Development of a Privacy-By-Design Speech Assistant Providing Nutrient Information for German Seniors

Andreas Seiderer, Hannes Ritschel, Elisabeth André
{seiderer,ritschel,andre}@hcm-lab.de
Augsburg University
86159 Augsburg, Germany

ABSTRACT

Elderly people could take benefit of speech assistants since they provide the most natural way to interact with assistive technology. Nevertheless, current speech assistants are mainly based on cloud-systems which are non-functional without a stable internet connection or introduce severe privacy-concerns. In this work we provide an overview of state-of-the-art available components for developing a privacy-by-design, open source speech assistant for seniors, as well as a fully functional implementation. We chose the health related task of getting nutrition information of packaged foodstuff for which we rely on an open data database. Our prototype was used in a workshop with two German seniors to gather further insights into the special requirements of elderly people, which can be taken into account for future speech assistants.

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; • **Applied computing → Consumer health**; • **Social and professional topics → Seniors**.

KEYWORDS

speech assistant, seniors, privacy, nutrition information, open data

ACM Reference Format:

Andreas Seiderer, Hannes Ritschel, Elisabeth André. 2020. Development of a Privacy-By-Design Speech Assistant Providing Nutrient Information for German Seniors. In *6th EAI International Conference on Smart Objects and Technologies for Social Good (GoodTechs '20), September 14–16, 2020, Antwerp, Belgium.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3411170.3411227

1 INTRODUCTION

Speech assistants are becoming more and more popular since in many situations they allow the easiest and most intuitive interaction with a computer system. Especially, for seniors such interfaces make it easier to interact with assistive modern technology. Nevertheless, seniors are a special user group since they face age related problems, are often less familiar with current technology and partly don't have access to a stable internet connection which is required for most commercial, cloud-based speech assistants. Since seniors are often using such systems for health-related tasks like nutrition or drug intake, privacy concerns arise.

In this paper we provide an overview of current open source solutions for developing privacy respecting speech assistants. For our implemented prototype we chose the task of receiving nutrition information of a packaged foodstuff since the labeling on packages is often hardly readable for seniors and often lack of healthiness ratings. Since we developed this system for German elderly users, it was a requirement that the German language was supported by the system. We use an enforced version of the term "privacy-by-design", since the speech assistant should not only be able to run offline but additionally its software should be open source (including machine learning models) and the required data should stem from open data providers. In this case the *open food facts*[1] database was used as data source. This prototype was demonstrated in a workshop with two seniors from which we gathered qualitative feedback in a semi-structured interview and discuss its outcomes. Those findings are relevant for future privacy-by-design speech assistants, which enable seniors using state-of-the-art assistive technology.

2 RELATED WORK

2.1 Speech Assistants for Seniors

In this section we provide an overview of recent literature about speech assistants for seniors. Examples of systems including robots related to nutrition information can be found in a following section.

Jesús-Azabal et al. [14] presented a combination of speech assistant and smartphone app for seniors to remind them about their daily drug dosage. Their voice assistant had two main requirements: (1) it should be able to work offline and (2) it needed to support appropriate speech for elders. The system used the open source speech assistant software *Snips* [7]. Shalini et al. [30] developed a voice assistant app for the Google Assistant and Amazon Echo platforms to manage seniors' personal health, using commercial devices with a screen. The system was used by older adults and family members and provided audiovisual, easy to understand health information on spoken queries. In the work by Cheng et al. [6] a speech assistant for elderly persons facing type 2 diabetes was developed and evaluated. It was based on Google Home and a web interface to show visualizations. The seniors in the preliminary study preferred the system against smartphone apps.

All authors of this literature selection chose assistive, health related topics for seniors. Additionally, their speech assistants were mainly intended to simplify the interaction with the systems. The prototypes usually combined a graphical and speech interface. This is also the case for our prototype. However, our system should be able to run offline and *Snips* is no longer an option (see 3.1.1).

