# **Original Paper**

# Using Speech Data From Interactions With a Voice Assistant to Predict the Risk of Future Accidents for Older Drivers: Prospective Cohort Study

Yasunori Yamada<sup>1</sup>, PhD; Kaoru Shinkawa<sup>1</sup>, ME; Masatomo Kobayashi<sup>1</sup>, PhD; Hironobu Takagi<sup>1</sup>, PhD; Miyuki Nemoto<sup>2</sup>, PhD; Kiyotaka Nemoto<sup>3</sup>, MD, PhD; Tetsuaki Arai<sup>3</sup>, MD, PhD

<sup>1</sup>IBM Research, Tokyo, Japan

<sup>2</sup>Department of Psychiatry, University of Tsukuba Hospital, Ibaraki, Japan

<sup>3</sup>Department of Psychiatry, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan

#### **Corresponding Author:**

Yasunori Yamada, PhD IBM Research Nihonbashi, Hakozaki-cho, Chuo-ku Tokyo, 103-8510 Japan Phone: 81 80 6706 9381 Email: <u>ysnr@jp.ibm.com</u>

# Abstract

**Background:** With the rapid growth of the older adult population worldwide, car accidents involving this population group have become an increasingly serious problem. Cognitive impairment, which is assessed using neuropsychological tests, has been reported as a risk factor for being involved in car accidents; however, it remains unclear whether this risk can be predicted using daily behavior data.

**Objective:** The objective of this study was to investigate whether speech data that can be collected in everyday life can be used to predict the risk of an older driver being involved in a car accident.

**Methods:** At baseline, we collected (1) speech data during interactions with a voice assistant and (2) cognitive assessment data—neuropsychological tests (Mini-Mental State Examination, revised Wechsler immediate and delayed logical memory, Frontal Assessment Battery, trail making test-parts A and B, and Clock Drawing Test), Geriatric Depression Scale, magnetic resonance imaging, and demographics (age, sex, education)—from older adults. Approximately one-and-a-half years later, we followed up to collect information about their driving experiences (with respect to car accidents) using a questionnaire. We investigated the association between speech data and future accident risk using statistical analysis and machine learning models.

**Results:** We found that older drivers (n=60) with accident or near-accident experiences had statistically discernible differences in speech features that suggest cognitive impairment such as reduced speech rate (P=.048) and increased response time (P=.040). Moreover, the model that used speech features could predict future accident or near-accident experiences with 81.7% accuracy, which was 6.7% higher than that using cognitive assessment data, and could achieve up to 88.3% accuracy when the model used both types of data.

**Conclusions:** Our study provides the first empirical results that suggest analysis of speech data recorded during interactions with voice assistants could help predict future accident risk for older drivers by capturing subtle impairments in cognitive function.

(J Med Internet Res 2021;23(4):e27667) doi: 10.2196/27667

# **KEYWORDS**

cognitive impairment; smart speaker; speech analysis; accident; prevention; older adults; prediction; risk; assistant

# Introduction

As the world's older adult population increases, car accidents involving older adults have become an increasingly serious

https://www.jmir.org/2021/4/e27667

social problem. While it has been reported that older drivers have an increased risk of car accident involvement per unit distance travelled [1-4], they also showed a substantially higher rate of serious injury than that of middle-age car drivers [5,6].

Even in normal aging, there is a decline in many cognitive abilities related to driving, and this cognitive decline is known to be one of the risk factors for older adults being involved in car accidents [7-9]. Associating cognitive assessment scores with either self-reported car accidents, crash records, or on-road driving measures has been investigated to identify predictors of driving safety (in previous empirical studies [7]). In particular, cognitive abilities such as visual attention, short-term memory, and executive functions (evaluated with neuropsychological tests) were consistently shown to have associations with driving safety [7,10-12]. In this respect, if cognitive impairments relevant to driving safety in older adults can be inferred solely from behavior data in everyday situations in a passive way, this would be beneficial for accident prevention.

Speech in daily life can be used a potential data sources for determining cognitive impairments related to driving safety. Speech involves multiple interacting cognitive abilities including attention, memory, and executive functions [13,14]. Many empirical studies have used speech data to identify cognitive impairments resulting from aging and diseases such as Alzheimer disease [15-18] and characterized speech changes related to cognitive impairments by extracting linguistic and paralinguistic features from speech data [19-25]. For example, difficulties with word finding and word retrieving have been quantified by tallying pronoun frequency and pause durations [19,20,24,26-28]. A reduction in speech expressiveness has also been quantified by measuring lexical diversity and speech rate [19,23,29-31]. Using a combination of these features, previous studies [19-25,29] have succeeded in differentiating individuals with cognitive impairments from healthy controls. Although no study has investigated the relationship between speech data and driving safety, it is reasonable to explore the possibility that speech data could be used for inferring ability to drive safely from changes in cognitive functioning in older drivers.

At the same time, there is growing interest in using speech data that can be collected in everyday situations for applications in health care owing to the popularity of voice-based interaction systems such as voice assistants in smart speakers and smartphones [32-34]. One approach is to provide various types of voice-based tests via a smart-speaker platform. For example, previous studies [35,36] have used mobile apps for collecting speech responses to neuropsychological tasks such as verbal fluency and picture description tasks; they showed accurate classification rates in detecting patients with Alzheimer disease [35] and dementia [36]. Another approach is to analyze

health-related insights from speech data collected during daily voice-based interactions. For example, vocal characteristics in speech data during typical tasks on smart speakers appeared to be associated with neuropsychological test scores [37], while linguistic features extracted from phone conversation data were significant indicators for differentiating patients with Alzheimer disease from older adults with normal cognition [38]. This approach, focusing on speech data that can be collected in everyday situations, would increase opportunities for frequent assessment by facilitating passive and unobtrusive monitoring.

