Hints of Uncanny Utterances in a Disrupted Interaction Continuum

Daniele Occhiuto[a] and Franca Garzotto[b]

Politecnico di Milano, DEIB Dipartimento di Elettronica, Informazione e Bioingegneria, 20133 Milano, Italy

Keywords: Natural Language Understanding, Uncanny Valley, User Modeling, Conversation Competence and Performance.

Abstract: Our work explores the relation between users and conversational agents from the HCI perspective in an interaction continuum linking humans and agents together. We highlight the need for a common representation space that we name "shared playground". In the shared playground users and agents coordinate through the linguistic notions of competence and performance to reach an "agreement" in order to communicate successfully. The human-agent coordination is possible only if both parties share some preliminary knowledge. we argue that natural language understanding alone is not sufficient to achieve a satisfactory conversation. We elicit the need for level(s) of representation in order to engage the user by ascribing human traits of the agent. We clarify the rise of an Uncanny Valley in conversations and propose possible solutions to mitigate its effects. Finally, we present a set of features to quantitatively describe the eeriness in conversations with the hope to temper distant conversational agents and consolidate closer conversational companions.

1 INTRODUCTION

Seamless communication with machines using natural language has been the ambition of several investigations in the HCI field. Since the formulation of the "imitation game" (Turing, 2009), natural language processing systems improved substantially, enhancing the quality of the conversations between users and machines. Today, intelligent agent exposing conversational capabilities are widespread both in research and industry. Companies are massively advertising their intelligent agents as actuators of tasks. In such cases the conversation ranges over a closed domain to achieve a specific goal.

First generations of conversational agents, known as spoken dialogue systems in the past, did not restrict the conversation to a singular topic. In 1966, Eliza (Weizenbaum, 1966), was the first attempt in carrying out an unrestricted dialogue with humans. While the communication with Eliza occurred through an aseptic interface, just a screen and a keyboard, users accepted the agent as a personal "confessor". Indeed when Weizenbaum, Eliza's creator, asked to publish the conversations with his agent, the users firmly disagreed pointing out the privacy implications.

Eliza is the first hint that a minimalist user interface may be enough to engage the user in meaningful conversations. Although today's agents have visual cues that support their utterances through lights, animations or even human representation as carefully explored by Cassel et. al. (Cassell et al., 2000), the conversations topics are not centered on the user. As a consequence, the user engagement diminishes and after playing around with the agent's "Easter eggs" they downgrade the agent to a task oriented assistant (Luger and Sellen, 2016). What made the difference in Eliza's dialogues is the underlying psychotherapy model inspired to Carl Roger's theory (Rogers, 1977).

Providing the agent with a personality seems the solution to trigger the user interest. This answer has been probed few years after Eliza's trial. Starting with Parry and giving rise to the series of prizes for passing the Turing Test (Bradeško and Mladenić, 2012). We do not want to debate about how the initial idea of the "imitation game" was twisted and misinterpreted in the Loebner prize as explained in (Hutchens, 1996). Our intent is to seek for the presence of humans traits in conversational agents, since, as soon as we admit that users tend to concede human traits to agents, we face the ascent of the Uncanny Valley (Mori et al., 2012).

In our work we highlight the importance of a "shared playground" where the user and the agent commit to unconditional utterances by agreeing on an

[a] https://orcid.org/0000-0001-6696-5057
[b] https://orcid.org/0000-0003-4905-7166

Occhiuto, D. and Garzotto, F.
Hints of Uncanny Utterances in a Disrupted Interaction Continuum.
DOI: 10.5220/0009157302250230
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 2: HUCAPP*, pages 225-230
ISBN: 978-989-758-402-2; ISSN: 2184-4321

explicit dialogue model. In our model the meaning is not driven by a task but conveyed through a set of common conceptual representations. We imagine an abstract playground bringing humans and machines to comparable competences bypassing the uncanny valley that breaks the interaction continuum (Figure 1).

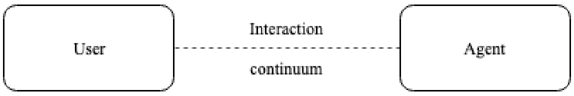

Figure 1: Interaction continuum high-level representation.

The work is organized as follows: in the next section we present the relevant results in the literature; we describe our approach in dealing with the rise of the Uncanny Valley, we discuss potential benefits and flaws of our method and we conclude with a list of improvement to extend our proposal.

