consent, test directions and transitions, and breaks between tests. Δ Different forms are used at pre-, mid-, and post-test. Break duration in parentheses is computed by a custom-developed Python code and instructed by the synthetic voice. † Newly added features in the modified protocol. fd Synthetic voice gives feedback in response to a key press if necessary. Mic Subjects' verbal responses are recorded on a microphone. Cam Subjects' paper works are observed from the required distance using a document camera. Batch 0 without informed consent is combined with Batch 1 at mid- and post-tests and comes after the biological sampling. Testing sequences of Batch 2 and 4 can be changed by automatic timing control by the Python codes (see Fig. 2 for more information).

3.2. Speaking avatar

We developed an avatar and presented him with his mouth moving along with the synthesized voices given that human speech can be better understood by seeing the face of a talker even when the speech is audible and intact [31]. We first created a series of still images of different facial expressions and compiled the images in short time intervals (less than 50 ms) to create an animation image that included moving eyes, eyebrows, jaws, and lips to imitate facial movements of human speech. We then put a face covering on the avatar as a means of requesting participants to do the same and building rapport with the participant during the testing session. The avatar was added and programmed as a movie component to the PsychoPy Builder [21] in time with the audio files of synthetic voices. Consequently, the synthetic voice was presented in synchrony with the avatar just like he was talking to participants on the computer monitor, so that our participants would better understand verbal instructions and be more engaged with the computer during the testing session.

3.3. Contact-free testing environment

We configured the hardware by connecting a desktop computer as the central processing unit (CPU) to two monitors, two keyboards, and two mice for an experimenter and a participant (see Fig. 1). We set up the dual monitors to display duplicate content, and accordingly, the experimenter was able to operate each test on the computer, observe a participant's performance on the monitor, and observe the behaviors/responses of a participant from farther away than the 6-ft required distance. For a participant to better interact with the computer, we set up a sound speaker and a microphone near the participant's monitor to provide the verbal instructions and audio-record their verbal responses.

Some of our paper/pencil tests (i.e., MoCA, TMT, and SDMT) required monitoring participants' hand drawing or writings from a close distance for instant evaluations and/or feedback [22,27,28,32]. We

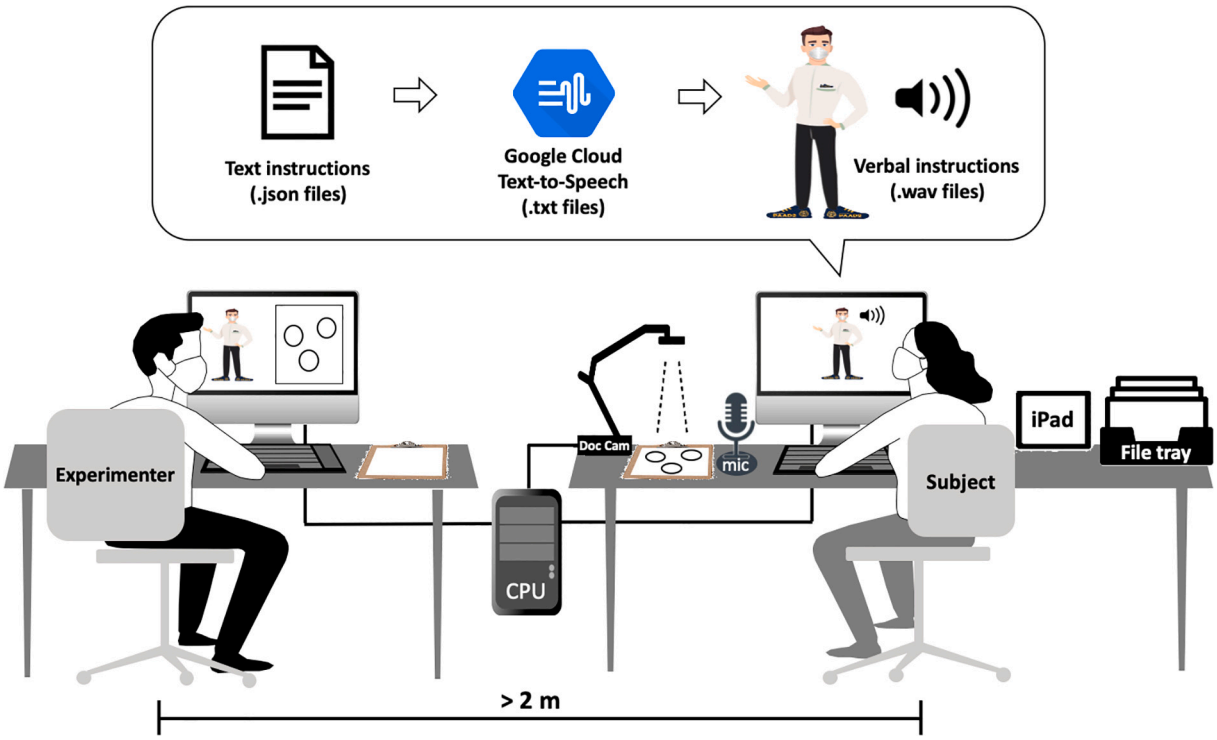

Fig. 1. A schematic overview of the avatar-directed contactless cognitive testing environment.

therefore set the paper forms on a clipboard along with a document camera above them, then connected the camera to the computer and developed a custom Python code based on an open-source computer vision package, the OpenCV library [33]. Using this method, the experimenter was able to see the camera view on the monitor to observe participant's performance on the paper forms and provide necessary evaluation/feedback through the synthetic voice by pressing designated keys on the keyboard (see Fig. 1 for a schematic overview).

