This article is available at: http://link.springer.com/article/10.1007/s13218-015-0357-0#page-1
DOI 10.1007/s13218-015-0357-0

Now All Together: Overview of Virtual Health Assistants Emulating Face-to-Face Health Interview Experience

Christine Lisetti, Reza Amini, and Ugan Yasavur

Received: date / Accepted: date

Abstract We discuss a large research project aimed at building socially expressive virtual health agents or assistants (VHA) that can deliver brief motivational interventions (BMI) for behavior change, in a communication style that individuals and patients not only accept, but also find emotionally supportive and socially appropriate. Because of their well-defined sequential structure, BMIs lend themselves well to automation, and are adaptable to address a variety of target behaviors, from obesity, to alcohol and drug use, to lack treatment adherence, among others. We discuss the advantages that VHAs provide for the delivery of health interventions. We describe components of our intelligent agent architecture that enables our virtual health agents to dialogue with users in realtime while delivering the appropriate intervention based on the patient's specific needs at the time. We conclude by identifying open research challenges in developing virtual health agents.

Keywords Intelligent virtual health agents · Embodied conversational agents (ECA) · Computational models of empathy and rapport · Spoken dialog systems for Health · Markov decision process (MDP) · Hidden Markov models (HMM) · Motivational interviewing · Brief interventions · Behavior change

Part of the research described in this article was developed with funding from the National Science Foundation, grant numbers HRD-0833093, IIP-1338922, IIP- 1237818, and IIS-1423260.

C. Lisetti
11200 SW 8th Street ECS 361, Miami FL, 33199, USA
Tel.: +1-305-348-6242
Fax: +1-305-348-6242
E-mail: lisetti@cis.fiu.edu

1 Introduction

In a spirit reminiscent of various preventive traditional Eastern medicines, the interests of Western medicine healthcare have recently started to move toward finding ways of preventively promoting wellness, rather than solely treating already established illnesses.

Indeed, our contemporary times are witnessing the rise of an epidemic of behavioral issues related to lifestyles. In 2013, the Word Health Organization reported that obesity, *worldwide*, has more than doubled since 1980. It found that 1.4 billion adults were overweight, of which 500 million were obese, and 42 million children under the age of five were overweight; overall more than 10% of the worlds's adult population was obese [58]. Overweight and obesity are risk factors for a number of diseases, such as cardiovascular troubles, diabetes, musculoskeletal disorders, some cancers (e.g. colon) and can be prevented by diet and exercise. Excessive alcohol use, on the other hand, is the third leading preventable cause of death (79,000 death annually) in the United States [34], and is responsible for a wide range of health and social problems (e.g. liver cirrhosis, risky sexual behavior, domestic violence). Other risks exist with use of tobacco or narcotics.

One of the main obstacles to conquer change of these unhealthy habits, is often 'just' a matter of finding the intrinsic motivation to change lifestyle. The recent direction in healthcare and medicine toward preventive interventions, therefore has given rise to a growing effort to develop effective evidence-based health promotion interventions to help people find motivation to change[1].

[1] We use the term evidence-based as most scholars in medicine and healthcare seem to agree that the evidence-based decision-making process integrates 1) best available research evidence, 2) practitioner expertise and other available

The tenet is that unhealthy lifestyles place people at risk of serious health problems, which can be prevented if addressed on time.

Behavior change interventions that use motivational interviewing (MI) techniques (described later) have been identified as some of the most successful interventions to help individuals' find and cultivate intrinsic motivation to change [29].

Although there exist computer-based behavior change interventions that successfully apply cognitive behavior therapy to help people suffering from anxiety and depression [50], motivational interviewing intervention and its adaptations differ in that they are focussed on lifestyle changes and specific target behaviors (e.g. smoking) rather than on depression or anxiety. Many challenges are involved in delivering motivational interviewing interventions to people in need, however, such as finding the time to deliver them in busy doctors' offices, obtaining the extra training that helps staff become comfortable providing these interventions, and managing the cost of delivering the interventions [35].

As we discuss in this article, intelligent virtual health agents (VHA) shown in Figure 1, that engage with and educate people about the benefits of behavior change, offer many advantages that complement traditional interventions. Intelligent virtual agents (IVA) (in this case designed for health) are digital systems with a physical anthropomorphic representation (be it graphical or robotic), and capable of having a conversation (albeit still limited) with a human counterpart, using artificial intelligence broadly referred to as an 'agent'. With their human-like features and capabilities, IVAs provide computer users a natural interface, in that they interact with humans using humans' natural and innate communication modalities such as facial expressions, body language, speech, and natural language understanding[2]. Virtual intelligent agents designed to deliver motivational interviewing health interventions, or VHAs, are the topic of this article.

We first provide a brief background on brief motivational interviewing behavior change interventions, and discuss the advantages that virtual health agents competent in delivering these can have for diverse populations in need of help. We then explain specifics of the design, implementation, and evaluation of our functional virtual health agent system, working on a variety of computer platforms, adaptable for a variety of target

Fig. 1 Your personal virtual health agent.

behaviors. We conclude with observations on challenges that exist to bring virtual health agents to their full potential of helping individuals find health and wellbeing.

2 Motivational versus Persuasive Technologies

Motivational interviewing (MI) is aimed at helping patients, or potentially at-risk healthy individuals, to find the *intrinsic* motivation to change their lifestyle with respect to a specific unhealthy pattern of behavior. MI has been defined by Miller and Rollnick [29] as a *directive, client-centered counseling style for eliciting behavior change by helping clients to explore and resolve ambivalence.* One of MI central goals is to *magnify discrepancies that exist between someone's goals and current behavior.* MI basic tenets are that 1) if there is no discrepancy, there is no motivation; 2) one way to develop discrepancy is to become ambivalent; 3) as discrepancy increases, ambivalence first intensifies; if discrepancy continues to grow, ambivalence can be resolved toward change.