In this study, we aimed to investigate the relationship between speech data and future driving experiences related to car accidents in healthy older adults by collecting speech data during interactions with a voice assistant with simulated tasks on smart speakers and smartphones. We hypothesized that these speech data could be used for predicting accident risk for older drivers.

# Methods

# **Participants**

We recruited healthy older adults aged 60 years or older through recruiting agencies and advertisements in the local community in Ibaraki, Japan. All examinations were conducted in Japanese. Older adults met the inclusion criteria if they were in good physical and mental health and had no serious diseases, disabilities, mental illness (eg, major depression, bipolar disorder, and schizophrenia), or neurodegenerative diseases (eg, Parkinson disease and dementia). This study was conducted with the approval of the University of Tsukuba Hospital Ethics Committee (H29-065). All participants provided written consent after the procedures of the study had been fully explained.

A total of 71 older individuals participated in the cognitive assessments and speech data collection (women: 38/71, 53.5%; age: range 61-80 years, mean 71.1, SD 4.9). Of the original 71 participants, 60 consented to the follow-up study about their driving experiences (women: 33/60, 55.0%; age: range 61-80 years, mean 70.8, SD 5.1; Table 1). They were contacted again approximately one-and-a-half years after the speech data collection (mean 17.3 months, SD 2.7) and answered a questionnaire on their driving experiences within the past year. The questionnaire included free-form questions about accidents and near accidents; *near accidents* were described as infractions and any other incidents while driving that they deemed to be dangerous regardless of severity and culpability.



Table 1. Demographic and assessment data for study participants.

| Variable   | Total (N=60) | Individuals without accident or<br>near-accident experiences (n=34) | Individuals with accident or near-<br>accident experiences (n=26) | P value |
|--|--------------|---|---|---------|
| Age (years), mean (SD)   | 70.8 (5.1)   | 70.5 (4.9)  | 71.3 (5.3)  | .45     |
| Education (years), mean (SD)   | 13.7 (2.2)   | 13.7 (2.2)  | 13.6 (2.1)  | .93     |
| Sex, n (%)   |              |   |   | .53     |
| Men  | 27 (45)      | 17 (50)   | 10 (38)   |         |
| Women  | 33 (55)      | 17 (50)   | 16 (62)   |         |
| Mini-Mental State Examination <sup>a</sup> , mean (SD)               | 27.6 (1.8)   | 27.4 (1.8)  | 27.9 (1.8)  | .28     |
| LM IA <sup>b</sup> , mean (SD)                                       | 9.6 (3.8)    | 9.1 (3.7)   | 10.2 (4.0)  | .43     |
| LM IIA <sup>c</sup> , mean (SD)                                      | 7.5 (3.6)    | 7.3 (3.7)   | 7.6 (3.6)   | .74     |
| Frontal Assessment Battery <sup>d</sup> , mean (SD)                  | 13.7 (2.7)   | 13.4 (2.7)  | 14.2 (2.7)  | .45     |
| Trail making test-part A (seconds), mean (SD)                        | 33.2 (9.8)   | 33.6 (9.7)  | 32.6 (10.1)   | .72     |
| Trail making test-part B (seconds), mean (SD)                        | 89.5 (49.7)  | 95.7 (60.9)   | 81.3 (28.2)   | .71     |
| Clock Drawing Test <sup>e</sup> , mean (SD)                          | 6.7 (0.8)    | 6.7 (0.7)   | 6.7 (1.0)   | .36     |
| Geriatric Depression Scale <sup>f</sup> , mean (SD)                  | 2.9 (2.4)    | 2.8 (2.4)   | 3.1 (2.4)   | .62     |
| Severity scores for atrophy in medial temporal structures, mean (SD) | 0.9 (0.6)    | 0.8 (0.4)   | 0.9 (0.7)   | .86     |

<sup>a</sup>The total possible score ranges from 0 to 30.

<sup>b</sup>LM IA: immediate recall of the logical memory-story A of the Wechsler memory scale-revised for episodic memory; the total possible score ranges from 0 to 25.

<sup>c</sup>LM IIA: delayed recall of the logical memory-story A of the Wechsler memory scale-revised for episodic memory; the total possible score ranges from 0 to 25.

<sup>d</sup>The total possible score ranges from 0 to 18.

<sup>e</sup>The total possible score ranges from 0 to 7.

<sup>f</sup>The total possible score ranges from 0 to 15.

#### **Cognitive Assessments**

Cognitive assessments and examinations were those typically used for the diagnosis of dementia and comprised 12 variables: age, sex, education, 7 neuropsychological test scores (Mini-Mental State Examination for global cognition; immediate and delayed recall of the logical memory-story A of the Wechsler memory scale-revised for episodic memory; the Frontal Assessment Battery for executive function; the trail making test-part A and B for executive function and attention; and the clock drawing test for visuospatial function), and 2 clinical scores (Geriatric Depression Scale and the severity of medial temporal lobe atrophy). The severity of medial temporal lobe atrophy was evaluated using structural magnetic resonance imaging (MRI) scans-1.5 T, T1-weighted images and a 3D gradient-echo sequence-with the following parameters: sagittal orientation with 1.2-mm thick sections; time repetition/time echo: 2400/3.52 milliseconds; flip angle: 8°; field of view: 192×192. We expressed the severity of medial temporal lobe atrophy as a Z score relative to cognitively healthy adults by using a standalone, voxel-based specific regional analysis system for Alzheimer disease [39]. Two psychiatrists (KN and TA) reviewed the results of the cognitive assessments and confirmed that participants did not meet the criteria for dementia based on those of the National Institute on Aging and Alzheimer's

RenderX

Association and Alzheimer disease Neuroimaging Initiative 2 [40].

