Toward a Semi-Automated Scoping Review of Virtual Human Smiles

Sharon Mozgai[1], Jade Winn[2], Cari Kaurloto[2], Andrew Leeds[1], Dirk Heylen[3], Arno Hartholt[1]
[1]USC Institute for Creative Technologies, [2]USC Libraries, [3]University of Twente
[1]Playa Vista, CA, USA [2]Los Angeles, CA, USA [3] Enschede, the Netherlands
[1]{mozgai, leeds, hartholt}@ict.usc.edu
[2]{jadewinn, cari.kaurloto}@usc.edu
[3]d.k.j.heylen@utwente.nl

Abstract

Smiles are a fundamental facial expression for successful human-agent communication. The growing number of publications in this domain presents an opportunity for future research and design to be informed by a scoping review of the extant literature. This semi-automated review expedites the first steps toward the mapping of Virtual Human (VH) smile research. This paper contributes an overview of the status quo of VH smile research, identifies research streams through cluster analysis, identifies prolific authors in the field, and provides evidence that a full scoping review is needed to synthesize the findings in the expanding domain of VH smile research. To enable collaboration, we provide full access to the refined *VH smile dataset*, key word and author word clouds, as well as interactive evidence maps.

Keywords: human language technologies, machine learning, embodied conversational agents, virtual humans, datasets

1. Introduction

Virtual humans (VHs) are digitally embodied characters designed to simulate face-to-face human interaction. In contrast to chatbots that primarily rely on text or language-based technologies, VHs can employ additional communicative modalities, including the paralinguistic aspects of the voice (e.g., prosody or voice quality), as well as gesture and facial expressions (Wu et al., 2018). A fundamental facial expression for successful human-agent communication, capable of impacting the interpretation of dialogue and modulating the relationship between interlocutors, is smiling (Heylen, 2003; Ochs et al., 2013).

Numerous publications explore the complex topic of VH smiles and span a broad range of research areas; for example, the modeling and generation of VH smiles, rapport building, mimicry, perception, and social signal processing (Obaid et al., 2010; Pelachaud, 2017; Gratch et al., 2006; Prepin et al., 2012; Ochs et al., 2017). This diverse literature, unsurprisingly, includes a wide variety of applications, study designs, and outcome variables. Previous scoping reviews have been inclusive of VH smiles as part of a larger research aim, such as surveying the use of VH facial expressions in prosocial design (Oliveira et al., 2021). However, no prior reviewers have specifically isolated and examined the breadth of extant VH smile literature.

A need remains to systematically bring together this multi-disciplinary research to (1) map the vast body of literature on VH smiles, (2) identify gaps to inform future research, and (3) guide VH design. Toward these aims, while managing the diffuse nature of this literature, we adopted a semi-automated scoping review methodology to rapidly mine an existing primary VH document dataset. This *VH dataset* was compiled by our research team and spans the previous 30 years of VH research. Here we present our first steps toward a semi-automated scoping review of the VH smile literature and make the following contributions:

- Introduce our primary *VH dataset* of 32,924 pieces of published primary research as well as the distillation of the *VH smile dataset* of 76 articles.
- Present our document mining approach to increase the speed of article identification and mapping of the domain.
- Discover and describe topic clusters within the *VH smile dataset.*
- Identify prolific researchers in the field of VH smile research.

The remainder of this paper is organized as follows. The next section describes the scoping review datasets included in the analysis, details the leveraged methodology to collect and collate the documents, and outlines the semi-automated approach employed to facilitate this review. The third section provides specific results for the *VH smile dataset*. The final section discusses the results of our work, defines the limitations of our research and outlines next steps. To enable collaboration, we provide full access to the *VH smile dataset* inclusive of paper titles, abstracts, authors, and document embeddings, as well as generated word clouds and interactive evidence maps.[1]

2. Methods

The here presented work is part of the Virtual Human Fidelity Coalition (VHFC), a collaboration between the University of Southern California Institute for Creative

[1]https://github.com/USC-ICT/VHFC

Name	Search Terms	Range	Found	Included
VH dataset	virtual human(s), embodied conversation agent(s), virtual agent(s), digital human(s)	1990-2021	60,640	32,934
VH smile dataset	(virtual human(s), embodied conversation agent(s), virtual agent(s), digital human(s)) AND smile)	1999-2021	498	76

Table 1: *VH* and *VH smile datasets*. Resources collected from these databases: ACM, ArXiv, Ebsco, Engineering Village, IEEE Xplore, Gale Computer, Proquest, PubMed, ScienceDirect, Scopus, Wiley, and Web of Science.