During the testing, the synthetic voice instructed participants to receive, complete, and submit all paper forms in a contact-free manner for safety management. A file tray along with all test forms in file folders were set up on the right side from the participants, at least 24 h before the testing session. For each test during the testing session, the avatar instructed participants to take out certain testing form(s) from a file folder in a particular color (e.g., yellow, purple, or red) from each shelf of the file tray and complete the necessary paper tests using a pencil or pen. Separate folders were necessary to keep the contents of certain forms of memory tests confidential before the administration (e.g., delayed recognition for the ROCFT and RAVLT). Upon completion of paperwork, participants were instructed to submit the completed forms into the bottom shelf of the file tray.

3.4. Programming platform

We developed and operated the entire testing protocol using Python programming language [34] along with an open-source Python-based experiment control software, PsychoPy [21,35,36]. PsychoPy (available at psychopy.org) was developed for designing and editing behavioral experiments based on a graphical user interface (GUI) called "Builder" and/or Python scripts [21,35,36]. PsychoPy Builder allows the researcher to generate a Python script for the developed experiment, which is easily executed as a Python program. PsychoPy allowed us to start and stop the synthetic voice and talking avatar as sound and movie components in synchrony, audio-record verbal responses using the microphone components, measure the duration of test performances, program the sequences of the entire testing procedure, and automatically execute the proper tests using the clock functions and code components based on sub-millisecond precision [37,38]. Detailed information on the components and function are publicly available in the PsychoPy reference manual [39].

3.5. Batched and automated test timing and sequence

Using the PsychoPy Builder interface, we developed Python programs to enable the computerized administration of the MoCA, TOPF, ROCFT, PASAT, RAVLT, TMT, SDMT, Paired Associates, and Logical Memory based on each test's administration manual. We then compiled all test programs into five (mid- and post-test) or six (pre-test) batches so that the Python programs would keep functioning to measure the timing when participants were working on tests one after another (see Table 1 and Fig. 2 for an overview). By collating the tests into batches, the programs were able to automatically execute the correct test at the exactly desired time, count the duration for a break, and provide instructions for a break of any duration. The Python batches continued its timer function when *E*-Prime tests were operated.

While test timing was not important and all directions were provided in a fixed order in Batch 0 and Batch 1, time measurement was critical for the 20- or 30-min delayed recall of the ROCFT, RAVLT, Paired Associates, and Logical Memory tests (hereafter called delayed memory tests) in Batches 2, 3, 4, and 5 to choose and administer the correct test at the exact right time. See Fig. 2 for the sequences and logic of Batch 2, 3, 4, and 5. Immediately after the copy/learning or initial recall trials for the delayed memory tests, a timer was programmed to start keeping track of time. When beginning the break routine, the remaining time to the 20- or 30-min delay was counted, and the avatar instructed the participants to take a break for the measured duration. Using the text components in the PsychoPy Builder, a countdown timer in minutes and seconds was displayed on the computer screen during the break. When the break was over, the avatar provided instructions for the delayed memory tests in each batch. When the time limit was exceeded, rarely but possibly by a few seconds or minutes for slow test takers, no break was offered, and the delayed recall trials started immediately.

