Ping Du
Department of Mechanical Engineering,
Iowa State University,
Ames, IA 50011
e-mail: pdu@iastate.edu

Erin F. MacDonald
Assistant Professor
Department of Mechanical Engineering,
Iowa State University,
Ames, IA 50011
e-mail: erinmacd@iastate.edu

Eye-Tracking Data Predict Importance of Product Features and Saliency of Size Change

Features, or visible product attributes, are indispensable product components that influence customer evaluations of functionality, usability, symbolic impressions, and other qualities. Two basic components of features are visual appearance and size. This work tests whether or not eye-tracking data can (1) predict the relative importances between features, with respect to their visual design, in overall customer preference and (2) identify how much a feature must change in size in order to be noticeable by the viewer. The results demonstrate that feature importance is significantly correlated with a variety of gaze data. Results also show that there are significant differences in fixation time and count for noticeable versus unnoticeable size changes. Statistical models of gaze data can predict feature importance and saliency of size change. [DOI: 10.1115/1.4027387]

1 Introduction

Product visuals are an important determinant of customer preference in almost all product categories. The preference for the overall visual design can be thought as based, wholly or in part, on the preferences for the visual design of individual product *features*, defined as visible product attributes or characteristics. We focus on two challenges in the visual design of products: (1) determining feature importance, which refers to the importance of a feature to a customer forming a preference for the whole product; and (2) how noticeable size changes of a feature are to the customer.

Both of these challenges, importance and size, are directly linked to the profitability of a design. Designers cannot spend equal amounts of time perfecting all visual features. They must focus on those that are most important to the customer. Likewise, production budgets for intricate molds, labor-intensive manufacturing processes, and expensive materials must be weighted toward investing in product features that are most likely to increase sales. Size concerns present budgetary constraints as well; for example, a company may have the opportunity to save 10% on production costs by reducing the size of a product feature by 2%, but may have worries that customers will notice this change and perceive it as a loss of quality or luxury. Visual appearance and size of features can both be constrained by product function and other product objectives, such as weight.

As compared with a survey approach for gathering such information, eye-tracking offers more information with less exposure to stimuli. The data offered by eye-tracking hardware/software systems include gaze fixation location, or where a subject is looking on a computer monitor, fixation duration, and fixation timing and ordering—referred to here as *gaze* data. Gaze data have been used to indicate attribute importance [1], but the relationship between gaze data and importance ratings has not been directly proven. This paper lays a foundation for future use of the gaze data to facilitate product design.

This paper proceeds as follows: Section 2 provides background information for eye-tracking research, attribute importance, and feature size; Sec. 3 contains research propositions and associated hypotheses for this paper; and Sec. 4 specifies the methodology. Results regarding feature importance are presented in Sec. 5, and results regarding size changes are provided in Sec. 6. Sections 7 and 8 discuss the results and present conclusions.

Contributed by the Design Theory and Methodology Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received June 5, 2013; final manuscript received March 26, 2014; published online June 2, 2014. Assoc. Editor: Jonathan Cagan.

2 Background

2.1 Eye-Tracking Research. Gaze data provide quantitative information on the visual acquisition of information. Eye-tracking devices and corresponding software collect, refine, and analyze gaze data. According to the "eye-mind" hypothesis, what people look at is an indication of what they are mentally processing [2,3]. Gaze data provide insights into human cognitive processes to facilitate the investigation of the origins of decisions or behaviors [4] and have been used in research areas, including psychology [5–7], marketing [8–10], human–computer interaction [4,11,12], and industrial engineering [13].

Eye-tracking has become one of the major process-tracing methods for information acquisition research [14]. Another major process-tracing method is computerized process tracing, which is usually conducted through the MOUSELAB software. The MOUSELAB software [15,16] displays information on the computer with covered boxes; people acquire the information by moving the mouse cursor over a box; in the end, the software provides details about which boxes have been visited, the time spent on each box, and so forth. This kind of output is similar to that from the eye-tracking process, but as Lohse and Johnson [14] identified, eye-tracking technology can monitor the process of how the information is acquired more completely and naturally.