[1]https://openfoodfacts.org accessed April 28 2020

GoodTechs '20, September 14–16, 2020, Antwerp, Belgium

2.2 Recognition of Foodstuff Type and Amount

Providing nutrient information requires at least the recognition of the type of food. For example, the type of non-packaged food and – to some extent – drink can be automatically detected by using photos [21], gas sensors [8] or other sensors capable of measuring unique physical or chemical properties. At least the food intake and consistency (e.g. crispy) of a foodstuff can be recognized by analyzing audio data [33] if a person is speaking during eating. For example a drinking activity can be detected by using accelerometer data of a smartwatch [10] or by observing vessels with a pressure matrix on a table [36]. Nevertheless, manual methods are still required since automatic recognition is not always accurately possible, e.g. if the foodstuff is self-prepared. In [28], a multi device system was developed to reduce the effort of manual logging by combining a smartwatch, smartphone and a mobile smartscale. The smartscale helps to estimate the foodstuff amount, which may play an important role in some use-cases.

For our speech assistant we decided to use the bar-codes on the products' packagings. One can obtain the nutrients and estimate the healthiness of the foodstuff from open databases like *open food facts*. This is also a very robust and widely used solution in food logging with smartphone apps. There are also automatic systems estimating the calorie amount [9] solely based on photos. Nevertheless, such systems can just provide hints since, e.g. the amount of contained sugar is usually invisible.

2.3 Nutrition Information / Recommendations

2.3.1 Understandable Nutrition Information. In Germany and some other countries, there is currently no compulsory indicator on the packaging allowing for a quick assessment of the healthiness of food. With our speech assistant we wanted to make it easier for seniors to identify healthy and unhealthy foodstuff while providing hints for the reason. For the rating we depend on the Nutri-Score [15], which is a rating from A (healthy) to E (unhealthy). This score is already introduced in some countries, such as in France, and at least discussed in several others. There are several other types of indicators available with different advantages and disadvantages, which could be researched in future work.

2.3.2 Robots. There are several examples for nutrient information systems that are based on robots. *Pillo* [32] is a commercial robot which acts as a health manager for its family and is able to answer questions about the amount of calories in certain food. In [26] a robot providing information about the calories of beverages was introduced. However, no direct interaction, such as asking questions, was possible. A concept of a robotic dietitian with adaptive linguistic style was presented in [24]. Its main idea was to make the robot more persuasive in a diet setting by using personalized language. Recommendations for nutrition were part of the assistive robot for seniors presented in [27]. These recommendations, intended to improve the seniors' wellbeing, stemmed from the *CARE* context-aware recommender system introduced in [29]. A dialog system for a robot to recommend dishes by determining the personal preferred texture and taste of food was presented in [35].

2.3.3 Conversational Agents. In [17], a task-oriented conversational system called *CH1* for diabetic patients was introduced, intended to calculate the amount of consumed carbohydrates in a dish. The system was implemented for the Italian language and interpreted textual user utterances and generated text outputs. The system *NESTORE* [3] was intended to support seniors to achieve a healthier lifestyle. It acts as a coach for several health domains, including nutrition. This recent system shares some goals of the CARE system. One possibility to interact with NESTORE is a chatbot whose natural language understanding was powered by the Rasa framework (see section 3.1.2). Casas et al. [5] presented a chatbot based on the commercial *Chatfuel* service, which is intended to improve the food lifestyle. The user defines the goals to reach, such as a reduced meat consumption. In [2] a conversational agent called *Foodie Fooderson* relied on IBM's Watson conversational services. It was designed to help to reduce food waste and to improve eating habits with recipe recommendations, which considered allergies and dietary goals.

3 DEVELOPMENT OF A SPEECH ASSISTANT

Usually, a speech assistant consists of automatic speech recognition (ASR), natural language understanding (NLU), a dialog manager, a natural language generator (NLG) and a text to speech (TTS) engine. Currently, available solutions from Amazon, Apple, Google or Microsoft heavily depend on cloud-based services which facilitate building speech assistants or chatbots while using the infrastructure of the specific provider. The use of online services for gathering or processing data is common for many tasks today. However, cloud-based data processing always goes hand in hand with privacy risks, with questions regarding the source of information or uncertainty with whom the data was shared. In particular, speech recognition comes with a high privacy risk. This is why we decided to develop a privacy-by-design speech assistant.

At first, we will provide an overview of components to build open source speech assistants. Afterwards we present our implementation of a speech assistant with the use case to provide nutrition information for a specific product.

3.1 Components

Several solutions exist for ASR, NLU, the implementation of a dialog manager, NLG and TTS. We present a selection of current open source options, with a focus on the German language.