# **Speech Data Collection**

We simulated conversations with a voice assistant on modern smart speakers and smartphones and collected the speech data while performing 3 typical task scenarios: information retrieval (asking for tomorrow's weather), shopping online (booking a movie ticket), and personal schedule management (creating a calendar event). The tasks began with a simple scenario and then advanced to the more complicated ones. Each task started with an initiating question from the system ("what can I help you with?"), with follow-up questions that asked for detailed information related to the task. The follow-up questions were presented in a fixed order. The questions consisted of four categories-open-ended, to which participants responded with a free-form sentence (Multimedia Appendix 1: Table S1); multiple choice, to which participants responded by choosing one of the options stated in the question; prepared input, to which participants responded with information (eg, passcode) specified by the experimenter; and confirmation, to which participants responded by accepting or rejecting a statement made by the system. The system presented at least 22 questions in total to each participant for the 3 tasks.

To simulate conversations, we took a Wizard-of-Oz [41] approach, in which the participants were told that they were talking with a computer system, though in fact the interaction was mediated by an experimenter (ie, the wizard). We chose this approach so that we could avoid uncertain factors such as errors in automatic speech recognition. During the tasks, the experimenter made the system present a question. After the participant responded, the experimenter prompted the system to move onto the next question if the response contained the necessary information corresponding to the question; otherwise, they would repeat the same question. Each open-ended, multiple choice, and prepared input question presented by the system was scripted in advance and the same for all participants. For confirmation questions, we prepared several variations for each question and the experimenter chose one, depending on the participant's previous response. For example, the experimenter

chose "you are purchasing one ticket, is it OK?" or "you are purchasing two tickets, is it OK?" to have the participant confirm the number of tickets to book.

The interface for speech data collection was implemented as a tablet-based app on an Apple iPad Air 2. In the experiment, participants sat down in front of the tablet and talked with the system (Figure 1a). During the tasks, the tablet showed a screen indicating whether it was speaking (Figure 1b) or listening (Figure 1c). The experimenter sat behind the participant and operated the system by using a separate interface hidden from the participants. Speech data were recorded in raw format with a sampling rate of 44.1 kHz through the embedded microphone in the tablet. Each experimental session took approximately 30 minutes per participant, including instructions and wrap-up. Additional details about our apparatus and procedure have been previously published [37].

Figure 1. Overview of experimental setup: (a) setup for collecting speech data, (b) screen showing participant's turn, and (c) screen showing the tablet's turn.



#### **Data Analysis**

From each participant's speech data, we automatically extracted 84 paralinguistic speech features used in previous studies on inferring cognitive impairments and detecting early signs of Alzheimer disease [19,20,23,27-29,31,42,43]. They consisted of 56 acoustic features and 28 prosodic features.

The acoustic features consisted of features related to mel-frequency cepstral coefficients (MFCCs), jitter, and shimmer. We used the mean and first-order derivatives of the first 12 MFCCs, which represent the short-term power spectrum of the speech signal. Jitter and shimmer features measure cycle-to-cycle variations of fundamental frequency and amplitude [44]. Prosodic features included speech rate, pitch variability, phonation time, number of phonemes needed for completing tasks, response time, total pause duration, and proportion of long pauses (pauses >0.8 seconds). Both acoustic and prosodic features were extracted from each task's speech data separately. We used Python (version 3.8) audio-processing libraries (librosa, version 0.8.0 [45]; Signal\_Analysis, version 0.1.26 [46]).

Statistical analyses were performed using Statistics and Machine Learning Toolbox (version 11.1) for MATLAB (version R2017a, The MathWorks Inc) environment. To assess the differences in each variable between participants with and without accident or near-accident experiences, we used 2-sided Mann-Whitney tests for continuous data and chi-square tests for categorical data. We did not correct for multiple comparisons, and P values <.05 were considered significant.

The prediction models for differentiating individuals with and without accident or near-accident experiences were built using multiple types of binary classifiers with automatic sequential forward selection of features. Model performance was evaluated with both leave-one-subject-out cross validation and 100 iterations of 10-fold cross-validation methods. The classifiers included k-nearest neighbors [47], random forest [48] and support vector machine [49]. The parameters that we studied were as follows: the number of neighbors for the k-nearest neighbors; the number and the maximum depth of trees for random forest; kernel functions, penalty parameter, and the parameter associated with the width of the radial basis function kernel for the support vector machine. We performed an exhaustive grid search to determine these parameters. The algorithms were implemented using the Python scikit-learn package (version 0.23.2).

# Results

For speech data collection (the 30-minute sessions), we obtained an average of 23.8 responses within 100.2 seconds (SD 28.6) from each participant. The average response duration of each task scenario ranged from 17.4 to 59.6 seconds (mean 33.41, SD 9.5). The average duration of a single response for each participant ranged from 1.1 to 7.6 seconds (mean 4.2, SD 1.1). At follow-up, 26 of the 60 participants (43.3%) reported car accident or near-accident experiences within the previous year.

