ViBE: Dressing for Diverse Body Shapes
Supplementary File

Wei-Lin Hsiao[1,2] Kristen Grauman[1,2]

[1]The University of Texas at Austin [2] Facebook AI Research

This supplementary file consists of:

- Sampled bodies from clustered types for tops dataset
- Details for user study on validating propagation of positive clothing-body-pairs
- Proposed ViBE's architecture details
- Implementation details for collaborative-filtering (CF) baselines
- Qualitative examples for tops recommendation
- All user study interfaces
- Examples of body-versatile and body-specific dresses judged by Turkers
- Example explanations for Turkers' dress selections

I. Clustered Body Types for Tops Data

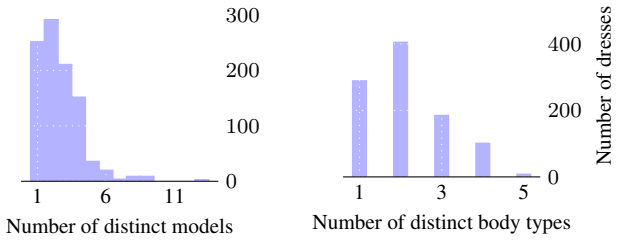
We use k-means [1] clustering (on features defined in main paper Sec.3.4) to quantize the body shapes in our dataset into five types. We do this separately for tops and dresses datasets. Fig. 1 shows bodies sampled from each cluster for the tops dataset, and the result for dresses are in the main paper in Fig. 4.

Figure 1: **Tops dataset**: columns show bodies sampled from the five discovered body types. Each type roughly maps to 1) average, 2) curvy, 3) tall, 4) slender, 5) curvy and tall.

II. User Study to Validate Label Propagation

In this Birdsnest dataset we collected, positive body-clothing pairs are directly obtained from the website, where fashion models wear a specific catalog item. Negative pairs are all the unobserved body-clothing pairings. Taking the dress dataset we collected as an example, we plot the histogram of the number of distinct models wearing the same dress in Fig. 2a. A high portion of false negatives can be observed . After propagating positive clothing pairs within each clustered type, the new histogram with the number of distinct body *types* wearing the same dress is in Fig. 2b. We see most dresses are worn by at least 2 distinct body types, which corresponds to at least 40% individual models being paired with each dress.

To validate whether pairing bodies with clothing worn by different body types gives us true negatives, and whether propagating positive clothing pairs within similar body types gives us true positives, we randomly sample ~ 1000 body-body pairs where each are from a different clustered type (*negatives*), and sample 50% of the body-body pairs within each clustered type (*positives*), and explicitly ask human judges on Amazon Mechanical Turk whether subject A and B have similar body shapes such that the same item of clothing will look similar on them. The instruction interface is in Fig. 7 and the question interface is in Fig. 8. Each body-body pair is answered by 7 Turkers, and we use majority vote as the final consensus. In total, 81% of the positive body-body pairs are judged as similar enough that the same clothing will look similar on them. When we break down the result by cluster types in Tab. 1, we can see that the larger clusters tend to have more similar bodies. On the other hand, 63% of the negative body-body pairs are judged as *not* similar enough to look similar in the same clothing, making them true negatives.



(a) (b)

Figure 2: **Dress dataset**: comparison of number of distinct models vs body types wearing the same dress. Left: initially, over 50% of the dresses are worn by fewer than 3% of the models, indicating a false negative problem. Right: using our discovered body types, most dresses are worn by 2 distinct body types (40% of the models).

Cluster type	1	2	3	4	5
Number of bodies	23	9	14	6	8
Agreement (%)	98	45	82	29	58

Table 1: **Dress dataset**: body-body similarity within the same type, as judged by humans.

III. Architecture Definition for ViBE

The architectures of our embedding model are defined as follows: Let `fck` denote a fully connected layer with k filters, using ReLU as activation function. h_{attr} is an MLP defined as `fcn`, `fc32`, `fc8`; h_{cnn} is defined as `fcn`, `fc256`, `fc8`; h_{meas} is defined as `fcn`, `fc4`, `fc4`; h_{smpl} is defined as `fcn`, `fc8`, `fc4`. `n` is the original features' dimensions, with $n = 64$ and 100 for dresses' and tops' attributes, $n = 2048$ for CNN feature, $n = 4$ for measurement of vital statistics, and $n = 10$ for SMPL parameters. f_{cloth} is defined as `fc8`, `fc4`; f_{body} is defined as `fc16`, `fc4`.