Brief motivational interviewing interventions (BMI) are adaptations of MI for primary care settings [15]: they are short, one-on-one counseling sessions, focused on specific aspects of a problematic lifestyle behavior. BMIs can be delivered in 3-5 minutes [32] and aim to moderate or eliminate an individual's problem behavior. BMIs are top ranked out of 87 treatment styles in terms of efficiency and health outcomes [28]. It is reported that even a few minutes of discussion about behavioral problems can be as effective as more extended counseling [4].

Whereas MI can be conducted without any formal assessment, BMIs are well structured and follow a determined sequential path: 1) *assess* an individual's patterns of problem behavior, 2) *tailor the feedback* on that assessment to raise individuals' awareness about their

resources, and 3) the characteristics, needs, values, and preferences of those who will be affected by the intervention [24].

[2] It is important to note that, unlike avatars that are representation of the user in a virtual environment controlled or tele-operated by the user, IVAs are autonomous entities capable of making decisions and of interacting with users.

specific problematic behavior, and 3) *personalize menus of option plans* to change behavior, based on an individual's readiness to change. If the individual is not ready to change, there is no point making plans to change, and BMI helps individuals get closer to being ready by magnifying internal discrepancies.

It is important to note that in MI and BMIs, the motivation to change is elicited *from* the client, and not imposed from without: other motivational approaches that use a paternalistic expert role to coerce or persuade are not as conducive for identifying and mobilizing client's intrinsic values and goals: confrontation usually reinforces the targeted behavior, and *direct persuasion generally increases client resistance and diminish the probability of change* [46,41].

Interestingly, although originally conceived as a preparation for further treatment, enhancing motivation and adherence, research has revealed that change often occurred soon after one or two sessions of MI, without further treatment, relative to control groups receiving no counseling at all [29]. As a result, motivational interviewing and other brief motivational interventions are quickly gaining popularity as alternative or adjunctive approaches to more traditional, more cost-intensive, less effective, and less accessible approaches to behavior change [57].

Furthermore, whereas initially developed to address addictive behaviors, such interventions have been adapted and implemented with success for a variety of behaviors ranging from diabetes self-management [14], to treatment adherence among psychiatric patients [53], to fruit and vegetable intake among African Americans [44], among other target behaviors.

3 Advantages of Virtual Health Agents and Assistants

Before we discuss the advantages of intelligent virtual agents that can deliver health interventions, we point out to advantages of computer-based health interventions in general.

3.1 Computer-based health interventions

Computer-based interventions (CBIs) that perform assessment and feedback have been shown to be well accepted by their users [49,13], and can be as effective as when the intervention is delivered by a person [20]. Next we discuss what CBIs provide, followed by our resulting choices for our virtual health agents functionalities:

1. *increase accessibility:* because CBIs can be developed to run on personal computers (PC), on servers accessed via Internet browsers, or on smart phones, they increase access in rural areas, and provide 24 hours/7 days a week support anywhere, without the need to schedule appointments with a physician:
   our VHAs work on all three platforms: Personal Computers [61], Internet browsers [27], smart phones, and humanoid robots [1];
2. *increase cost-effectiveness:* one or two BMI sessions often yield greater change than no counseling at all, and even though follow-up BMI sessions can increase positive outcome, they are not always offered in medical and public health settings for lack of trained human personnel:
   we work on VHAs that can be massively reproduced or that are available via the Internet, thereby reducing costs of delivering BMIs;
3. *increase self-disclosure:* patients that engage in behavior that put them at risk (e.g. excessive drinking, unsafe sex, over-eating) tend to report more information to a computer interviewer than to its human counterpart [47]. The knowledge that a computer system does not have an intrinsic moral value system to judge the patient favors the self-disclosure of sensitive information; since the more is known about a behavioral problem, the easier it can be solved, CBIs can be at an advantage and know more of the patient's problem behavior than a human interviewer:
   we study how different delivery modes of BMIs affect an individual's self-disclosure about at-risk behavior;
4. *tailor information:* communication strategies intended to reach one specific person about characteristics unique to that person can be derived from assessment, and are superior to commonly used generic communication such as brochures:
   we build VHAs that create and remember user profiles for the delivery of (repeatable and adaptive) tailored interventions over the long term.

Whereas current BMIs delivered by computers have been found effective for individuals who complete them, high drop-out rates due to their users' low level of engagement during the text-based interaction they provide, limit their long-term adoption and impact [40, 56,16]. The full success of CBIs is therefore dependent upon their ability to retain their users.

3.2 Virtual Health Agents

With their ability to converse in natural languages while displaying socially appropriate non-verbal messages (e.g. gestures, facial expressions) that are congruent with verbal utterances, intelligent virtual agents (also referred to as embodied conversational agents [9]) are known to increase users' engagement, and are predicted to revolutionize human-computer interaction as we know it. Predictably, intelligent virtual agents specialized on health intervention delivery are expected to revolutionize healthcare delivery [31].

We asked ourselves the questions: *Could virtual agents capable of delivering health interventions - referred to as virtual health assistants (VHA) - help individuals whose lifestyle places them at-risk of serious or chronic diseases? Could VHAs be more effective than existing computer text-based CBIs? What should VHAs' abilities be, and how can we create them?*

We have implemented VHAs with the multimodal architecture shown in Figure 4 (discussed later), and found that VHAs offer additional advantages compared with text-based CBIs, because virtual health agents can:

1. *act as social actors:* research has shown that people (unconsciously) respond in social ways to computers that provide them with appropriate social cues (e.g. politeness, humor) [43], and that people can develop personal relationship with artificial agents over the long term [6,9] (e.g. coach, social companion):
   we work on building VHAs that can build rapport and emulate a patient-physician long-term relationship;
2. *conduct spoken dialogs:* large portions of BMIs are based on assessment questionnaires and these can easily be presented in menu forms (as most CBIs are). Yet, given that they are always targeted to a specific target behavior (e.g. over-eating, excessive alcohol consumption), BMI linguistic domain is semi-restricted, rendering spoken dialog feasible; recent progress on synthetic voices and text-to-speech engines (TTS) enables VHAs to speak in natural languages, thereby shrinking the literacy and techno-phobia divide for underserved population:
   we develop VHAs that can not only deliver BMIs using menus of multiple choice questions selected with a mouse [27], but also speak and understand the user's spoken answers in two modes: free spoken utterances, or spoken utterances read from menu of possible answers [61]; users can choose which mode of delivery they prefer;
3. *express empathy:* one of the main factors for success of a BMI is the ability for an clinician to exhibit empathy (e.g. tone of voice, reflective listening, non-verbal messages) [29,36]; latest progress on intelligent virtual agents research has led to the development of computational models of empathy [23, 39,37,27,8]:
   we study how a VHA's non-verbal communicative abilities can be fine-tuned to build rapport and communicate empathy at appropriate times during the dialog, and we synchronize the VHAs' utterances, voice, gestures, facial expressions and head movements (cf. Fig. 2) for natural communication during BMI [27,3];

Fig. 2 Virtual health agent empathic facial expressions.