3.1.1 Speech Assistant Frameworks. Several frameworks combine components to simplify the development of speech assistants. Table 1 lists a selection of open source frameworks. They support non-English language, run offline and still are in active development. Only the open source solutions for ASR and TTS are listed.

Snips [7] was a popular open source speech assistant, following the private-by-design principle and provided a German language model. This software would have been an interesting option. Nevertheless, Snips was recently acquired by Sonos, which closed down an essential part of the system (Snips Console) which was not open source. Thus, it was not considered as an option in this overview.

All solutions currently face the same problem: the missing support for non-English languages for their ASR and TTS. For example, it is hard to find generic, good working models for German. A recent

ASR based on deep learning is Mozilla's DeepSpeech[2], but currently (April 2020) there is not yet a good working German model. For Kaldi we found an appropriate model we will describe in a later section. TTS is – in our case – less critical as long as the generated German speech is reasonably good understandable.

Table 1: Active open source speech assistant frameworks supporting non-English languages and with offline support.

Name	ASR	TTS
Jasper[3]	Pocketsphinx, Julius	Festival TTS, Flite, pico TTS, MaryTTS
Kalliope[4]	CMUSphinx	eSpeak, MaryTTS, pico TTS
Leon[5]	DeepSpeech	Flite
LinTO[6]	LinSTT (Kaldi based)	pico TTS
Mycroft[7]	DeepSpeech or Kaldi	Mimic, eSpeak, MaryTTS
Open Assistant[8]	CMUSphinx	eSpeak

3.1.2 Chatbot Frameworks. Some solutions already incorporate NLU and dialog management. Nevertheless, these components can also be found in chatbot frameworks, which usually just lack the ASR and TTS components. One solution is the Microsoft Bot Framework[9]. It is open source but mainly intended to be used with Microsoft's Azure cloud services. Another framework is Rasa [4], which already incorporates offline NLU for non-English languages and dialog management. There are several other chatbot frameworks available, but especially Rasa is a quite sophisticated solution thanks to its integrated NLU and dialog manager.

3.1.3 Speech Recognition. Deep Neural Networks in speech recognition are usually superior to older techniques (e.g. HMMs) in terms of word error rates. Thus, we will focus on ASR frameworks based on modern algorithms. For two of them we could find German models. DeepSpeech currently lacks an official German language model. Recently, a German model was published by [1], which has not been available during our development phase. For our prototype we relied on a model for Kaldi [23], which was described in [20].

Automatic activation on voice was not provided by our system to allow full transparency when audio recording is conducted and since an interaction with the GUI was required to take a picture of a bar-code. Nevertheless, an implementation would be possible by adding an open source keyword spotter like mycroft-precise[10] and a voice activity detector to start and stop recording.

3.1.4 NLP, Dialog Management and NLG. In this section selected open source frameworks for NLP, dialog management and NLG are presented. For **NLP** we found three popular solutions with existing German models: Apache OpenNLP[11], Stanford CoreNLP [18] and SpaCy[12]. SpaCy is a more recent software and applies deep learning methods by using Tensorflow. It is used as part of the NLU of the Rasa framework. For **dialog management** there are several open source solutions and their usage heavily depends on the complexity and purpose of the system. In our case we required a task-oriented solution without the need for more complex dialogs (e.g. small talk). Thus, the integrated template based solution of the Rasa framework was appropriate. From Rasa we also used the integrated **NLG** although it is currently quite limited. A more sophisticated option is for example SimpleNLG [11] or RosaeNLG[13], which provide German language models. Deep learning techniques are currently not widely applied in open source solutions.

3.1.5 Speech Synthesis. For TTS there are several older open source solutions. For our prototype we decided to use MaryTTS[14] for which reasonable German language models are available. The generated audio can be easily received from its integrated server. Other TTS software with German models are listed in Table 1. In the future Mozilla's TTS[15] using deep learning methods could be an appropriate choice. Together with their ASR these components are intended to be integrated in web browsers as part of the Web Speech API[16].

3.2 Implementation for Nutrient Information

An overview of our system is shown in Figure 1. We installed all components in different docker containers in a virtual machine (VM) running Xubuntu 19.04 x86-64. Eight cores of the host system's CPU (Intel Xeon E5-2690 v4 CPU), 8 GiB RAM and 80 GiB HDD storage were dedicated to the VM. The communication between the components was provided by the program logic on the client device (simplified in the figure).