There are a number of eye-tracking studies related to the work presented here. Pieters and Warlop [9] used eye-tracking technology to study visual attention during brand choice. They found that, on average, subjects had longer fixation times on the brand they eventually chose compared with other alternatives; neither time pressure nor task motivation altered this relationship. Gofman et al. [17] found that the first gaze location on food packages was correlated with both the total amount of time spent on the packages and the purchase decisions.

Koivunen et al. [18] analyzed gaze path data to study how people perceived product designs with different given tasks: memorizing the product, evaluating its aesthetics, usability, and durability. They also tested how the products were evaluated when no instructions were given. They observed that gaze paths and fixation times varied for the different tasks. Reid et al. [19] used both the gaze data and survey data to elucidate how customer judgments were affected by different representations of product design.

Fig. 1 The Tobii T120 eye tracker (left) and the associated control computer (right)

2.1.1 Eye-Tracking Equipment and Data. Eye-tracking equipment can be used while investigating 3D surroundings, but it is most typically used in conjunction with a computer monitor or screen. The screen presents different visual stimuli to a subject as he or she proceeds through an experimental session. This study used a Tobii T120 commercial eye tracker, shown in Fig. 1. The 17-in. thin-film transistor monitor displayed the experiment with a 1024×768 resolution. The Tobii hardware was used in conjunction with IMOTIONS' ATTENTION TOOL software [20] on a control computer, which managed the gaze data for further analysis. The software can also be configured to record survey data, for example, as in Qualtrics [21].

Prior to data analysis, gaze data are parsed into areas of interest (AOIs), or areas of a given stimulus related to the research hypothesis [11]. Fixation time and fixation count are commonly analyzed types of gaze data. Fixations are "eye movements that stabilize the retina over a stationary object of interest" [22]; fixation time refers to the duration of one fixation, and count refers to the number of fixations. Data on percentage-fixation time and first-located time are also used in this paper. The percentage-fixation time is the fixation time spent on an AOI divided by the total fixation time spent on the stimulus. Compared with the fixation time, the absolute measure of the gaze attention, the percentage-fixation time is a relative measurement. It takes into account of the fact that different stimuli attract different total fixation times, and different people have varied evaluation speeds. The first-located time for an AOI is a measurement of the time between initial exposure to a stimulus and first fixation on that AOI. Information about additional eye-tracking measurements can be found in Refs. [11,12].

2.2 Attribute Importance. Addressing feature importance, specifically the importance of the visual design of features, can be thought of as studying a particular type of product attribute importance. Relative attribute importance identifies product attributes that are most likely to change customer preference through variation in attribute configuration. Bettman et al. define customer decision-making rules, such as compensatory and lexicographic decision rules [23] for which attribute importance can either directly or indirectly determine product choice. They model customer preference decisions (choices) by assigning different importance to different product attributes. In this model, differences in attribute importance cause each attribute of a choice option to have a weighted subjective value. These values are added together to get the total utility for the option. In this model, the customer's final choice decision largely depends on the attributes that are most important to the customer, due to the larger weights on these attributes.

Attribute importance has been assessed in different ways. The most direct way of estimating attribute importance is to ask subjects why they choose a product option. By collecting the attributes indicated in an interview, the relative importance of an attribute can be estimated by the number of times the attribute is mentioned [24]. Attribute importance can also be estimated by establishing relationships between attributes and preference decisions or other evaluations [24]. Banks [25] applied linear discriminant functions to relate preference ratings on attributes to preference of overall products; functions' coefficients were then "converted to units of the standard deviation of the corresponding variable" to indicate the relative importance of the attributes.