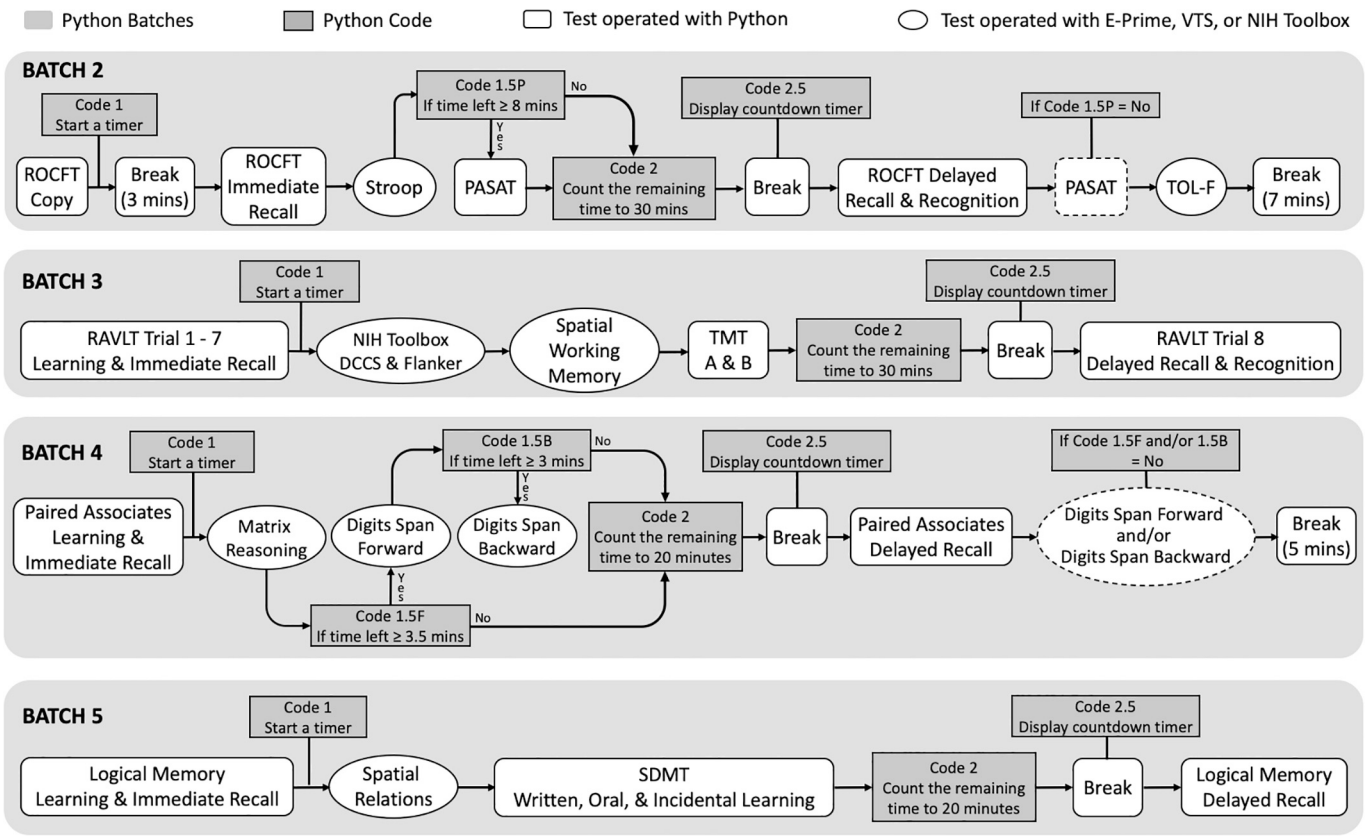

Fig. 2. Flowcharts of the Python batches 2–5.
Codes are available in the appendix. Batch 0 and 1 have no code components and are excluded in this figure but depicted in Table 1. Abbreviation: DCCS, Dimensional Change Card Sort Test; PASAT; Paced Auditory Serial Addition Test; RAVLT, Rey Auditory Verbal Learning Test; ROCFT, Rey-Osterrieth Complex Figure Test; SDMT, Symbol Digits Modalities Test; TMT, Trail Making Test; TOL-F, Tower of London - Freiburg version; VTS, Vienna Test System.

3.6. Automation of program execution

Our modified protocol is further equipped with automatic execution of all computer tests and batch programs with Python or *E*-Prime 3.0 following the manual implementation of the initial program. We enabled the automatic administration by using a customized code from an open-source operating system interface package, OS module, in the Python standard library [40]. We specifically added the function 'os. startfile (path)' to the code component in the PsychoPy Builder in order to start a file with its associated application, which acted like double-clicking the designated test files or batch programs. This technique allowed the experimenter to efficiently administer the entire testing protocol by eliminating any chance of wasting time to locate and manually start a test or batch file or making errors by executing a wrong program. Below we further describe how each testing protocol is specifically programmed in the different batches.

3.7. Informed consent and screening

In Batch 0 at the pre-test, the avatar read out the summary of safety precautions and informed consent. Participants were instructed to press spacebar on the keyboard to move on to the next paragraph when they fully understood the information. They were also encouraged to further read over the hard copies of the consent form or ask any questions to the experimenter. At the end of the consent, the avatar asked the participants to sign the form and submit the signed document into the file tray. The avatar then asked the participant to pick up a pencil and work on the paper form on a clipboard for the first three questions of the MoCA and verbally respond to the other questions. The MoCA was the screening tool of cognitive impairment, so the experimenter, who was trained and certified for the administration of the MoCA, carefully evaluated participants' verbal responses and drawings through the document camera and entered scores for each item on the keyboard. An algorithm was written to score the MoCA responses so that an indication of inclusion or exclusion could be provided. Based upon this, the avatar instructed the cognitively intact individuals to move on to the biological sampling session or people suspected of cognitive impairment to see the experimenter for further instructions. The experimenter discontinued the testing for excluded people and provided appropriate clinical referrals based on the PAAD-2 protocol [5].

3.8. Operational procedure of testing batches

Batch 1 at the pre-test provided general directions and then asked the participant to pick up the word list card from a colored folder for TOPF. The experimenter evaluated and audio-recorded the verbal responses. After that, the avatar directed the participants to move on to the next two NIH Toolbox tests on the iPad. When the NIH Toolbox tests were finished, the next two tests were sequentially executed with *E*-Prime 3.0. At the mid- and post-test, the avatar started with safety precautions and general instructions. Then without the TOPF, Batch 1 at the mid-test continued the testing with the NIH Toolbox and E-Prime tests in the same order as the pre-test, while Batch 1 at the post-test started with the MoCA and continued to the NIH Toolbox and E-Prime tests in the same order. See Table 1 for an overview.

In Batch 2, the avatar instructed the participant to take out paper forms from a colored folder for the ROCFT copy trial. Once finished, the program started a 30-min timer and instructed the participant to take a 3-min break with a countdown timer shown on the screen. After that, participants were asked to take out a paper form for the ROCFT

immediate recall. Then, the Stroop Color-Word Task, the same version with a similar clinical trial [41], was executed with *E*-Prime 3.0, and then the PASAT was followed. After that, the avatar instructed a break for the remaining time to 30 min. After the break, the participant was instructed to take out paper forms from colored folders, complete and submit them one after another for the ROCFT delayed recall and recognition. If less than 8 min was left for the PASAT, a break was given and the PASAT was administered after the delayed recognition. Next, the TOL-F [42] was administered, then the avatar gave the half-way break for 7 min. The duration of ROCFT trials was measured by a key press. See Table 1 and Fig. 2 for an overview.

After the break, Batch 2 was closed and followed by the Batch 3, in which the RAVLT learning and immediate recalls were administered. Then, the 30-min timer started, and the avatar directed participants to the iPad for the NIH Toolbox tests and then to the computer for the *E*-Prime test. Then, the TMT was followed, for which participants' drawings on the paper forms were observed via the document camera and feedback was given through the synthetic voice by pressing the designated prompt keys by the experimenter, which was programmed based on the TMT protocol [32]. Afterwards, the remaining time was counted, the avatar instructed a break for the measured duration, and a countdown timer was presented. After the break, the avatar gave directions for RAVLT delayed recall and recognition on a paper form, which was obtained from a colored folder and submitted to the file tray. See Table 1 and Fig. 2 for an overview.

