Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 449–458
Marseille, 11–16 May 2020
© European Language Resources Association (ELRA), licensed under CC-BY-NC

BLISS: An Agent for Collecting Spoken Dialogue data about Health and Well-being

Jelte van Waterschoot* ✉, **Iris Hendrickx**†, **Arif Khan**†, **Esther Klabbers**‡, **Marcel de Korte**‡, **Helmer Strik**†, **Catia Cucchiarini**† and **Mariët Theune***
*Human Media Interaction, University of Twente
Drienerlolaan 5, 7522 NB, Enschede, The Netherlands
{j.b.vanwaterschoot, m.theune}@utwente.nl

†Centre for Language Studies/Centre for Language and Speech Technology, Radboud University,
Erasmusplein 1, Nijmegen, The Netherlands
{i.hendrickx, a.khan, w.strik, c.cucchiarini}@let.ru.nl

‡ReadSpeaker
Dolderseweg 2A, 3712 BP, Huis ter Heide, The Netherlands
{esther.judd, marcel.korte}@readspeaker.com

Abstract

An important objective in health-technology is the ability to gather information about people's well-being. Structured interviews can be used to obtain this information, but these are time-consuming and not scalable. Questionnaires provide an alternative way to extract such information, yet they typically lack depth. In this paper, we present our first prototype of the Behaviour-based Language-Interactive Speaking Systems (BLISS), an artificial intelligent agent which intends to automatically discover what makes people happy and healthy. The goal of BLISS is to understand the motivations behind people's happiness by conducting a personalized spoken dialogue based on a happiness model. We built our first prototype of the model to collect 55 spoken dialogues, in which the BLISS agent asked questions to users about their happiness and well-being. Apart from a description of the BLISS architecture, we also provide details about our dataset, which contains mentions of over 120 activities and 100 motivations and is made available for usage.

Keywords: conversational, spoken dialogue system, happiness model, natural language generation, healthcare, spoken Dutch

1. Introduction

Recent projections show that in the near future the health sector will deal with a growing demand for healthcare, an increasing number of vacancies, and higher expenditures. Amongst others, this has led to a paradigm shift in healthcare that emphasizes prevention, citizen empowerment and self-management and in which citizens are increasingly required to assume an independent, self-determining position. Along with these changes, there has been a critical analysis of the current definition of health adopted by the World Health Organization (WHO) that describes health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity."

Huber et al. (2011) discuss the shortcomings of this definition and suggest an alternative definition of health which is defined as "the ability to adapt and to self manage".

The view of health adopted in this paper is in line with Huber et al. (2011) and with a Positive Psychology view (Seligman, 2002; Seligman, 2011) in which positive experiences play a central role. In Dutch health care systems, this view on positive health and happiness has been widely embraced. Caretakers are trained to focus on the broad definition of health, including physical, mental and social well-being, and more holistic topics such as quality of life and self-management (Ministerie van Volksgezondheid, Welzijn en Sport, 2016).

This new definition of health requires operationalizations and appropriate instruments for measuring positive health dimensions such as functional status, quality of life and sense of well-being (Huber et al., 2011). Professionals attempt to gain insight into these dimensions through questionnaires and interviews with people who receive long-term or structural health care. This leads to insights, assessments and opportunities for positive health for the clients and caretakers. In addition, in-depth qualitative interviews identify opportunities for happiness improvements.

In-depth interviews provide the most insights, but require a serious time investment, both for the actual interview and the analysis and reporting. The research reported in this paper is couched in a larger project, Behaviour-based Language-Interactive Speaking Systems (BLISS), that attempts to offer a solution by developing an intelligent, personalized system that communicates with clients in spoken language to facilitate their self/joint-management of health, wellness and psychological well-being - and collects those measurements and insights at the same time. Razavi et al. (2019) developed a similar system and found that communication skills of older adults could be improved through such a system.

In this paper we report on the first steps undertaken to develop the BLISS agent for self-management of health, wellness and psychological well-being, in particular the initial phase of data collection. The aim was to start with available language resources for the Dutch language to develop a first version of the system that could be used to collect initial data.

2. Background

In the early 1990's, Defense Advanced Research Projects Agency (DARPA) launched the Airline Travel Information System (ATIS) project (Price, 1990), which sparked research on spoken dialogue system (SDS). The first SDS for the Dutch language was the OVIS system (Strik et al., 1997), which was a train timetable information system. The Continuous Speech Recognition (CSR) module used acoustic models (HMMs), language models (unigram and bigram), and a lexicon. The output of the CSR, a word graph with acoustic likelihoods, was processed by the Natural Language Processing (NLP) module using syntactic unit counts and concept bigram values. The concepts were defined in a stochastic attributed context-free grammar (ACFG). The spoken language generation part of OVIS consisted of a template-based language generation module linked to a speech synthesis system (Theune et al., 2001; Theune et al., 1997). The OVIS system was developed using a bootstrapping method.

As a follow-up to OVIS, the IMIX project (van den Bosch and Bouma, 2011) developed a multimodal question answering system for Dutch, combining speech and visual modalities. One of its use cases was answering questions about repetitive strain injury; however, this was only for demonstration purposes.