IV. Implementation Details for CF-based Baseline

The collaborative filtering (CF) based baselines consist of a global bias term $b_g \in \mathbb{R}$, an embedding vector $x_u \in \mathbb{R}^d$ and a corresponding bias term $b_u \in \mathbb{R}$ for each user u, and an embedding vector $y_i \in \mathbb{R}^d$ and a corresponding bias term $b_i \in \mathbb{R}$ for each item i. The interaction between user u and item i is denoted as:

$$p_{ui} = \begin{cases} 1, & \text{if } u \text{ observed with } i \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The goal of the embedding vectors and bias terms is to factor users' preference, meaning

$$\hat{p_{ui}} = {x_u}^T y_i + \sum_{*=u,i,g} b_*. \quad (2)$$

Figure 3: **Tops dataset**: example recommendations for two subjects by all methods. Subjects' images and their estimated body shapes are shown on the top of the tables. Each row gives one method's most and least recommended tops. Discussion in Sec. V.

The model is optimized by minimizing the binary cross entropy loss of the interaction:

$$\min_{x_*, y_*} \sum_{u,i} p_{ui} \log(\hat{p_{ui}}) + (1 - p_{ui}) \log(1 - \hat{p_{ui}}). \quad (3)$$

For body-AWARE-CF, we augment the users' and items' embeddings with body and clothing features, $v_u, v_i \in \mathbb{R}^n$: ${x_u}' = [x_u, v_u]$, ${y_i}' = [y_i, v_i]$. These augmented embeddings of users and items, together with the bias terms, produce the final prediction $\hat{p_{ui}}$. We found using $d = 20$ and $n = 5$ to be optimal for this baseline. We train it with SGD with a learning rate of 0.0001 and weight decay 0.0001, decay it by 0.1 at the last 20 epoch and the last 10 epoch, and train until epoch 60 and 80 for the body-agnostic and body-aware CF variants, respectively.

V. Qualitative Figures for Tops

We show qualitative recommendation examples on unseen people (heldout users) for dresses in Fig. 9 in the main paper, and for tops in Fig. 3 here. Each row is a method, and we show its most and least recommended garments for that person. As the tops are less body-specific (in this dataset), either body-AGNOSTIC-CF, AGNOSTIC-EMBED or AWARE-CF fails to recommend garments adapting to subjects with very different body shapes, and most/least recommended garments are almost the same for the two subjects. ViBE recommends cardigans and sweaters with longer hems for the average body shape user, which could create a slimming and extending effect, and it recommends sleeveless, ruched tops for the slender user that shows off her slim arms while balancing the volume to her torso.

VI. User Study Interfaces

In total, we have 4 user studies. Aside from the self-evaluation, each question in a user study is answered by 7 Turkers in order to robustly report results according to their consensus.

Body-similarity user study. This study is to decide whether two subjects (in the same cluster) have similar body shapes such that the same piece of clothing will look similar on them. The instructions for this user study are in Fig. 7, and the question interface is in Fig. 8. This user study validates our positive pairing propagation (see results in Sec. 3.2 in the main paper and Sec. II in this supplementary file).

Dress type user study. This study is to decide whether a dress is body-versatile or body-specific. The instructions for this user study are in Fig. 9, and the question interface is in Fig. 10. We show the most body-versatile and body-specific dresses as rated by the Turkers in Fig. 4. Dresses rated as most body-versatile are mostly solid, loose, shift dresses, and those rated as most body-specific are mostly sleeveless, tight or wrapped dresses with special neckline designs. This is because dresses that cover up most body parts would not accentuate any specific areas, which "play it safe" and are suitable for most body shapes. Dresses that expose specific areas may flatter some body shapes but not others. In total, 65% of the dresses are annotated as more body-versatile than body-specific. This user study is for better analyzing garments in our dataset, as a body-aware clothing recommendation system offers more impact when garments are body-specific. (See results in Sec. 4.1 in the main paper.)

Complementary subject-dress user study. This study is to decide which dress complements a subject's body shape better. The instructions for this user study are in Fig. 11, and the question is in Fig. 12. This user study is for creating a human-annotated benchmark for clothing recommendation based on users' body shapes. (See results in Sec. 4.3 of the main paper.)

Self evaluation. This study is to collect user feedback on which dress complements one's own body shape better. The instructions for this user study are the same as the complementary subject-dress user study above. The interface for users to select the body shape that best resembles them is in Fig. 13, and the question is in Fig. 14. We ask participants to select a 3D body shape directly, as opposed to providing their own photos, for the sake of privacy. This user study is for more accurate clothing recommendation evaluation, as each person knows her own body best. (See results in Sec. 4.3 of the main paper.)

