Body Talk User Studies

Vinod Ramesh, Amber Horvath, Griffin Gonsalves

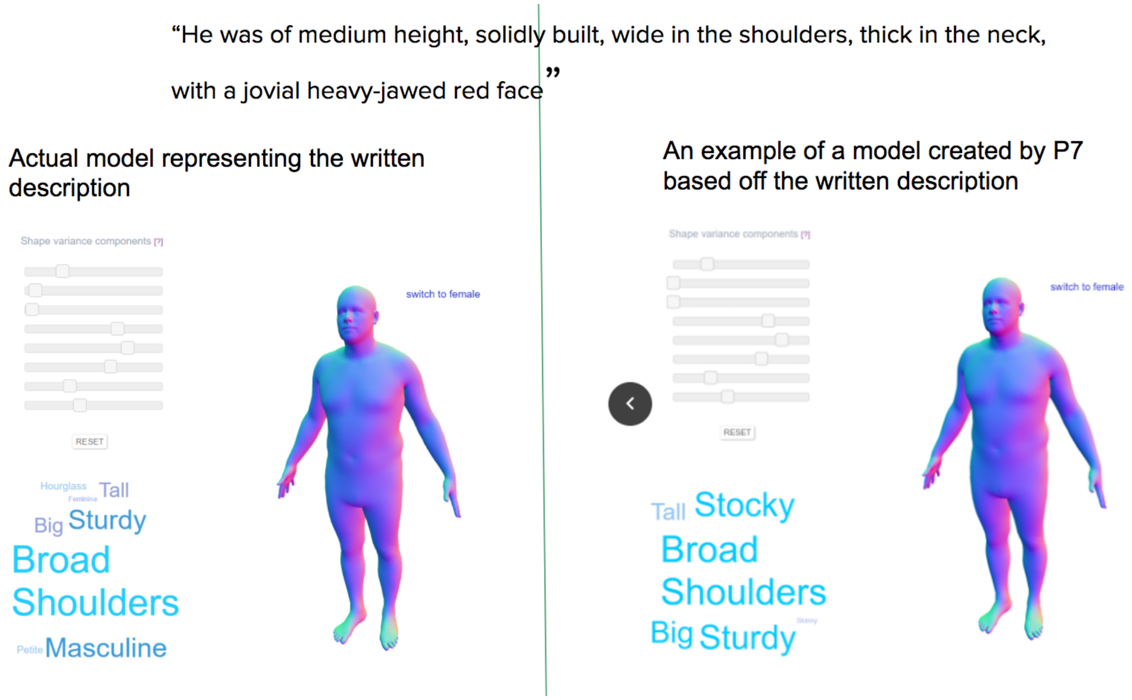

Fig. 1. An example of a model created both by the algorithm (left) and a participant (right) based on the written description provided at the top.

Abstract—Visualizations are only as good as the information they succeed in conveying. When generating human body models, they must accurately create the human body they are attempting to visualize. In our evaluation of the BodyTalk model creation software, we found that all our 7 participants were able to create a model of themselves they felt accurately represented their body shape. Furthermore, participants also were able to make models that matched a textual description provided during the study. In conclusion, we found that the BodyTalk software is usable and successful in creating human shaped models and succeeded in bypassing the "uncanny valley" effect.

1 Introduction

Currently, creating 3D visualizations of human bodies is very limited. The only software readily available are expensive 3D body scanners, which are restricted to specialized set ups and hard for the average consumer to access or character creators in video games, which are restricted to those game worlds. With computers becoming more powerful every year and more capable of producing and rendering 3D models, more specialized 3D modeling software that is not restricted to experts could prove useful.

One field which would benefit from the ability to create 3D models is the field of avatar creation. Avatars are commonly used in 2D applications to provide a visual representation of ones self to the other users of the application. However, while most hardware allows for the usage of 3D modeling, there is little applications making use of 3D models as avatars. This could be useful in applications such as teleconferencing where a visual representation of the person talking may allow for a higher level of engagement with the conversation. Furthermore, 3D avatars have been shown to be successful as mediums for online therapy rooms as they allow for high engagement and empathy between the caretaker and the participant [3]. An important aspect of these avatar-based systems is that they rely heavily upon identification with the avatar that is created. In order for a person to gain the benefits of empathy and engagement through avatars, they must see themselves be represented through the avatar [2].