Without a break, Batch 4 started for the Paired Associates learning and immediate recall. Next, the 20-min timer started, and the avatar gave instructions for the *E*-Prime test. After then, E-Prime tests were executed. The break was verbally instructed by the synthetic voice and provided for the remaining time on the 20-min timer. If insufficient time was left in 20 min delay for a test, the test was skipped, and a break was given for the remaining time. The skipped test was administered after the delayed recall. Then, Batch 4 was closed, and Batch 5 was executed. See Table 1 and Fig. 2 for an overview.

Batch 5 started with Logical Memory learning and immediate recall while audio-recording verbal responses, which was followed by a 20-min timer starting with the *E*-Prime test. Subsequently, the SDMT written, oral, and incidental learning trials [28] were administered. For the SDMT practice trials, the experimenter observed the participants' drawings on the paper forms via the document camera and provided appropriate feedback through the synthetic voice by pressing the designated prompt keys. After the SDMT, the program counted the remaining time, the avatar instructed a break for the measured duration with a countdown timer displayed; if no time was left, no break was offered. After the break, the delayed recall followed along with audio-recordings, and upon completion the avatar instructed the end of testing and our appreciation for participants' efforts.

3.9. Troubleshooting

Batched and automated operation of the test programs are efficient and convenient. Based on our pilot tests, we found that the programs occasionally crashed for an unknown reason, especially when the batch program was continuing its timer during another *E*-Prime 3.0 test. We therefore included in our protocol the use of a manually operated backup timer of 20- and 30-min to ensure the precise administration of the delayed recall trials in the event of a crash. We also placed each program file in the same directory as the batch program, so the experimenter could manually execute a test program at the right time if the automatic execution was not working.

3.10. Participants and COVID-19 safety precautions

Participants in the PAAD-2 trial are middle-aged (40–65 years) adults with a family history of Alzheimer's disease, who are cognitively normal, healthy enough for exercise, not otherwise clinically impaired, and identified as sedentary based on American College of Sports Medicine (ACSM)'s physical activity guidelines [43]. The inclusion criteria did not specifically include any criteria that put participants into the Centers of Disease Control (CDC)'s high-risk category when originally defined. However, as of June 25, 2020, the CDC removed the specific age threshold of >65 years and replaced that with a statement that risk increases with increasing age [44]. Prior to scheduling participants, we discuss the CDC's risk guidance to ensure that they are aware of their own personal risk category classification. Within 24 h of scheduled visits, the experimenter and participants are required to complete a COVID-19 screening form. This allowed for the reporting of any COVID-19 symptoms or positive diagnosis, any contacts with people having COVID-19 symptoms or positive diagnosis, and/or any travel(s) outside the state in the past 14 days. We also used this screening to identify if participants had additional factors that would put them at increased risk for serious health consequences when contracting COVID-19 [45]. For participants identified as having high risk of serious consequences of a COVID-19 infection [46], we discussed this with the participant prior to scheduling.

We also used additional safety precautions including ensuring that they were the first or only person to complete cognitive testing or that they were scheduled for testing more than one hour following a previous participant. For all participants, when the experimenter or a participant entered the testing room, they were first required to sanitize their hands. After each testing session, all devices on the desks were wiped off. We covered the participant's keyboard with a transparent plastic slip and exchanged it with a new one after each testing session. We wiped off the file tray, the pencils and pen, and the experimenter's and participant's chairs after each testing session, and also switched all of these pieces of equipment with another set of equipment after each participant.

4. Protocol evaluation

4.1. Participants

We have safely and efficiently operated the modified protocol since July 2020 for the pre-test ($n = 37$), mid-test ($n = 7$), and post-test ($n = 15$). We compared the pre-test data of the modified protocol with that of the original protocol ($n = 32$) in terms of the test duration (Table 1) and test performance (Table 2). We describe demographics for participants completing the original and modified protocol in Table 2.

4.2. Data analyses

We conducted a series of one- or two-sample *t*-tests and equivalence tests (ET) to detect significant differences or equivalence between data collected using the original and modified protocols. The goal of ET is to examine whether the null hypothesis that there is significant difference between two parties can be actually rejected, which is exactly opposite to the traditional comparative study (e.g., *t*-test) examining a null hypothesis that there is no meaningful difference between two approaches [47,48]. Significant equivalence is determined with the equivalence margin (δ), the maximum acceptable range of values in which the subtle difference must fall to be considered equivalent [47].

The ET complements the traditional hypothesis testing and vice versa. For example, when the null hypothesis of a traditional *t*-test is accepted, the absence of a true effect is supported but not statistically verified; ET can statistically uphold this case [49]. ET can also identify significantly greater than zero but negligible effect, when it is smaller than the meaningful effect size by falling in the equivalence margin [47]. To examine whether the presence of meaningful differences between two protocols can be rejected, we followed two one-sided tests (TOST) procedure, an established method of ET [48,49], with an upper and lower equivalence margin at ±0.7, which was determined in consideration of the sample size [49]. All statistical analyses were conducted with R 4.0.3 [50].

Table 2
Comparisons of performance between the original and modified PAAD-2 cognitive testing protocol.