A look at the healthcare applications employing spoken dialogues, reveal that they have been developed mainly for limited domains such as breast cancer screening (Beveridge and Fox, 2006) or military mental healthcare (Morbini et al., 2012). An example from the field of persuasive technology is (Meschtscherjakov et al., 2016), focusing on support for speedy recovery or taking up regular exercise or medication. The Council of Coaches (COUCH) project uses multiple virtual agents to provide support for users who have for example diabetes or COPD (op den Akker et al., 2018). In the health domain we do not want to give the wrong information to patients. Therefore instead of dealing with free speech as input, Bickmore and Picard (2005) suggest to use a menu of options or limited text input, to both make the dialogue smoother and prevent the system from making crucial errors such as giving users the wrong answer to their questions, because of mishearing the user. Especially in health applications where a high intent accuracy is required, often no free speech is used (Bickmore and Giorgino, 2006, p. 563). Similarly, the virtual agents from COUCH use speech, but the users interact with them using input selected from a menu.

Chatbots are becoming more popular to offer 24h customer support, and we can also see this trend in healthcare (Kardol, 2015). For example, Chantal[1] and Bibi[2] are both virtual general practitioner assistants who can chat in Dutch (written communication) about health care issues and practical questions like making an appointment to speak with the GP. Also personal assistants such as Anne[3], and robots like Tessa[4] or Zora[5], have been put into elderly homes to help older adults (Martinez-Martin and del Pobil, 2018).

All of these systems are designed to answer domain-specific user questions. Our focus in BLISS is on long-term interaction, asking engaging questions (instead of answering questions) and learning a user happiness model through normal spoken conversation. ELIZA, one of the first chatbots, was rule-based and designed as a therapeutic chatbot that could ask questions to users (Weizenbaum, 1966). Users talking to ELIZA disclosed personal information and were engaged with her. This kind of interaction is very different from how people interact with smart devices nowadays. As noted by Radlinski et al. (2019), communication with smart devices is often very command-like in style and not similar to human-human communication. They set up a Wizard-of-Oz experiment to collect a dataset of more spoken natural conversations in the context of movie recommendation and found indeed that these conversations contain far more complex information than what smart devices are capable of now. Similarly, with BLISS we want to have people speaking with the agent in a natural conversational way. Specifically for obtaining natural conversation data about personal topics, Zhang et al. (2018) collected PERSONA-CHAT, a dataset containing text-based chitchat between two people recruited via crowdsourcing. They trained a chatbot on the dataset and indeed found that the chatbot was more engaging to talk to for people than chatbots trained on other resources such as Twitter. More importantly for our research, the profiles the chatbots generated from the conversations with users contained valuable information about the users' personal lives.

3. Architecture

In BLISS, we use the classical spoken dialogue system architecture for our agent, consisting of five main components: the Automatic Speech Recognition (ASR), Text-to-Speech Synthesis (TTS), the Dialogue Manager (DM), the Natural Language Generation (NLG) and Natural Language Understanding (NLU). Another component is Apache ActiveMQ[6], a message broker service that the other components use for communication. Figure 1 shows how these components interact with each other in further detail. In the current implementation of the BLISS agent, the NLU, NLG and DM components run locally on the device, whereas the ASR and TTS components are off-the-shelf products and run as cloud-services. We designed the system in such a way that in the future we will be able to add an embodiment (e.g. a virtual character) to the TTS service or to use another speech recognition server for the ASR component (e.g. with ap ersonalized recognition model).

The whole interaction process can be briefly described as follows. Whenever the ASR receives audio from the microphone connected to the DM, the ASR creates a transcription and sends it to the DM. The DM forwards the transcription to the NLU component, which returns an intent of the user. The DM then calls on the NLG component to match the

[1] https://zaurus.nl/chantal/
[2] http://virtueledoktersassistent.nl/
[3] https://anne4care.nl/
[4] https://www.tinybots.nl/
[5] http://zorarobotics.be/
[6] https://activemq.apache.org/

intent of the user to an intent and behaviour of the agent. Once the behaviour of the agent has been selected, the TTS receives a message from the DM to realize the agent's behaviour by generating the speech. The generated speech is sent to the DM, which plays the audio.

3.1. Automatic Speech Recognition

For the ASR, the Corpus Spoken Dutch (CGN)[7] was used to train the Acoustic Model (AM) and Language Model (LM) for the speech recognition component. For training the AM, the KALDI (Povey et al., 2011) framework was used. We implemented a cloud-based, speech recognition server and used the neural network based online decoding with iVectors to set up the speech recognition server [8]. The server listens to the audio, which it receives over the internet and sends the decoded transcription.
The user speaks into the microphone of a headset using a laptop and this audio is recorded and sent to the ASR server using a websocket. The ASR can detect the end-of-sentence in the speech signal.

