



Multimodal Cues of the Sense of Presence and Co-presence in Human-Virtual Agent Interaction

Magalie Ochs, Jeremie Bousquet, Philippe Blache

► To cite this version:

Magalie Ochs, Jeremie Bousquet, Philippe Blache. Multimodal Cues of the Sense of Presence and Co-presence in Human-Virtual Agent Interaction. ACM International Conference on Intelligent Virtual Agent (IVA), Jul 2019, Paris, France. pp.218-220, 10.1145/3308532.3329438 . hal-02355255

HAL Id: hal-02355255

<https://hal.science/hal-02355255>

Submitted on 8 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multimodal Cues of the Sense of Presence and Co-presence in Human-Virtual Agent Interaction

Magalie Ochs

Aix Marseille Univ, Université de
Toulon, CNRS, LIS, Marseille, France

Jeremie Bousquet

Aix Marseille Univ, Université de
Toulon, CNRS, LIS, Marseille, France

Philippe Blache

Laboratoire Parole & Langage,
Aix-Marseille Université, France

ABSTRACT

A key challenge when studying human-agent interaction is the evaluation of user's experience. In virtual reality, this question is addressed by studying the sense of "presence" and "co-presence", generally assessed thanks to well-grounded subjective post-experience questionnaires. In this article, we aim at exploring behavioral measures of presence and co-presence by analyzing multimodal cues produced during an interaction both by the user and the virtual agent. In our study, we started from a corpus of human-agent interaction collected in a task-oriented context: a virtual environment aiming at training doctors to break bad news to a patient (played by a virtual agent). Based on this corpus, we have used machine learning algorithms to explore the possibility of predicting user's sense of presence and co-presence. In particular, we have applied and compared two techniques, Random forest and SVM, both showing very good results in predicting the level of presence and co-presence.

KEYWORDS

Virtual Reality, Presence, Co-presence, Multimodal behavior, Virtual patient

ACM Reference format:

Magalie Ochs, Jeremie Bousquet, and Philippe Blache. 2019. Multimodal Cues of the Sense of Presence and Co-presence in Human-Virtual Agent Interaction. In *Proceedings of ACM International Conference on Intelligent Virtual Agents, PARIS, France, July 2–5, 2019 (IVA '19)*, 3 pages. <https://doi.org/10.1145/3308532.3329438>

1 INTRODUCTION

A key challenge when studying human-agent interaction, is the evaluation of user's experience. Most of existing methods make use of subjective evaluations based on *questionnaires* filled by the users after their interaction with the virtual agent [2, 17, 18]. These questionnaires are used to assess the perception by the user of the virtual agent, the task, the virtual environment, her/his global satisfaction, engagement, etc. In the virtual reality domain, user's experience is usually evaluated through the measure of the *sense of presence*, in other words the feeling of being present in the virtual environment. Besides this notion, when the virtual environment is populated by virtual agent or avatars, the sense of *co-presence* (also

commonly designated as *social presence*), can be evaluated to assess "the sense of being and acting with others in a virtual space" [13]¹.

However, in spite of their interest, questionnaires are faced with an important drawback: the subjectivity of the approach consisting in asking users to self-report their feeling. Previous works have tried to find objective measures by hypothesizing that different levels of the sense of presence and co-presence may be correlated with different verbal and non-verbal users' behaviors [6, 7]. As highlighted in [12], we can distinguish "subjective presence" from "behavioral presence", subjective presence being measured through a questionnaire and "behavioral presence" corresponding to bodily responses. In this article, we focus on **behavioral measures of presence**. Only few works have been conducted on the behavioral measures of presence and even less for co-presence. Moreover, most of studies have focused on a particular user's behavior related to the task (for instance the navigation path [16]), or on specific behavioral cues such as the interpersonal distance with a virtual agent [1] or the body movements [12, 14]. We propose to analyze a range of behavioral measures of presence and co-presence by studying both verbal and non-verbal behavioral cues simultaneously.