Of those, 23 participants reported a near-accident experience, 2 reported accidents, and 1 reported both. The near-accidents consisted of near-misses with a car or pedestrian resulting in a sense of fear and anxiety (eg, from failure to notice a crossing pedestrian), errors in operation (eg, stepping on the accelerator instead of the brake), and unintentional violations (eg, entering the opposite lane).

In comparisons between individuals with and without accident or near-accident experiences, there were no significant differences in any cognitive assessment variables (age: P=.45; education year: P=.93; sex: P=.53; Mini-Mental State Examination: P=.28; immediate and delayed recall of the logical memory-story A of the Wechsler memory scale-revised: P=.43, P=.74; the Frontal Assessment Battery: P=.45; the trail making test-part A and B: P=.72, P=.71; the clock drawing test: P=.36; Geriatric Depression Scale: P=.62; severity scores for atrophy in medial temporal structures: P=.86; Table 1); however, we 10 speech features significant found with differences— $\Delta$ MFCC<sub>1</sub>: *P*=.005,  $\Delta$ MFCC<sub>4</sub>: *P*=.043,  $\Delta$ MFCC<sub>5</sub>: *P*=.011, ΔMFCC<sub>7</sub>: *P*=.035, ΔMFCC<sub>12</sub>: *P*=.023; jitter: *P*=.034; response time: P=.040; proportion of long pauses: P=.044; speech rate: P=.048; and number of phonemes needed for completing tasks: P=.049 (Figure 2; Multimedia Appendix 1: Table S2). Those with accident or near-accident experiences showed decreased speech rate and jitter as well as increased response time and long pauses. These speech features were reported in previous studies as significant indicators of changes in cognitive function, and the trends in their changes were consistent with those observed in individuals with cognitive impairments and patients with Alzheimer disease and mild cognitive impairment (for speech rate [23,27,31]; for jitter [42,43]; for response time [20,27]; for proportion of long pause [27,28]).

To visualize whether the variance seen among a variable set is capable of discriminating between individuals with and without potential future accident or near-accident experiences, we performed principal component analysis on 2 variable sets: the 12 cognitive assessment variables and 10 speech features (Figure 3). The cognitive assessment variable set had little capability to differentiate the groups; there was considerable overlap and no clear separation. In contrast, the speech variable set enabled some separation of the groups.

Input variables for the classification models were either or both the 12 cognitive assessment variables and 10 speech features. When model performance was evaluated with leave-one-subject-out cross-validation, with only the cognitive assessment variables, we obtained 75.0% accuracy (65.4% sensitivity, 82.4% specificity, and 69.4% F1 score; Figure 4a), with only the speech features, the model accuracy increased to 81.7% accuracy (65.4% sensitivity, 94.1% specificity, and 75.6% F1 score; Figure 4b), and with speech features and cognitive assessment variables combined, performance improved further (88.3% accuracy, 88.5% sensitivity, 88.2% specificity, and 86.8% F1 score; Figure 4c). When we evaluated the model using 10-fold cross validation, the results showed similar trends (Multimedia Appendix 1: Table S3): the model using the cognitive assessment variables achieved 75.5% accuracy (95% CI 75.1-75.9), the model using speech features achieved 80.1% accuracy (95% CI 79.7-80.5), and the model using both types of features achieved 85.5% accuracy (95% CI 85.1-85.9).

Figure 2. Box plots (line and diamond represent median and mean, respectively) for speech features with significant differences between individuals with and without accident or near-accident experiences—jitter: P=.034; response time: P=.040; speech rate: P=.048.



Figure 3. Principal component analysis plots using (a) cognitive assessment variables and (b) speech features, with confidence interval ellipsoid set to 0.95. PC: principal component.



**Figure 4.** Confusion matrixes for predicting future accident risks of older drivers obtained using leave-one-subject-out cross-validation for models with (a) cognitive assessment variables, (b) speech features, and (c) cognitive assessment variables and speech features combined. The number in parentheses indicates the number of participants. wo: without; w: with.



# Discussion

# **Principal Results**

In light of the increasing demand for preventing car accidents involving older adults, we investigated the possibility that future accident risk related to cognitive impairments could be automatically predicted with passive unobtrusive monitoring. To this end, we focused on speech data because many previous studies have succeeded in quantifying and detecting cognitive impairments from speech data [19,20,23,27-29,31,42,43], speech data are becoming more accessible, and voice-based interaction systems such as voice assistants are becoming more popular [32-34].

The statistical analysis showed that the speech data collected during typical tasks on smart speakers and smartphones had statistically discernible speech features between older drivers with and without accident or near-accident experiences. These speech features indicated that older drivers with these experiences tended to show decreased speech rate and jitter as well as increased response time and long pauses. These changes in speech features were reported as statistically significant signatures for cognitive impairments by previous studies on patients with Alzheimer disease and mild cognitive impairment [19,20,23,27-29,31,42,43]. The results suggest that speech features could capture subtle impairments of cognitive function in older drivers. On the other hand, we found no differences in any cognitive assessment variables, but this could be explained by the criteria for driving risks that differed from those in previous studies [7,10,12,50-53]. While previous studies compared older drivers with and without car-accident experiences regardless of having near-car-accident experiences and reported significant differences in cognitive assessment scores between them [7,10,12,50-53], we focused on both accident and near-accident experiences, and the majority of the high-risk group in our study were individuals with near-accident experiences but without actual car accidents. Speech data and cognitive assessment results suggest that eliciting discernible changes relevant to future near-accident experiences may require cognitive assessment for subtle impairments, such as, test batteries used for screening preclinical Alzheimer disease [54,55]. Even so, if speech data during interactions with voice assistants can be used for predicting future accident risk, it would greatly increase the accessibility of early screening with a relatively low burden.