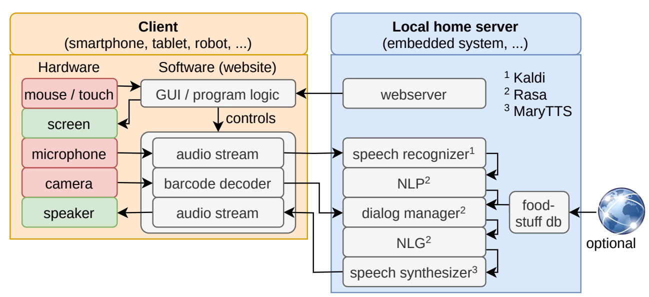

Figure 1: Simplified system overview.

The **GUI** (cf. Figure 2) and **program logic** were provided by an HTML5 website with Javascript code. At first a user would photograph the bar-code, which was decoded by the Javascript code and sent to the Rasa instance running on the server via a websocket connection. A textual response from Rasa with the product name was sent back. The text was displayed and speech was generated

[2]https://github.com/mozilla/DeepSpeech accessed April 28 2020
[2]https://jasperproject.github.io/ accessed April 28 2020
[3]https://kalliope-project.github.io/ accessed April 28 2020
[4]https://getleon.ai/ accessed April 28 2020
[5]https://linto.ai/ accessed April 28 2020
[6]https://mycroft.ai/ accessed April 28 2020
[7]https://openassistant.org/ accessed April 28 2020
[9]https://dev.botframework.com/ accessed April 28 2020
[10]https://github.com/MycroftAI/mycroft-precise accessed April 28 2020
[11]https://opennlp.apache.org/ accessed April 28 2020
[12]https://spacy.io/ accessed April 28 2020
[13]https://rosaenlg.org accessed April 28 2020
[14]http://mary.dfki.de/ accessed April 28 2020
[15]https://github.com/mozilla/TTS accessed April 28 2020
[16]https://wicg.github.io/speech-api/ accessed April 28 2020

from it by sending the text to MaryTTS and playing back the resulting audio data on the client. Now the user could ask a question by using text input or ASR. For the ASR the audio recording could be started and stopped with a record button. After stopping, the audio data was sent to the ASR server and the transcribed text was received and displayed by the GUI. Additionally, the text was sent to the Rasa instance, which generated its response. Like before, the response was made visible by the GUI and TTS was applied.

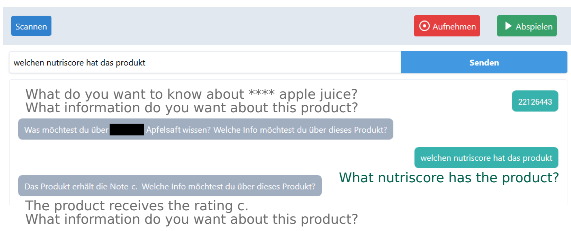

Figure 2: A screenshot of the GUI allowing scanning a bar-code, starting / stopping the speech recording, text input and replaying the audio of the TTS. Translations were added.

For **ASR** we integrated Kaldi with the model tuda_swc_mailabs_-voc400k[17], which was the best and also the largest model provided by [20]. For an integration we set up the suggested Kaldi GStreamer server, which allows sending uncompressed, mono audio data with a sample rate of 16 kHz to Kaldi and receiving its response via a web-socket connection. For a German sentence like "Tell me how much sugar is contained in the foodstuff." a time duration of about three to five seconds between sending the audio data to the server and receiving the recognized text was observed. This was an acceptable delay and includes all major system components: ASR, NLP, dialog manager, NLG and TTS (generation time). Some components' latencies like of the ASR depend on the word count. In our use case just short questions are asked where the ASR requires most processing time. Thus, for longer sentences it could be useful to reduce its latency by doing optimizations (e.g. model, GPU acceleration). The recognition accuracy in our tests was reasonable with laptop and smartphone microphones in relatively quiet environments. In some cases even flawed transcribed texts could be correctly interpreted by the NLU. Optionally, the system allows to directly input text via keyboard e.g. when the ASR would fail completely.