Participants' characteristics		Original ($n = 32$)	Modified ($n = 37$)	ET (p)	t- or z-test (p)
	Age, mean (SD)	56.9 ± 1.2	56.1 ± 0.85	< 0.05	0.60
	Female gender (%)	31 (96.9%)	31 (83.8%)	< 0.01	0.16[†]
	Race/ethnicity, non-Hispanic white (%)	28 (87.5%)	35 (94.6%)	< 0.01	0.54[†]
	Years of education, mean (SD)	16.0 ± 0.33	16.9 ± 0.36	0.13	0.09
Testing batches and components		Comparisons of test scores (%, $M \pm SE$)		ET (p)	t-test (p)
0	MoCA	28.2 ± 0.25	28.4 ± 0.22	< 0.01	0.76
1	Test of Premorbid Functioning, %	86.2 ± 1.78	82.1 ± 1.73	0.11	0.11
	NIH TB List Sort Working Memory Score	53.8 ± 1.33	54.5 ± 1.25	< 0.01	0.73
	NIH TB Picture Sequence Memory Score	55.1 ± 2.33	57.0 ± 1.56	< 0.05	0.48
	MST - Lure Discrimination Index, %	12.5 ± 2.30	21.1 ± 3.22	0.24	<0.05
	MST - Old Discrimination Index, %	84.2 ± 1.90	81.1 ± 1.50	0.06	0.20
	PDT - Perceptual Discrimination Index, %	83.8 ± 2.01	83.2 ± 2.39	< 0.05	0.83
2	ROCFT - Copy, %	95.7 ± 1.38	98.2 ± 0.51	0.13	0.10
	ROCFT - Immediate Recall, %	53.4 ± 5.96	70.5 ± 5.23	0.23	< 0.05
	Stroop Effect, %	21.8 ± 3.17	23.8 ± 2.42	< 0.01	0.62
	PASAT - 3 s, %	79.2 ± 3.33	81.6 ± 2.87	< 0.05	0.59
	PASAT - 2 s, %	57.0 ± 2.93	64.3 ± 2.50	0.16	0.06
	ROCFT - Delayed Recall	51.4 ± 6.20	67.2 ± 5.81	0.15	0.07
	ROCFT - Delayed Recognition	86.5 ± 0.95	88.1 ± 1.24	< 0.05	0.32
	TOL - Planning Score	47.5 ± 6.23	54.3 ± 4.89	< 0.05	0.38
3	RAVLT - Learning Summary, %	59.8 ± 2.06	64.5 ± 1.87	0.12	0.09
	RAVLT - Immediate Recall, %	64.6 ± 3.78	71.1 ± 3.11	0.06	0.19
	NIH TB Dimensional Change Card Sort Score	55.5 ± 2.31	55.1 ± 1.83	< 0.01	0.90
	NIH TB Flanker Inhibitory Control Score	42.9 ± 1.46	42.5 ± 1.62	< 0.01	0.86
	Spatial Working Memory Score	89.0 ± 3.85	92.7 ± 3.36	< 0.05	0.47
	Trail Making - A, *sec*	32.0 ± 1.44	35.5 ± 1.58	0.10	0.11
	Trail Making - B, sec	52.3 ± 3.00	52.2 ± 2.48	< 0.01	0.98
	RAVLT - Delayed Recall, %	64.2 ± 4.27	69.5 ± 3.20	< 0.05	0.32
	RAVLT - Delayed Recognition, %	85.2 ± 2.97	83.6 ± 2.01	< 0.01	0.66
4	Paired Associates - Learning Summary, %	43.1 ± 4.30	49.1 ± 4.42	< 0.05	0.34
	Matrix Reasoning, %	49.2 ± 3.12	51.2 ± 2.90	< 0.01	0.63
	Digits Span Forward Score, %	58.7 ± 2.50	55.6 ± 2.25	< 0.05	0.36
	Digits Span Backward Score, %	38.2 ± 2.60	36.3 ± 2.79	< 0.01	0.63
	Paired Associates - Delayed Recall, %	32.6 ± 4.51	36.0 ± 4.67	< 0.05	0.60
5	Logical Memory - Learning Summary, %	60.1 ± 2.47	61.2 ± 1.79	< 0.01	0.72
	Spatial Relations, %	44.4 ± 3.85	50.4 ± 3.41	< 0.05	0.24
	SDMT - Written, %	48.6 ± 1.52	47.1 ± 1.03	< 0.05	0.43
	SDMT - Oral, %	55.7 ± 1.84	55.4 ± 1.67	< 0.01	0.89
	SDMT - Incidental Learning, %	63.1 ± 5.01	68.3 ± 4.46	< 0.01	0.89
	Logical Memory - Delayed Recall, %	56.2 ± 2.90	56.2 ± 1.9	< 0.01	0.99

Abbreviation: MoCA, Montreal Cognitive Assessment; MST, Mnemonic Similarity Task; PASAT, Paced Auditory Serial Addition Test; PDT, Perceptual Discrimination; RAVLT, Rey Auditory Verbal Learning Test; ROCFT, Rey-Osterrieth Complex Figure Test; SDMT, Symbol Digits Modalities Test.
Note: Two-sample t- or z-tests or equivalence tests (ET) were conducted to compare the demographics and cognitive performance between two groups of participants of the original and modified protocols. Results indicate that participants of two protocols are significantly equivalent and not significantly different in their ages and 94% of test scores. Equivalence margin (Cohen's d, δ) was ±0.7. [†]Results of Fisher's exact probability tests with Yates continuity correction.

4.3. Results

As expected, safety precautions have been effective such that none of the experimenters or participants have contracted COVID-19 in our testing environment. All participants have expressed a clear understanding of the instructions and of the feedback from the synthetic voice. We asked the participants to press a key to repeat the synthesized instructions or request clarification when unclear, yet they rarely pressed the key to replay any instructions or ask clarifying questions (< 1% of the total instructions). Automatic control of the testing sequence and timing functioned well without causing any significant error or delay during the testing. All data including audio recordings of verbal responses, hand-written and -drawn responses on paper forms, and keyboard and mouse responses on the computer have been safely acquired.