The classification model using speech features achieved 81.7% accuracy, which is 6.7% higher than that using cognitive assessment data, and models achieved up to 88.3% accuracy

with both combined. Dimensional reduction and visualization using principal component analysis, an unsupervised method, showed that the feature space with speech data was better able to separate those with and without accident or near-accident experiences than the feature space with cognitive assessment variables. These results and those of the statistical analysis indicate that speech data during typical tasks with voice assistants could have comparable (or possibly more) information for predicting future accident risks of older drivers compared with the standard cognitive assessments.

Our results show paralinguistic speech characteristics were useful for predicting future accident risks of older drivers. Previous user-interface studies reported that voice input was effective and was preferable as an input modality for older adults [56-58], while other studies reported that the performance of automatic speech recognition tended to be worse in older adults than in other age groups [59,60]. From this perspective, our results suggest that models for predicting future accident risks of older drivers can be made robust against errors of automatic speech recognition by exploiting paralinguistic features.

Our results highlight the possibility that cognitive impairments related to future car accident risks could be detected using speech data collected in everyday life. Assistive and automated driving systems are promising technologies that may help older adults with cognitive challenges to safely continue driving [61]. Recent studies suggested the importance of individual differences in cognitive abilities for assistive and automated driving technologies for older adults [62,63] because literature has suggested that cognitive abilities affect both performance with automated technology and perceptions of automation (ie, trust) [64,65]. Hence, our approach to detect cognitive impairments associated with driving risks might provide useful information for the personalization of assistive and automated driving systems based on the cognitive abilities of older adults.

#### Limitations

Our work had several limitations. First, we collected speech data in a lab setting. The controlled setting might affect the way people interact with a voice assistant. In future work, data collection in free-living situations using voice assistants would be needed along with additional interaction scenarios. Second, the sample size was limited. In spite of this limitation, our statistical analysis of speech features showed consistent trends indicating subtle cognitive impairments in older adults with future accident or near-accident experiences, and the prediction performance (to predict independent future accidents) using speech features was as high as 88%, even when the classifier was trained on a subsample. From these perspectives, we believe

that our results can be confirmed by future studies. Third, our definition of future car accident risks was based on self-reports of accident and near-accident experiences. In future work, we need to consider obtaining more objective measures for accident risks by combining self-reports with on-road driving assessments, informant reports, or drive recorder videos.

# Conclusion

Given the increasing demand for car accident prevention involving older adults, we explored the possibility of predicting future accident risks associated with cognitive impairments by using behavioral data that can be collected in everyday life. To this end, we focused on speech data collected during interactions with voice assistants in smart speakers and smartphones and investigated the associations with future accident risks by following up with older drivers. We found that (1) older drivers with accident or near-accident experiences had statistically discernible changes in speech features, implying cognitive impairments, and (2) the machine learning model using speech features could predict future accident or near-accident experiences with up to 88.3% accuracy. Although further studies with speech data collected in everyday life and objective data for near-accidents are needed, our study provides the first empirical results suggesting that speech data during interactions with voice assistants in smart speakers and smartphones could help predict future accident risks of older drivers by capturing subtle impairments in cognitive function. We believe that our results can be used in future efforts toward preventing driving accidents of older adults through continuous passive unobtrusive monitoring.

# Acknowledgments

This work was supported by Japan Society for the Promotion of Science KAKENHI (grants 19H01084 and 18K18164).

# **Conflicts of Interest**

None declared.

# **Multimedia Appendix 1**

Supplementary tables. [PDF File (Adobe PDF File), 194 KB-Multimedia Appendix 1]