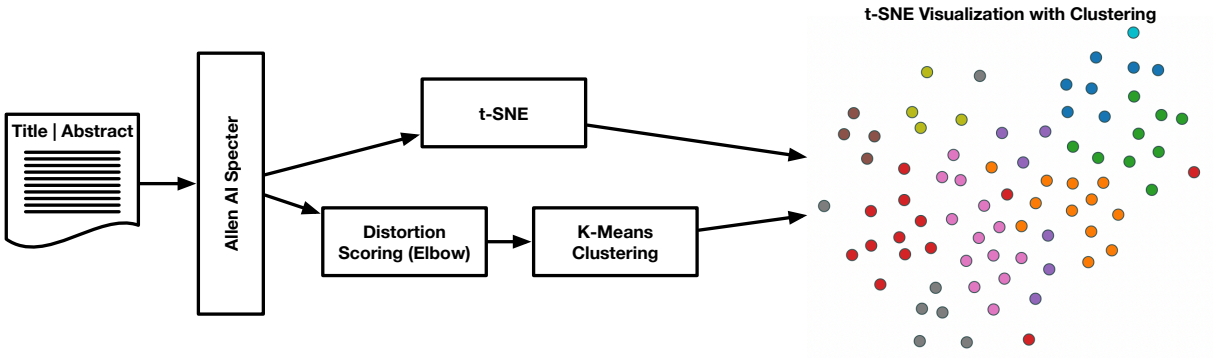

Figure 1: Approach overview. A collection of all papers' titles and abstracts are processed through Specter to derive high-dimensional semantic embeddings of each paper. To visualize the data in two-dimensional representation we leverage t-SNE, while the high-dimensional embeddings are evaluated and clustered on a separate path. Lastly, the cluster assignments are used to color each data point in the two-dimensional representation.

Technologies and USC Libraries. The overarching goal of the VHFC is to explore and catalog research on virtual human fidelity across multidisciplinary domains to drive the efficiency of future VH design while maximizing the efficacy of VH intervention outcomes.

As a first step in the project, comprehensive literature searches were conducted by two expert information specialists in consultation with our research team. Together, we searched 12 electronic databases from 1990 to 2021 to create a primary *VH dataset* of the previous 30 years of VH research. Initial inclusion criteria for this review considered: journal articles, conference proceedings, and grey literature such as dissertations and theses, and review articles published in English. Articles that centered on robots or conversational agents without embodiment were excluded. The search strategy was not limited by study design. Multiple search terms for VHs were employed: VH, virtual agent, and embodied conversational agents are popular terms in the academic literature, while digital human is often the term of choice in industry. A summary of databases, search terms, date ranges, total number of works found (i.e., before removing duplicates or incomplete entries) and total number of included publications is provided in Table 1.

A total of 60,640 resources were retrieved and uploaded into the online systematic review software, Covidence. The software's automatic de-duplicating feature removed 26,487 resources. Additionally, we cleaned the data of any missing data points, leaving 32,934 papers in the *VH dataset*. A subset of this *VH dataset* was created using all possible tenses of the search terms *smile* to mine document titles and abstracts. Following the same de-duping process, a dataset of 141 VH smile articles was collected. Level 1 screening (i.e., titles and abstracts) of the *VH smile dataset* was conducted by two trained reviewers. Articles were excluded that did not include (1) VHs, ECAs, virtual agents, or digital humans and (2) smiles or smiling, resulting in a final *VH smile dataset* of 76 documents. We present our semi-automated document-mining approach of these 76 articles to rapidly map the field of VH smiles and provide evidence that this domain warrants a full scoping review.