4. *emulate patient-physician concordance:* research revealed that patient-physician race and ethnic concordance is linked with higher patient satisfaction and better health care processes. Yet, ethnic minorities are poorly represented among physicians and other health professionals [11,12]. Since research also revealed that people respond to the ethnicity of intelligent agents similarly to how they respond to humans' ethnicity [33] (e.g. the same words mean different things when coming from a human or virtual agent that has similar ethnicity to the human interlocutor's), VHAs might be able to emulate some level of cultural competence and concordance:
   we create VHAs (shown in Figure 3) that can match the user's ethnicity physical features, and study ways for VHAs to be culturally competent and able to establish rapport with patients of different ethnicities;
5. *diminish variability:* there exists wide variability (from 25% to 100%) in different counselor's rates of improvement among their patients [29], and the counselor's empathic abilities play a large role in this variability:
   our VHAs diminish variability by learning from the best human counselors: we record and annotate videos of effective human coun-

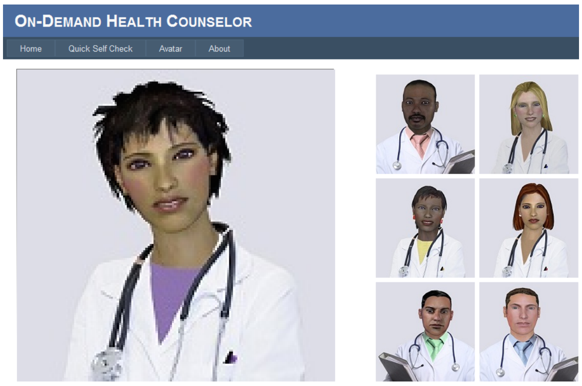

Fig. 3 Virtual health agent of various ethnicity.

selors' verbal and non-verbal behaviors during an intervention session, and use data-driven machine learning to animate in real-time the VHA's learned behaviors;

6. *demonstrate infinite patience:* one trap for counselors is to try to move the patient toward change more quickly than [s/]he is ready for. Respecting the patient's six stages of change – pre-contemplation, contemplation, determination, action, maintenance, and either relapse or exit [41] – can be challenging for therapists who work within the traditional biomedical model of counseling and have the *righting reflex*, or the tendency to 'set things right' using direct advocacy – thereby acting out patients' ambivalence toward change and increasing resistance: **we design VHAs with infinite patience, able to present and repeat relevant BMI material at the pace the patient is able to receive it, thereby at an advantage with respect to human helpers' righting reflex**.

4 Virtual Health Assistant Implementation

In an effort to address the limitations of current computer-based interventions – namely users' loss of interest over the long term and dropouts [62], which are also problematic in traditional face-to-face interventions – our approach is to leverage users' acceptance of IVAs which are documented as increasing enjoyment and engagement with computer systems [9]: by developing expressive 3D animated characters with speech understanding and empathic abilities, which are knowledgeable about motivational interviewing interventions.

As we describe next, our virtual health agents (shown in Figures 1-4) are able to deliver the content of an existing[3] evidence-based BMI [51, 19], in an empathic communication style.

[3] It is important to note that our system can be adapted to any type of brief motivational intervention, e.g. against obesity.

4.1 Virtual health assistant architecture overview

Our intelligent agent architecture is composed of the main modules shown in Figure 4 (described in details in [27]).

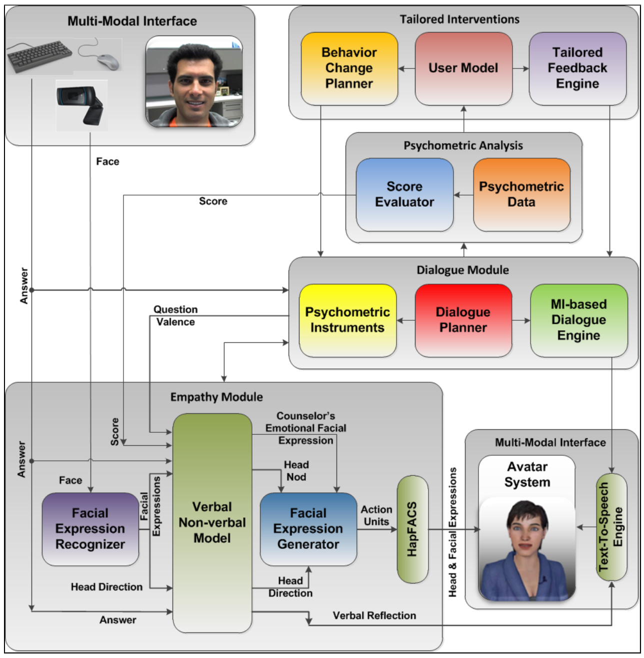

Fig. 4 Virtual Health Agent for Brief Motivational Interventions (VHA-4-BMI).