Results indicate that the modified protocol is significantly equivalent to the original protocol in terms of its duration (Table 1) and participants' age and performance (Table 2). No significant differences were found in gender ($p = .16$), race/ethnicity ($p = .54$), or years of education ($p = .09$) between the two groups. Although years of education was not significantly equivalent ($p = .13$), gender ($p = .01$) and race/ethnicity ($p = .01$) were significantly equivalent. As can be seen in Table 1, the time control of the modified protocol was robust and the 20- and 30-min time delays were accurately maintained. The only tests for which duration of the modified protocol was significantly different from the original protocol was the ROCFT copy ($p < .001$), immediate recall ($p < .01$), and delayed recall ($p < .05$), for which the task time is completely determined by the participant with no time limit. For all of these, participants in the modified protocol took significantly longer to complete the task than those who used the original protocol. This longer duration of task completion is likely reflected in the marginally higher performance during the modified protocol for copy ($p = .10$), immediate recall ($p < .05$), and delayed recall ($p = .07$). The only other performance difference was found for the MST lure discrimination index ($p < .05$). All other test scores were not significantly different and, in most cases, significantly equivalent between two protocols (see Table 2).

5. Discussion

In this modified protocol for the PAAD-2 cognitive testing, we describe the specific methods of the implementation of TTS synthesis and computer programming techniques and their benefits for safety and the integrity of cognitive assessment. The adoption of the techniques from AI and computer vision packages enabled us to provide standardized instructions and feedback without human speech and to closely view paper copies of documents from a safe distance (farther than 6 ft)

for an extensive cognitive testing protocol during a pandemic. According to the feedback from experimenters and the consistently positive responses from participants, the modified testing procedures have provided a safe and pleasant environment for cognitive assessment for both the experimenter and participants. This is critical for the prevention of COVID-19 but also for the provision of accurate and standardized verbal instructions compared with speaking under a face covering from the required social distance.

We also evaluated the integrity of the modified protocol and substantiated that the modified protocol is robust and generally equivalent to the original protocol in terms of its duration and participants' performance. The automated control of test timing and sequence we developed functioned flawlessly and required less training for the experimenter than traditional human-led administration of cognitive tests. Our interpretation of the marginal differences from the original protocol for the ROCFT test duration is that the experimenter in the original protocol often asked whether the drawing tasks were finished from a close distance, which could have functioned as a prompt to stop the task. By contrast, in the modified protocol, participants self-initiated and finished the drawing task without any prompt and the experimenter was a distance away and not directly observing their behaviors. The longer duration of drawing tasks could be associated with the learning and memory performance.

We acknowledge the limitation of comparing two different groups of participants for the evaluation of protocol legitimacy. Although no significantly different demographic characteristics were detected between the two groups, it is possible that there may be marginal differences in terms of other unmeasured variables between the two groups that could affect cognitive performance. We will carefully consider any marginal differences between the two protocols as the study continues and when analyzing the study outcomes.

Employing speech synthesis technique for neurocognitive testing in the pandemic has the clear advantage of mitigating the risk of the transmission of the virus. Recent studies have revealed that small speech droplets generated by ordinary speaking could remain airborne for extended periods of time and therefore it is highly possible that normal talking causes airborne viral transmission of the COVID-19 virus in confined environments [7,8]. In addition to wearing a face covering, having a computer speak for all necessary instructions and feedback further contributed to eliminating the production of speech droplets and thus substantially decreased the risks for COVID-19 in a confined laboratory environment. We asked about 15 participants to provide either positive or negative feedback on our new testing session and received only positive comments: “Loved the avatar Dr. Shin [Dr. Shin Park] created limiting amount of talking person to person.” and “Everything was extremely safe to the point of over safe. But very much appreciated.”

Moreover, using a synthetic voice facilitates the testing procedure by limiting the extent to which the experimenter must speak while wearing a face covering. Recent evidence indicates that wearing face coverings during professional and essential activities increased the perception of vocal fatigue and discomfort, difficulties in understanding speech, auditory feedback, and difficulties in coordinating speech and breathing [11]. Using synthetic voice eliminates such difficulties and thus the reduces the chance of the experimenter being fatigued and making errors in testing instructions and feedback during an extensive testing session. In this regard, the computerized testing is more rigorous and standardized than the experimenter-led version once properly administered and more easily trainable across administrators at different levels.

Ethical and practical challenges must be considered relative to human research activities during the COVID-19 pandemic [2]. Such challenges include but are not limited to: What level of risk for disease transmission is acceptable to resume in-person human subject research? What safety precautions are mandatory? To address this challenge, researchers developed a risk-benefit framework to prioritize studies in tiers (0,1,2,3) based on a combination of the incremental risk of COVID-19 transmission (high, medium, low, or none) introduced by the research activity and the potential benefits of study participation (1–4) at an individual level [2]. The framework considers contact distance, contact duration, number of contacts per day, personal protective equipment, and participant characteristics (e.g., age, medical condition, risk of contracting COVID-19). Our study would be considered as a tier 2 study with a low risk based on the contact-free administration with the state-of-the-art technologies and safety precautions. Further discussion is needed how to consider our safety precautions into the risk-benefit framework and how to efficiently utilize the modern technology we employed in other research settings.