Another unexplored medium of creating 3D models is the use of words. When a person describes a certain shape, different words can be used to create a mental representation of the object the person is describing. Words like "short" or "stocky" may be used to describe a character like George Costanza from Seinfeld or an actor like Danny Devito. Humans have a generally shared understanding of how these words may be used to describe a person. By applying the shared understanding of how textual descriptions may be used to describe and create a mental

- *Vinod Ramesh is with Oregon State University. E-mail: rameshv@oregonstate.edu.*
- *Amber Horvath is with Oregon State University. E-mail: horvatha@oregonstate.edu.*
- *Griffin Gonsalves is with Oregon State Univeristy. E-mail: gonsalvg@oregonstate.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

picture of a body shape, Streuber et al. created Body Talk [8]. Using text-based descriptions to create models may be useful in the field of criminology when using victim and witness testimonies to try and find a culprit during a case. Text-based descriptions may also be used when attempting to visualize characters from novels or other text-based mediums.

Body Talk allows for the creation of 3D models on a website [1] with word descriptor sliders being used to change the dimensions and shape of the 3D model. However, Streuber et al. did not evaluate the effectiveness of the model's created despite the author's intended goal of having these new visualizations be used in other applications by users. In our study, our participants attempted to create models using the Body Talk website that they felt represented themselves. We are interested in whether these visualizations accurately represent themselves and could be used in avatars in other settings. We also evaluated our participant's ability to create models based purely on a text description to represent the use case of visualizing a character from a novel or a crime victim report.

1.1 Background

1.1.1 Algorithm

Body Talk was developed by a group of researchers at the Max-Planck Institute for Intelligent Systems. The authors originally ran a series of studies using Amazon Turk to evaluate a series of 17 photographs on how much they corresponded to a set of descriptive words. Using this crowdsourcing, attribute ratings are generated based off of standard linguistic descriptions of 3D shape. In total, the survey was able to gather 15 ratings for each word descriptor over a total of 256 bodies(128 Male, 128 Female). The next major step was to generate a linear function that relates the ratings of the words to the 3D model. Additionally some descriptors are correlated meaning that words like "skinny" and "petite" should influence each other. The algorithm takes this into account by adjusting these particular ratings by their standard deviation, and conditioning for a vector that is used in the 3D model.

Fig. 2. Survey on Amazon MTurk

1.1.2 Body Talk Website

The website allows for the user to manipulate sliders corresponding to different words that may be used to describe a person. These 30 words were selected via previous research done by Hill et al. [4]. Figure 4 shows what the Body Talk website looks like to a user. The sliders under Figure 4.1 show the various word descriptors. Note that some of

Consider a single gender. Let the shape of body $i \in 1, \ldots, 128$ be a vector $\mathbf{y}_i = [\beta_1, \ldots, \beta_8]^T$ where the β's are the linear coefficients that represent body shape in the PCA space. Let the vector of ratings for each rater k and body i be a vector $[r_{1,i,k}, \ldots, r_{W,i,k}]^T$, where $W = 30$ words. The individual ratings are noisy and we found it useful to average the ratings for a body over the raters, giving 128 rating vectors that we denote $\mathbf{x}_i = [\bar{r}_{1,i}, \ldots, \bar{r}_{W,i}]^T$. We also tested median rating vectors with similar results.

Our observation matrix is then

$$X = \begin{bmatrix} 1 & \mathbf{x}_1^T \\ \vdots & \vdots \\ 1 & \mathbf{x}_{128}^T \end{bmatrix} \quad (1)$$

and the bodies are represented in $Y = [\mathbf{y}_1, \ldots, \mathbf{y}_{128}]^T$ with one body per row. Assuming a linear relationship between ratings and shape coefficients, we solve for the regression coefficients B in

$$Y = XB + \epsilon \quad (2)$$

using least squares.