The **NLU, dialog management and NLG** was provided by the Rasa framework. Our system can give nutrient information, such as the amount of sugar, fat, calories, carbohydrates and protein in grams per 100 gram of food and drink. To allow an easier interpretation of the sugar and fat amounts, they are shortly explained by telling the user, whether this is a low, moderate or high amount. Similar to *open food facts* we use the thresholds from [12]. Additionally, the Nutri-Score and allergens can be output. So-called "intents" are detected by the NLU for recognizing user requests. Specific information, such as "sugar" is called an "entity". In total 43 sentences were provided as training data to allow the NLU to learn to detect the required intents and entities. The base model for SpaCy (used by Rasa) was "de_core_news_sm-2.1.0". A fallback, if the user input couldn't be understood, was included as an intent. In such cases the system could ask again. After detecting the intent and getting the entities, Rasa was able to acquire the required information with custom "action" functions in Python and to generate system responses by using templates. The Python functions received all nutrient information from a local *open food facts* sqlite database based on the bar-code. The database included 1,110,884 products, all without missing data, and could easily be updated with the most recent online version. The usage of a local instance of the database improves the latency of data requests and prevents tracking.

Finally, as **TTS**, MaryTTS with the German model "dfki-pavoque-styles" was used. It usually provided acceptable results but there might be better models. Especially, the names of products and their producers can introduce problems.

[17] https://github.com/uhh-lt/kaldi-tuda-de accessed April 28 2020

4 EVALUATION

Our implementation showed that it is currently technically possible to implement a German open source speech assistant with privacy-by-design on regular hardware. Additionally, we used this prototype to demonstrate the application of a speech assistant for nutrition information during a workshop with seniors.

The workshop was conducted by one researcher, which first explained the study procedure, what data was recorded (written notes) and for what purposes the data was used. Afterwards, the seniors were asked about their age and experience with speech assistants, smart home systems and smartphones. As a next step the system was demonstrated and the seniors could also try it out on their own. Finally, a semi-structured interview with both participants was conducted. The workshop took place in the seniors' home and altogether lasted around 60 minutes.

We acquired a senior couple (one female, one male) in the age between ca. 60 to 70 (avg. 67.5), which was living independently in their own home. They already had experience with an installed smart home system, which was mainly intended to provide information (e.g. data visualizations) accessed via web browser from a PC. Additionally, the system fulfilled some automation tasks. They owned a smartphone, but used it only occasionally. Both participants did not have any experience with speech assistants for smart home or smartphone, but were aware of this technology due to TV reports, advertisements and news.

The speech assistant was presented to the seniors by the researcher. For demonstration purposes one drink carton of orange and grapefruit juice were chosen, which were included in the local copy of the *open food facts* database. A Microsoft Surface 4 Pro (Intel m3 CPU) tablet with Windows 10 was used to run the web interface (front-end). The back-end VM was running on a compact desktop PC with AMD Ryzen 5 3600 CPU. The tablet and desktop PC shared the same network. With the carton of orange juice it was explained to the participants how to interact with the system and which information about a product can be obtained. Afterwards both seniors were free to test the system with a package of grapefruit juice.

Finally, a semi-structured interview was conducted. It was asked, whether the participants noticed problems or had ideas how to extend or improve the prototype's functionality. Additionally, we were interested in which situations and places the participants could envision to use such an assistant providing nutrition information. The acceptance to use a smartphone or a smartdisplay as interaction devices were asked, too. Since privacy played an important role in the development of the prototype we also questioned the seniors whether they would agree to share data (especially audio) with commercial companies when they would benefit from a speech assistant with better ASR, TTS and more functionality. We were also interested, which nutrition data was relevant to the seniors and whether the Nutri-Score and rating of the sugar amount were helpful to them.

5 RESULTS AND DISCUSSION

Technical setup. During the evaluation the speech recognition worked reasonably well in a silent environment and recognized most of both senior's questions. Nevertheless, ASRs heavily depend on the audio equipment and DSP. Optimized open hardware / source platforms are also available (e.g. ReSpeaker for Raspberry Pi[18]) which might improve the ASR's performance and could be considered for future prototypes.