# References

- 1. Cicchino JB, McCartt AT. Trends in older driver crash involvement rates and survivability in the United States: an update. Accid Anal Prev 2014 Nov;72:44-54. [doi: 10.1016/j.aap.2014.06.011] [Medline: 25003969]
- Cicchino JB, McCartt AT. Critical older driver errors in a national sample of serious U.S. crashes. Accid Anal Prev 2015 Jul;80:211-219. [doi: <u>10.1016/j.aap.2015.04.015</u>] [Medline: <u>25916662</u>]
- 3. Ma L, Yan X. Examining the nonparametric effect of drivers' age in rear-end accidents through an additive logistic regression model. Accid Anal Prev 2014 Jun;67:129-136. [doi: <u>10.1016/j.aap.2014.02.021</u>] [Medline: <u>24642249</u>]
- 4. Lombardi DA, Horrey WJ, Courtney TK. Age-related differences in fatal intersection crashes in the United States. Accid Anal Prev 2017 Feb;99(Pt A):20-29 [FREE Full text] [doi: 10.1016/j.aap.2016.10.030] [Medline: 27855312]
- Ryan GA, Legge M, Rosman D. Age related changes in drivers' crash risk and crash type. Accid Anal Prev 1998 May;30(3):379-387. [doi: <u>10.1016/s0001-4575(97)00098-5</u>] [Medline: <u>9663297</u>]
- 6. Meuleners LB, Harding A, Lee AH, Legge M. Fragility and crash over-representation among older drivers in Western Australia. Accid Anal Prev 2006 Sep;38(5):1006-1010. [doi: <u>10.1016/j.aap.2006.04.005</u>] [Medline: <u>16713982</u>]
- Anstey KJ, Wood J, Lord S, Walker JG. Cognitive, sensory and physical factors enabling driving safety in older adults. Clin Psychol Rev 2005 Jan;25(1):45-65. [doi: <u>10.1016/j.cpr.2004.07.008</u>] [Medline: <u>15596080</u>]
- 8. Bryan J, Luszcz MA. Measurement of executive function: considerations for detecting adult age differences. J Clin Exp Neuropsychol 2000 Feb;22(1):40-55. [doi: 10.1076/1380-3395(200002)22:1;1-8;FT040] [Medline: 10649544]
- 9. McGwin, Jr G, Brown DB. Characteristics of traffic crashes among young, middle-aged, and older drivers. Accid Anal Prev 1999 May;31(3):181-198. [doi: 10.1016/s0001-4575(98)00061-x] [Medline: 10196595]
- Odenheimer GL, Beaudet M, Jette AM, Albert MS, Grande L, Minaker KL. Performance-based driving evaluation of the elderly driver: safety, reliability, and validity. J Gerontol 1994 Jul;49(4):M153-M519. [doi: <u>10.1093/geronj/49.4.m153</u>] [Medline: <u>8014389</u>]
- 11. Sims RV, McGwin G, Allman RM, Ball K, Owsley C. Exploratory study of incident vehicle crashes among older drivers. J Gerontol A Biol Sci Med Sci 2000 Jan;55(1):M22-M27. [doi: <u>10.1093/gerona/55.1.m22</u>] [Medline: <u>10719769</u>]
- 12. Daigneault G, Joly P, Frigon J. Executive functions in the evaluation of accident risk of older drivers. J Clin Exp Neuropsychol 2002 Apr;24(2):221-238. [doi: 10.1076/jcen.24.2.221.993] [Medline: 11992205]
- Gibson MT, Mangat D, Gagliano G, Wilson M, Fletcher J, Bulman J, et al. Evaluation of the efficacy of a redox agent in the treatment of chronic periodontitis. J Clin Periodontol 1994 Nov;21(10):690-700. [doi: 10.1111/j.1600-051x.1994.tb00788.x] [Medline: 7852614]
- Kirshner HS. Primary progressive aphasia and Alzheimer's disease: brief history, recent evidence. Curr Neurol Neurosci Rep 2012 Dec;12(6):709-714. [doi: <u>10.1007/s11910-012-0307-2</u>] [Medline: <u>22932755</u>]

- 15. Taler V, Phillips NA. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. J Clin Exp Neuropsychol 2008 Jul;30(5):501-556. [doi: 10.1080/13803390701550128] [Medline: 18569251]
- Garrard P, Lambon Ralph MA, Watson PC, Powis J, Patterson K, Hodges JR. Longitudinal profiles of semantic impairment for living and nonliving concepts in dementia of Alzheimer's type. J Cogn Neurosci 2001 Oct 01;13(7):892-909. [doi: <u>10.1162/089892901753165818</u>] [Medline: <u>11595093</u>]
- 17. Oulhaj A, Wilcock GK, Smith AD, de Jager CA. Predicting the time of conversion to MCI in the elderly: role of verbal expression and learning. Neurology 2009 Nov 03;73(18):1436-1442 [FREE Full text] [doi: 10.1212/WNL.0b013e3181c0665f] [Medline: 19794124]
- Forbes-McKay KE, Venneri A. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. Neurol Sci 2005 Oct;26(4):243-254. [doi: <u>10.1007/s10072-005-0467-9</u>] [Medline: <u>16193251</u>]
- 19. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. J Alzheimers Dis 2016;49(2):407-422. [doi: 10.3233/JAD-150520] [Medline: 26484921]
- 20. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst) 2015 Mar;1(1):112-124 [FREE Full text] [doi: 10.1016/j.dadm.2014.11.012] [Medline: 27239498]
- Roark B, Mitchell M, Hosom J, Hollingshead K, Kaye J. Spoken language derived measures for detecting mild cognitive impairment. IEEE Trans Audio Speech Lang Process 2011 Sep 01;19(7):2081-2090 [FREE Full text] [doi: 10.1109/TASL.2011.2112351] [Medline: 22199464]
- 22. Sajjadi SA, Patterson K, Tomek M, Nestor PJ. Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. Aphasiology 2012 Jun;26(6):847-866. [doi: 10.1080/02687038.2012.654933]
- Themistocleous C, Eckerström M, Kokkinakis D. Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls. PLoS One 2020;15(7):e0236009 [FREE Full text] [doi: 10.1371/journal.pone.0236009] [Medline: 32658934]
- 24. Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno-Tempini M. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; 2014 Presented at: Association for Computational Linguistics; June 27; Baltimore, Maryland, USA p. 27-37. [doi: 10.3115/v1/w14-3204]
- 25. Satt A, Hoory R, König A, Aalten P, Robert P. Speech-based automatic and robust detection of very early dementia. 2014 Presented at: Fifteenth Annual Conference of the International Speech Communication Association; Sep 14-18; Singapore.
- 26. Ahmed S, Haigh AF, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. Brain 2013 Dec;136(Pt 12):3727-3737 [FREE Full text] [doi: 10.1093/brain/awt269] [Medline: 24142144]
- Khodabakhsh A, Yesil F, Guner E, Demiroglu C. Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. J Audio Speech Music Proc 2015 Mar 25;2015(1):1-15. [doi: 10.1186/s13636-015-0052-y]
- Pastoriza-Dominguez P, Torre I, Dieguez-Vide F, Gomez-Ruiz I, Gelado S, Bello-Lopez J. Speech pause distribution as an early marker for Alzheimer's disease. medRxiv. Preprint posted online on January 4, 2021. [doi: 10.1101/2020.12.28.20248875]
- 29. Hernández-Domínguez L, Ratté S, Sierra-Martínez G, Roche-Bergua A. Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. Alzheimers Dement (Amst) 2018;10:260-268 [FREE Full text] [doi: 10.1016/j.dadm.2018.02.004] [Medline: 29780871]
- 30. Aramaki E, Shikata S, Miyabe M, Kinoshita A. Vocabulary size in speech may be an early indicator of cognitive impairment. PLoS One 2016;11(5):e0155195 [FREE Full text] [doi: 10.1371/journal.pone.0155195] [Medline: 27176919]
- Hoffmann I, Nemeth D, Dye CD, Pákáski M, Irinyi T, Kálmán J. Temporal parameters of spontaneous speech in Alzheimer's disease. Int J Speech Lang Pathol 2010 Feb;12(1):29-34. [doi: <u>10.3109/17549500903137256</u>] [Medline: <u>20380247</u>]
- 32. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. NPJ Digit Med 2019;2:9 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
- 33. Manfredi C, Lebacq J, Cantarella G, Schoentgen J, Orlandi S, Bandini A, et al. Smartphones offer new opportunities in clinical voice research. J Voice 2017 Jan;31(1):111.e1-111.e7. [doi: 10.1016/j.jvoice.2015.12.020] [Medline: 27068549]
- 34. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020;3:122 [FREE Full text] [doi: 10.1038/s41746-020-00332-0] [Medline: 33015374]
- Konig A, Satt A, Sorin A, Hoory R, Derreumaux A, David R, et al. Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. Curr Alzheimer Res 2018;15(2):120-129. [doi: 10.2174/1567205014666170829111942] [Medline: 28847279]
- 36. Hall AO, Shinkawa K, Kosugi A, Takase T, Kobayashi M, Nishimura M, et al. Using tablet-based assessment to characterize speech for individuals with dementia and mild cognitive impairment: preliminary results. AMIA Jt Summits Transl Sci Proc 2019;2019:34-43 [FREE Full text] [Medline: <u>31258954</u>]