In recent decades, speech synthesis technology has been widely applied to commercially-available mobile devices and computers [14,15] and also efficiently utilized in the fields of healthcare [51] and education [52]. For example, synthesized speech has been used for an interactive medication reminder and tracking on wrist devices [53], as a clinical assistant for visually impaired people [54], and other assistive device, speech-based healthcare apps, websites, and/or emergency call centers [51]. A comprehensive review on the use of speech technology for healthcare was recently published [51]. Nonetheless, the application of speech technology for behavioral experiments or psychological assessment remains at a rudimentary level. To our knowledge, this innovative protocol for the PAAD-2 is the seminal attempt for synthesized voices to completely replace human speech for informed consent and verbal instructions and feedback for a comprehensive battery of laboratory-based cognitive assessment. This technique can be used for older adults with sensory or cognitive impairments by adjusting the pace of speech to help with their understanding. Other low risk means of behavioral experiments are available such videoconference [55], telephone [56], or web-based software [57]. Such online-based methodologies are beneficial in its mobility and accessibility but limited in its level of precision compared with lab-based systems and have slightly more variability in its measures [57].

We plan to use this methodology in the future even after the pandemic for its benefits for safety management, standardization of test directions, precise control over test timing, automation of test sequences and execution, efficiency of data collection procedures, and the integrity of data obtained. Our employment of the AI-based methods may be informative for other researchers interested in employing safe, rigorous, and automated laboratory tests during and after the pandemic.

Funding

This work has been completed as part of a phase 2 clinical trial (ClinicalTrials.gov NCT03876314), “The Effect of Physical Activity on Cognition Relative to APOE Genotype (PAAD-2)”, which is funded by the National Institutes of Health (R01AG058919). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Ethics approval and consent to participate

Approval for this study was obtained from the Institutional Review Board of the University of North Carolina at Greensboro (IRB number 18–0228). Informed consent was obtained from all individual participants included in the study at the first in-person visit at the pre-test.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Consent for publication

Not applicable.

Authors' contributions

KSP initiated the use of speech synthesis techniques, configured the contactless test settings, computerized the entire protocol, operated and tested the protocol for data collection, and wrote the draft and final versions of the manuscript, along with significant contributions from JLE who designed the original protocol, gave input into the conception of the revised protocol, assisted with pilot testing and the provision of protocol modifications, and edited and finalized the manuscript. All authors read and approved the final manuscript and accept personal responsibility for the accuracy and integrity of the presentation of this protocol.

Declaration of Competing Interest

The authors declare no commercial, financial or any other conflict of interest in this research.