This defines our words-to-shape model (**w2s**). Given a new rating vector $\mathbf{x}$, we multiply by B to obtain the body shape coefficients $\mathbf{y}$, which define the shape in the SMPL PCA space.

Fig. 3. Linear function created from dataset

these word descriptors are paired, i.e. if a user slides the "Big" slider to the right, the system will automatically slide sliders such as "Small" to the left or "Heavyset" to the right to prevent the creation of unrealistic body models. This "enforced word correlation" may be turned off, but, since we were partially evaluating the effectiveness of enforced word correlation through our creation of models using purely text-based descriptions, we did not inform our participants of the existence of this slider during the course of the task. After every change in the slider's values, the model shown in Figure 4.2 is updated. While the model is automatically generated with the change of a slider's value, the word cloud seen in Figure 4.3 is also updated. These words should match how a person would describe this model's body shape. They can also be used as a way to validate a textual description provided of a person.

2 Related Work

CESAR Dataset Will add links from last paper...

3 Methodology

4 Results

Data was gathered from the user studies to determine whether Body Talk would be viable in other applications. All 7 Oregon State University students who took part in these user studies were able to make models they felt accurately depicted their body shape. On a 1 to 5 scale, the average student gave their generated model a 4 on how closely the model resembled them. These results show that Body Talk could indeed be useful in generating realistic body types from words (students used sliders for each word in the interface).

However, aside from these impressive stats, the students gave the word descriptions generated from their models a 3.6 out of 5 on whether they believed the words described their body types accurately. This score is okay but not great like the previous score. This is no surprise because Body Talk generated around 6 - 16 words for different models. With that many body type descriptions generated, there is bound to by some words that a students just wont agree with. Rightfully so, some of the word traits would probably not match their body types at all. Aside from Body Talk generating inaccurate body type word descriptions,

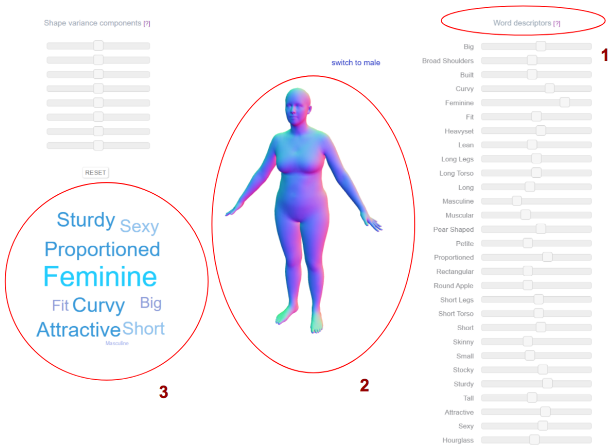

Fig. 4. A screenshot of the Body Talk website. (1) shows the Word Descriptor sliders; (2) shows the model generated by the current configuration of the sliders; (3) shows the outputted word cloud generated based upon the current configuration of sliders

it also generated very accurate ones as well. This would result in a higher score for this test. But just because a student gives a high score, it doesn't mean that it is accurate data. For example, a student may get a "fit" word description generated by Body Talk based off the model the student made of themselves. Even if the person really isn't fit, the student may like to see this trait anyways. Because the program gave them a positive trait, they will give the test a higher score resulting in inaccurate data. The same thing can happen the other way around. A student may find a bad trait that may be true but they don't like, so they give the test a lower score than it should be. All these factors most likely played a factor in the user tests which led to a passable average score.

5 Future Work

Can include improvements to user studies (ex. set time, more participants, more generated models, more tests)

5.1 Lorem ipsum

Lorem ipsum dolor sit amet (see Table 1), consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.

