Parkinsonism and Related Disorders 99 (2022) 43–46

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

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

Short Communication

Can AI make people happy? The effect of AI-based chatbot on smile and speech in Parkinson's disease

Mayuko Ogawa [a,b,1], Genko Oyama [a,b,c,d,e,f,*,1], Ken Morito [g], Masatomo Kobayashi [h], Yasunori Yamada [h], Kaoru Shinkawa [h], Hikaru Kamo [a], Taku Hatano [a,b,d], Nobutaka Hattori [a,b,c,d,e,f,i]

[a] Department of Neurology, Juntendo University Graduate School of Medicine, Tokyo, Japan
[b] Department of Neurodegenerative and Demented Disorders, Juntendo University Graduate School of Medicine, Tokyo, Japan
[c] Department of Home Medical Care System Based on Information and Communication Technology, Juntendo University Graduate School of Medicine, Tokyo, Japan
[d] Department of Drug Development for Parkinson's Disease, Juntendo University Graduate School of Medicine, Tokyo, Japan
[e] Department of Patient Reported Outcome Based Integrated Data Analysis in Neurological Disorders, Juntendo University Graduate School of Medicine, Tokyo, Japan
[f] Department of Research and Therapeutics for Movement Disorders, Juntendo University Graduate School of Medicine, Tokyo, Japan
[g] GLORY Ltd., Tokyo, Japan
[h] IBM Research, Tokyo, Japan
[i] Neurodegenerative Disorders Collaborative Laboratory, RIKEN Center for Brain Science, Saitama, Japan

A R T I C L E I N F O

Keywords:
Parkinson's disease
Telemedicine
Facial expression
Speech

A B S T R A C T

Introduction: Approaches for objectively measuring facial expressions and speech may enhance clinical and research evaluation in telemedicine, which is widely employed for Parkinson's disease (PD). This study aimed to assess the feasibility and efficacy of using an artificial intelligence-based chatbot to improve smile and speech in PD. Further, we explored the potential predictive value of objective face and speech parameters for motor symptoms, cognition, and mood.

Methods: In this open-label randomized study, we collected a series of face and conversational speech samples from 20 participants with PD in weekly teleconsultation sessions for 5 months. We investigated the effect of daily chatbot conversations on smile and speech features, then we investigated whether smile and speech features could predict motor, cognitive, and mood status.

Results: A repeated-measures analysis of variance revealed that the chatbot conversations had a significant interaction effect on the mean and standard deviation of the smile index during smile sections (both $P = .02$), maximum duration of the initial rise of the smile index ($P = .04$), and frequency of filler words ($P = .04$), but no significant interaction effects were observed for clinical measurements including motor, cognition, depression, and quality of life. Explorative analysis using statistical and machine-learning models revealed that the smile indices and several speech features were associated with motor symptoms, cognition, and mood in PD.

Conclusion: An artificial intelligence-based chatbot may positively affect smile and speech in PD. Smile and speech features may capture the motor, cognitive, and mental status of patients with PD.

1. Introduction

Telemedicine and telehealth using video-conferencing systems afford a solution to improve access to specialists for patients with Parkinson's disease (PD). A large portion of telemedicine visits is composed of conversation [1]. Therefore, technologies for the objective measurement of facial expressions and voice may enhance clinical and research evaluation in telemedicine.

Automated recognition of facial expressions and speech is a cutting-edge technology that has recently emerged. Facial expression recognition technology is widely used in various situations, such as security systems, although ethical concerns have been raised [2]. Automatic speech recognition (ASR) and natural language processing (NLP) have evolved based on the extensive development of machine-learning