RenderX

- Kobayashi M, Kosugi A, Takagi H, Nemoto M, Nemoto K, Arai T. Effects of age-related cognitive decline on elderly user interactions with voice-based dialogue systems. 2019 Presented at: IFIP Conference on Human Computer Interaction (Springer); September 2-6; Cyprus p. 53-74. [doi: 10.1007/978-3-030-29390-1\_4]
- Yamada Y, Shinkawa K, Shimmei K. Atypical repetition in daily conversation on different days for detecting Alzheimer disease: evaluation of phone-call data from regular monitoring service. JMIR Ment Health 2020 Jan 12;7(1):e16790 [FREE Full text] [doi: 10.2196/16790] [Medline: 31934870]
- Matsuda H, Mizumura S, Nemoto K, Yamashita F, Imabayashi E, Sato N, et al. Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer Disease. AJNR Am J Neuroradiol 2012 Jun;33(6):1109-1114 [FREE Full text] [doi: 10.3174/ajnr.A2935] [Medline: 22300935]
- 40. Procedures manual. Alzheimer's Disease Neuroimaging Initiative 2 (ADNI 2). 2008. URL: <u>https://adni.loni.usc.edu/</u> wp-content/uploads/2008/07/adni2-procedures-manual.pdf [accessed 2021-02-02]
- 41. Dahlbäck N, Jönsson A, Ahrenberg L. Wizard of Oz studies why and how. Knowl Based Syst 1993 Dec;6(4):258-266. [doi: 10.1016/0950-7051(93)90017-n]
- 42. Alhanai T, Au R, Glass J. Spoken language biomarkers for detecting cognitive impairment. 2017 Presented at: Proc IEEE Workshop Autom Speech Recognit Underst; December 16-20; Okinawa, Japan p. 409-416. [doi: 10.1109/asru.2017.8268965]
- 43. Farrús M, Codina-Filbà J. Combining prosodic, voice quality and lexical features to automatically detect Alzheimer's disease. arXiv. Preprint posted online November 18, 2020 [FREE Full text]
- 44. Farrús M, Hernando J, Ejarque P. Jitter and shimmer measurements for speaker recognition. 2007 Presented at: Eighth annual conference of the international speech communication association; August 27-31; Antwerp, Belgium.
- 45. McFee B, Raffel C, Liang D, Ellis D, McVicar M, Battenberg E. librosa: Audio and music signal analysis in python. 2015 Presented at: Proceedings of the 14th python in science conference; July 6-12; Austin, Texas. [doi: 10.25080/majora-7b98e3ed-003]
- 46. Signal–Analysis. GitHub. URL: https://brookemosby.github.io/Signal\_Analysis/ [accessed 2021-02-26]
- 47. Goldberger J, Hinton G, Roweis S, Salakhutdinov R. Neighbourhood components analysis. Adv Neural Inf Process Syst 2004;17:513-520.
- 48. Breiman L. Random forests. In: Boehmke B, Greenwell B, editors. Hands-On Machine Learning With R. Boca Raton: Chapman and Hall/CRC; 2001:5-32.
- 49. Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. 1992 Presented at: Fifth Annual Workshop on Computational Learning Theory; July 27-29; Pittsburgh, Pennsylvania p. 144-152. [doi: 10.1145/130385.130401]
- 50. Stutts JC, Stewart JR, Martell C. Cognitive test performance and crash risk in an older driver population. Accid Anal Prev 1998 May;30(3):337-346. [doi: 10.1016/s0001-4575(97)00108-5] [Medline: 9663293]
- 51. McKnight AJ, McKnight AS. Multivariate analysis of age-related driver ability and performance deficits. Accid Anal Prev 1999 Sep;31(5):445-454. [doi: 10.1016/s0001-4575(98)00082-7] [Medline: 10440541]
- 52. Owsley C, Ball K, Sloane ME, Roenker DL, et al. Visual/cognitive correlates of vehicle accidents in older drivers. Psychol Aging 1991 Sep;6(3):403-415. [doi: 10.1037//0882-7974.6.3.403] [Medline: 1930757]
- 53. Payyanadan R, Gibson M, Chiou E, Ghazizadeh M, Lee J. Contextual design for driving: developing a trip-planning tool for older adults. Transp Res Part F Traffic Psychol Behav 2017 Apr;46:462-476. [doi: <u>10.1016/j.trf.2016.08.005</u>]
- Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol 2014 Aug;71(8):961-970 [FREE Full text] [doi: 10.1001/jamaneurol.2014.803] [Medline: 24886908]
- Duke Han S, Nguyen CP, Stricker NH, Nation DA. Detectable neuropsychological differences in early preclinical Alzheimer's disease: a meta-analysis. Neuropsychol Rev 2017 Dec;27(4):305-325 [FREE Full text] [doi: 10.1007/s11065-017-9345-5] [Medline: 28497179]
- Smith AL, Chaparro BS. Smartphone text input method performance, usability, and preference with younger and older adults. Hum Factors 2015 Sep;57(6):1015-1028. [doi: <u>10.1177/0018720815575644</u>] [Medline: <u>25850116</u>]
- Liu Y, Chen C, Lin Y, Chen H, Irianti D, Jen T, et al. Design and usability evaluation of mobile voice-added food reporting for elderly people: randomized controlled trial. JMIR Mhealth Uhealth 2020 Sep 28;8(9):e20317 [FREE Full text] [doi: 10.2196/20317] [Medline: 32985999]
- Stigall B, Waycott J, Baker S, Caine K. Older adults' perception and use of voice user interfaces: A preliminary review of the computing literature. In: Proceedings of the 31st Australian Conference on Human-Computer-Interaction.: Association for Computing Machinery; 2019 Presented at: 31st Australian Conference on Human-Computer-Interaction; December 3-5; Fremantle Australia p. 423-427. [doi: 10.1145/3369457.3369506]
- 59. Werner L, Huang G, Pitts B. Automated speech recognition systems and older adults: a literature review and synthesis. 2019 Presented at: Proc Hum Factors Ergon Soc Annu Meet; October 28-November 1; Seattle, Washington, USA p. 42-46.
- 60. Portet F, Vacher M, Golanski C, Roux C, Meillon B. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Pers Ubiquit Comput 2011 Oct 2;17(1):127-144. [doi: 10.1007/s00779-011-0470-5]