(a) **Body-versatile**

(b) **Body-specific**

Figure 4: **Dress data**: top 10 body-specific and -versatile dresses voted by human annotators.

VII. Explanations for Turkers' Dress Selections

In our complementary subject-dress user study, we ask Turkers to select which dress complements a given subject's body shape better, and to briefly explain reasons for their selections, in terms of the fit and shape of the dresses and the subject (see Sec. 4.3 in the main paper). The provided explanations are utilized as a criterion for evaluating whether the Turker has domain knowledge for answering this task; we do not adopt responses from those that fail this criterion.

Example explanations for adopted responses on 6 different subjects are shown in Fig. 5 and Fig. 6. The reason for why a dress is preferred (or not) are usually similar across multiple Turkers, validating that their selections are not arbitrary nor based on personal style preferences. We believe that including these explanations in our benchmark further enriches its usage. For example, one could utilize it to develop models that provide natural-language-explanations in clothing recommendation.

References

[1] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl. Constrained K-means Clustering with Background Knowledge. In *ICML*, 2001. 1

Dress A just looks like there will be too much fabric on the hips and will make the model look very heavy on the bottom when she is not. The waist of Dress B is not terribly nipped in but will still look better than dress A

The model looks like a block without waist or bust definition. The stripes and baggy hip look of A would only make her look at 10 lbs heavier. The long sleeves and bust definition of B will give the model a slimming effect. The color of B is also more flattering to rectangular torso shape. B is also more feminine with the flouncy sleeves and floral pattern. The feminine touches add to the slimming effect.

Horizontal stripes are not flattering to many body types. Dress A would look great with the subject's thin body type and toned legs.

(a) **Subject 1**

Dress A would better showcase her natural curviness.

The overlapping cloth *(of dress A)* will cover waist better!

The model's body silhouette fits best with a looser dress *(A)* that masks body imperfections.

(Dress A) better for her big build.

(Dress) A looks more slimming on an ample body. It will flow over her curves better and not draw too much attention to a large bust.

Dress A is better for this subject, because this subject have large hips and bust.

The drop waist *(of dress A)* would be flattering to the fuller figure subject. The shoulder peek a boo of dress B is not ideal for this body type.

(b) **Subject 2**

B is probably too tight.

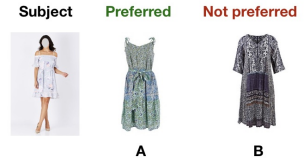
With the body she has, the dress *(A)* will highlight her curves and look perfect.

The shape and dark color *(of dress A) would* look better on a bigger woman.

Dress A is fitted in the waistline, which would complement a curvier figure.

Dress A seems to be a bit more larger than dress B to fit her large hips.

(c) **Subject 3**



B is too baggy, would make her look huge.

Dress B may be too large for the subject while it will most likely fit, Dress A would fit without any issues.

(Dress A) would allow for her to show off her arms.

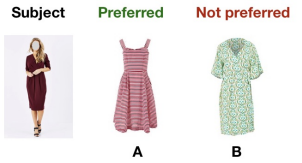
Dress A draws the eyes towards the waist and hugs the subject. Dress A is better here.

Tighter dress *(A)* better for skinnier build.

The model's body silhouette fits best with a waisted dress *(A)* that values the body.

(d) **Subject 4**

Figure 5: **Dress data**: examples of Turkers' explanations for their selections for four subjects. Two more examples are in Fig. 6.



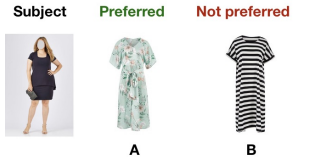
Stripes *(on dress A)* compliment the shape of the subject, accentuating the hips and bust.

The subject would look better in Dress A with it's tied in waist and flared bottom providing more shape to the subjects body compared to Dress B which with it's baggy nature and long sleeves would billow out on the subjects narrow frame making her appear to weigh more than she actually does.

This one *(i.e., dress A)* is cinched at the waist that would accentuate her curves. The other looks too much like a house dress and is too shapeless for her figure.

Although dress A has horizontal stripes, it would be flattering to the subject because she is thin and youthful.

(a) **Subject 5**



The left *(i.e., dress A)* would be better because it would help to cover her dress.

She is fuller-figured so the v neck and the style of the dress *(A)* will look better on her than horizontal stripes.

I think the first dress *(A)* would fit a full figured hour glass better

The adjustable waistband of Dress A would work well with a curvier figure.