* Corresponding author. Department of Neurology, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo, 113-8421, Japan.
E-mail address: g_oyama@juntendo.ac.jp (G. Oyama).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.parkreldis.2022.04.018
Received 22 February 2022; Received in revised form 27 April 2022; Accepted 28 April 2022
Available online 5 May 2022
1353-8020/

approaches, which enable use in various situation, such as creating translation engines. In medicine, application of speech analysis technologies has been harnessed for the early detection of dementia [3], and speech analysis in PD [4–6]. In PD, an increase in the frequency of filler words, pitch variability, and Mel-Frequency Cepstrum Coefficients (MFCCs) have been reported to be associated with disease severity [4]. Decreased speech rate associated with disease duration [5]. Shimmer and jitter increased with the Unified Parkinson's Disease Rating Scale (UPDRS) score, representing a reduction in voice quality measured by the cycle-to-cycle variability in amplitude and pitch, respectively [6]. Artificial intelligence (AI)-based chatbots, a technology using ASR and NLP to simulate conversations with users in natural language, are widely employed (e.g., Apple's Siri and Amazon's Alexa). Chatbots may offer scalability and 24-h availability to plug the gaps between patients and clinicians by gathering patients' health-related information during daily life for chronic diseases, including PD [7].

In this study, we developed an AI-based chatbot app for iPad to assist telemedicine for PD that collects patients' health information, including motor and non-motor problems and general information regarding their daily life (e.g., their hobbies, favorite foods, weekend activities, and private topics) through daily conversations, and a video-conferencing app that collects face and voice data remotely. This study aimed to assess the feasibility and efficacy of using an artificial intelligence-based chatbot to improve smile and speech in PD. We hypothesized that chatbots would have a positive effect on smile and speech. In addition, we aimed to explore the potential predictive value of objective face and voice parameters for motor symptoms, cognition, and mood.

2. Methods

Twenty patients with PD (11 men and 9 women) were recruited from the outpatient clinic of Juntendo University Hospital. Inclusion criteria were: (1) a diagnosis of clinically established or probable PD according to the Movement Disorder Society (MDS) clinical diagnostic criteria for PD; (2) native Japanese speaker; and (3) patients aged between 20 and 80 years who signed a written consent form after receiving a complete explanation of the research. Exclusion criteria were: (1) cognitive impairment, operationalized as a Mini-Mental State Examination (MMSE) score of <20; (2) severe speech problems undetectable by a tablet microphone; and (3) individuals who were unable to complete the study for any reason.

This study comprised a trial phase and a randomized phase (Fig. 1A). In the trial phase, the content of the chatbot was revised based on feedback from the participants and neurologists, and finalized after conclusion of the trial phase. In the 5-month randomized phase, participants were randomized at a 1:1 ratio to an intervention group that received both daily chatbot and weekly video-conferencing sessions or a control group that received weekly video-conferencing sessions only. Simple randomization was performed using a random number table generated by a computer.

For each chatbot session, participants had a multi-turn conversation with the chatbot app comprising at least five pairs of questions and