2.1. Semi-Automated Data Mining Approach

To visualize the complex relationships between papers and to discover prolific authors and topic clusters within this unstructured document dataset, we employed a multi-step process visualized in Figure 1. We leveraged the state-of-the-art document-level representation learning method SPECTER pre-trained directly on paper titles and abstracts as well as their citation-relationships to derive dense high-dimensional numeric representations for each document (Cohan et al., 2020). Next, we employed t-SNE, a dimensionality reduction algorithm to render the high-dimensional embeddings on a two-dimensional interactive mapping, enabling the visual inspection of the relationships between papers (Van der Maaten and Hinton, 2008). As it is difficult for

ID	Cluster Label	Keywords	# Pubs
0	Socio-Emotional State Displays	stance, straight face, facial changes, social context, state display, alignment	12
1	Signal Processing	emotion recognition, social signals, audiovisual fusion, feature, turn	4
2	Social Effects	social, emotion, study, user impressions, trust, mimicry, interaction feedback	13
3	Modeling Social Behavior	behavior, human-agent social interactions, emotion, personality, behavior	13
4	Deployed VHs	museum guide, recommender, pedagogical agent, application, friendliness	7
5	3D Facial Modeling	vectors, 3D, facial expression synthesis, mapping, model, parameter	7
6	Rubbish Bin	N/A	4
7	Virtual Patient	virtual patient, clinical, photorealistic, incisor	1
8	Animation	animation, genuine smile generation, facial, motion, temporal, dynamic	9
9	Deception	lying, truth, deceit, cooperation, trustworthy, lie, deceivers, truth tellers	6

Table 2: *VH smile dataset* clusters derived by word cloud analysis and manual coding of paper titles and abstracts.

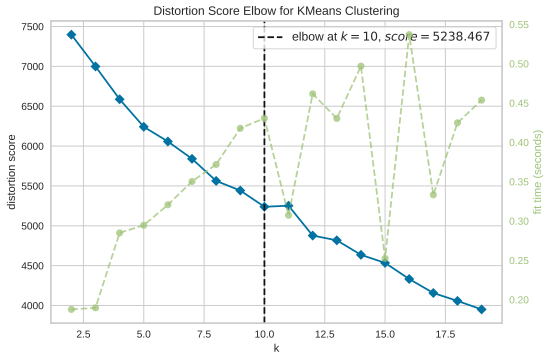

Figure 2: Visualization of k-Elbow distortion metric for optimal k in k-Means clustering.

the human mind to derive meaning and relationships of a high-dimensional representation of the document-embeddings, t-SNE enables two-dimensional visualization of the data while maintaining complex non-linear relationships between the datapoints.[2] To identify the number of research topics and their cluster entries within the vast field of VH smile research we employed the elbow method (Fig. 2) to optimally identify k for the k-means clustering (Kodinariya and Makwana, 2013; Ahmed et al., 2020). In our analysis, we ran the clustering for k $\in$ [2,20] in the *VH smile dataset*. For each value of k we calculate the sum of squared errors (SSE) as the distortion score and selected the elbow, or optimal number of clusters, as the trade-off value between an optimal SSE and a small k.

Following this we identified the topic of each cluster leveraging word cloud analysis (Cui et al., 2010). Before running the algorithm[3] we removed common words known as stopwords (e.g., a, do, get, she, I, etc.) to render the word cloud plots more meaningful and focused on the actual topic rather than just common English words.[4] Additionally, we removed words that were likely to be common to all clusters due to our search strategy (i.e., search terms in Table 1). Once the word clouds (Fig. 3) were rendered, two coders independently (1) reviewed each plot to identify keywords and (2) conducted a manual review of the titles and abstracts in each cluster. Meaningful cluster labels were derived through team discussion and reconciliation of the independently derived codes (see Table 2). While the process of naming the clusters may be somewhat subjective, the access to a reproducible, digestible, and quantitative algorithm such as the word cloud algorithm renders this process transparent and efficient, while double coding and team discussion aims to decrease individual bias. Finally, we utilized word cloud analysis to visualize prolific authors in the *VH smile dataset* and in each of the topic clusters.

3. Results

To map the domain of VH smile research, we determined the optimal number of clusters to be k = 10 for the *VH smile dataset*. Manual review of word clouds for each cluster (Fig. 3) as well as the coding of the associated paper titles and abstracts were synthesized to derive representative cluster names and related key terms (Table 2). The highest populated clusters include papers in the following research streams: Cluster 3 *Modeling Social Behavior* (n=13), Cluster 2 *Social Effects* of VH smiles in agent-human interaction (n=13), and Cluster 1 *Socio-Emotional State Displays* (n=12). Of the ten derived clusters, Cluster 6, affectionately labeled our "rubbish bin" captured errors in the manual coding of the possible 141 articles in the *VH smile dataset* aggregating duplicate and irrelevant articles (e.g., robotics). Additionally, Cluster 7 *Virtual Patient* only contains one paper. Manual review of this cluster determined this paper could be incorporated into Cluster 4 *Deployed VHs*, resulting in eight

[2]We utilize the SciKitLearn implementation of t-SNE with the default parameter setting and a random seed of 0 for reproducibility: `https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html`

[3]We use a common Python word cloud package: `https://github.com/amueller/word_cloud`

[4]We use the standard stopword dictionary that accompanies the Python implementation of the word cloud library.