3.2. Natural Language Understanding, Dialogue Management and Natural Language Generation

The DM is responsible for responding to the user behaviour perceived via the input component (ASR) and for generating the agent behaviour that is realized via an output component (TTS). It also controls the NLU and NLG components. The DM of the BLISS agent is based on the dialogue engine Flipper (van Waterschoot et al., 2018). It uses an information-state based approach, keeping track of all user information and of what agent behaviours have been performed. The DM is connected to a PostgreSQL database, where we store all of the dialogue information per user.

intent	example keyword(s)
question	what do you mean
inform	-
confirm	yes
disconfirm	no
salutation	hello
valediction	goodbye
stalling	ehm
auto-feedback	uh-huh

Table 1: List of user intents that can be recognized by the prototype with examples (translated to English). The default intent is an inform.

The NLU component in our first prototype, used to collect the data described in Section 4., uses keyword-spotting for intent recognition. We selected relevant intents from the DIT++ taxonomy for our prototype (Bunt et al., 2010). User intents can be classified as question, salutation, inform, valediction, confirm, disconfirm, stalling and auto-feedback (see Table 1). Additionally, we use the Dutch Pattern[9] library to extract emotion from the transcripts of the ASR and for retrieving verbs and nouns from user responses (De Smedt and Daelemans, 2012). Stopwords are filtered with spaCy's[10] default stopword list for Dutch. In our prototype, the nouns and verbs represent the activities of a particular user.
The NLG component is rule- and template-based. In our first prototype, the agent follows a script of small-talk after which it starts asking the user questions. For the generation of these questions we use templates, with placeholders for activities users talked about. The placeholders are filled with the verbs extracted from the user utterance by the NLU component, after lemmatizing them to fit in the template.

3.3. Text-to-Speech Synthesis

The TTS component in our prototype is provided by ReadSpeaker[11]. The current commercially available voices from ReadSpeaker are based on Unit Selection Synthesis (USS). The USS method (Hunt and Black, 1996) relies on a large acoustic database recorded by a professional voice talent, which is searched at synthesis time to find small audio segments which are concatenated to produce a smooth, natural-sounding utterance. This utterance is sent to the dialogue manager for playback.

4. Data Collection

To the best of our knowledge, our project is the first to work with a non-task oriented Dutch SDS that tries to elicit user information in the health and well-being domain. We require spoken conversational data, specifically about health and well-being. Public corpora such as the CGN unfortunately do not contain this type of data. Therefore we decided to first create a prototype of the BLISS agent with limited capabilities, able to collect this type of data. This first version of the system was tested at several venues with users, with the following two aims:

a: To find out how people interact with a computer when talking about their daily activities and underlying motivations for these activities.

b: To collect data that could be used for further improvements of the system.

In this section we describe our set-up for the data collection, together with an example of the dialogue flow, our pre-processing steps and the meta-data of participants.

4.1. Set-up

We tested our prototype at three different conferences with a predominantly Dutch-speaking audience. At each of these conferences, we used a separate room where users participated in an interaction with the BLISS agent. Our setup required an internet connection for the ASR and TTS cloud-services, a laptop for running the BLISS agent and a headset for speaking and listening to the agent. Users were given an information brochure before participating

[7] http://lands.let.ru.nl/cgn/
[8] https://github.com/opensource-spraakherkenning-nl/Kaldi_NL
[9] https://github.com/clips/pattern
[10] https://spacy.io/
[11] https://www.readspeaker.com