In short (technical details follow later), during each interaction:

– the sensing *Multi-Modal Interface*, (vs. the expressive *Multimodal Interface 3D character system* includes (1) a microphone that captures the user's spoken answers, (2) a mouse that captures the answers that the user clicks on if speech is not recognized, and (3) a webcam that capture facial expressions and head movements (which are in turn interpreted by the facial expression recognizer in the *Empathy Module*);
– the *Dialogue Module* directs the sessions, elicits information from the user and record the answers; three versions of the Dialogue Module exist: one implements a text-based dialog offering multiple choices to click on [27], one version implements a speech-enabled dialog offering multiple choices to be read out loud [2], and one version implements a spoken dialog manager enabling users to speak freely with their VHA [61] (discussed later);
– *Psychometric Analysis* is performed based on the information collected by the Dialogue Module to calculate normative data that will be provided to the

user during the Tailored Feedback portion of the intervention (more below), data that can challenge his or her beliefs about lifestyle and magnify his or her internal discrepancies about the problem behavior;
- the *User model* is updated with user's data; past answers are available for gauging progress during follow-up sessions over the long-term; users of the web-based version remain anonymous by providing a self-made ID and password;
- a *Tailored Intervention* is offered by the VHA as: (1) sensitive *Tailored Feedback* delivered with empathic personalized message scripts (e.g. "It can be surprising and (for some) discouraging to see they fall higher on the alcohol use scale than they expected. Some people think there might be a mistake in the way results are calculated. Feel free to check your answers again on the feedback report that I just printed for you..."; and (2) *Behavior Change Plans* plans are discussed, summarized by the VHA, and printed out as a friendly personalized informative and supportive feedback; the plans respect the user's readiness for change given by Prochaska's six stages of change (1997), and are also based on the user's preferences for replacement behaviors – which were identified during the earlier dialogue stage (e.g. if the user identified that the main reason to drink is to relax, and that walking in a park is relaxing, plans are made around scheduling walking activities).
- the *Empathy Module* can, in real time: (1) interpret the user's facial expressions in terms of the user's most probable emotion, and (2) convey an ongoing sense of empathy by controlling the 3-dimensional semi-realistic VHA character's verbal and nonverbal social communication messages.

4.2 Modeling VHA's empathic communication

Discussing at-risk behaviors such as heavy drinking, binge eating, or drug use, can be emotional for people to talk about (e.g. shame, discouragement, fear, anger, as well as hopefulness, satisfaction, pride). In MI or BMI sessions, what is crucial is the ability of the therapist to *establish rapport and to express accurate empathy* by applying "a skillful reflective listening to clarify and amplify the [user's] own experiencing and meaning" [29].

So, *what* are *rapport and empathy? Can they be modeled computationally, and to what extent?*

Empathy can be of two (non-exclusive) kinds: (1) *affective (or parallel) empathy* [21, 5] is a visceral emotional reaction when one perceives that another is experiencing, or about to experience an emotion [60]. It involves raw (vs. deliberate) emotional contagion or resonation. People with high capacity to empathize with others exhibit unconscious mimicry (or mirroring) of people's facial expressions and bodily postures [10, 25], which also improves *rapport* or the feeling of flow and connection during conversations [54]; and

(2) *Cognitive empathy* is an intellectual reaction involving an understanding (rather than a feeling) of another's experiences and concerns, combined with the capacity to communicate that understanding [22]. Reflective listening (RL) is one of the main communication strategies used by counselors to express cognitive empathy, which involves two key steps: seeking to understand a speaker's thoughts or feelings, then conveying the idea back to the speaker to confirm that the idea has been understood correctly [45][4].

How to model empathy computationally?

Whereas our virtual agent does *not* experience nor understand the subjective experience associated with the user's emotions, it *does* perceive some of the user's emotions with computer vision[5], and it *does* react to them with 3D realtime animations and RL, conveying a welcomed sense of empathy and rapport, confirmed by users' evaluation of our agent (discussed later).

Our *Empathy Module* (cf. Figure 4) partially emulates both kinds of empathy: (1) *affective empathy* is modeled by partial mirroring the user's non-verbal behaviors, including user's head nods and facial expressions perceived by the sensing *Multimodal Interface* shown in Fig. 4, and (2) *cognitive empathy* is modeled by implementing two simple RL techniques to help the user become aware of his or her behavior: *simple reflections* paraphrase the user's answers (e.g. Client: "I only have a couple of drinks a day." VHA: "Ok, so you can have as many as 14 drinks per week."), whereas *double-sided reflections* magnify discrepancies found in the user's answers (e.g. VHA: "So you told me earlier that your drinking doesn't affect your performance at work; on the other hand you're telling me that you miss work at least twice a week because you're hangover, is that right?").

We explored two main approaches to model empathy: *rule-based* and *data-driven*.

Our first approach was **rule-based** and is fully described in [27]: based on pre-defined rules that we generated from discussions with an MI expert practitioner, the decision tree shown in Fig. 5 decides, in real time, when and what expressions the agent should ex-

[4] RL raises an individuals' awareness about their at-risk behavior(s) – an initial important step toward behavior change.

[5] We use the SightCorp off-the-shelf face reader for detecting emotional facial expressions.

press (e.g. eyebrow expressions, smiles), when the agent should nod, and what verbal reflections to utter.

When the counselors question valence is positive or the user's emotional facial expression is positive (e.g. we observed that users often express the happy or surprise emotions), the decision tree instructs the VHA to mirror the user's expression with a positive facial expression (and vice versa). Furthermore, if psychometrics analysis (described earlier) reveals that the user's answers places him or her at a low risk level, or if the overall score (history of answers) is low, the decision tree generates more positive expressions (and vice versa). For example, as shown in Figure 5 traversing the left most branch of the tree: if (1) the counselor asks "Does drinking help you to relax?", which is a positive valence question, and (2) the user expresses a happy facial expression, and (3) the user answers "No" (i.e., low risk answer), and (4) the user has a low overall score calculated from normative psychometrics, then the decision tree returns a happy face with a large smile and raised eyebrows. On the other hand, if, when traversing the leftmost branch of the tree, (1) the VHA asks a question with a negative valence"How often have you had a feeling of remorse?", and (2) the user shows a sad expression, and (3) the user answers yes, and (4) the overall score places the user at high-risk of alcohol dependence and abuse, then the VHA displays a concerned expression and nods her head.