2 BACKGROUND

To successfully communicate, humans set up a model holding the information necessary to interact with their peers. To achieve a successful conversation, an agreement between the two interlocutors is necessary, psychololinguistics refers to this agreement with the term *alignment*(Branigan et al., 2003). For us, the consequence of such alignment is reflected by the user engagement in the conversation. Humans tend to align to other human interlocutors, but they also adapt to the non-humans agents when interacting with them (Luger and Sellen, 2016) (Pearson et al., 2006).

If humans try to coordinate with the computer, the same cannot be said for the agent. Usually, the agent has poor to no knowledge about the user and is projected to accomplish all and only the tasks it is programmed for. Accordingly, early investigations such as (Kobsa, 1990) discuss the pros and cons of generating a model to represent the user. Discipline known under the term "user modeling". The authors argue that user modeling is a prerequisite for a cooperative dialogue, likewise they identify potential misuses: e.g. the user awareness of being tracked, her/his consent to be monitored, and the applicable restrictions on personal data. Other methods center on the user expertise by exploiting a collection of users stereotypes (Chin, 1989). The limitation of stereotypes occur in the narrow set of user models available to the system, it is hard to determine the agent's behavior when a human interlocutor does not fall in one of the stereotypes. Moreover, the system may require some time to identify the right user model to be implemented in the conversation.

An alternative path consists in modeling the user throughout the interaction. The latter leads to an adaptive interface where the user model gets built automatically by means of machine learning based algorithms (Langley, 1999). The early attempts in user modeling are reviewed by Fischer (Fischer, 2001). In more recent studies, the concept of stereotypes and the machine learning techniques are unified under the notion of personas applied to neural network architectures (Li et al., 2016). The personas capture the user characteristics and speaking style in order to generate tailored responses.

The goal of user modeling is to lighten the user alignment effort, still, we believe that the agreement to capitalize on the conversation must come from both parties. In fact, flattening the user endeavor to reach an alignment may reduce her/his engagement in the conversation. We believe that a "perfect" (overfitting) user model would mirror the user so well that there would be no point in talking with the agent. Luckily enough, a similar undertaking was examined in the robotics field by Bartneck et. al. (Bartneck et al., 2009). The authors built a robotic clone of one of the authors confirming that there is no significant interest in interacting with the "clone" user model. While the work is critical on the Uncanny Valley hypothesis and advocates that it should not limit the development of robots resembling humans. The authors propose that either there is no difference in likeability of humans and of androids, either the concept of likability is a more complex phenomenon. One limitation of the work resides in the interaction with the robot that are described as "short" and "simple". We agree with the fact that the extent to which users give human traits to agents - anthropomorphism - may be a multi-dimensional construct that is not suited to the bi-dimensional projection depicted in Mori's theory (Mori et al., 2012).

The notion of Uncanny Valley has been investigated for conversational interfaces as well. An empirical test to demonstrate the existence of the "eeriness" perceived by the user has been undertaken for embodied conversational agents in (Stein and Ohler, 2017). The authors set up a virtual environment to stage a conversation happening between two fictional characters. The result of their investigation shows that humans ascribe emotions to peers and not to the agent even when the virtual scene is scripted by it. Emotions seem to remain basic human qualities that are not attributable to agents. The likability defined in the Uncanny Valley hypothesis could be partitioned in different levels of human likeness depending on the nature of the human trait. Moving away from human representations, avoiding embodied representations in particular, some studies searched for the Uncanny equivalent that we emphasized in the aseptic settings in the interaction with Eliza: a screen and a keyboard. Toady such settings are known as “chatbots” or more generally “bots”. One fundamental aspect disclosed by Ciechanowski et. al. (Ciechanowski et al., 2018) dwells in observing that textual chatbots were perceived as “less weird” than the embodied avatars. In akin experiments (Gillespie and Corti, 2016), authors remark that chatbots which generate utterances spoken by humans in a real conversation are not perceived as “eerie”. Demonstrating by means of *echo-borgs* that the Uncanny Valley is closely coupled with the nature of the interlocutor.

To our knowledge no model has been proposed to demystify the Uncanny Valley according to features of the conversation by keeping in mind a common conceptual representation of both parties involved in the dialogue. The equilibrium necessary to preserve the interaction continuum is the core contribution of our work and will be explored in the next sessions.