In this paper, we have explored **multimodal behavioral measures of presence and co-presence**. In order to analyze precisely this aspect but also to *automatically predict* the sense of presence and co-presence, we have used machine learning techniques on a human-agent interaction corpus. To this aim, we have used different tools to annotate automatically verbal and non-verbal behavior of the user and the agent. Moreover, we have collected questionnaires assessing the sense of presence and co-presence of the user after the interaction. Technically, different virtual reality displays - known to generate different degrees of immersion - have been used to collect different experiences in terms of sense of presence (the corpus is described in more details in [9]). Starting from this dataset, we have learned different models to analyze the importance of verbal and non-verbal cues as features to predict different levels of presence and co-presence. The accuracy of the model shows that multimodal cues of the user but also of the agent can be used to predict presence and co-presence.

2 PREDICTION OF THE SENSE OF PRESENCE BASED ON MULTIMODAL CUES

The corpus has been collected in the context of a project aiming at developing a *virtual reality environment* to train doctors to *break bad news to a virtual patient* (for details on the project, see [10]). A platform has been developed with a virtual patient with which the doctors can interact in natural language. In total, the corpus is composed of interactions of 38 persons (28 males, 10 females) with

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IVA '19, July 2–5, 2019, PARIS, France

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6672-4/19/07.

<https://doi.org/10.1145/3308532.3329438>

¹Note that no consensus exists on the notion of co-presence. A detailed discussion on the different definitions can be found in [2]

a mean age of 29 years (SD:10.5), each of them interacting three times in three different environment (PC, VR headset, and CAVE) (for more details on the collected corpus see [9]). Each interaction is described by a set of features characterizing user’s and agent’s verbal and non-verbal cues and described in details in [8].

Our goal is to predict automatically users’ sense of *presence* and *co-presence* based on behavioral cues. In our context, we consider *two classification problems* making it possible to predict separately the level of the sense of presence and that of co-presence. For each interaction, these levels have been assessed through two questionnaires [2, 11]. The resulting values are continuous in $[0, 5]$. We propose to approach this question as a multi-class problem by clustering the scores into three classes representing: a *low level* of presence and co-presence for the values ranging in $[1; 2, 5]$, a *medium level* for values in $]2, 5; 3, 5[$ and a *high level* for values in $[3, 5; 5]$. Each human-agent interaction is then associated to a value indicating the level of presence and co-presence to predict.

Practically, we compared two machine learning techniques: *SVM* and *Random Forest*. These methods, among the best classifiers [4], have the advantage, compared with other statistical models such as RNN, to handle high-dimensional data with a high generalization power [5, 15]. In the present work, Random Forest (RF) and SVM classifiers models are learned on two tasks, one for each notion. In both cases, we predict the class corresponding to the level of presence or co-presence (low, medium, high). In order to avoid the prediction classification problem due to unbalanced data-set, we have done over-sampling on minority classes by duplicating unbalanced samples to obtain equivalent number of samples in each class, that finally contains 35 samples.

Concerning the random forest algorithm, in order to minimize the generalization error to avoid over-fitting [3], we have evaluated beforehand the optimal number of decision trees on the presence prediction task by considering the performance of the classifiers (F-score, Precision, Recall) and the out-of-bag (OOB) estimated accuracy expected to provide a relevant cue on generalization performances of the RF. Based on the results, we used 300 trees (few improvements is observed with a larger number of trees). For the SVM and the RF, the following methodology has been applied: (1) we first performed tuning of hyper-parameters of each classifier, on each task (prediction of presence and co-presence from a given set of features) (using *GridSearchCV* from *sklearn*); (2) with the identified best hyper-parameters, we performed a k-Fold cross-validation (with $k=10$). As a consequence, the entire data was split into ten parts and performance metrics were evaluated ten times, each time considering one part as the test set and the other nine as the training set; (3) to reduce variance in results given the size of data-set, we repeated the k-Fold cross-validation ten times: resulting scores values are the average over these hundred tests (10×10 folds). To determine the optimal hyper-parameters, we considered those maximizing the F1-Score.

In order to evaluate the importance of user’s and agent’s behavior to predict presence and co-presence, we have compared the performance of different models learned on different set of features: (1) user’s multimodal cues only, (2) agent’s multimodal cues only, (3) both of them, jointly. Multimodal cues correspond to the features characterizing the verbal behavior (i.e. lexical richness, linguistic complexity, average sentence length) and the non-verbal one (i.e.