During the try-outs a problem occurred, we were already aware of: the vocabulary of the ASR did not include the very special word "Nutri-Score". Instead, similar sounding German words were recognized. Thus, it would be necessary to train a new model with appropriate speech samples to get better results, which currently is a non-trivial task. Nevertheless, it was possible to get the Nutri-Score rating by asking whether a foodstuff is healthy or what the rating of it is. Since the NLU was limited in its answer possibilities it often guessed correctly the meaning of a question, even if it was not transcribed correctly. A domain specific ASR, such as applied in *Snips* [7], would have been preferable at least for this task oriented dialog system. The problem with unknown vocabulary also existed with the TTS engine. Especially, trademarks or company names of a foodstuff could be pronounced incorrectly. Nevertheless, the audio output could be understood by our workshop participants.

The combination of a graphical touch and speech interface worked well for this task. It was possible to see what the system understood and allowed an overview of what was already asked by the users. The possibility to start / stop the audio recording with a GUI button press on the touchscreen allowed the seniors to have more control over the system. Additionally, due to the use of bar-codes it was helpful to see the view of the camera to position the product correctly. In future prototypes the GUI could also be used for visualizations e.g. to compare nutrition facts like it was implemented for a different task in [30].

Seniors' interaction with speech assistants. Seniors face several age-related and partly health-related limitations that have to be considered for the interaction design. Their sight, hearing and motor abilities are reduced. Foodstuff packagings already include nutrition information, which were also provided by our system. Nevertheless, our workshop participants confirmed, that the text on the packagings are partly very hard to read for them due to being too small or printed with a bad contrast. Thus, they often required a reading aid and a place with good lighting conditions. Thus, our developed system would be useful for them. The seniors could imagine using such an assistant in their kitchen to get a better idea about the healthiness of specific foodstuff.

To provide a good usability of our system we ensured, that the text and GUI elements on the screen were big enough for an easy interaction. The literature reports, that direct manipulation with touchscreens can reduce the barriers for elderly people to interact with computer systems [22]. To simplify the usage of our system the GUI required only minimal interaction. The volume of the speech output was adjusted to be well understandable in front of the tablet. Its touchscreen worked well for both participants.

A recent overview of voice user interfaces for elderly can be found in [31]. Currently, there is relatively little research, which takes age into account when building voice interfaces, as confirmed by Stigall et al. In a study presented in [16] it was researched how age-related cognitive decline influenced the use of voice interfaces. Several implications for speech assistants, like longer pauses between words, are mentioned, which should be considered in future systems. Werner et al. [34] also indicate, that there might be age-related differences in the word error rates of ASR systems, which would need to be considered when building ASR models.

Privacy. The privacy-by-design aspect of our system was appreciated by the seniors. Especially, they would not have allowed sharing voice data with a cloud service and already tried not to do so with their smart TV, smartphone and PC with Windows 10. Severe privacy concerns related to ASR were also expressed by a senior in a study described in [25]. In general several privacy issues are known for different commercial products (e.g. [13]) and through media were partly known to the seniors of our workshop. In fact the perceived privacy risk seems to be an important hindering factor for the usage of speech assistants, which could also be observed in a study presented in [19] with 724 voice assistant users. Both workshop participants agreed, that they even prefer a partly worse working offline speech assistant over a cloud based solution with better performance. Our participants mentioned, that a commercial product being able to run completely offline could be a compromise.

Nutrition information and system's extensions. The workshop participants noticed the value of the system in delivering nutrition information of a specific product. They described it as more comfortable and appreciated the interpretation of the healthiness by the Nutri-Score and the amount of sugar in comparison. Additionally, they wished to receive information and an interpretation about included vitamins and minerals of a foodstuff. Unfortunately, such information is usually not provided by the producers and could, at least in Germany, only be received from non-open databases for some products like the "Bundeslebensmittelschlüssel"[19]. As a useful addition to the system the management of their foodstuff inventory was mentioned, so that they could get informed when products are near the expiration date or when certain foodstuff is running out.

6 CONCLUSION

In this paper we developed and evaluated a prototype of a privacy-by-design open source speech assistant for seniors providing nutrition information. We provided an overview of currently available open source solutions to develop speech assistants, which are also usable for non-English speakers, in this case for the German

[18] https://github.com/respeaker/seeed-voicecard accessed April 28 2020

[19] https://www.blsdb.de/bls?background accessed April 28 2020

language. Next to related work we gathered further insights by conducting a workshop with two seniors where the developed prototype could be tested. It became clear that senior's could greatly profit from speech assistants, but their special needs would have to be addressed in the design and their privacy concerns should be respected, which can especially be assured with open source software and open data.