(b) **Subject 6**

Figure 6: **Dress data**: examples of Turkers' explanations for their selections for two more subjects. See text for discussion.

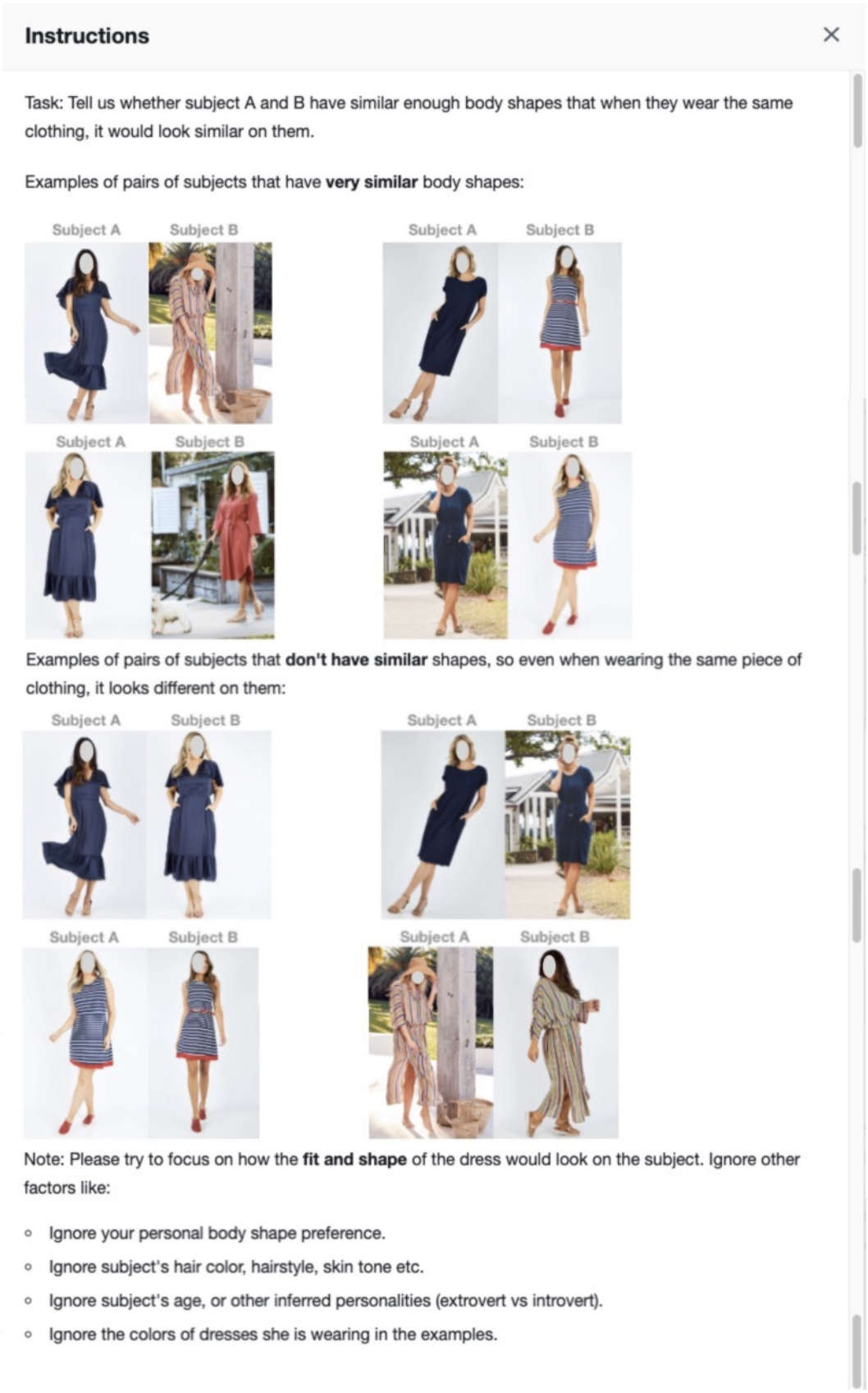


Figure 7: **Body similarity** user study: instructions for judging whether two subjects have similar body shapes such that the same piece of clothing will look similar on them.