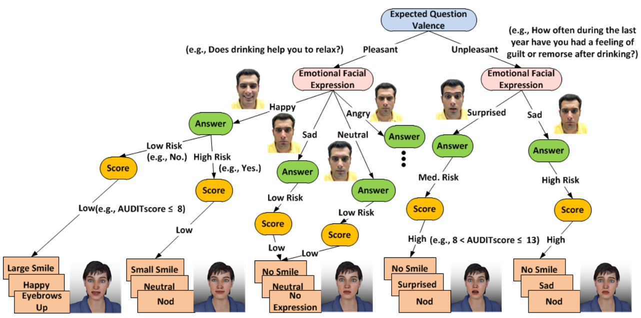

Fig. 5 A sample portion of the empathy decision tree.

Our evaluation results (described in [27]) indicate that VHA's rapport-enabled (vs. textual only) delivery improves users' acceptance of CBIs in terms of users' attitude, perceived enjoyment, perceived usefulness, and trust. *Most importantly for medical and healthcare experts, users' intention to reuse our VHA-delivered intervention increased by 31% compared to the same computer-based intervention delivered with text-only* [27]. This is significant because research showed that computer-based interventions are effective for users who complete them (including follow-up sessions over the long-term). Whereas text-only interventions tend to loose users' interest, our VHA appears to increase users' interest in repeating the interaction.

The major limitation of rule-based models is that rules must be constructed from clinical psychology literature or consultants, neither of which providing formal specifications needed for computer implementations; the quality of the derived rules therefore solely depends on the designer's intuition. Even more problematic is that, as the number of non-verbal input features to be modeled increases (e.g. adding hand gestures, leaning forward, as we did in our data-driven approach described next), the number of rules increases exponentially: e.g. for 10 input features which can take 2 values each, up to 2^{10} combinations must be considered, making scalability of rule derivation difficult, at best.

Our second approach to modeling VHAs' nonverbal empathic communication (fully detailed in [2]) is **data-driven**, based on annotated videos of human clinician-patient intervention sessions which we recorded. We use Hidden Markov Model (HMM) machine learning techniques to learn the empathy model [42]. Our system input is a *sequence of feature vectors* representing consecutive words. Since HMM is a statistical model that is used for learning patterns where a *sequence* of observations is given, the sequential property of our problem led us to choose HMMs as in [26] to predict non-verbal signals.

Our data-driven model is the first to (1) use a large number of real-time interactive features (see Fig. 6), (2) model agent's behavior in *two* roles: when it speaks, *and* when it listens, as non-verbal communication patterns differ in these two modes, and (3) evaluate the agent's non-verbal model both *objectively* in terms of accuracy, precision, recall, and F-measure, and *subjectively* with users of our behavior change interventions. Results are fully provided in [2], and show that the data-driven approach is superior to the rule-based approach in terms of users' experience of rapport and enjoyment interacting with the virtual health agent.

5 Enabling VHAs to conduct spoken sessions

As mentioned earlier, we implemented three versions of the Dialogue Module shown in Figure 4: (1) one version is based on multiple choice entries to be clicked on with a mouse [27]; (2) one version asks the user to read out loud multiple choices answers displayed on the screen next to the VHA; and (3) one version enables the agent to speak and understand natural speech using a **data-driven approach based on Reinforcement Learning (RL) and Markov Decision Processes (MDP)**, summarized below and described fully in [61];

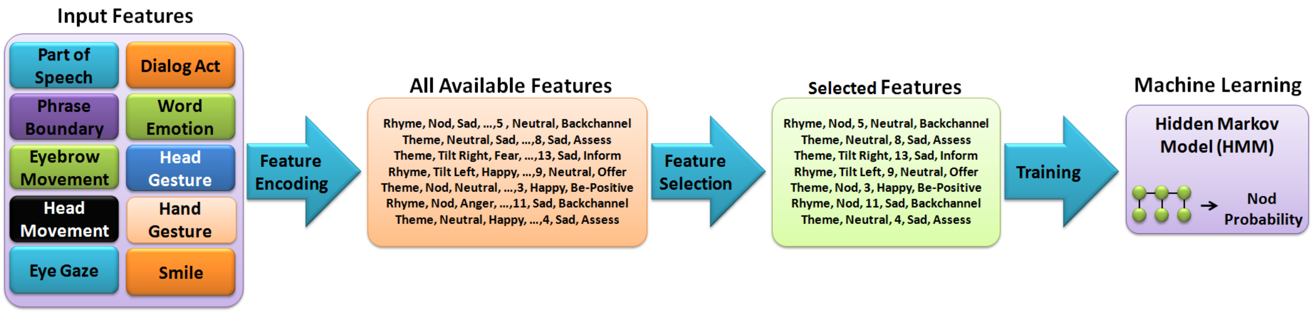

Fig. 6 Input features selected to model non-verbal behaviors

According to the clinician's guide for conducting brief interventions from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) [35], a brief intervention for alcohol-related health problems can be delivered in three sequential steps: Step 1: *Screening* about alcohol *use*; Step 2: *Assessing* for alcohol use *disorders* with 2 sub-steps assessing *abuse* and *dependence*; Step 3: *Advising* and assisting according to degree of alcohol problem with 2 sub-steps for advising drinkers *at-risk* or with *alcohol use disorder*. A sample dialogue of the interaction during Step 1 is provided in Table 1.

Dialog systems can be classified into two main categories based on their dialog management technique, which can be either based on machine learning (e.g. based on reinforcement learning), or hand-crafted. Systems based on RL are popular in the SDS community and are reported to work better than hand-crafted ones for *speech*-enabled systems [64,18] against noisy speech recognition. *Hand-crafted systems*, on the other hand, can be divided into three subcategories, with dialog management approaches using finite states [52], plans and inference rules [17,7] or information states [55].

RL-based dialog systems can learn dialog strategies in a given dialog state from their prior experiences. The idea of having a dialog manager that can learn interactively from its experience is a cost effective methodology given the alternative approaches: crafting system responses to all possible user's input using rules and heuristics [38]. The RL-based approach provides the opportunity to automate the design of dialog management strategies by having the system learn these strategies from received reward signals.