In conjoint or discrete choice analysis, attribute importance can be interpreted from the estimated part-worths of attribute/levels (configurations). Orsborn et al. [26] apply this specifically to visual product features. They estimated customer preferences for quantified aesthetic forms using a logit model, and mentioned that attribute importance was indicated by the magnitude of the estimated part-worths. MacDonald et al. [27] studied importance of product attributes, but not visual features. They refined the definition of attribute importance in product design in order to perform statistical tests on this metric in a discrete choice study. Importance was defined as the percentage of customer choice that is determined by a specific attribute, in a hypothetical market where a full factorial combination of products is available. Jaccard et al. [28] conducted an information search task and gained insight on how customers searched for information about different attribute dimensions while making automobile purchasing decisions. Subjects evaluated a choice with available product profiles. Each profile had nine attribute dimensions, each with associated information available to the subjects. The authors calculated two indices of importance for each subject: the order and number of pieces of information collected by subjects. In a mobile phone purchasing case study, Reisen et al. [29] used eye-tracking to test the relationship between attribute importance rankings and the frequency of evaluating related text. They found that the two variables were highly correlated, but there was potential bias in the nonrandom ordering of the related text. Warell and Nåbo [30] proposed a "design format modeling" method to capture and describe the visual form of products. With their method, important design features for a collection of products can be identified by comparing the weighted occurrence frequencies for the features. In a case study of home electronics by Bang and Olufsen, they found that features like "metal finishes," "black surfaces," and "geometrical forms" appeared more frequently compared with the others.

There are other studies that address importance of components of alternatives, that do not study products per se. Jaccard and King [31] estimated attribute importance by comparing two conditional probabilities, defined as the absolute difference between the probability of an intention, such as the intention to vote for a candidate, with a presence of an attribute and without. Schkade and Johnson [32] used the MOUSELAB system to investigate how people evaluated two-payoff gambles in two response modes, pricing, and choice, separately. They used the duration of time spent on an attribute as a measure of attention and indirectly demonstrated that the amount of attention that an attribute attracts may be an indication of its salience or importance.

2.3 Feature Size. Designers determine feature size using customer preference, technical requirements, and other sources of input. For example, a large grille on a car promotes engine cooling and better performance, but may look ugly to customers. Designers want a size change to be noticed when it has positive effects on customer preference, for example, "30% more free" in a detergent bottle. But designers work to hide, conceal, or diminish size changes that could have negative effects on customer preference, for example, a decrease in car trunk size versus last year's model.

Noticeable versus unnoticeable difference is referred to as "saliency," and this term is used in a binary sense (salient or not).

The relationship between attribute size and customer preference has been studied in the marketing literature. Michalek et al. [33] observed that large number size on a dial-readout scale was preferred as it indicated easy readability. Coelho do Vale et al. [34] discussed that package sizes of tempting products, small versus large, could affect customer choices through the activation of self-regulation. With self-regulation activated, customers were more likely to approach small packages, believing the packages would help regulate consumption. Chandon and Ordabayeva [35] found that compared with supersizing a product in three dimensions (height, width, and length), supersizing a product in only one dimension largely increased its choice share. This relationship was not affected by the fact that volume increase was clearly marked. A product downsized in three dimensions would have larger choice share than that downsized in one dimension. These results were due to the visual bias that the same amount of volume change through three dimensions was considered smaller than that through one dimension. Yang and Raghubir [36] have found that elongated containers for frequently purchased goods were considered to have a larger volume which could lead to decreased purchase quantity. Krider et al. [37] studied the perception of container shape and showed that a rectangular cream cheese container was considered larger than a round one even though they actually had the same volume, leading people to buy a lower quantity if packaged as a rectangle. Krider et al. also discovered that customers initially relied on a single dimension, which was most salient, to compare areas.

3 Research Propositions and Associated Hypotheses

People use different viewing strategies for evaluating stimuli sequentially (Seq) and side-by-side (SBS), so both are tested here. In the experiment, all stimuli are evaluated in pairs. Within a SBS given pair, "product A" refers to the left-side stimulus and "product B" to the right-side stimulus. Within a given Seq pair, "product A" refers to the stimulus shown first and "product B" to the stimulus shown subsquently, on the next screen.

PROPOSITION 1. *Feature importance is correlated with gaze data in preference choices between two products.* This proposition is inferred from and supported by the literature presented in Secs. 2.1 and 2.2; see Refs. [1,9,17,24,28,32]. The proposition is tested by the following hypotheses. This first set of hypotheses is accompanied by explanations in plain English to assist in understanding. For explanation of terms mentioned in the hypotheses below, refer to Sec. 2.1.1:

- *Hypothesis 1a: There is a positive correlation between feature importance and the feature's fixation time.* It is hypothesized that subjects spend a longer time looking at more important features during the choice task, and that the longer they look, the more important the feature.
- *Hypothesis 1b: There is a positive correlation between feature importance and the feature's percentage-fixation time.* It is hypothesized that subjects spend a larger percentage of a product stimulus' total fixation time looking at important features.
- *Hypothesis 1c: There is a positive correlation between feature importance and the feature's fixation count.* It is hypothesized that subjects look more frequently at important features than other features.
- *Hypothesis 1d: There is a negative correlation between feature importance and the feature's first-located time.* It is hypothesized that subjects look at important features first.