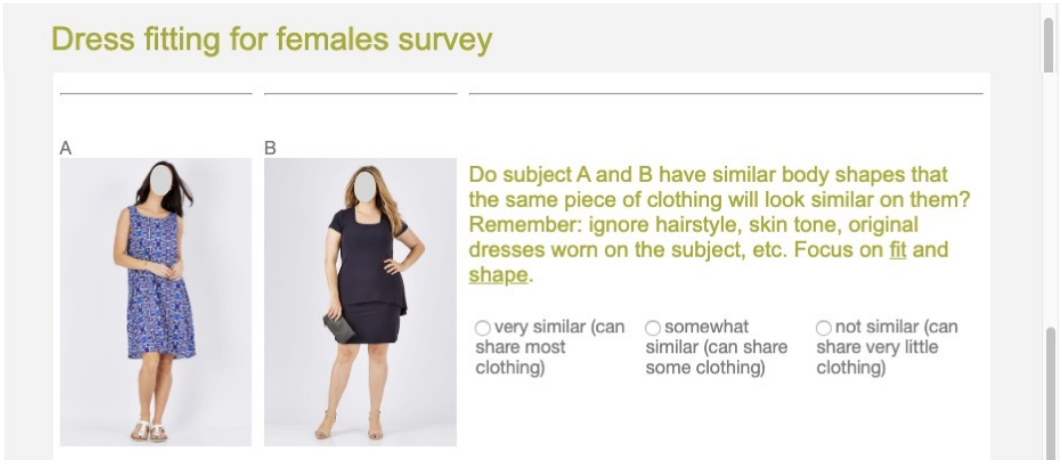


Figure 8: **Body similarity** user study: question to Turkers for judging whether two subjects have similar body shapes such that the same piece of clothing will look similar on them.