Approaches for dialog systems based on reinforcement learning (RL) use Markov decision processes (MDP) [48] or partially observable Markov decision processes (POMDP) frameworks [63,59] to develop robust dialog managers [18,64]. We used the MDP approach to avoid the scalability problems associated with POMDPs, and our system design was inspired by the NjFun system [48].

Within the RL framework, dialog strategies are represented as mappings from states to actions within Markov decision processes (MDP), i.e. a dialog strategy specifies, for each system state, what is the next action to be taken by the system. To avoid the data sparsity problem during training, we divided the whole system into 5 sections according to the NIAAA BI guide steps, resulting in 5 corresponding MDPs with local goals and reward functions (see Fig. 7). Unlike earlier dialog systems, the flow of our dialog does not depend on unfilled slots, but is rather decided dynamically during the interaction.

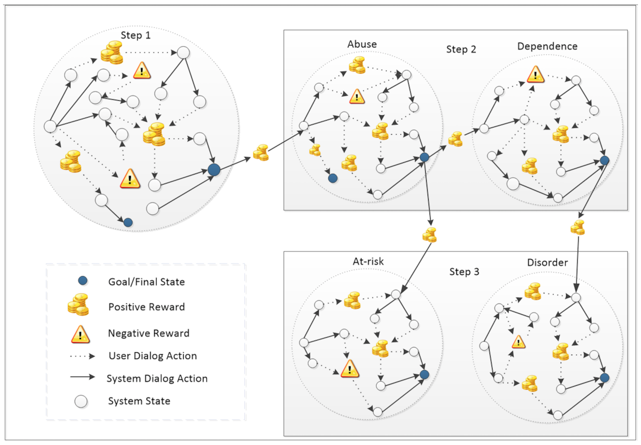

Fig. 7 Markov Decision Processes for BMI Spoken Dialogs

We evaluated the system *objectively* and found that its task completion rate is high under noisy conditions and against diverse client population who has no prior experience using the system. Our *subjective* evaluation of the system demonstrated that users' acceptability of the system is very high in terms of *Ease of Use* (81 Yes versus 8 No) and *Intention to Reuse* (63 Yes versus 26 No) the system. Further *subjective* evaluation is described in [61].

Table 1 Sample Dialog

S1:	Hi, My name is Lola, I am a virtual character created to have conversations with you and to teach you about healthy lifestyles. Today, I'm here to talk about drinking to figure out if you have any issues. May I ask you few questions about your drinking?	
C1:	Yes, sure you can.	
S2:	Do you sometimes drink beer, wine, or other alcoholic beverages?	
C2:	Well... yes, I drink wine or beer sometimes!	
S3:	How many times in the past year have you had 5 or more drinks in a day?	
C3:	I think around once a week.	
S4:	Actually, consuming more than 4 drinks on any single day is classified as heavy drinking. So it seems that you are sometimes going above heavy drinking limits? Is that right?	
C4:	Yes, that's correct.	
S5:	How frequently do you have an alcoholic beverage each week?	*Step 1*: *Screening* for *Use*
C5:	I think at least 3 days a week.	
S6:	On a typical drinking day, how many drinks you have?	
C6:	I just drink... maybe half a bottle of whisky.	
S7:	Sorry, I could not understand your answer, can you please repeat?	
C7:	I said half a bottle of scotch.	
S8:	Ok, thanks for the information you have provided to me about your drinking. Next, I will try to determine some important effects of your drinking.	

6 Challenges

Although our first results are promising, intelligent virtual health assistants' social realism and speech recognition still need to be improved. Further research is needed to bring VHAs to their full potential. Challenges to address include:

- studying social communication and clinical psychology literature to understand best clinician practices in terms of psycho-linguistic communication; this can not only help identify potential rules for rule-based or hybrid approaches to modeling non-verbal/verbal communication, but also can help designing well-crafted evaluation studies for VHAs;
- perfecting machine learning data-driven approaches to model both non-verbal and spoken communication: whereas rule-based approaches to modeling rely on the developer's intuition to derive rules from social communication literature, data driven approaches enable a developer to extract features based on what is actually happening in real counseling sessions between the counselor and the client; our experience revealed data-driven approaches to be more promising, and therefore generating data corpora is essential;
- collecting and annotating large corpora of physician-patient interactions;
- generating (and sharing) tools for automatic annotations of these corpora to speed up data preprocessing steps of data-driven approaches;
- studying how to increase robustness of automatic speech recognizers, and providing users multiple modalities to communicate their answers (e.g. mouse, microphone);
- avoiding the uncanny valley with respect to VHA's physical/graphical realism [30], and establishing what kind of embodiment appeals to users via user studies (e.g. 2D cartoon characters might be more engaging for children than photorealistic 3D characters; photorealistic 3D characters might confuse elderly);
- ensuring that the level of social realism achieved from modeling empathy and rapport is appropriate with the physical/graphical realism of the virtual agent, and well accepted by users.

7 Conclusion

In this article, we described our research project on Virtual Health Agents for Brief Motivational Interventions (VHA-4-BMI), and discussed their advantages. The increasing demand for health promotion interventions that are accessible for population at large, and tailored to an individual's specific needs is real. The technological advances that we describe are an attempt to address that urgent need. We hope that our work and its positive results will encourage researchers to continue our efforts toward providing access to quality and effective health interventions via engaging and supportive virtual health agents.