3 METHOD

The Uncanny Valley hypothesis can be extended to the conversational domain. In our case we wish to set apart the features of interest to ensure the alignment between the user and the agent. We seek to define the “agreement” under which the interaction continuum does not break. Figure 2 illustrates Mori’s hypothesis. In his theory, the affinity reports the degree of perceived eeriness: a toy robot relates to the healthy person more than an industrial robots does. Therefore humans tend to ascribe to the toy robot a greater number of human traits than to the the industrial counterpart. Human likeness, on the other hand, relates mainly to the agent appearance, again a toy robot is more similar to humans than an industrial one.

In the conversation with an incorporeal agent where the interaction is characterized by the aforementioned aseptic setting, by which means do users perceive the human likeness of their interlocutors?

Our research question investigates how the x-axis in the Uncanny Valley graph is translated to a domain where there is no human likeness: the user senses have no access to the agent representation. The agent is treated as a black-box where the only contact point is natural language. The original human likeness can be split into features that characterize the conversation giving rise to a set local maxima (Figure 3). The maxima portray the eeriness of the specific conversation. The z-axis still reflects the affinity, while x-axis and y-axis are the first and second principal component of our feature set. We transfer a generic qualitative measurement from the agent down to the level of a single instance of the agent’s performance (the conversation). We can now sum up all the agent’s performances to obtain the agent anthropomorphic traits by weighting the diverse interactions with it. Our approach enables a quantitative measure of the agent eeriness.

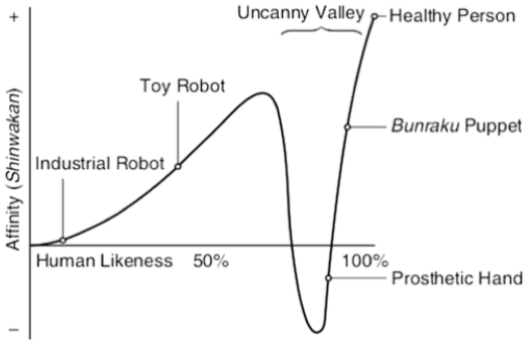

Figure 2: Uncanny Valley hypothesis bi-dimensional graphical representation.

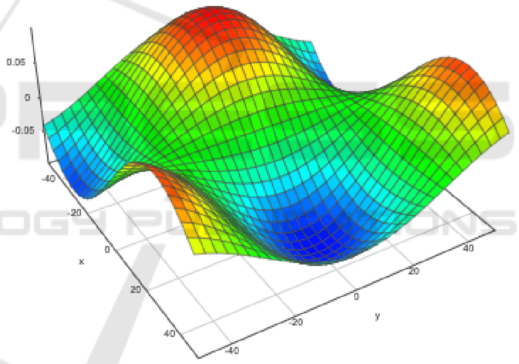

Figure 3: Uncanny Valley hypothesis tri-dimensional graphical representation - Uncanny Landscape.

We group the set of features based on their scope and we distinguish between features that apply to:

- The conversation;
- The fragment;
- The utterance.

The term fragment is borrowed from linguistic theories, where it is used to describe the parts of a sentence (Lappin and Benmamoun, 1999). Here we exploit it to exemplify the parts of a conversation. Fragments originate from points in the conversation where the discourse can be truncated. We may decide to separate fragments on the basis of a length threshold or on the basis of the active context. The granularity of the fragments is not explored in the current study, as we consider fragments as the measure

Table 1: Conversation features as principal component descriptions for human-likeness.

Feature	Conversation	Fragment	Utterance
duration	x	x	x
turn_taking_count	x	x	-
utterance_repetition_count	x	x	-
known_entities_count	x	x	x
unknown_entities_count	x	x	x
known_properties_count	x	x	x
unknown_properties_count	x	x	x
parsable_interpretable_count	-	-	x

between the whole conversation and the single utterance. Keeping the conversation versus the utterances alone would produce a less fine-grained dialogue partitioning.

Table 1 reports the features of interest and the scope to which they apply, marked with "x". The first feature that we consider is the duration in seconds. The total duration of the conversation is a rough indicator of the type of the interaction. We could expect that richer conversations correspond to longer duration. However, in rare cases the user may remain silent and not terminate the interaction; we would obtain a considerable duration but very poor utterances or no utterance at all (just plain silence). For this reason we pair the duration with "turn taking" count. The number of turn taking is directly related to the number of utterances, yet not the same. If the agent is the only one talking, we would collect a high number of utterances but practically no turn takings. The number of turn taking reveals how often the conversation control is transferred from an interlocutor to the other. Hence we calculate the percentage of time where the agent leads the conversation with the user and vice-versa.