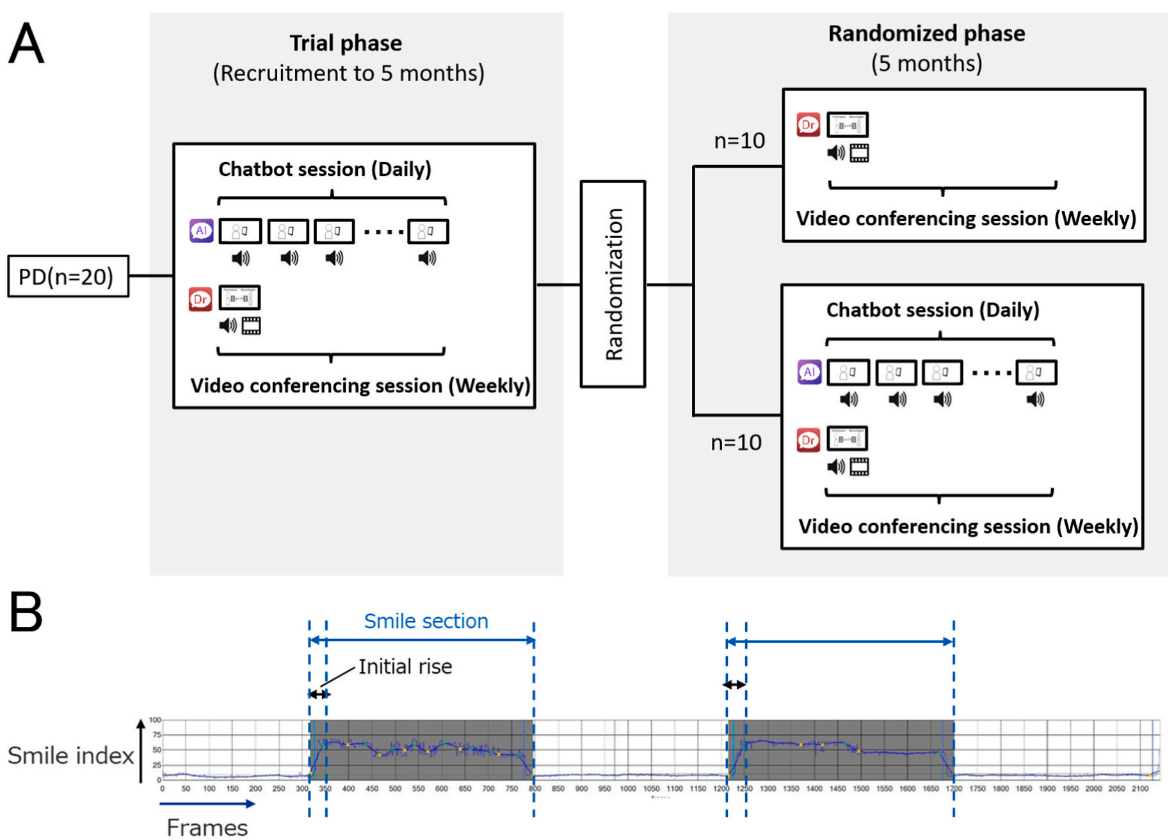

Fig. 1. Study design. A) Study protocol. In the trial phase, participants used an AI-based chatbot app daily for 1–4 months after providing written consent to participate. During the trial phase, participants participated in weekly video-conferencing sessions with a neurologist and daily conversations with an AI-based chatbot at least once. In the 5-month randomized phase, participants were randomized at a 1:1 ratio to an intervention group that received both daily AI-based chatbot and weekly video-conferencing sessions or a control group that received weekly video-conferencing sessions only. Simple randomization was performed using a random number table generated by a computer. During each chatbot session, audio samples were recorded, and video and audio samples were recorded during each video-conferencing session.

B) The smile index. The smile index was calculated from the degree of deference from baseline based on the GLORY CO Ltd facial recognition library. The system calculated the total smile index based on the differences from the references, which were also calculated from the database. The smile index ranged from 0 (straight face) to 100 (smile). We determined the "smile section" based on differential calculus of the time-series data of the smile index during video-conferencing sessions.

responses. The conversation content simulated a typical teleconsultation and included general conditions, changes in symptoms, and problems recently experienced in daily life. The chatbot also asked about participants' hobbies, favorite foods, and relevant topics to create a favorable atmosphere. Finally, the chatbot generated a report about the session on the dashboard for neurologists. Detail of apps is available in supplementary data.

In each video-conferencing session with a neurologist, the participant had at least 5 min of conversation. The neurologist asked patients about general conditions, changes in symptoms, and problems recently experienced in daily life. This session did not include any clinical decisions, such as changing medications. For the intervention group, the neurologist could see the dashboard screen depicting an overview of the participants' chatbot sessions.

Participants underwent an in-person clinical assessment by a neurologist at the time of recruitment, before and after the intervention (visit 1–3, respectively). The following scales were administered when the participants were on medications: the MDS-sponsored revision of the UPDRS (MDS-UPDRS), MMSE and Japanese version of Montreal Cognitive Assessment (MoCA-J), Beck Depression Inventory-II (BDI-II), and Parkinson's Disease Questionnaire-39 (PDQ-39). Facial expressions were videotaped using an iPad Air 2 (Apple Inc., Cupertino, CA, USA) and c922 pro stream webcam (Logitech International S.A., Switzerland) as reference data for subsequent analysis, which were placed in front of the patients at 50–55 cm apart at face height. We recorded two sets of "straight face" and "smile face", which were obtained by instructing participants to imitate the sample picture of a smile. Participants could see their faces during the recording with the iPad monitor.