References

1. Abeyruwan, S., Baral, R., Yasavur, U., Lisetti, C., Visser, U.: Humanoid robots and spoken dialog systems for brief health interventions. In: 2014 AAAI Fall Symposium Series (2014)
2. Amini, R.: Learning data-driven models of non-verbal behaviors for building rapport using an intelligent virtual agent. Ph.D. thesis, Florida International University (2015)

3. Amini, R., Lisetti, C.: HapFACS : an Open Source API / Software to Generate FACS-Based Expressions for ECAs Animation and for Corpus Generation. In: The fifth biannual Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII 2013). IEEE Computer Society, Geneva, Switzerland (2013)
4. Babor, T.F., Grant, M.: Programme on substance abuse : project on identification and management of alcohol-related problems. report on phase ii, an randomized clinical trial of brief interventions in primary health care (1992)
5. Bavelas, J.B., Black, A., Lemery, C.R., Mullett, J.: Empathy and its developement. In: N. Eisenberg, J. Strayer (eds.) Motor mimicry as primitive empathy, pp. 317–338. Cambridge University Press, Cambridge, UK (1987)
6. Bickmore, T., Gruber, A., Picard, R.: Establishing the computer – patient working alliance in automated health behavior change interventions. Patient Education and Counseling **59**, 21–30 (2005)
7. Bohus, D., Rudnicky, A.I.: The ravenclaw dialog management framework: Architecture and systems. Computer Speech & Language **23**(3), 332 – 361 (2009)
8. Boukricha, H., Wachsmuth, I.: Empathy-Based Emotional Alignment for a Virtual Human: A Three-Step Approach. KI - Künstliche Intelligenz **25**(3), 195–204 (2011)
9. Cassell, J., Sullivan, J., Prevost, S., Churchill, E.F.: Embodied Conversational Agents. Social Psychology **40**(1), 26–36 (2000)
10. Chartrand, T.L., Bargh, J.a.: The chameleon effect: the perception-behavior link and social interaction. Journal of personality and social psychology **76**(6), 893–910 (1999)
11. Cooper, L., Powe, N.: Disparities in patient experiences, health care processes, and outcomes: the role of patient-provider racial, ethnic, and language concordance. Health Care **753** (2004)
12. Cooper, L.A., Roter, D.L., Johnson, R.L., Ford, D.E., Steinwachs, D.M., Powe, N.R.: Patient-centered communication, ratings of care, and concordance of patient and physician race. Annals of Internal Medicine **139**(11), 907–915 (2003)
13. Cunningham, J.A., Humphreys, K., & Koski-Jannes, A.: Providing personalized assessment feedback for problem drinking on the Internet. In: Paper presented at the 33rd annual convention of the Association for the Advancement of Behavior Therapy, Toronto. (1999)
14. Doherty, Y., Hall, D., James, P.T., Roberts, S.H., & Simpson, J.: Change counselling in diabetes: The development of a training programme for the diabetes team. Patien. Education & Counseling **40**, 263–278 (2000)
15. Emmons, K.M., Rollnick, S.: Motivational interviewing in health care settings. Opportunities and limitations. American journal of preventive medicine **20**(1), 68–74 (2001)
16. Evers, K.E., Cummins, C.O., Prochaska, J.O., Prochaska, J.M.: Online health behavior and disease management programs: are we ready for them? Are they ready for us? Journal of medical Internet research **7**(3), e27 (2005)
17. Ferguson, G., Allen, J., et al.: Trips: An integrated intelligent problem-solving assistant. In: Proceedings of the National Conference on Artificial Intelligence, pp. 567–573. JOHN WILEY & SONS LTD (1998)
18. Frampton, M., Lemon, O.: Recent research advances in reinforcement learning in spoken dialogue systems. Knowledge Engineering Review **24**(4), 375–408 (2009)
19. Hester, R.K., Squires, D.D., Delaney, H.D.: The Drinker's Check-up: 12-month outcomes of a controlled clinical trial of a stand-alone software program for problem drinkers. Journal of Substance Abuse Treatment **28**(2), 159–169 (2005)
20. Hester, R.K., & Delaney, H.: Behavioral Self-Control Program for Windows: Results of a controlled clinical trial. Journal of Consulting and Clinical Psychology **65**, 685–693 (1997)
21. Hoffman, M.L.: Empathy and Moral Development: Implications for Caring and Justice. Cambridge University Press (2000)
22. Hojat, M.: Empathy in patient care: antecedents, development, measurement, and outcomes. Springer Verlag (2007)
23. Huang, L., Morency, L.p., Gratch, J.: Virtual Rapport 2.0. In: International Conference on Intelligent Virtual Agents, Intelligence, Lecture Notes in Artificial Intelligence., pp. 68–79. Springer-Verlag Berlin Heidelberg (2011)
24. Jacobs, J.A., Jones, E., Gabella, B.A., Spring, B., Brownson, C.: Tools for Implementing an Evidence-Based Approach in Public Health Practice The Need for Evidence-Based Public Health. Preventing Chronic Disease **9**(1), 1–9 (2012)
25. Lakin, J.L., Chartrand, T.L.: Using Nonconscious Behavioral Mimicry to Create Affiliation and Rapport. Psychological Science **14**(4), 334–339 (2003)
26. Lee, J., Marsella, S.C.: Learning a Model of Speaker Head Nods using Gesture Corpora. In: Decker, Sichman, Sierra, Castelfranchi (eds.) 8th Int'l Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009) (2009)
27. Lisetti, Christine L., Amini, R., Yasavur, U., Rishe, N.: I can help you change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. ACM Transactions on Management Information Systems **4**(4) (2013)
28. Miller, W.R., Wilbourne, P.L.: Mesa grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. Addiction **97**(3), 265–277 (2002)
29. Miller, W. R. & Rollnick, S.E.: Motivational interviewing: 2nd Ed. Preparing people to change addictive behavior. New York: Guilford Press. (2002)
30. Mori, M.: The Uncanny Valley. Energy **7**(4), 33–35 (1970)
31. Morrow, T.: Virtual Health Assistants Poised to Revolutionize Healthcare Delivery. American Journal of Managed Care pp. 13–14 (2012)
32. Moyer, A., Finney, J.W., Swearingen, C.E., Vergun, P.: Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction **97**(3), 279–292 (2002)
33. Nass, C., Isbister, K., and Lee, E.: Truth is beauty: Researching embodied conversational agents. In J. Cassells (Ed.),. In: J. Cassell (ed.) Embodied conversational agents. Cambridge, MA: MIT Press. (2000)
34. National Center for Chronic Prevention and Health Promotion: Excessive Alcohol Use At a Glance: Addressing a Leading Risk for Death, Chronic Disease, and Injury. Tech. rep., National Center for Chronic Prevention and Health Promotion (2011)
35. NIAAA: Niaaa alcohol alert no. 66: Brief interventions (2006). URL http://pubs.niaaa.nih.gov/publications/AA66/AA66.pdf
36. Norfolk, T., Birdi, K., Walsh, D.: The role of empathy in establishing rapport in the consultation: a new model. Medical education **41**(7), 690–7 (2007)