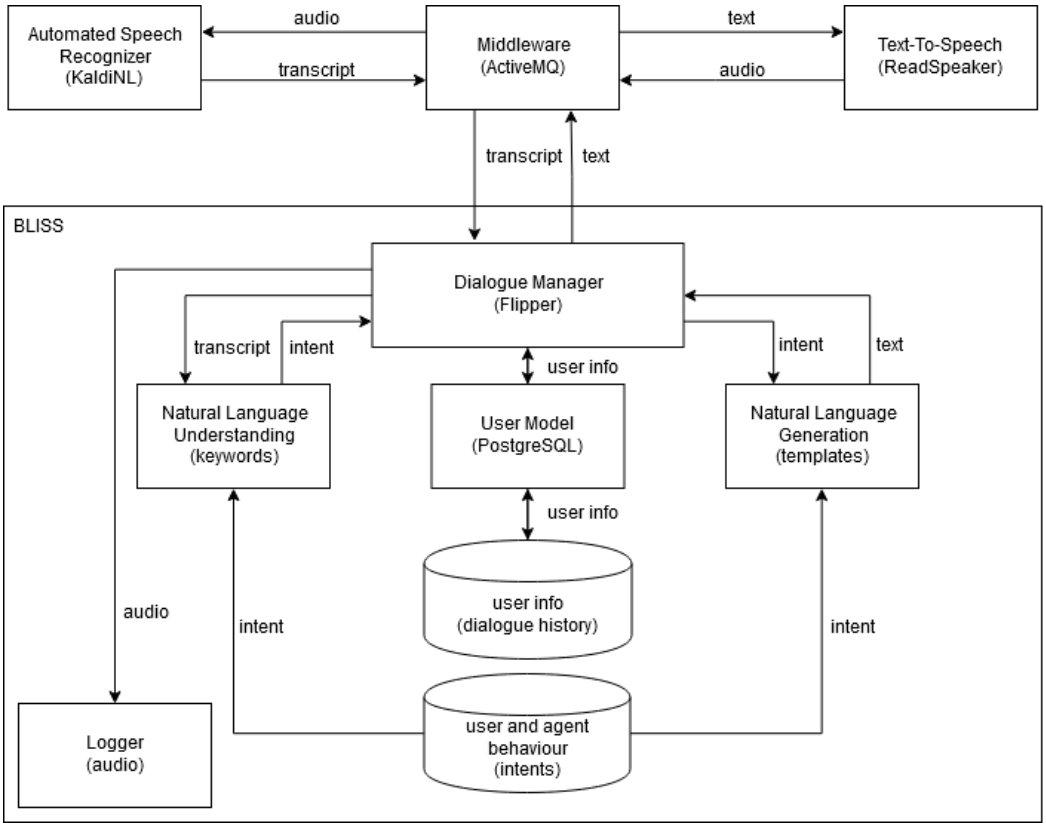

Figure 1: Visualization of the BLISS architecture. The BLISS agent consists of a dialogue manager, a natural language understanding and natural language generation component. A database is used to store user information and to retrieve intents for the agent and user. All communication with the TTS and ASR is handled by the ActiveMQ component.

and were asked to sign a written consent form and provide general demographic information (age range, gender and place of growing up)[12]. If the participant had no further questions, the agent initiated the conversation with the participant. Participants who did not respond to the agent's questions were instructed to repeat themselves. After the interaction, we debriefed participants about part of the workings of the BLISS agent.

4.2. Dialogue flow

In Table 2, an example dialogue of a participant with the agent is shown. The agent initiated each conversation with some introductory social dialogue to collect information about the familiarity of the participant with conversational agents, and their sentiment towards them. After this introductory part, the agent asked the participant three times to mention an activity they liked to do, each time with a follow-up question about the motivation behind the activity. Any user utterances following a question by the agent that were classified as an *inform intent* (see Table 1) were assumed to be the answer to that question. The agent ignored responses that were only transcribed as "ehm" (*stalling intent*), and waited for another user response instead. If the user spoke at the same time as the agent, the user's speech was recorded, but ignored by the agent. The agent only listened to what the user said after assigning the turn to the user. At the end of the dialogue, the agent asked whether participants had changed their opinion about talking to a computer, and finished the conversation.

We implemented two small forms of user adaptation in the first BLISS prototype. First, after the agent asks how participants feel about talking with a computer, the sentiment of the user's response is analyzed with Pattern. The agent then reacts differently to positive (>0.5), negative (<-0.5) or neutral (-0.5<= & <= 0.5) responses. Second, after asking the users about the activities they would like to do, the agent uses Pattern to extract the verbs from the user's answer (e.g. "cycling" from "I think that would be cycling."). Then the agent fills this in the template of the follow-up question about the user's motivation for doing the activity. If no non-stopword verb is detected in the user utterance, the default placeholder "that" is used.

4.3. Pre-processing

In total we recorded 59 sessions. We decided to include a session in our dataset if at least one of the questions by the agent had been answered by the participant. We discarded 4 conversations in which none of the questions had been answered, after which 55 sessions remained. Of the 55 sessions, 9 sessions had latency issues because the reply by the agent was (too) slow. Of these 9 sessions, 4 were incomplete because the participant could not answer all the questions. We decided to keep these 9 sessions in our dataset, because they contain answers to some of the questions and we are able to use the user speech as well, except for learning response times. Our total dataset thus

[12]These demographic labels match with the labels for CGN.

#	Speaker	Turn (translated corrected transcript)
1	Agent	Hello, have you ever talked with a computer before?
2	User	Yes
3	Agent	Okay, which ones have you talked with?
4	User	With Google Home
5	Agent	What do you think about talking with me?
6	User	It is pretty okay
7	Agent	I'm looking forward to knowing what you think about me at the end.
8	Agent	Now, I'm curious about you.
9	Agent	If you could choose one thing you want to do this weekend, what would it be?
10	User	Reading a bit.
11	Agent	Sounds good. Why do you like that?
12	User	It's very relaxing.
13	Agent	And if you could choose a second activity, what would that be?
14	User	I think that would be cycling.
15	Agent	What would you miss most if you wouldn't do cycling?
16	User	The physical activity, but then I would resort to walking.
17	Agent	Looking back at last week, which activity did you enjoy most?
18	User	Visiting my friends.
19	Agent	Great! What kind of feeling did it give you?
20	User	A lot of fun
21	Agent	Thanks! That's it for now. Do you still think the same about me as before?
22	User	Yes.
23	Agent	Thanks for participating, bye!

Table 2: Example dialogue of a user with the BLISS agent.

Region	Users
Dutch	42
Flemish	4
Other	9

(a) Region

Gender	Users
Female	33
Male	22
Undisclosed	0

(b) Gender distribution

Age	Users
18 –30	23
31 –45	16
46 –60	13
61 –110	3

(c) Age bins

Experience	Users
Yes	27
No	23
Unclear	5

(d) Familiarity

Table 3: We asked the participants for their region (a), gender (b), age (c), and familiarity with conversational virtual agents (d), where the age is divided into bins for life phases in line with the bins of the CGN. The region represents where participants grew up for most of their life between the ages of 4 and 16.