RenderX

- 61. Knoefel F, Wallace B, Goubran R, Sabra I, Marshall S. Semi-autonomous vehicles as a cognitive assistive device for older adults. Geriatrics (Basel) 2019 Nov 16;4(4):63 [FREE Full text] [doi: 10.3390/geriatrics4040063] [Medline: 31744041]
- 62. Rovira E, McLaughlin AC, Pak R, High L. Looking for age differences in self-driving vehicles: examining the effects of automation reliability, driving risk, and physical impairment on trust. Front Psychol 2019;10:800 [FREE Full text] [doi: 10.3389/fpsyg.2019.00800] [Medline: 31105610]
- 63. Depestele S, Ross V, Verstraelen S, Brijs K, Brijs T, van Dun K, et al. The impact of cognitive functioning on driving performance of older persons in comparison to younger age groups: a systematic review. Transp Res Part F Traffic Psychol Behav 2020 Aug;73:433-452. [doi: 10.1016/j.trf.2020.07.009]
- 64. Rovira E, Pak R, McLaughlin A. Effects of individual differences in working memory on performance and trust with various degrees of automation. Theor Issues Ergon Sci 2016 Nov 21;18(6):573-591. [doi: 10.1080/1463922x.2016.1252806]
- Pak R, McLaughlin AC, Leidheiser W, Rovira E. The effect of individual differences in working memory in older adults on performance with different degrees of automated technology. Ergonomics 2017 Apr;60(4):518-532. [doi: 10.1080/00140139.2016.1189599] [Medline: 27409279]

# Abbreviations

**MFCC:** mel-frequency cepstral coefficients **MRI:** magnetic resonance imaging

Edited by G Eysenbach; submitted 02.02.21; peer-reviewed by A Hall, V Stara; comments to author 25.02.21; revised version received 08.03.21; accepted 15.03.21; published 08.04.21 <u>Please cite as:</u> Yamada Y, Shinkawa K, Kobayashi M, Takagi H, Nemoto M, Nemoto K, Arai T Using Speech Data From Interactions With a Voice Assistant to Predict the Risk of Future Accidents for Older Drivers: Prospective Cohort Study LMad Literaret Bag 2021;22(4):e27667

J Med Internet Res 2021;23(4):e27667 URL: <u>https://www.jmir.org/2021/4/e27667</u> doi: <u>10.2196/27667</u> PMID:

©Yasunori Yamada, Kaoru Shinkawa, Masatomo Kobayashi, Hironobu Takagi, Miyuki Nemoto, Kiyotaka Nemoto, Tetsuaki Arai. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.04.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