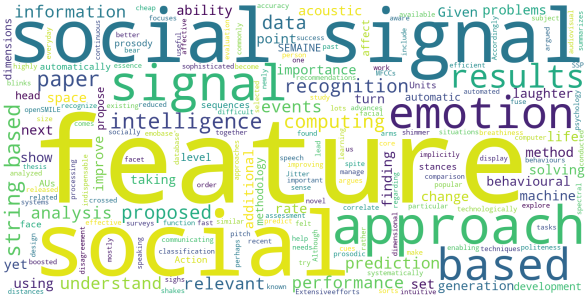

Figure 3: Visualization of keywords found in Cluster 1 *Signal Processing* abstracts.

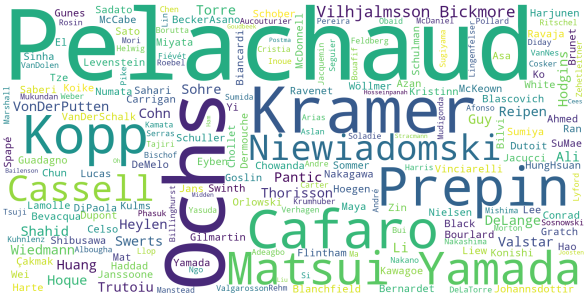

Figure 4: Visualization of prolific authors found in the full *VH smiles dataset*.

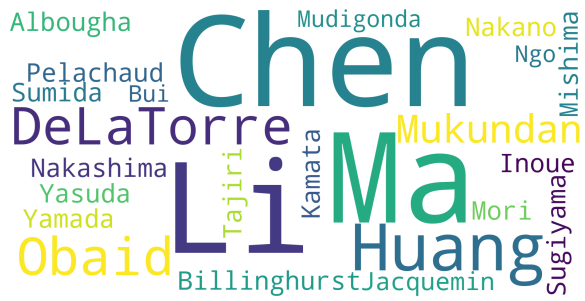

Figure 5: Visualization of prolific authors in Cluster 5 *3D Facial Modeling*.

distinct research streams. After defining the clusters, we moved on to word cloud visualization of prolific authors in the overarching *VH smile dataset* (Fig. 4) and topic-specific clusters (Fig. 5) to further map the field and identify the major contributors in each area.

4. Discussion

This semi-automated review leveraged the resources in the preexisting *VH dataset* to expedite the first steps toward the mapping of VH smile literature. This initial investigation (1) identifies and catalogues research streams concentrated in multidisciplinary topic clusters, (2) brings to the forefront key themes and prolific authors within each topic cluster, and (3) provides evidence that a full scoping review is warranted to further map the field, aggregate research findings, and identify

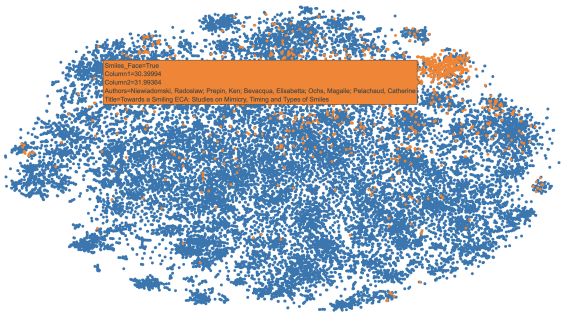

Figure 6: Interactive map of all documents(orange) when the search terms *smile(s)(ed)(ing)* and *facial expression(s)* are applied to the *VH dataset*(blue).

gaps in the current research.