Activity Class	Frequency
Hobby	57
Social circle	30
Outdoor	22
Social activity	22
Rest	16
Work	14
Total	161

Table 4: This table shows the clusters of activities people talked about. For example, hobbies includes watching TV, but also walking and sports.

consists of 55 sessions.

4.4. Details

The sessions in our dataset have an average length of 2 minutes and 34 seconds (σ = 60 seconds). If we look at the 46 sessions that did not have any latency issues, the average length is 2 minutes and 18 seconds (σ = 24 seconds). The ASR module transcribed 662 utterances in total.

Table 3 shows that most of the participants were in the younger age categories (μ = 33,53, σ = 14,28). Of the 55 participants, 40% were male. Most of our participants (75%) were from the Netherlands, a few participants were Flemish (7%), while the remainder of the participants had a different country of origin. Around 50% of the users had talked to a dialogue system before, such as Siri, Google Assistant, or Alexa.

5. Qualitative Data Analysis

We performed an explorative qualitative analysis of our dataset. We are mainly interested in how people talk to the BLISS agent and which information we could extract from the conversation. We did a preliminary thematic analysis on the dataset to structure the information about what people said. We also analyzed some dialogue aspects, such as hesitations and repetitions and summarize the impressions participants had of the agent.

5.1. Activities & Motivations

The BLISS agent wants to learn what makes people happy and healthy. Part of people's happiness and health is determined by the activities they undertake. Therefore the agent needs to learn which activities make people happy and why they choose these particular activities, their motivations. In this section we search for common themes in the user's answers and cluster them. In Table 4 we show the clusters of activities mentioned during the dialogues. For the clustering we extracted the noun and verb phrases from the automatically transcribed answers and grouped them together under the common themes we identified. We excluded all answers that were incomprehensible (incorrect and incomplete transcripts), missing (system error), irrelevant (questions about the system) and answers in which users said

that they could not think of another activity. The classes in Table 4 are not mutually exclusive, as some of the users included multiple activities in an answer to an activity question. For example, "walking in nature" can be classified as both a hobby and outdoor activity. After filtering the answers, activities were clustered manually, which resulted in six classes in total.

1. *Hobby.* Activities such as watching TV, reading, travelling and doing sports.
2. *Outdoor.* Mentions of an outdoor location, such as the beach, the forest or specific cities.
3. *Resting.* Sleeping, doing nothing or just relaxing.
4. *Social circle.* Often people described not only the activities, but also with whom they wanted to do this activity, such as with their partner, friends or family.
5. *Social activities.* Activities such as eating out, going to a party or having coffee.
6. *Work.* Work-related activities, such as attending a conference or volunteering.

Answers to the third activity question the BLISS agent asked (Table 2, line 17) included more specific activities than the answers to the first and second questions (Table 2, line 9 and 13). For example, activities such as "celebrating a birthday at the office" or "I received my diploma yesterday" were all answers to the third question. To the first and second question, people generally responded with their hobbies like "reading" or "walking". Table 5 shows the categorization of the motivations, in which we clustered them similarly to the clustering of the activities. However, instead of deriving themes from the data, we used the dimensions of the dialogue tool[13] of the Institute of Positive Health (IPH), based on the work of Huber et al. (2011). We first filtered the motivations by excluding motivations that were incomprehensible, missing or irrelevant and excluding replies in which the users could not think of a motivation. Table 5 shows that most participants mentioned motivations related to feeling good and wanting to do something, because this makes them feel happy (quality of life).

1. *Quality of life.* Motivations related directly feeling good and happy and doing things you love.
2. *Daily functioning.* Motivations related to taking time for yourself and knowing what you need.
3. *Participation.* Motivations which include social contacts, such as family or friends and helping out others.
4. *Physical health.* Motivations related to wanting to exercise, have a regular sleeping pattern and feeling fit.
5. *Meaningfulness.* Motivations related to finding a purpose, being excited and wanting to learn.
6. *Mental well-being.* Motivations related to mental health and feeling in control of your life.

Motivation class	Frequency
Quality of life	43
Daily functioning	31
Participation	18
Physical health	12
Meaningfulness	8
Mental well-being	4
Total	116

Table 5: If we map the motivations to the positive health model of the IPH, based on the work of Huber et al. (2011), we can see that most users mention motivations related to their general quality of life and daily functioning.

The first and second motivation question asked by the BLISS agent (Table 2, line 11 and 15) often received responses in the dimension of daily functioning, whereas the third question (Table 2, line 19) mainly received responses in the quality of life dimension. Mental well-being and meaningfulness were not often mentioned as motivations. The second motivation question was less suited for the dialogue, because it asked about what people would miss, instead of asking directly why people liked a certain activity. It would sometimes lead to people repeating the activity they mentioned or saying "I wouldn't really miss anything". The third motivation question was often answered with different variations of "a good feeling".
In Table 6 we show the combinations of activities and motivations per question. For each of the activity questions, we combined it with the corresponding follow-up motivation question. For example, if the answer to the activity question was "to go for a walk with friends" (activity classes: *hobby*, *outdoor* and *social circle*) and the reason for this was "it is great to be in nature" (motivation class: *quality of life*), this would add 1 to each of the following combinations: *hobby - quality of life*, *outdoor - quality of life* and *social circle - quality of life*.
As a result, we see in Table 6 that most people who mentioned a hobby, often gave a motivation related to daily functioning or quality of life. Additionally, people could mention multiple motivations for one activity, or have one reason for multiple activities, hence the number totals in Table 6 are different from those in 4 and 5.