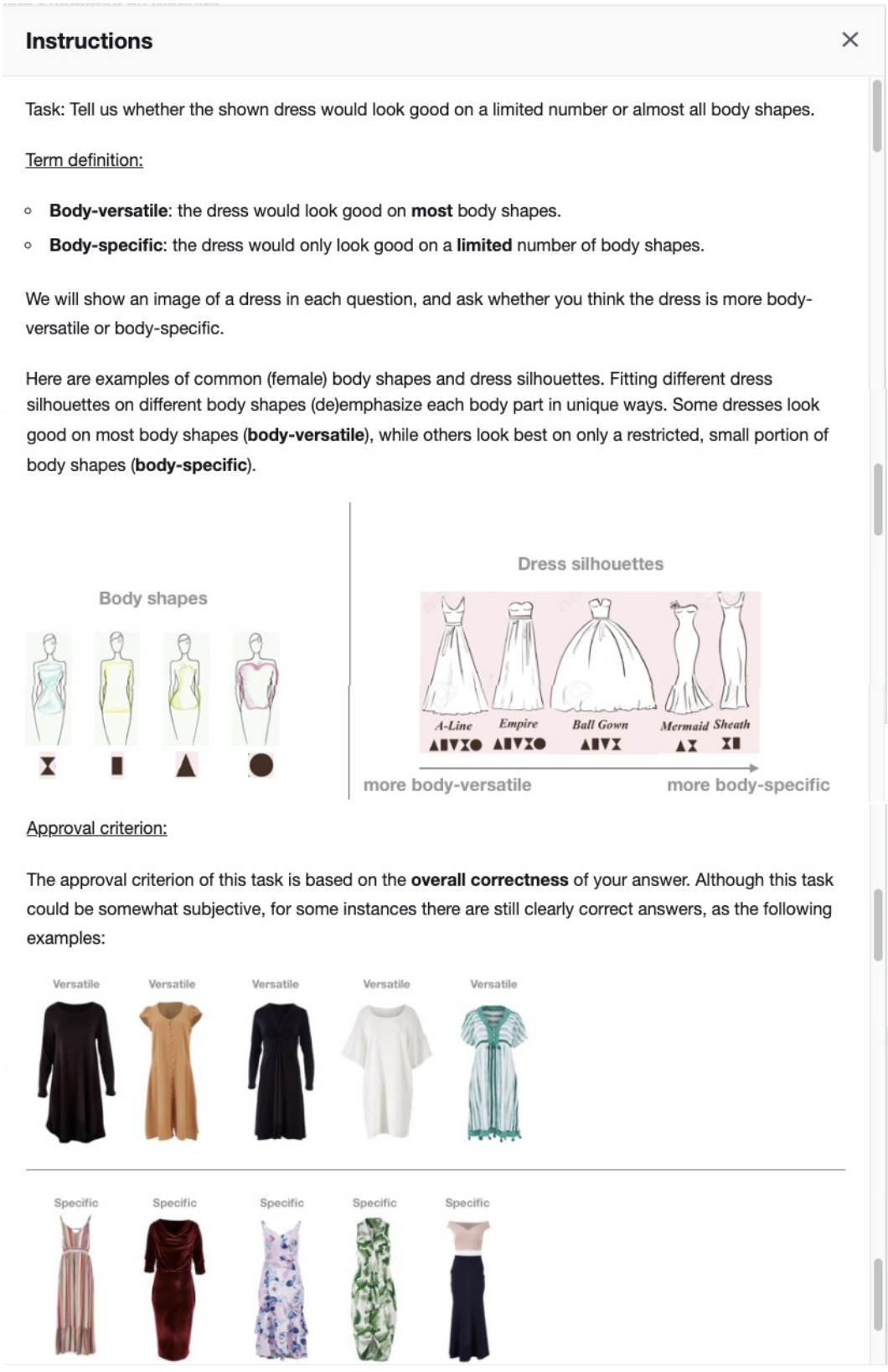


Figure 9: **Dress type** user study: instructions for deciding whether a dress is body-versatile or body-specific.

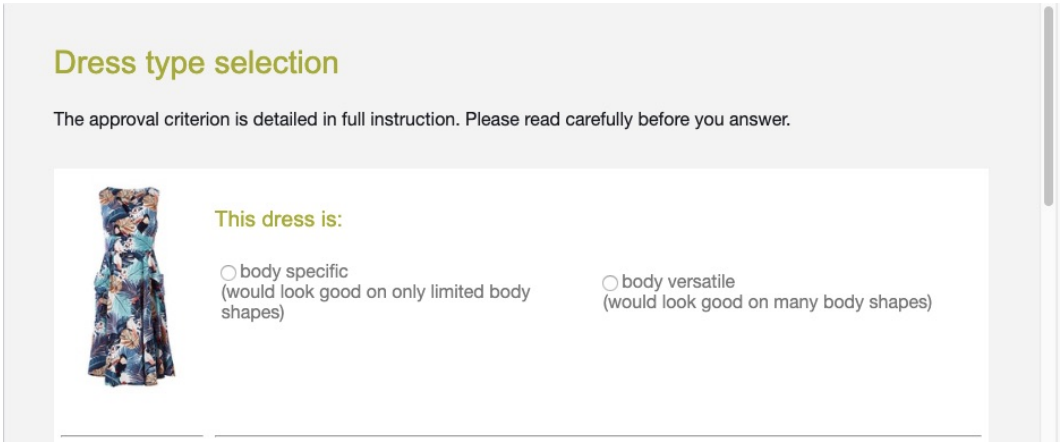


Figure 10: **Dress type** user study: question for deciding whether a dress is body-versatile or body-specific.

Instructions

Task: Tell us whether dress A or dress B would look better on the subject, in terms of complementing her body shape.

We will show images of each subject in two or three different outfits, as below left.

Note: Please try to focus on whether the **fit and shape** of the dress would accentuate the subject's good features. Ignore other factors like:

- Ignore your personal style preference.
- Ignore subject's hair color, hairstyle, skin tone etc.
- Ignore subject's age, or other inferred personalities (extrovert vs introvert).
- Ignore the colors of dresses she is wearing in the examples.

Here are examples of common (female) body shapes and dress silhouettes. Fitting different dress silhouettes on different body shapes (de)emphasize each body part in unique ways. Some dresses look good on most body shapes (**body-versatile**), while others look best on only a restricted, small portion of body shapes (**body-specific**).

When making these decisions, a good explanation for your choice, using the above example, could be like: *Dress A is better for this subject, because the subject seems to have hips that are larger than bust, and dress B deemphasizes the shoulders while dress A accentuates them. To create a well proportioned look, dress B seems better.*

Figure 11: **Complementary subject-dress** user study: instructions for deciding which dress complements a subject's body shape better.