This study was conducted in accordance with the ethical standards of the Declaration of Helsinki. This study was approved by the Institutional Review Board of Juntendo University Hospital (#19-005).

The detail of video and audio data analysis is available on supplementary data. To assess facial expression features, we developed a "smile index." We calculated nine facial expression features including the mean, maximum, and standard deviation of all smile indices during each smile section and the mean, maximum, and minimum duration of smiles and the initial rise of the smile index (Fig. 1B).

We extracted primary and exploratory sets of speech features from audio samples. The primary set comprised five speech features associated with PD severity in the literature as potential measures of the effects of interventions on PD, including the frequency of filler words, speech rate, pitch variability, jitter, and shimmer. Filler words such as "uh" were automatically detected using IBM Watson Speech to Text. The exploratory set comprised 75 speech features associated with motor, cognitive, and mood disorders, including PD, dementia, and depression (Supplementary data).

We applied repeated-measures analysis of variance (ANOVA) for clinical, facial expression, and speech measurements with a 2×2 mixed design, with group (intervention and control) as the between-subject factor and time (pre- and post-intervention) as the within-subject factor, after assessing normality of the data using the Shapiro-Wilk test ($P > .05$). Speech and facial expression measures were grouped into former and latter groups, which were considered pre- and post-intervention, respectively. The level of statistical significance, P, was set at 0.05 (two-sided). All analyses were performed with R 4.0.5, Python 3.6.6, SciPy 1.1.0, and scikit-learn 0.23.2. The detail of exploratory analysis is available on supplementary data.

3. Results

Supplemental Table 1 shows the clinical characteristics of the participants. None of the participants had medical conditions which could affect speech and facial expression. Except for one participant that could not attend visit 3, all participants completed the clinical assessments at all three visits. During the randomized phase, participants in both groups completed 13–20 video-conferencing sessions. Each participant in the intervention group completed 58 ± 155 chatbot sessions at home (Supplemental Table 2). Video and audio data from 396 video-conferencing sessions were collected. Video and audio samples that failed to meet the criteria were excluded, and 323 video samples and 298 audio samples were included in the analysis. In addition, 39 samples from the in-person clinical assessment, and 356 samples from chatbot sessions were included in the explorative analysis.

Repeated-measures ANOVA revealed no significant main effects of *group* or *time* and no significant interaction effect of *group* × *time* for all clinical measures ($P > .05$ for MDS-UPDRS part I to IV, MMSE-J, MoCA-J, BDI-II, and PDQ-39 summary index). Repeated-measures ANOVA revealed a significant interaction effect of *group* × *time* on the mean smile index during the smile section ($F_{(1,18)} = 5.96$, $P = .02$), standard deviation of smile index during the smile section ($F_{(1,18)} = 5.39$, $p = .02$), and maximum duration of the initial rise of smile index ($F_{(1,18)} = 4.44$, $P = .04$). Specifically, in the intervention group, these features increased by 11.0%, 10.2%, and 67.7%, respectively, while in the control group, they decreased by 7.9%, 6.7%, and 36.8%, respectively. There were no significant effects of *group* or *time* on the other facial expression features ($P > .05$).

Repeated-measures ANOVA revealed a significant interaction effect of *group* × *time* on filler words ($F_{(1,18)} = 4.98$, $P = .04$). Specifically, the frequency of filler words decreased by 8.6% and increased by 22.8% in the intervention and control groups, respectively. Significant main effects of *time* on pitch variability ($F_{(1,18)} = 7.2$, $p = .02$), shimmer ($F_{(1,18)} = 11.4$, $P = .003$), and jitter ($F_{(1,18)} = 7.0$, $P = .02$) were observed. Specifically, these speech features decreased by 4.2–5.4% in the latter sessions. No significant main effect of *group* was noted for the other speech features ($P > .05$).