PROPOSITION 2. *Saliency of size change can be predicted by gaze data.* Sütterlin et al. [1] used gaze data to examine how customers evaluated pairs of options, which were described by text information and were shown sequentially. Some information provided in a pair was the same between options while the other information in the two options was different. In this way, the two options in a pair had both shared and unique information. They observed that the shared information between the two options was evaluated normally when it appeared in the first option but almost ignored when it appeared again in the second option. Based on their findings, we expect that this phenomenon will appear when the size change of a feature for a pair is unnoticeable. Gaze data, therefore, could be used to detect the saliency of size changes by testing whether features are ignored or not.

Two more measurements, Δ fixation time and Δ fixation count, are defined to test the proposition. Δ fixation time/count represents the difference in these quantities for features appearing on product B versus product A. The proposition is tested by a number of hypotheses summarized in Table 1. Note that the blank cells in Table 1 are also tested in the analysis, for completeness. For further explanation of terms and calculations mentioned in the hypotheses below, refer to Sec. 6.1.

As compared with its unnoticeable size-change counterpart:

- Hypothesis 2a: A noticeable size change of a feature in product B has a longer fixation time (Seq).
- Hypothesis 2b: A noticeable size change of a feature in product B has a higher fixation count (Seq).
- Hypothesis 3a: A noticeable-size-change feature pair (for example, two car grilles of noticeable unequal size) has a longer total fixation time (SBS).
- Hypothesis 3b: A noticeable-size-change feature pair has a higher total fixation count (SBS).
- Hypothesis 4a: Δ fixation time of a noticeable-size-change feature is different (Seq).
- Hypothesis 4b: Δ fixation count of a noticeable-size-change feature is different (Seq).

4 Method and Procedure

To test the hypotheses, a computer-based experiment was designed for 72 subjects and implemented using a Tobii eye-tracker and IMOTIONS' ATTENTION TOOL software, introduced in Sec. 2.1.1. Two product categories, cars, and electric bicycles, were used, described in Sec. 4.1. Table 2 provides an overview of the experiment design, described in detail in Sec. 4.2. Subjects for the experiment are introduced in Sec. 4.3. Data preparations, conducted before the results analysis, are introduced in Sec. 4.4.

4.1 Stimuli. Sample stimuli used in the experiment are shown in Fig. 2. The 2012 Chevy Cruze from the Chevrolet website [38] and a Shanyang electric bicycle model from the Global-tradekey website [39] were used as the base digital photographs of the stimuli. Only one base photograph for each of the products was used; different perspectives of the products were not shown. This car was selected because it provided a basic sedan model that was familiar to customers. This electric bicycle was selected because it was transformed from a bike model, which made for a product that was familiar in some ways, but unfamiliar in others. The car and the electric bicycle are both vehicles and are both durable goods, but they differed in their familiarity to customers in the U.S. This allows for explorations as to how product familiarity influences the relationships between feature importance and gaze data. For example, while brand may complicate car evaluations, it is unlikely to affect electric bicycle evaluations as brand is not yet strongly identified with physical features for this new category of bikes.

Sets of stimuli were created from the base photographs in Adobe Photoshop. The features that are varied are called "varied features." The design permutations of these varied features are referred to as "design variants" or "feature design variants," and the size permutations are referred to as "size variants" or "feature size variants." The car stimuli included four design variants each for four varied features: headlights, grille, side mirrors, and wheels. The electric bicycle stimuli included four design variants each for four varied features: handlebars, footrest, seat, and cargo

Table 1 An illustration of the test for proposition 2 about size changes

		Condition	
Associated AOI	Fixation metric	Seq	SBS
Size-changed feature in product B	Time	H2a: noticeable, longer	—
	Count	H2b: noticeable, higher	—
Feature pair with size change	Time	—	H3a: noticeable, longer
	Count	—	H3b: noticeable, higher
Feature pair with size change	Δ time	H4a: noticeable, longer	—
	Δ count	H4b: noticeable, higher	—

Table 2 An overview of the experiment design. "*H*" refers to the associated hypotheses tested.