5.2. Interaction

Around 10% of all user transcripts contains a word that was explicitly not recognized by the ASR (labeled as 'unknown' by the ASR). The ASR recognizes nonverbal utterances like "uh", "uhm" and "mmm". Our dataset contains 55 utterances (on the total of 662) that only consisted of such non-verbal reactions. A common type of ASR error is a mistranscription leading to an incomprehensible utterance, like the one shown below, where it can be observed that the ASR wrongly transcribed the user's speech.

[13] https://iph.nl/download/dialogue-tool/

	Relaxation	Hobby	Social presence	Social activities	Outdoor	Work	Total
Physical health	3	6	0	1	4	0	14
Daily functioning	9	14	2	1	7	1	34
Mental well-being	1	0	1	1	0	1	4
Participation	1	5	8	5	2	1	22
Meaningfulness	1	2	0	0	0	0	3
Quality of life	1	16	10	6	8	5	46
Total	16	43	21	14	21	8	123

Table 6: Mentions of activities linked together with the motivations derived from the IPH model. There are six classes for both activities and motivations. All activities and motivations are derived from the transcripts of the ASR.

ASR *dat ik mijn ouders dan weer energie*
that I my parents than again energy

Speech *dat ik mijn ouders dan weer eens zie*
that I my parents then again a time see

In some cases the dialogue took a very long time to complete due to practical technical issues. This influenced the overall dialogue flow. We show an example of a problematic dialogue in Table 7, which contains the relevant answers mixed with backchannel (indicated with *) and more meta-level (indicated with **) utterances where the participant gives explicit feedback on the system quality. Note that the current version of the BLISS agent does not contain mechanisms to cope with explicit problem signals from users. For example, as seen in Table 7, the agent does not "understand" that the answer of the user (Line 2), is not an answer to the question it asked and continues asking the next question (Line 3).

We asked users for explicit feedback on their conversation with the agent. We started with a direct question to establish their familiarity with conversational agents, followed by an open question to determine their stance towards dialogue systems. Around 50% of the users already had talked to a dialogue system before such as Siri, Google Assistant, Google Home, Alexa. Most users (43.6%) were positive about engaging in a conversation with the agent ("amusing", "nice", "interesting"), 20% had a more negative, cautious attitude ("weird", "ill at ease") and in 36.4% of the answers we had a neutral answer ("it depends") or we could not determine the sentiment of the user answer. At the end of the conversation we asked whether their stance had changed after speaking to the system. This was the case in 16% of the cases. At the end of the dialogue, the vast majority of people were positive about the system ("yes, i still like you"); and some users gave constructive feedback about the voice quality ("well to be honest, I find your voice a bit forced"), dialogue flow ("yes I rather like you but you are a bit slow"), and level of comprehension ("you still need to learn to have a conversation").

#	Speaker	Turn (translated transcript)
1	Agent	Sounds good. Why do you like that?
2	User	he faltered already with the question.**
3	Agent	And if you could choose a second activity, what would that be?
4	User	uhm then I would go have a drink with my friends in the city
5		yes*
6		yes I do still have a little bit of patience**
7		because he took my text also this time**
8	Agent	What would you miss most if you wouldn't do take?
9	User	if you take, what do you mean with take?**
10	Agent	Looking back at last week, which activity did you enjoy most?
11	User	uh last week i found too active, i liked the best to uh on Sunday go to hockey with the kids
12		mmm.*
13		are we getting any other questions?**
14	Agent	Great! What kind of feeling did it give you?
15	User	uh yes*
16		that makes me happy

Table 7: Part of a problematic dialogue of a user with the BLISS agent. For example, the user is asking questions about the quality of the system (Line 2, 7 and 13).

6. Discussion

BLISS is a work in progress, and a new version is planned for the near future. While using the system for data collection, we noticed several issues with different components of the system. Some of these issues are related to technical implementation (hardware and software), while others are more related to the usage of the system. One of the important components is the NLG, which generates the questions for the user based on the initial answer. We observed that sometimes the quality of the generated follow-up motivation question (second or third question) was not good and sometimes the question did not even make sense. For example, a follow-up question of the agent would be "Why do you like going?", as it used the verb "going" from the previous user sentence: "I like going to the cinema or going out for dinner." In this case the NLU component did not extract the complete activity for the NLG component. In such cases users often responded with a meta-question about the

system (e.g. "What are you saying?"), which broke the flow of the dialogue.