Explorative analysis revealed that the multiple regression models using smile features predicted MoCA-J ($r = 0.41$, $P = .010$), MMSE ($r = 0.45$, $P = .004$), part IV of MDS-UPDRS ($r = 0.50$, $P = .001$), MDS-UPDRS I.1 (cognition; $r = 0.53$, $P < .001$), and MDS-UPDRS III.14 (general bradykinesia; $r = 0.62$, $P < .001$). The classification accuracies of the machine-learning models using video-conferencing speech were >80%. The accuracies using AI-based chatbot speech were ≥75% for all three aspects. The detail of the result of explorative analysis is in supplementary data.

4. Discussion

This study demonstrated that an AI-based chatbot had significant positive effects on smile parameters as well as speech features representing the frequency of filler words but did not significantly affect clinical measurements in patients with PD. Among speech features, pitch variability, shimmer, and jitter decreased regardless of intervention. Furthermore, explorative analysis revealed that facial expression and speech parameters were associated with motor symptoms, cognition, and mood in patients with PD.

Chatbot itself may positively affect facial expressions in PD, and our analysis identified significant effect of the AI-based chatbot intervention on filler word frequency among previously reported speech features in PD [4–6]. This imply that quantitative evaluation might capture the small changes that cannot be detected by conventional scales that physicians or patients rated. Our results also imply that regular weekly talking sessions with doctors also resulted in positive effects and imply that daily conversations may improve emotion regardless of the mode of delivery. These results suggest that using chatbots may enhance patients' smile and speech without the need for healthcare workers to allocate substantial time resources, although the clinical significance of these changes should be validated in future studies.

A question is whether smile and speech can detect or predict motor symptoms, cognition, and depression. The smile index in the in-person facial expression test was significantly associated with motor symptoms and cognition, which is in line with previous reports [8,9]. Nevertheless, the mechanisms of mimetic expression are complex and multifactorial. They involve facial muscle bradykinesia as well as the associated reduction in facial emotion recognition. Patients with PD had decreased global facial expressions, especially anger, disgust, fear, and neutral expressions; in contrast, surprise, sadness, and happiness were relatively preserved [10]. Moreover, facial expressions can differ among cultures [11]. As such, this hypothesis remains controversial, and further investigations are warranted. Machine-learning models based on video-conferencing speech samples achieved accuracies of 80–90% for all motor, cognitive, and mental aspects. Our models could predict all of these aspects using a single source of conversational speech. This may be because conversations in teleconsultations contain specific speech elements capable of capturing each aspect by nature. For future perspectives, using predictive models from facial and speech features obtained remotely may enhance telehealth to detect or predict subtle changes in clinical symptoms.

This was a single-center pilot study with a small sample size without ethnic diversity. Therefore, our result cannot be generalized and needs further studies. However, our data may suggest that an AI-based chatbot had a positive effect on patients' smile and speech, and that the evaluation of facial expressions and speech features remotely may provide information on motor, cognitive, and mental status of patients with PD. Collectively, our findings highlight AI-based chatbots as promising tools in telehealth.

Study funding

This study was supported by Grants-in Aid from the Research Committee of CNS Degenerative Diseases; Research on Policy Planning and Evaluation for Rare and Intractable Diseases; Health, Labour and Welfare Sciences Research Grants; and the Ministry of Health, Labour and Welfare, Japan (20FC1049), and grants from the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C) (#21K12711). The funders had no role in the study design, data collection or analysis, decision to publish, or manuscript preparation. The Department of Neurodegenerative and Demented Disorders was supported by grants from GLORY Ltd., Kirin Company Ltd., Mitsubishi UFJ Lease & Finance Company Ltd.

Author contributions

The conception and design of the study, or acquisition of data, or analysis and interpretation of data: GO, MO, KM, MK, YY, KS, HK, TH, and NH, (2) drafting the article or revising it critically for important intellectual content: MO, GO, KM, MK, YY, KS, HK, TH, NH. (3) final approval of the version to be submitted: All authors.