5.2 Mezcal Head

Lorem ipsum dolor sit amet (see Fig. 5), consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Table 1. VIS/VisWeek accepted/presented papers: 1990–2015.

year	Vis/SciVis	SciVis conf	InfoVis	VAST	VAST conf	TVCG @ VIS	CG&A @ VIS	VIS/VisWeek incl. TVCG/CG&A	VIS/VisWeek w/o TVCG/CG&A
2015	33	9	38	33	14	17	15	159	127
2014	34		45	33	21	20		153	133
2013	31		38	32		20		121	101
2012	42		44	30		23		139	116
2011	49		44	26		20		139	119
2010	48		35	26				109	109
2009	54		37	26				117	117
2008	50		28	21				99	99
2007	56		27	24				107	107
2006	63		24	26				113	113
2005	88		31					119	119
2004	70		27					97	97
2003	74		29					103	103
2002	78		23					101	101
2001	74		22					96	96
2000	73		20					93	93
1999	69		19					88	88
1998	72		18					90	90
1997	72		16					88	88
1996	65		12					77	77
1995	56		18					74	74
1994	53							53	53
1993	55							55	55
1992	53							53	53
1991	50							50	50
1990	53							53	53
sum	**1515**	**9**	**595**	**277**	**35**	**100**	**15**	**2546**	**2431**

5.2.1 Duis Autem

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est. Lorem ipsum dolor sit amet.

5.2.2 Ejector Seat Reservation

Duis autem [6][1] vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat,[2] vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Confirmed Ejector Seat Reservation Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat [7]. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum

[1] The algorithm behind Marching Cubes [6] had already been described by Wyvill et al. [9] a year earlier.

[2] Footnotes appear at the bottom of the column.

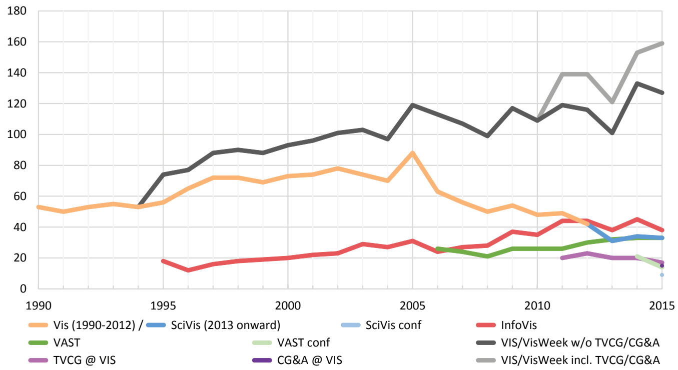

Fig. 5. A visualization of the data from Table 1. The image is from [5] and is in the public domain.

dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Rejected Ejector Seat Reservation Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie

6 Conclusion

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.

Acknowledgments

The authors wish to thank A, B, C. This work was supported in part by a grant from XYZ.

References

[1] Mpi bodytalk. `http://bodytalk.is.tue.mpg.de/`. Accessed: June 12th, 2017.

[2] J.-F. Belisle and H. O. Bodur. Avatars as information: Perception of consumers based on their avatars in virtual worlds. *Psychology and Marketing*, 27(2), July 2010.

[3] A. J. . C. W. Biljana van Rijn, Mick Cooper. Avatar-based therapy within prison settings: pilot evaluation. *British Journal of Guidance and Counseling*, 45(2), Mar. 2015.

[4] S. S. H. C. A. B. M. J. Hill, M. Q. and A. J. O'Toole. Exploring the relationship between body shapes and descriptions by linking similar spaces. *J Vis.*, 15(12), 2015.

[5] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics*, 23, 2017. To appear. doi: 10.1109/TVCG.2016.2615308

[6] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Computer Graphics*, 21(4):163–169, Aug. 1987. doi: 10.1145/37402.37422

[7] G. M. Nielson and B. Hamann. The asymptotic decider: Removing the ambiguity in marching cubes. In *Proc. Visualization*, pp. 83–91. IEEE Computer Society, Los Alamitos, 1991. doi: 10.1109/VISUAL.1991.175782

[8] M. Q. H. C. A. H. S. Z. A. O. . M. J. B. Stephan Streuber, M. Alejandra Quiros-Ramirez. Body talk: Crowdshaping realistic 3d avatars with words. *SIGGRAPH*, July 2016.

[9] G. Wyvill, C. McPheeters, and B. Wyvill. Data structure for *soft* objects. *The Visual Computer*, 2(4):227–234, Aug. 1986. doi: 10.1007/BF01900346