ACKNOWLEDGMENTS

We want to thank the students Andreas Lenhardt and Thomas Stöhr for their help implementing the system prototype.

REFERENCES

[1] Aashish Agarwal and Torsten Zesch. 2019. German End-to-end Speech Recognition based on DeepSpeech. In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Long Papers*. German Society for Computational Linguistics & Language Technology, Erlangen, Germany, 111–119.
[2] Prashanti Angara, Miguel Jiménez, Kirti Agarwal, Harshit Jain, Roshni Jain, Ulrike Stege, Sudhakar Ganti, Hausi A Müller, and Joanna W Ng. 2017. Foodie fooderson a conversational agent for the smart kitchen. In *Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering*. IBM Corp., 247–253.
[3] Leonardo Angelini, Elena Mugellini, Omar Abou Khaled, Christina Röcke, Sabrina Guye, Simone Porcelli, Alfonso Mastropietro, Giovanna Rizzo, Noemi Boqué, Josep M del Bas, et al. 2019. The NESTORE e-coach: accompanying older adults through a personalized pathway to wellbeing. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '19)*. 620–628.
[4] Tom Bocklisch, Joey Faulkner, Nick Pawlowski, and Alan Nichol. 2017. Rasa: Open source language understanding and dialogue management. *arXiv preprint arXiv:1712.05181* (2017).
[5] Jacky Casas, Elena Mugellini, and Omar Abou Khaled. 2018. Food diary coaching chatbot. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 1676–1680.
[6] Amy Cheng, Vaishnavi Raghavaraju, Jayanth Kanugo, Yohanes P Handrianto, and Yi Shang. 2018. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1–5.
[7] Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, et al. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190* (2018).
[8] Chi Tai Dang, Andreas Seiderer, and Elisabeth André. 2018. Theodor: A Step Towards Smart Home Applications with Electronic Noses. In *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*. ACM, 11.
[9] Takumi Ege and Keiji Yanai. 2017. Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*. 367–375.
[10] Simon Flutura, Andreas Seiderer, Ilhan Aslan, Chi-Tai Dang, Raphael Schwarz, Dominik Schiller, and Elisabeth André. 2018. DrinkWatch: A Mobile Wellbeing Application Based on Interactive and Cooperative Machine Learning. In *Proceedings of the 2018 International Conference on Digital Health*. ACM, 65–74.
[11] Albert Gatt and Ehud Reiter. 2009. SimpleNLG: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*. 90–93.
[12] Kornelia Hagen. 2010. Nutritional information: traffic light labelling is the best way to reach consumers. *Weekly Report* 6, 19 (2010), 141–151.
[13] Matthew B Hoy. 2018. Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly* 37, 1 (2018), 81–88.
[14] Manuel Jesús-Azabal, Javier Rojo, Enrique Moguel, Daniel Flores-Martin, Javier Berrocal, José García-Alonso, and Juan M Murillo. 2019. Voice Assistant to Remind Pharmacologic Treatment in Elders. In *International Workshop on Gerontechnology*. Springer, 123–133.
[15] Chantal Julia, Serge Hercberg, World Health Organization, et al. 2017. Development of a new front-of-pack nutrition label in France: the five-colour Nutri-Score. *Public Health Panorama* 3, 04 (2017), 712–725.
[16] Masatomo Kobayashi, Akihiro Kosugi, Hironobu Takagi, Miyuki Nemoto, Kiyotaka Nemoto, Tetsuaki Arai, and Yasunori Yamada. 2019. Effects of Age-Related Cognitive Decline on Elderly User Interactions with Voice-Based Dialogue Systems. In *IFIP Conference on Human-Computer Interaction*. Springer, 53–74.
[17] Bernardo Magnini, Vevake Balaraman, Mauro Dragoni, Marco Guerini, Simone Magnolini, and Valerio Piccioni. 2018. Ch1: A conversational system to calculate carbohydrates in a meal. In *International Conference of the Italian Association for Artificial Intelligence*. Springer, 110–122.
[18] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*. 55–60.
[19] Graeme McLean and Kofi Osei-Frimpong. 2019. Hey Alexa… examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior* 99 (2019), 28–37.