Although most of the participants were native speakers of Dutch, sometimes they would use *code-switching*, which means that they would use English words while talking to the system in Dutch. Because the ASR did not contain those words in the acoustic and language model, it produced recognition errors for such words. At the conferences we also had latency issues with the ASR and TTS. Since these components are both cloud-based and dependent on a stable internet connection, the delays would produce some discomfort to the participants and they would try to repeat their answer, or were too quick in answering. If a stable connection exists, the ASR has about a second of delay and the TTS works almost immediately. Also, for the "yes" or "no" answers of the participants, recognition was poor, and manual interference (transcribing the user speech for the BLISS agent) was required to resume the dialogue.

Our prototype typically does not yet have the ability to deal appropriately with situations in which the user doesn't respond with an answer to the question. Sometimes users repeated the question before answering, hesitated or requested some elaboration. Often the agent interpreted the user's response as the answer to the question it had just asked. For example, the agent could ask: "What would you like to do this weekend?" and the user would respond "This weekend. Let me think", after which the agent could ask "What kind of feeling did thinking give you?". In future versions of the BLISS agent we want to improve the detection of actual answers.

Even though the ASR did not recognize all of the sentences by the users correctly, we found that even with errors, relevant information about users can be extracted and for each of the users we did find at least one activity. This means that even though the speech recognition is not perfect, it is very well usable for retrieving this type of user information. This is an important finding because it indicates that this procedure is robust against ASR errors and therefore realistic, as perfect ASR can never be assumed under real-life conditions.

7. Future Work

One of the goals of BLISS is to provide users with personalized dialogues. For starters, the users' speech is recorded together with the transcripts obtained through ASR. These speech recordings can be used to adapt the ASR so that it can better recognize the users' speech and to improve and personalize the dialogue. Additionally, we will use the data for creating a more personalized happiness model. This means that the agent should be able to detect full user answers more appropriately, such that it waits until the user is done answering. We will also improve the agent's activity and motivation extraction to create a correct user model. With the collected data we can create a simple method for this purpose. However, if we want to use more reliable machine learning methods, much more data is still required. This could be accomplished by extracting information from health records or by incorporating conversational dialogues between humans that may be available. At the moment we are contacting several health organizations and companies to see whether such data can be made available for specific case studies. An important aspect is of course that this is done under conditions of security and privacy, with approval from the ethics committee and in agreement with GDPR regulation. Thus far people have only interacted once with the BLISS agent. We will also deploy the BLISS agent in a long-term experiment to see if the personalized happiness model helps increase self-management on the part of users.

With regards to the TTS, we note that one of the drawbacks of the current TTS module is that the USS method is not very flexible, which limits the extent to which personalization and expressiveness can be accomplished. To tackle this issue, a neural TTS speech synthesis system is currently being developed by ReadSpeaker. Such a neural speech synthesis system can provide high-quality speech and add much more flexibility than is possible with USS systems (see (Habib et al., 2019) for an example). In particular, the aim is to personalize speech output to the users by being able to flexibly modify the input to the model (i.e. through pronunciation adaptation or changing emphasis for certain words (see e.g. (Shechtman and Sorin, 2019))). That way, new voices are likely to sound more appropriate for the conversational setting of a dialogue and will sound more pleasant for the users.

In this paper, we have shown the results of our first data collection with the first prototype of the BLISS agent, which already gives some useful insights into aspects of people's well-being, such as how people tend to describe their activities and how the agent should cope with a variety of responses given during a spoken dialogue. For the future, we intend to extend the scope of BLISS by incorporating more specific health contexts. Moreover, since self-management is a crucial element in the definition of health adopted in our research (Huber et al., 2011), our future work will investigate how the knowledge obtained through the dialogues can best be employed in the context of BLISS to realise a system that is capable of supporting users self-manage their health and well-being.

8. Resource availability

The collected dialogues, both the transcripts and the audiofiles are available for research. Access to this data set is granted after signing a Data Use Agreement for academic research purposes. We refer to the BLISS website[14] for the contact details.

Acknowledgements

This work is part of the research programme Data2Person with project number *628.011.029*, which is (partly) financed by the Dutch Research Council (NWO). We thank all the participants who volunteered to use the system and allowed us to record their data. We also thank Game Solutions Lab for their valuable contributions to this project.

[14]BLISS website: `https://bliss.ruhosting.nl/`

Bibliographical References

Beveridge, M. and Fox, J. (2006). Automatic generation of spoken dialogue from medical plans and ontologies. *Journal of Biomedical Informatics*, 39(5):482–499.

Bickmore, T. and Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5):556–571.

Bickmore, T. W. and Picard, R. W. (2005). Establishing and Maintaining Long-term Human-computer Relationships. *ACM Transactions on Computer-Human Interaction*, 12(2):293–327.

Bunt, H., Alexandersson, J., Carletta, J., Choe, J.-W., Fang, A. C., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., and Traum, D. (2010). Towards an ISO Standard for Dialogue Act Annotation. In Nicoletta (Conference Chair) Calzolari, et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 2548–2555, Valetta, Malta. European Language Resources Association (ELRA).

De Smedt, T. and Daelemans, W. (2012). Pattern for python. *Journal of Machine Learning Research*, 13:2063–2067.