Presence			F-Score	Precision	Recall
Doctor	Low	RF	0.90	0.86	0.97
		SVM	0.93	1	0.88
	Medium	RF	0.70	0.69	0.75
		SVM	0.78	1	0.68
	High	RF	0.70	0.82	0.82
		SVM	0.82	0.70	0.70
	Macro-avg.	RF	0.77	0.79	0.79
		SVM	0.85	0.90	0.86
Agent	Low	RF	0.89	0.85	0.97
		SVM	0.94	1	0.9
	Medium	RF	0.71	0.7	0.76
		SVM	0.78	0.9	0.72
	High	RF	0.51	0.68	0.46
		SVM	0.86	0.78	1
	Macro-avg.	RF	0.71	0.74	0.73
		SVM	0.86	0.89	0.87
Doctor + Agent	Low	RF	0.94	0.92	0.97
		SVM	0.95	1	0.92
	Medium	RF	0.79	0.77	0.84
		SVM	0.82	0.97	0.76
	High	RF	0.69	0.82	0.62
		SVM	0.86	0.78	0.97
	Macro-avg.	RF	0.81	0.84	0.81
		SVM	0.88	0.91	0.88

Table 1: Performance metrics of the Random Forest (RF) and System Vector Machine (SVM) for the prediction of the presence level considering only the doctor’s multimodal cues, or only the virtual patient’s multimodal cues or both of them.

entropy of the head and the arms movements). The duration is a common feature of the models and the expertise of the participant is considered as a feature of the doctor (details on the set of features described in

Globally, the performance measures reveal an accurate capacity of the models to predict the sense of presence of the user based on multimodal cues. The results show that both the doctor’s and the agent’s multimodal cues have to be considered for predicting of presence. Indeed, the the classifiers (SVM and RF) perform better when taking into account the features characterizing the verbal and non-verbal behavior of the virtual patient and the doctor than considering only the doctor’s or the agent’s behavior. The SVM generally appears as a better classifier on this task compared to the random forest, though it seems to overfit in some cases (precision equalling one). Comparatively, random forest probably provide more accurate estimations of the generalization capabilities of our models. The performances for each class (low, medium, average level of presence) are slightly different with better performance for the low level compared to the medium and high level. This result may be explained by the unbalanced data-set with a larger number of interactions rated with a low level of presence in the initial data-set before the oversampling.

Concerning the co-presence, the results show that this notion can be predicted accurately based on multimodal cues. Interestingly, the comparison of the performance metrics shows that the classification

Co-presence			F-Score	Precision	Recall
Doctor	Low	RF	0.92	0.97	0.92
		SVM	1	1	1
	Medium	RF	0.65	0.64	0.69
		SVM	0.86	0.77	1
	High	RF	0.68	0.73	0.66
		SVM	0.77	1	0.66
	Macro-avg.	RF	0.75	0.78	0.76
		SVM	0.88	0.92	0.89
Agent	Low	RF	0.97	0.95	1
		SVM	1	1	1
	Medium	RF	0.68	0.82	0.62
		SVM	0.89	0.82	1
	High	RF	0.74	0.71	0.8
		SVM	0.83	1	0.74
	Macro-avg.	RF	0.80	0.82	0.80
		SVM	0.90	0.94	0.91
Doctor + Agent	Low	RF	0.95	0.92	1
		SVM	0.99	1	0.97
	Medium	RF	0.69	0.72	0.69
		SVM	0.85	0.74	1
	High	RF	0.72	0.79	0.69
		SVM	0.77	1	0.65
	Macro-avg.	RF	0.79	0.81	0.79
		SVM	0.87	0.91	0.87

Table 2: Performance metrics of the Random Forest (RF) and System Vector Machine (SVM) for the prediction of the co-presence level considering only the doctor’s multimodal cues, or only the virtual patient’s multimodal cues or both of them.

provides better results considering only the agent’s multimodal cues than only the doctor’s multimodal cues or both of them. This result confirms that co-presence is strongly related to the virtual agent behavior compared to the sense of presence. The same remarks concerning the performance of the classifier (SVM versus random forest) and the difference of performances for each class can be applied for co-presence as for presence.