Section	Survey Questions	Stimuli	Condition	*H*	Gaze data	Survey data used?
I	(1) Indicate preferences	Feature design variants	Seq and SBS	1a	Fixation time	No
	(2) Rate satisfaction of (1)			1b	%-Fixation time	
	(3) Rate product A			1c	Fixation count	
	(4) Rate product B			1d	First-located time	
II	Indicate preferences	Feature size and design variants	Seq and SBS	1a	Fixation time	No
				1b	%-Fixation time	
				1c	Fixation count	
				1d	First-located time	
III	Identify and write down the features that are different between two products	Feature size and design variants	Seq	2a	Fixation time	Size-changed features that subjects mentioned labeled as "noticeable size change"; otherwise as "unnoticeable size change"
				2b	Fixation count	
				4a	Δ Fixation time	
				4b	Δ Fixation count	
			SBS	3a	Fixation time	
				3b	Fixation count	
IV	Rate importance for different features	—	—	1a to 1d	—	Ratings used to test correlations
V	Demographic questions	—	—	—	—	Yes

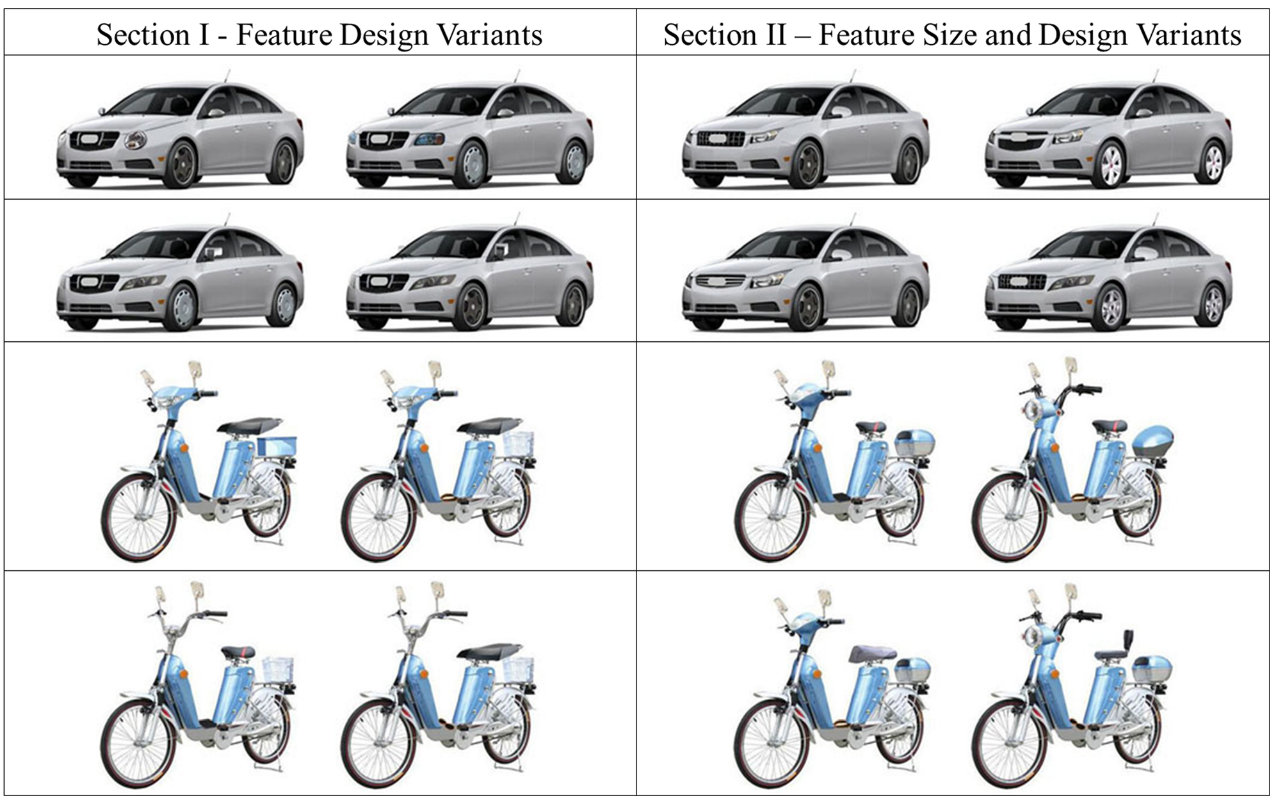

Fig. 2 Sample pairs used in the experiment (section II size variants are headlight (15%), side mirror (20%), seat (15%), and cargo box (10%), from top to bottom)