Declaration of competing interest

The Department of Neurodegenerative and Demented Disorders is a joint-research course supported by GLORY Ltd., Kirin Company Ltd., Mitsubishi UFJ Lease & Finance Company Ltd. The Department of Home Medical Care System based on Information and Communication Technology is a joint-research course supported by Sunwels Co., Ltd. The Department of Drug Development for Parkinson's Disease, Juntendo University Faculty of Medicine is a course supported by Ohara Pharmaceutical Co., Ltd. and PARKINSON Laboratories Co., Ltd. GO has received speaker honoraria from Medtronic, Boston Scientific, Otsuka Pharmaceutical Co. Ltd., Sumitomo Dainippon Pharma Co. Ltd., Eisai Co., Ltd., Takeda Pharmaceutical Company Ltd., Kyowa Hakko Kirin Co. Ltd., and AbbVie, Inc. NH received speaker honoraria from AbbVie GK, EA Pharma, Eisai Co., Ltd., Otsuka Pharmaceutical Co., Ltd., Ono Pharmaceutical Co., Ltd., OHARA Pharmaceutical Co., Ltd, Kyowa Kirin Co., Ltd., Senju Pharmaceutical Co., Ltd., Sumitomo Dainippon Pharma Co., Ltd., Takeda Pharma Co., Ltd., Medtronic, Inc., Novartis Pharma K. K.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.parkreldis.2022.04.018.

References

[1] S. Sekimoto, G. Oyama, S. Chiba, M. Nuermaimaiti, F. Sasaki, N. Hattori, Holomedicine: proof of the concept of interactive three-dimensional telemedicine, movement disorders, Off. J. Mov. Disord. Soc. 35 (10) (2020) 1719–1720.
[2] D. Castelvecchi, Is facial recognition too biased to be let loose? Nature 587 (7834) (2020) 347–349.
[3] N. Clarke, P. Foltz, P. Garrard, How to do things with (thousands of) words: computational approaches to discourse analysis in Alzheimer's disease, Cortex, J. Devoted Stud. Nervous Syst. Behav. 129 (2020) 446–463.
[4] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig, Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity, J. R. Soc. Interface 8 (59) (2011) 842–855.
[5] S. Skodda, Aspects of speech rate and regularity in Parkinson's disease, J. Neurol. Sci. 310 (1–2) (2011) 231–236.
[6] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig, Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests, IEEE Trans. Biomed. Eng. 57 (4) (2010) 884–893.
[7] D. Ireland, J. Liddle, S. McBride, H. Ding, C. Knuepffer, Chat-bots for people with Parkinson's disease: science fiction or reality? Stud. Health Technol. Inf. 214 (2015) 128–133.
[8] T. Maycas-Cepeda, P. López-Ruiz, C. Feliz-Feliz, L. Gómez-Vicente, R. García-Cobos, R. Arroyo, P.J. García-Ruiz, Hypomimia in Parkinson's disease: what is it telling us? Front. Neurol. 11 (1775) (2021).
[9] L. Ricciardi, M. Bologna, F. Morgante, D. Ricciardi, B. Morabito, D. Volpe, D. Martino, A. Tessitore, M. Pomponi, A.R. Bentivoglio, R. Bernabei, A. Fasano, Reduced facial expressiveness in Parkinson's disease: a pure motor disorder? J. Neurol. Sci. 358 (1–2) (2015) 125–130.
[10] S. Argaud, M. Vérin, P. Sauleau, D. Grandjean, Facial emotion recognition in Parkinson's disease: a review and new hypotheses, Movement disorders, Off. J. Mov. Disord. Soc. 33 (4) (2018) 554–567.
[11] W. Sato, S. Hyniewska, K. Minemoto, S. Yoshikawa, Facial expressions of basic emotions in Japanese laypeople, Front. Psychol. 10 (2019) 259.