3 CONCLUSION

In this article, we have explored SVM and random forest algorithms to automatically predict the sense of presence and co-presence of users based on objective multimodal behavioral measures. The algorithms has been trained and tested on a human-agent interaction corpus collected in the specific context of a virtual environment developed to train doctors to break bad news to a virtual patient. Specific verbal and non-verbal behavioral cues have been computed. We have defined high-level features to characterize the user’s multimodal behavior. These features describe the frequency of head and arms movements of the user as well the lexical richness and the linguistic complexity of her/his verbal behavior. To analyze the relevancy of the features as behavioral measures of presence and co-presence, they have been correlated, based on machine learning methods, to the sense of presence and co-presence assessed by users

at the end of her interaction with the virtual agent with specific subjective questionnaires.

ACKNOWLEDGMENTS

This work has been funded by the French National Research Agency project ACORFORMED (ANR-14-CE24-0034-02) and supported by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A*MIDEX), STIC-AMSUD Program for the "Empatia" Project.

REFERENCES

- [1] Jeremy N Bailenson, Eyal Aharoni, Andrew C Beall, Rosanna E Guadagno, Aleksandar Dimov, and Jim Blascovich. 2004. Comparing behavioral and self-report measures of embodied agents’s social presence in immersive virtual environments. In *Proceedings of the 7th Annual International Workshop on PRESENCE*. 1864–1105.
- [2] Jeremy N Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments* 14, 4 (2005), 379–393.
- [3] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
- [4] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. 2014. Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research* 15, 1 (2014), 3133–3181.
- [5] George Forman and Ira Cohen. 2004. Learning from little: Comparison of classifiers given little training. In *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer, 161–172.
- [6] Wijnand A Ijsselstein. 2002. Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates. *Proceedings of PRESENCE 2002* (2002), 245–259.
- [7] Jari Laarni, Niklas Ravaja, Timo Saari, Saskia Böcking, Tilo Hartmann, and Holger Schramm. 2015. Ways to measure spatial presence: Review and future directions. In *Immersed in Media*. Springer, 139–185.
- [8] Magalie Ochs, Sameer Jain, and Philippe Blache. 2018. Toward an Automatic Prediction of the Sense of Presence in Virtual Reality Environment. In *6th International Conference on Human-Agent Interaction (HAL-2018)*.
- [9] Magalie Ochs, Daniel Mestre, Grégoire de Montcheuil, Jean-Marie Pergandi, Jorane Saubesty, Evelyne Lombardo, Daniel Francon, and Philippe Blache. 2018. Training doctors’ social skills to break bad news: Evaluation of the impact of virtual environment displays on the sense of presence. *Journal on Multimodal User Interfaces (JMUUI)* to appear (2018).
- [10] M Ochs, G. Montcheuil, J-M Pergandi, J. Saubesty, B. Donval, C. Pelachaud, D. Mestre, and P. Blache. 2017. An architecture of virtual patient simulation platform to train doctor to break bad news. In *International Conference on Computer Animation and Social Agents (CASA)*.
- [11] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators and virtual environments* 10, 3 (2001), 266–281.
- [12] Mel Slater, John McCarthy, and Francesco Maringelli. 1998. The influence of body movement on subjective presence in virtual environments. *Human Factors* 40, 3 (1998), 469–477.
- [13] Mel Slater, Amela Sadagic, Martin Usoh, and Ralph Schroeder. 2006. Small-group behavior in a virtual and real environment: A comparative study. *Small-Group Behavior* 9, 1 (2006).
- [14] Mel Slater and Anthony Steed. 2000. A virtual presence counter. *Presence: Teleoperators & Virtual Environments* 9, 5 (2000), 413–434.
- [15] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9, 1 (2008), 1.
- [16] Martin Usoh, Kevin Arthur, Mary C Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P Brooks Jr. 1999. Walking> walking-in-place> flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 359–364.
- [17] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 497–503.
- [18] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments* 7, 3 (1998), 225–240.