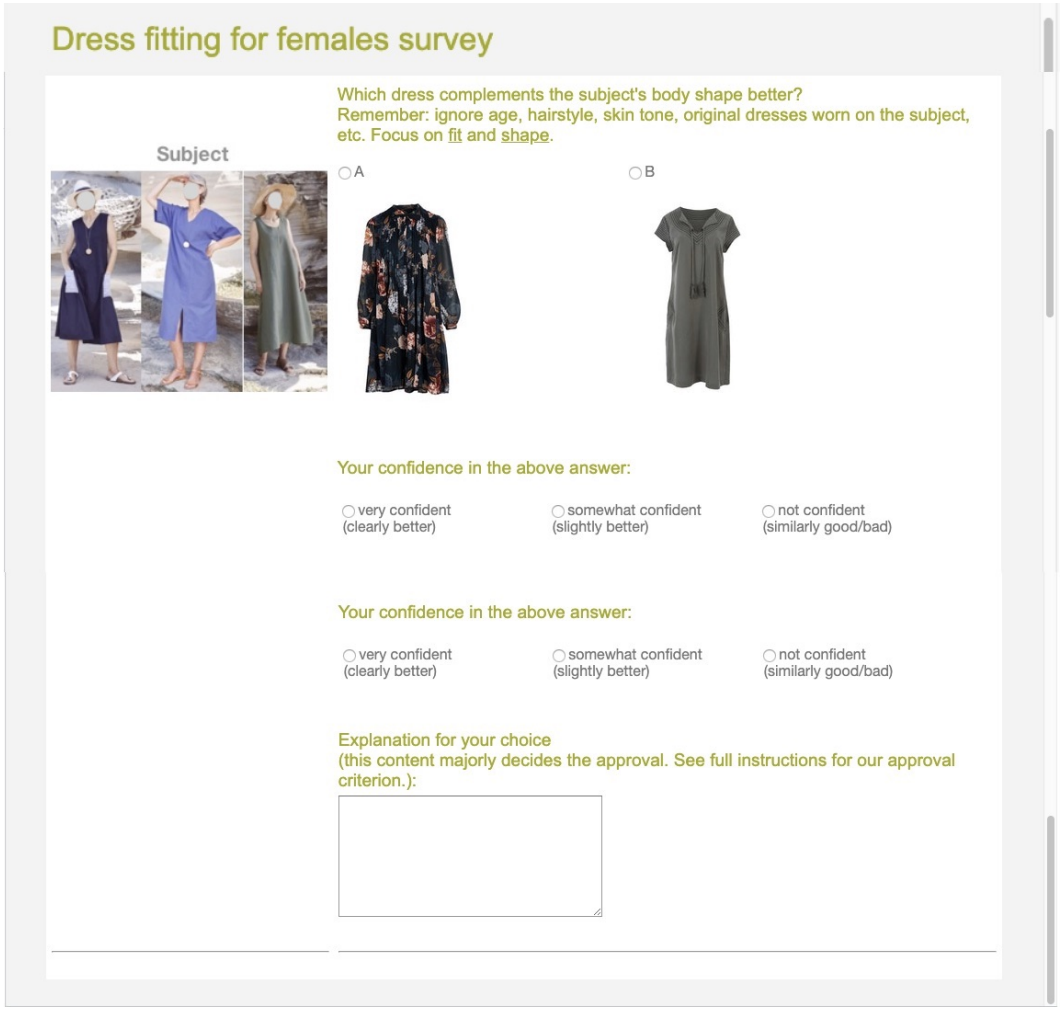


Figure 12: **Complementary subject-dress** user study: question for deciding which dress complements a subject's body shape better.