Yet another feature is the utterance repetition count, i.e. how often the same utterance appears, applicable to the whole conversation or to a specific fragment. The repetition of the same sentence suggests a stationary conversation where some utterances appear unclear either to the user or to the agent. As a consequence, the agreement that enables the conversation to thrive in the shared abstract playground is violated, the dialogue is broken and the user engagement is undone.

Before listing the remaining features we digress on the shared playground established between the user and the agent. By "shared playground" we refer to the necessary level(s) of representation to support the natural language understanding. In fact, natural language sentence structure alone, is not sufficient to grasp the meaning of an utterance; we argue that the knowledge extraction is not purely syntactic dependant. This does not mean that we do not believe in syntax-centered theories (see Chomsky (Chomsky, 2014)). But we consider to complement them with auxiliary constructs such as ontologies or data-driven knowledge bases. In truth we base our features on the concepts of *competence* and *performance* defined by Chomsky. We stress the need to bring users and agents under the same competence condition: the competence (i.e. the knowledge model of the language) must be shared between the user and the agent prior to the conversation undertaking. The performance (i.e. the actual use of the language) can then take oversee the conversation unrolling.

Now that we introduced the concepts of competence and performance we take care of adding the last features pinpointing the agent's eeriness. The shared playground brings both parties to comparable level(s) of representation meaning that:

- The agent knows all the entities in the user lexicon;
- The agent can represent the entities by means of the properties that characterize them;
- The agent can parse the user utterances;
- The agent can ask to describe an unknown lexicon-external entity by means of a set of known properties;
- The agent adopts the user lexicon throughout the conversation(s).

Conversely:

- The user knows all the entities in the agent lexicon;
- The user can query the entities by means of the properties that characterize them;
- The user can interpret the agent utterances;
- The user can describe an unknown lexicon-external entity by means of a set of known properties;
- The user adopts the agent lexicon throughout the conversation(s).

We can add the known and unknown entity count to the human likeness partition. At the same time we keep track of known and unknown properties count. Finally, we consider the number of utterances that were not parsable/interpretable. Notice that known entities and properties are not complemented by unknown entities and properties: a system could know very few entities (competence) and still very few unknown entities may appear in the actual dialogue (performance). And so we need to track both known and unknown entities and properties in our relevant features.

Lastly, the closing point in the enumerations above implies the reach of an implicit agreement between the user and the agent. Both adapt to one-another. We believe that if the agent is able to explain to the user that the user too must commit to the effort to reach the alignment, then we could preserve the interaction continuum and thus the user engagement. If we reach such a "concious" user behaviour, we may end up with conversational companions rather than conversational agents.

4 DISCUSSION

Our approach is new from the ones present in the literature in several ways: (i) we introduce the fragment granularity; (ii) we highlight entities and properties to be shared as common competence between parties and (iii) we partition the human likeness in quantitative features to identify the potential eeriness - from which the Uncanny Valley generates.

Although we introduced the concept of fragments, the exploitation of the conversation and the utterances is not new in our field, at least for what concerns qualitative evaluation and empirical research. The whole conversation can be qualitatively evaluated through questionnaires as in (Goh et al., 2007), whilst other investigation embrace qualitative evaluations of utterances based on statistical models (Alès et al., 2012).

One limitation of our method, could be the establishment of the user-agent alignment. Indeed we exploit common entities and properties to support quantitative information about the conversation. However, we did not specify how the shared playground is constructed. Exploring the development of the representation level(s) that characterize our shared knowledge surpasses the purpose of this work. But we will tackle the rise of the agreement between the user and the agent in a further study. For now we recognize that the entities building could capitalize on generative approaches based on data-driven learning algorithm as done in (Lowe et al., 2016) to predict new utterances.

Moreover, we wrote about level(s) of representation without specifying how many of them are necessary on purpose. Again, this is a limitation that we are aware of: the number of representation level could become so large that our method would not be scalable. We will investigate this aspect as well in a dedicated experiment. Finally, supervised learning techniques could be exploited to cluster the conversation as "uncanny". Indeed we did set up a model (Uncanny Landscape) per conversation based on the features listed in Table 1, we did not yet calculate the conversations' average to identify the "uncanny" agent.