References

[1] M.B. Omary, et al., The COVID-19 pandemic and research shutdown: staying safe and productive, J. Clin. Invest. 130 (6) (2020) 2745–2748.
[2] J.C. Lumeng, et al., Opinion: a risk–benefit framework for human research during the COVID-19 pandemic, Proc. Natl. Acad. Sci. 117 (45) (2020) 27749–27753.
[3] K.R. Myers, et al., Unequal effects of the COVID-19 pandemic on scientists, Nat. Hum. Behav. 4 (9) (2020) 880–883.
[4] N.S. Wigginton, et al., Moving academic research forward during COVID-19, Science 368 (6496) (2020) 1190–1192.
[5] K.S. Park, et al., The effect of physical activity on cognition relative to APOE genotype (PAAD-2): study protocol for a phase II randomized control trial, BMC Neurol. 20 (1) (2020) 231.
[6] C. Rothe, et al., Transmission of 2019-nCoV infection from an asymptomatic contact in Germany, N. Engl. J. Med. 382 (10) (2020) 970–971.
[7] V. Stadnytskyi, et al., The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission, Proc. Natl. Acad. Sci. 117 (22) (2020) 11875–11877.
[8] P. Anfinrud, et al., Visualizing speech-generated oral fluid droplets with laser light scattering, N. Engl. J. Med. 382 (21) (2020) 2061–2063.
[9] L. Morawska, J. Cao, Airborne transmission of SARS-CoV-2: the world should face the reality, Environ. Int. 139 (2020) 105730.
[10] L. Morawska, D.K. Milton, It is time to address airborne transmission of coronavirus disease 2019 (COVID-19), Clin. Infect. Dis. 71 (9) (2020) 2311–2313.
[11] V.V. Ribeiro, et al., Effect of wearing a face mask on vocal self-perception during a pandemic, J. Voice (2020). In press.
[12] Y. Wang, et al., Tacotron*: towards end-to-end speech synthesis*, in: INTERSPEECH, 2017.
[13] J. Shen, et al., Natural TTS synthesis by conditioning wavenet on MEL spectrogram predictions, in: 2018 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.
[14] Pew Research Center, *Nearly half of Americans use digital voice assistants, mostly on their smartphones*. 2017 Dec 12th Sep 23, 2020, Available from: http://www.pewrsr.ch/2kquZ8H.
[15] V. Petrock, *Voice Assistant Use Reaches Critical Mass*. 2019 Aug 15th Sep 23, 2020, Available from: https://www.emarketer.com/content/voice-assistant-use-reaches-critical-mass.
[16] A. van den Oord, et al., *Wavenet: A generative model for raw audio.* arXiv preprint. arXiv:1609.03499, 2016.
[17] A. van den Oord, et al., Parallel wavenet: fast high-fidelity speech synthesis, in: D. Jennifer, K. Andreas (Eds.), Proceedings of the 35th International Conference on Machine Learning, 2018, pp. 3918–3926. Editors. PMLR: Proceedings of Machine Learning Research.
[18] P. Wagner, et al., Speech Synthesis Evaluation - State-of-the-Art Assessment and Suggestion for a Novel Research Program, 2019.
[19] J. Cambre, et al., Choice of voices: a large-scale evaluation of text-to-speech voice quality for long-form content, in: *Proceedings of the* 2020 *CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, Honolulu, HI, USA, 2020, pp. 1–13.
[20] Google Cloud, Text-to-Speech Documentation, Available from: https://cloud.google.com/text-to-speech/docs, 2020.
[21] J. Peirce, et al., PsychoPy2: experiments in behavior made easy, Behav. Res. Methods 51 (1) (2019) 195–203.
[22] Z.S. Nasreddine, et al., The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment, J. Am. Geriatr. Soc. 53 (4) (2005) 695–699.
[23] D. Wechsler, *Test of Premorbid Functioning*. UK Version (TOPF UK), Pearson Corporation, UK, 2011.
[24] P.A. Osterrieth, Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. [Test of copying a complex figure; contribution to the study of perception and memory.], Arch. Psychol. 30 (1944) 206–356.
[25] A. Rey, L'examen clinique en Psychologie, Presses universitaires de France, Paris, 1964.
[26] D.M. Gronwall, Paced auditory serial-addition task: a measure of recovery from concussion, Percept. Mot. Skills 44 (2) (1977) 367–373.
[27] C.R. Reynolds, Comprehensive Trail-Making Test Examiner's Manual, PRO-ED, Inc., Austin, Texas, 2002.
[28] A. Smith, Symbol Digit Modalities Test Manual (W-129C), Western Psychological Services, Torrance, CA, 2011.
[29] Psychology Software Tools Inc., E-Prime 3.0, 2016. Retrieved from https://www.pstnet.com: Pittsburgh, PA.
[30] S. Weintraub, et al., Cognition assessment using the NIH Toolbox, Neurology 80 (11 Suppl 3) (2013) S54–S64.
[31] P. Arnold, F. Hill, Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact, Br. J. Psychol. 92 (2) (2001) 339–355.
[32] C.R. Bowie, P.D. Harvey, Administration and interpretation of the Trail Making Test, Nat. Protoc. 1 (5) (2006) 2277–2281.
[33] G. Bradski, The OpenCV Library. Dr Dobb's J. Software Tools 25, 2000, pp. 120–125.
[34] G. Van Rossum, F.L. Drake, The Python Language Reference Manual, Network Theory Ltd., 2011.
[35] J. Peirce, Generating stimuli for neuroscience using PsychoPy, Front. Neuroinform. **2** (10) (2009).
[36] J. Peirce, PsychoPy—psychophysics software in Python, J. Neurosci. Methods 162 (1) (2007) 8–13.
[37] D. Bridges, et al., The timing mega-study: comparing a range of experiment generators, both lab-based and online, PeerJ 8 (2020), e9414.
[38] P. Garaizar, M.A. Vadillo, Accuracy and precision of visual stimulus timing in PsychoPy: no timing errors in standard usage, PLoS One 9 (11) (2014), e112033.
[39] J. Peirce, PsychoPy - Psychology Software for Python, 2020, p. 540.
[40] Python Software Foundation, OS - miscellaneous operating system interfaces, in: Python 3.9.0 Documentation - The Python Standard Library, Python Software Foundation, 2020.
[41] K.I. Erickson, et al., Investigating gains in neurocognition in an intervention Trial of Exercise (IGNITE): protocol, Contemp Clin. Trials 85 (2019) 105832.
[42] C.P. Kaller, et al., Vienna test system manual, in: Tower of London - Freiburg Version, Schuhfried GmbH, Mödling, Austria, 2011.
[43] G. Liguori, ACSM, in: G. Liguori (Ed.), ACSM's Guidelines for Exercise Testing and Prescription, 11th ed, Lippincott Williams and Wilkins, Philadelphia, PA, 2021.
[44] CDC, *CDC Updates, Expands List of People at Risk of Severe COVID-19 Illness*. 2020 June 25, Available from: https://www.cdc.gov/media/releases/2020/p0625-update-expands-covid-19.html, 2020.
[45] CDC, *Coronavirus Disease 2019 (COVID-19)*. People at Increased Risk 2020, Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-increased-risk.html.
[46] CDC, *People with Certain Medical Conditions*. 2021 Mar. 29, Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html, 2021.
[47] E. Walker, A.S. Nowacki, Understanding equivalence and noninferiority testing, J. Gen. Intern. Med. 26 (2) (2011) 192–196.
[48] D. Lakens, A.M. Scheel, P.M. Isager, Equivalence testing for psychological research: a tutorial, Adv. Methods Pract. Psychol. Sci. 1 (2) (2018) 259–269.
[49] D. Lakens, Equivalence tests, Soc. Psychol. Personal. Sci. 8 (4) (2017) 355–362.
[50] R. Core Team, The R Project for Statistical Computing, The R Foundation, Vienna, Austria, 2020.
[51] S. Latif, et al., *Speech technology for healthcare: opportunities, challenges, and state of the art. IEEE Rev. Biomed. Eng.* 14, 2020, pp. 342–356.
[52] Z. Handley, Is text-to-speech synthesis ready for use in computer-assisted language learning? Speech Comm. 51 (10) (2009) 906–919.
[53] A.S. Mondol, I.A. Emi, J.A. Stankovic, MedRem: an interactive medication reminder and tracking system on wrist devices, in: 2016 IEEE Wireless Health (WH), 2016.
[54] K.-C. Liu, et al., Voice Helper: a mobile assistive system for visually impaired persons, in: 2015 IEEE International Conference on Computer and Information Technology, IEEE, 2015.
[55] J.E. Chapman, et al., Comparing face-to-face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke, J. Telemed. Telecare (2019), 1357633X1989078.
[56] M.J. Katz, et al., T-MoCA: A valid phone screen for cognitive impairment in diverse community samples, Alzheimer's Dementia: Diagn. Assess. Dis. Monitor. **13** (1) (2021).
[57] D. Bridges, et al., The timing mega-study: comparing a range of experiment generators, both lab-based and online, PeerJ 8 (2020) e9414.