[20] Benjamin Milde and Arne Köhn. 2018. Open Source Automatic Speech Recognition for German. In *Speech Communication; 13th ITG-Symposium*. VDE, 1–5.
[21] Zhao-Yan Ming, Jingjing Chen, Yu Cao, Ciarán Forde, Chong-Wah Ngo, and Tat Seng Chua. 2018. Food photo recognition for dietary tracking: System and experiment. In *International Conference on Multimedia Modeling*. Springer.
[22] Lilian Genaro Motti, Nadine Vigouroux, and Philippe Gorce. 2014. Interaction techniques for older adults using touchscreen devices: a literature review from 2000 to 2013. *Journal d'Interaction Personne-Système* 3, 2 (2014), 1–26.
[23] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society.
[24] Hannes Ritschel, Kathrin Janowski, Andreas Seiderer, and Elisabeth André. 2019. Towards a Robotic Dietitian with Adaptive Linguistic Style. In *Proceeding of the Poster and Workshop Sessions of AmI-2019, the 2019 European Conference on Ambient Intelligence*.
[25] Hannes Ritschel, Kathrin Janowski, Andreas Seiderer, Stefan Wagner, and Elisabeth André. 2019. Insights on usability and user feedback for an assistive robotic health companion with adaptive linguistic style. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 319–320.
[26] Hannes Ritschel, Andreas Seiderer, Kathrin Janowski, Ilhan Aslan, and Elisabeth André. 2018. Drink-O-Mender: An Adaptive Robotic Drink Adviser. In *Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction*. ACM, 3.
[27] Hannes Ritschel, Andreas Seiderer, Kathrin Janowski, Stefan Wagner, and Elisabeth André. 2019. Adaptive Linguistic Style for an Assistive Robotic Health Companion Based on Explicit Human Feedback. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Rhodes, Greece) *(PETRA '19)*. ACM, 247–255.
[28] Andreas Seiderer, Simon Flutura, and Elisabeth André. 2017. Development of a mobile multi-device nutrition logger. In *Proceedings of the 2nd ACM SIGCHI International Workshop on Multisensory Approaches to Human-Food Interaction*. ACM, 5–12.
[29] Andreas Seiderer, Stephan Hammer, Elisabeth André, Marcus Mayr, and Thomas Rist. 2015. Exploring Digital Image Frames for Lifestyle Intervention to Improve Well-being of Older Adults. In *Proceedings of the 5th International Conference on Digital Health 2015 - DH '15*. ACM Press, Florence, Italy, 71–78. 00009.
[30] Shradha Shalini, Trevor Levins, Erin L Robinson, Kari Lane, Geunhye Park, and Marjorie Skubic. 2019. Development and Comparison of Customized Voice-Assistant Systems for Independent Living Older Adults. In *International Conference on Human-Computer Interaction*. Springer, 464–479.
[31] Brodrick Stigall, Jenny Waycott, Steven Baker, and Kelly Caine. 2019. 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*. 423–427.
[32] Vincent Tao, Kristi Moy, and Vibhuti Arya Amirfar. 2016. A little robot with big promise may be future of personalized health care. *Pharmacy Today* 22, 9 (2016).
[33] Johannes Wagner, Andreas Seiderer, Florian Lingenfelser, and Elisabeth André. 2015. Combining Hierarchical Classification with Frequency Weighting for the Recognition of Eating Conditions. *INTERSPEECH* (2015), 5.
[34] Lauren Werner, Gaojian Huang, and Brandon J Pitts. 2019. Automated Speech Recognition Systems and Older Adults: A Literature Review and Synthesis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications Sage CA: Los Angeles, CA, 42–46.
[35] Jie Zeng, Yukiko I Nakano, Takeshi Morita, Ichiro Kobayashi, and Takahira Yamaguchi. 2018. Eliciting user food preferences in terms of taste and texture in spoken dialogue systems. In *Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction*. 1–5.
[36] Bo Zhou, Jingyuan Cheng, Mathias Sundholm, Attila Reiss, Wuhuang Huang, Oliver Amft, and Paul Lukowicz. 2015. Smart table surface: A novel approach to pervasive dining monitoring. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 155–162.