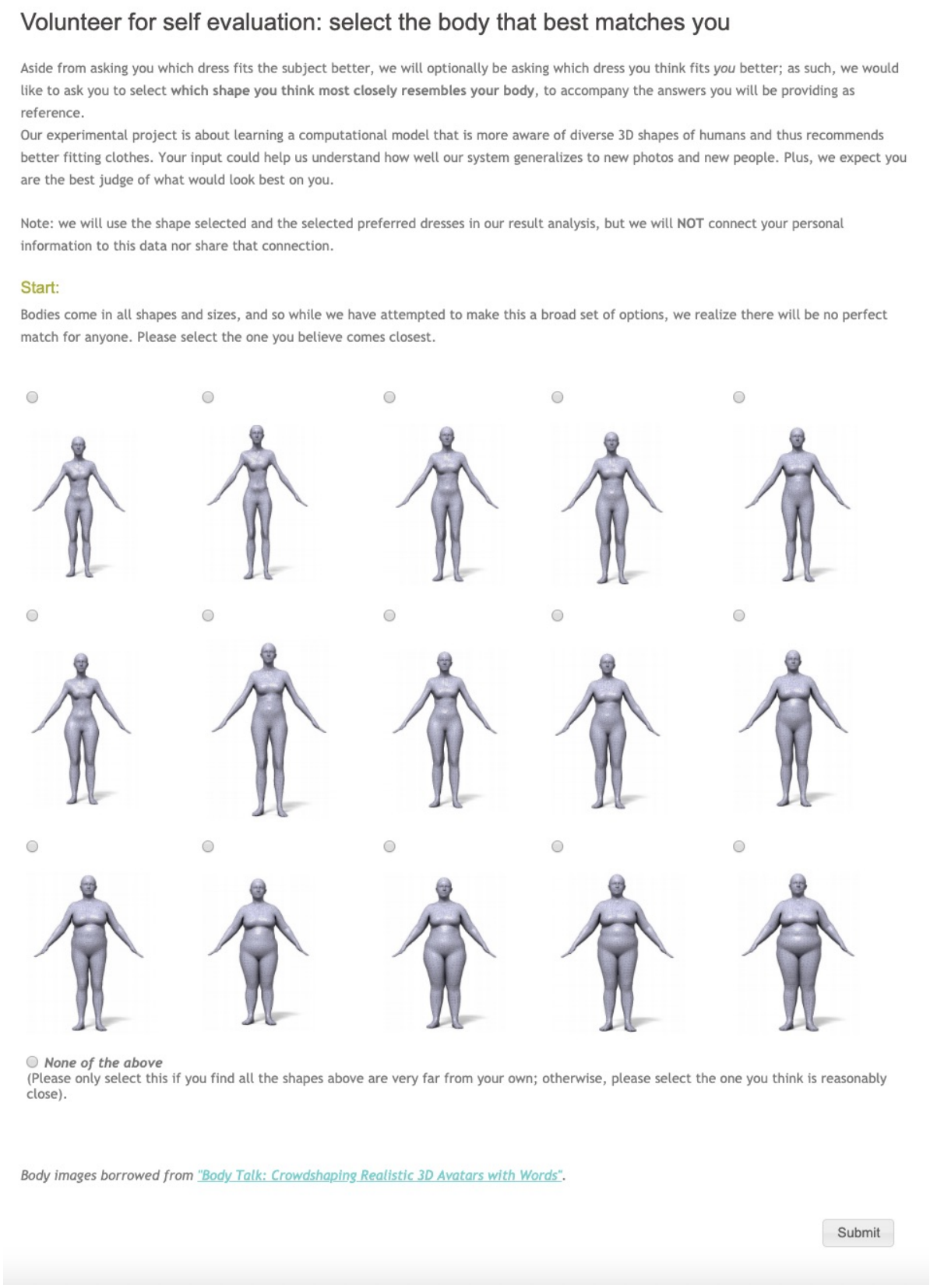


Figure 13: **Self evaluation**: interface for selecting the body shape that best resembles one's self.

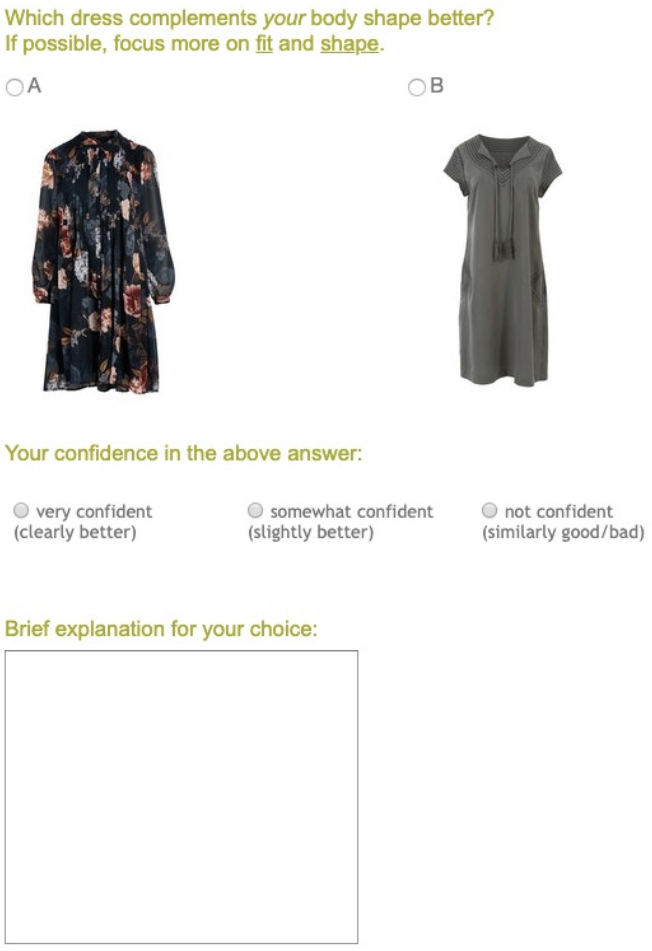


Figure 14: **Self evaluation**: question for deciding which dress complements one's own body better.