37. Ochs, M., and Sadek, D., and Pelachaud, C.: A formal model of emotions for an empathic rational dialog agent. Autonomous Agents and Multi-Agent Systems **24**(3), 410–440 (2012)
38. Paek, T., Pieraccini, R.: Automating spoken dialogue management design using machine learning: An industry perspective. Speech communication **50**(8), 716–729 (2008)
39. Pelachaud, C.: Modelling multimodal expression of emotion in a virtual agent. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **364**(1535), 3539–48 (2009)
40. Portnoy, D.B., Scott-Sheldon, L.a.J., Johnson, B.T., Carey, M.P.: Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Preventive medicine **47**(1), 3–16 (2008)
41. Prochaska, J.O., Velicer, W.F.: The transtheoretical model of health behavior change. American Journal Of Health Promotion **12**(1), 38–48 (1997)
42. Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE, vol. 77(2), pp. 257–286 (1989)
43. Reeves, B. and Nass, C., Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. University of Chicago Press, New York, NY (1996)
44. Resnicow, K., Soler, R., Braithwaite, R. L., Ahluwalia, J. S., & Butler, J.: Cultural sensitivity in substance use prevention. Journal of Community Psychology **28**, 271–290 (2000)
45. Rogers, C.R.: A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework. In: S. Koch (ed.) Psychology: the Study of a Science, vol. 3, chap. 3, pp. 184–256. McGraw-Hill, New York (1959)
46. Rollnick S, & Miller, W.: What is motivational interviewing? Behavioural and Cognitive Psychotherapy **23**, 325–334. (1995)
47. Servan-Schreiber, D.: Artificial intelligence and psychiatry. Journal of Nervous and Mental Disease **174**, 191–202 (1986)
48. Singh, S., Litman, D., Kearns, M., Walker, M.: Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. Journal of Artificial Intelligence Research **16**, 105–133 (2002)
49. Skinner, H.A.: Computerized Lifestyle Assessment Manual. Toronto: Multi-Health Systems (1994)
50. Spek, V., Cuijpers, P., Nyklícek, I., Riper, H., Keyzer, J., Pop, V.: Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychological medicine **37**(November 2006), 319–328 (2007)
51. Squires, D.D., Hester, R.K.: Using technical innovations in clinical practice: the Drinker's Check-Up software program. Journal of clinical psychology **60**(2), 159–69 (2004)
52. Sutton, S., Cole, R., De Villiers, J., Schalkwyk, J., Vermeulen, P., Macon, M., Yan, Y., Kaiser, E., Rundle, B., Shobaki, K., et al.: Universal speech tools: The cslu toolkit. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP), pp. 3221–3224. Sydney, Australia. (1998)
53. Swanson, A. J., Pantalon, M. V., & Cohen, K.R.: Motivational interviewing and treatment adherence among psychiatric and dually diagnosed patients. Journal of Nervous & Mental Disease, **187**, 630–635 (1999)
54. Tickle-Degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. Psychological Inquiry **1**(4), 285–293 (1990)
55. Traum, D., Larsson, S.: Current and New Directions in Discourse and Dialogue Text, Speech and Language Technology, chap. The Information State Approach to Dialogue Management, pp. 325–353. Kluwer Academic Publishers (2003)
56. Vernon, M.: A review of computer-based alcohol problem services designed for the general public. Journal of substance abuse treatment **38**(3), 203–211 (2010)
57. Walitzer, K.S., Dermen, K.H., Connors, G.J.: Strategies for preparing clients for treatment. A review. Behavior modification **23**(1), 129–151 (1999)
58. WHO World Health Organization: Obesity and Overweight Fact sheet 311 (2014)
59. Williams, J.D.: The best of both worlds: unifying conventional dialog systems and pomdps. In: INTERSPEECH, pp. 1173–1176 (2008)
60. Wispé, L.: History of the concept of empathy. In: N. Einsenberg, J. Strayer (eds.) Empathy and its development, chap. 2, pp. 17–37. Cambridge University Press (1987)
61. Yasavur, U., Lisetti, C., Rishe, N.: Let's talk! speaking virtual counselor offers you a brief intervention. Journal of Multimodal User Interfaces **8**, 381–398 (2014)
62. Yevlahova, D., Satur, J.: Models for individual oral health promotion and their effectiveness: A systematic review. Australian Dental Journal **54**(3), 190–197 (2009)
63. Young, S., Gašić, M., Keizer, S., Mairesse, F., Schatzmann, J., Thomson, B., Yu, K.: The hidden information state model: A practical framework for pomdp-based spoken dialogue management. Computer Speech & Language **24**(2), 150–174 (2010)
64. Young, S., Gašić, M., Thomson, B., Williams, J.: Pomdp-based statistical spoken dialog systems: A review (2013)

Authors' Bios

Dr. Christine Lisetti is an Associate Professor in the School of Computing and Information Sciences at Florida International University, USA, where she directs the Affective Social Computing Laboratory (http://ascl.cis.fiu.edu). She is one of the founders of the field of Affective Computing, and a member of the Editorial Board of the *IEEE Transactions on Affective Computing*. Dr. Lisetti's long-term research goal is to create digital engaging socially intelligent agents that can interact naturally with humans via expressive multimodal communication, in a variety of contexts involving socio-emotional content (e.g. health coach, social companion, educational games).

Reza Amini is a PhD candidate in computer science at Florida International University. His research interests are intelligent virtual agents, data driven approaches to modeling non-verbal behaviors.

Dr. Ugan Yasavur is a Senior Researcher at iPSoft. His research interests are spoken dialog systems, dialog management, and intelligent virtual agents.