A limitation of our rapid semi-automated review was the use of the specific search term *smile* within the *VH dataset* ultimately yielding only 76 relevant articles for analysis. For the second phase of this work, a planned scoping review, we will expand our investigation to include the search term *facial expression(s)*. An initial review of the *VH dataset* with these added terms revealed 1,206 articles to be included in level 1 screening (Fig. 6). Additionally, while the methodology presented above provided a quick and useful snapshot of the field and a database of relevant papers organized by topic cluster, a full scoping review following the guidelines outlined by Arksey and O'Malley, would take the important next steps in systematically synthesizing empirical results and reporting on aggregate findings (Arksey and O'Malley, 2005).

The adoption and deployment of VHs across multiple contexts such as education, healthcare, military, real estate, customer service, marketing, and sales to automate and innovate tasks is at an all-time high and continues to rise due to the contributions of major game engines, accessibility to the 5G network, and the rise of the metaverse (Ludusan and Wagner, 2021; Hartholt et al., 2019; Ma et al., 2019; Burden and Savin-Baden, 2019; Martha and Santoso, 2019; Endicott, 2021). Studies of human interaction often consider smile dynamics, however, this feature is frequently lacking in complexity and intentional design in VHs, presenting an opportunity to provide evidence-based recommendations for future research and design informed by a full scoping review of the extant VH smile literature.

5. Acknowledgements

Part of the efforts depicted were sponsored by the US Army under contract W911NF-14-D-0005. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

6. Bibliographical References

Ahmed, M., Seraj, R., and Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8):1295.

Arksey, H. and O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, 8(1):19–32.

Burden, D. and Savin-Baden, M. (2019). *Virtual humans: Today and tomorrow*. Chapman and Hall/CRC.

Cohan, A., Feldman, S., Beltagy, I., Downey, D., and Weld, D. S. (2020). Specter: Document-level representation learning using citation-informed transformers. *arXiv preprint arXiv:2004.07180*.

Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M. X., and Qu, H. (2010). Context preserving dynamic word cloud visualization. In *2010 IEEE Pacific Visualization Symposium (PacificVis)*, pages 121–128. IEEE.

Endicott, M. L. (2021). Virtual human systems: A generalised model.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., and Morency, L.-P. (2006). Virtual rapport. In *International Workshop on Intelligent Virtual Agents*, pages 14–27. Springer.

Hartholt, A., Fast, E., Reilly, A., Whitcup, W., Liewer, M., and Mozgai, S. (2019). Ubiquitous virtual humans: A multi-platform framework for embodied ai agents in xr. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 308–3084. IEEE.

Heylen, D. (2003). Facial expressions for conversational agents. In *N. Suzuki and C. Bartneck Subtle Expressivity for Characters and Robots. CHI 2003 Workshop (www. bartneck. de/workshop/chi3003)*.

Kodinariya, T. M. and Makwana, P. R. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95.

Ludusan, B. and Wagner, P. (2021). Knock-knock! who's there? the laughter-enhanced virtual real-estate agent. In *Elektronische Sprachsignalverarbeitung 2021. Tagungsband der 32. Konferenz.*

Ma, T., Sharifi, H., and Chattopadhyay, D. (2019). Virtual humans in health-related interventions: A meta-analysis. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6.

Martha, A. S. D. and Santoso, H. B. (2019). The design and impact of the pedagogical agent: A systematic literature review. *Journal of Educators Online*, 16(1):n1.

Obaid, M., Mukundan, R., Billinghurst, M., and Pelachaud, C. (2010). Expressive mpeg-4 facial animation using quadratic deformation models. In *2010 Seventh International Conference on Computer Graphics, Imaging and Visualization*, pages 9–14. IEEE.

Ochs, M., Pelachaud, C., and Prepin, K. (2013). Social stances by virtual smiles. In *2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 1–4. IEEE.

Ochs, M., Pelachaud, C., and Mckeown, G. (2017). A user perception–based approach to create smiling embodied conversational agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(1):1–33.

Oliveira, R., Arriaga, P., Santos, F. P., Mascarenhas, S., and Paiva, A. (2021). Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Computers in Human Behavior*, 114:106547.

Pelachaud, C. (2017). Conversing with social agents that smile and laugh. In *INTERSPEECH*, page 2052.

Prepin, K., Ochs, M., and Pelachaud, C. (2012). Mutual stance building in dyad of virtual agents: Smile alignment and synchronisation. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 938–943. IEEE.

Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Wu, J., Ghosh, S., Chollet, M., Ly, S., Mozgai, S., and Scherer, S. (2018). Nadia: Neural network driven virtual human conversation agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 173–178.