5 CONCLUSION

We presented a methodology to measure quantitative features that influence the human likeness. We linked the set of features to the concept of Uncanny Valley in conversations, stressing out the need of a common competence (shared playground) between users and agent to boost their performance during the dialogue.

We advocate that the agent affinity for the user - how the user ascribes human traits to the agent - is linked to the level(s) of representation explicit in the model. In the end we are confident that there is much to conversational agents than task oriented dialogues. We encourage the community to pay attention to interaction continuum disruption to achieve a seamless interaction with conversational companions.

ACKNOWLEDGEMENTS

Supported by TIM S.p.A. under the PhD grant "Internet of Products".

REFERENCES

Alès, Z., Duplessis, G. D., Şerban, O., and Pauchet, A. (2012). A methodology to design human-like embodied conversational agents. In *International Workshop on Human-Agent Interaction Design and Models (HAIDM'12)*, pages online–proceedings.

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2009). My robotic doppelgänger - a critical look at the uncanny valley. *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 269–276.

Bradeško, L. and Mladenić, D. (2012). A survey of chatbot systems through a loebner prize competition. In *Proceedings of Slovenian Language Technologies Society*

Eighth Conference of Language Technologies, pages 34–37.

Branigan, H., Pickering, M., Pearson, J., McLean, J., and Nass, C. (2003). Syntactic alignment between computers and people: the role of belief about mental states. In Alterman, R. and Kirsh, D., editors, *Proceedings of the 25th Annual Conference of the Cognitive Science Society, July 31 - August 2 2003, Boston, Massachusetts*, pages 186–191. Lawrence Erlbaum Associates.

Cassell, J., Sullivan, J., Churchill, E., and Prevost, S. (2000). *Embodied conversational agents*. MIT press.

Chin, D. N. (1989). Knome: Modeling what the user knows in uc. In Kobsa, A. and Wahlster, W., editors, *User Models in Dialog Systems*, pages 74–107, Berlin, Heidelberg. Springer Berlin Heidelberg.

Chomsky, N. (2014). *Aspects of the Theory of Syntax*, volume 11. MIT press.

Ciechanowski, L., Przegalinska, A., Magnuski, M., and Gloor, P. A. (2018). In the shades of the uncanny valley: An experimental study of human-chatbot interaction. *Future Generation Comp. Syst.*, 92:539–548.

Fischer, G. (2001). User modeling in human–computer interaction. *User Modeling and User-Adapted Interaction*, 11(1):65–86.

Gillespie, A. and Corti, K. (2016). The body that speaks: Recombining bodies and speech sources in unscripted face-to-face communication. *Frontiers in Psychology*, 7:1300.

Goh, O. S., Ardil, C., Wong, W., and Fung, C. C. (2007). A black-box approach for response quality evaluation of conversational agent systems. *International Journal of Computational Intelligence*, 3(3):195–203.

Hutchens, J. L. (1996). How to pass the turing test by cheating. *School of Electrical, Electronic and Computer Engineering research report TR97-05. Perth: University of Western Australia*.

Kobsa, A. (1990). User modeling in dialog systems: Potentials and hazards. *AI & SOCIETY*, 4(3):214–231.

Langley, P. (1999). User modeling in adaptive interfaces. In *Proceedings of the Seventh International Conference on User Modeling*, UM '99, pages 357–370, Secaucus, NJ, USA. Springer-Verlag New York, Inc.

Lappin, S. and Benmamoun, E. (1999). *Fragments: Studies in ellipsis and gapping*. Oxford University Press.

Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., and Dolan, B. (2016). A persona-based neural conversation model. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., and Pineau, J. (2016). On the evaluation of dialogue systems with next utterance classification. *arXiv preprint arXiv:1605.05414*.

Luger, E. and Sellen, A. (2016). Like having a really bad pa: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5286–5297. ACM.

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100.

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., and Nass, C. I. (2006). Adaptive language behavior in hci: How expectations and beliefs about a system affect users' word choice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1177–1180, New York, NY, USA. ACM.

Rogers, C. R. (1977). *Carl Rogers on personal power.* Delacorte.

Stein, J.-P. and Ohler, P. (2017). Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160:43–50.

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test*, pages 23–65. Springer.

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.