Habib, R., Mariooryad, S., Shannon, M., Battenberg, E., Skerry-Ryan, R., Stanton, D., Kao, D., and Bagby, T. (2019). Semi-supervised generative modeling for controllable speech synthesis. *arXiv preprint arXiv:1910.01709*.

Huber, M., André Knottnerus, J., Green, L., Van Der Horst, H., Jadad, A. R., Kromhout, D., Leonard, B., Lorig, K., Loureiro, M. I., Van Der Meer, J. W., Schnabel, P., Smith, R., Van Weel, C., and Smid, H. (2011). How should we define health? *BMJ (Online)*, 343(7817):1–3.

Hunt, A. J. and Black, A. W. (1996). Unit selection in a concatenative speech synthesis system using a large speech database. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, volume 1, pages 373–376. IEEE.

Kardol, T. (2015). Zora, een zelflerende robot in de praktijk. *Geron*, 17(2):43–45, jun.

Martinez-Martin, E. and del Pobil, A. P. (2018). Personal Robot Assistants for Elderly Care: An Overview. In Angelo Costa, et al., editors, *Personal Assistants: Emerging Computational Technologies*, chapter 5, pages 77–91. Springer International Publishing, Cham, 1st edition.

Meschtscherjakov, A., Gärtner, M., Mirnig, A., Rödel, C., and Tscheligi, M. (2016). The Persuasive Potential Questionnaire (PPQ): Challenges, Drawbacks, and Lessons Learned. In Boris De Ruyter, et al., editors, *Persuasive Technology*, pages 162–175, Salzburg, Austria, apr. Springer International Publishing.

Ministerie van Volksgezondheid, Welzijn en Sport. (2016). Het Nederlandse zorgstelsel. Brochure, available at: https://www.rijksoverheid.nl/documenten/brochures/2016/02/ 09/het-nederlandse-zorgstelsel.

Morbini, F., Forbell, E., DeVault, D., Sagae, K., Traum, D. R., and Rizzo, A. A. (2012). A Mixed-Initiative Conversational Dialogue System for Healthcare. In Gary Beunbae Lee, et al., editors, *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 137–139, Seoul, South Korea. Association for Computational Linguistics.

op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Bedsted, B., Pease, A., Pelachaud, C., Salcedo, V., Kyriazakos, S., and Hermens, H. (2018). Council of coaches a novel holistic behavior change coaching approach. In M. Ziefle, et al., editors, *ICT4AWE 2018 - Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, volume 2018-March, pages 219–226, Portugal. SciTePress.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlíček, P., Qian, Y., Schwarz, P., Silovský, J., Stemmer, G., and Veselý, K. (2011). The Kaldi speech recognition toolkit. In *Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE.

Radlinski, F., Balog, K., Byrne, B., and Krishnamoorthi, K. (2019). Coached Conversational Preference Elicitation: A Case Study in Understanding Movie Preferences. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 353–360.

Razavi, S. Z., Schubert, L. K., Kane, B., Ali, M. R., Van Orden, K. A., and Ma, T. (2019). Dialogue Design and Management for Multi-Session Casual Conversation with Older Adults. *Joint Proceedings of the ACM IUI 2019 Workshops*, page 9, March.

Seligman, M. E. P. (2002). Positive Psychology, Positive Prevention, and Positive Therapy. In Charles R. Snyder et al., editors, *Handbook of Positive Psychology*, chapter 1, pages 3–12. Oxford University Press, 2nd edition.

Seligman, M. E. P. (2011). *Flourish: A Visionary New Understanding of Happiness and Well-being*. Free Press, New York, NY, USA, 1st edition.

Shechtman, S. and Sorin, A. (2019). Sequence to Sequence Neural Speech Synthesis with Prosody Modification Capabilities. In *Proc. 10th ISCA Speech Synthesis Workshop*, pages 275–280.

Strik, H., Russel, A., Van Den Heuvel, H., Cucchiarini, C., and Boves, L. (1997). A spoken dialog system for the Dutch public transport information service. *International Journal of Speech Technology*, 2(2):121–131.

Theune, M., Klabbers, E., Odijk, J., and De Pijper, J. (1997). Computing prosodic properties in a data-to-speech system. In *Proceedings of the ACL/EACL'97 Workshop on Concept to Speech Generation Systems*, pages 39–45.

Theune, M., Klabbers, E., de Pijper, J.-R., Krahmer, E., and Odijk, J. (2001). From data to speech: a general approach. *Natural Language Engineering*, 7(1):47–86.

van den Bosch, A. and Bouma, G. (2011). *Interactive Multi-modal Question-Answering*. Theory and Applications of Natural Language Processing. Springer, Berlin, Heidelberg.

van Waterschoot, J., Bruijnes, M., Flokstra, J., Reidsma, D., Davison, D., Theune, M., and Heylen, D. (2018). Flipper 2.0: A Pragmatic Dialogue Engine for Embodied Conversational Agents. In *Proceedings of the 18th*

International Conference on Intelligent Virtual Agents, IVA '18, pages 43–50, Sydney, NSW, Australia. ACM.

Weizenbaum, J. (1966). Eliza−−a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, January.

Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., and Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213, Melbourne, Australia, July. Association for Computational Linguistics.