Table 3 Pairs of car stimuli that involve size variants of headlights, and design variants for other features (numbers “1”–“4” represent different design variants, percentages represent enlargements)

		Headlight	Grille	Wheel	Side mirror
Pair 1	Car A	1	4	2	1
	Car B	1 at 115%	1	1	4
Pair 2	Car A	1	3	3	2
	Car B	1 at 120%	2	4	3
Pair 3	Car A	1	1	2	3
	Car B	1 at 125%	2	1	1

box. The design variants were taken directly or modified from existing cars and bicycles, as shown in Fig. 3. To form product stimuli, the variants were “pasted” onto the base photographs and carefully blended. A pilot study was conducted to evaluate response to the stimuli. Subjects found that the stimuli looked natural and did not distract them from evaluating the product designs—the variant components were not noticeably different from the rest of the stimuli.

For size variants, the headlights, grille, and side mirrors of the car and the handlebars, seat, and cargo box of the electric bicycle were proportionally resized from the base photographs. Each of these features had three size variants. The levels of size variants for the features of the car were 15%, 20%, and 25%, and the electric bicycle were 10%, 15%, 20%. Each feature size variant appeared in one stimulus, which was paired with a stimulus that had the base size of the feature. Table 3 shows the three pairs of stimuli formed for the size variants of the headlights. Numbers “1” through “4” in the table stand for different feature design variants (see Fig. 3), and the percentages represent size variants (enlargements).

4.2 Experiment Design. The experiment was composed of instruction screens, name tag screens that indicated the name(s) of the upcoming stimulus(or stimuli) like “car A,” product stimulus screens that were used to collect gaze data, and survey question screens used to collect information, typically presented after product stimulus screens. A “screen” refers to the information presented on the computer screen, sometimes called a “page” or “slide.” There was no time limit for each screen. In this paper, the gaze data from the product stimulus screens are analyzed (not the gaze data from survey question screens). The experiment flow and an example set of screens are illustrated in Figs. 4 and 5, respectively.

The experiment had five sections, as summarized in Table 2, which took about 20 minutes to complete. Each section began with instructions and a practice question. Sample survey questions asked in the experiment are provided in Table 4. The experiment sections served different goals. Sections I and II were associated with the importance ratings collected in section IV to test proposition 1/hypotheses: 1a–1d. Section III tested proposition 2/hypotheses: 2a–4b. Product stimuli for section I of the experiment had only feature design variants (no size variants). Stimuli used in sections II and III had both size and design variants.

Six pairs of cars and six pairs of electric bicycles were prepared for section I (for a total of 24 stimuli). In this section, two randomly determined pairs of cars and then two randomly determined pairs of electric bicycles were shown to each subject for preference evaluations (eight stimuli per subject). As a result, each of the 12 prepared pairs was seen by four subjects in an experimental condition. After evaluating each pair of stimuli, subjects completed four questions in one screen: (1) indicate preferences using an eight-level scale, which ranged from “strongly prefer product A” to “strongly prefer product B;” (2) rate their satisfaction with the preference decision they just made using an eight-level scale, ranging from “very unsatisfied” to “very satisfied;” and (3)/(4) rate products A and B using an eight-level scale, ranging from “very bad” to “very good.” These scales are adapted from Houston and Sherman [40]. Using eight-level scales in these questions forces preference for either product A or B; data that will be analyzed for cancelation/focus behavior [40] in related work.

Nine pairs of cars and nine pairs of electric bicycles were prepared for section II (36 stimuli). In this section, three randomly determined pairs of cars and then three randomly determined pairs of electric bicycles were shown to each subject for preference evaluations (12 stimuli per subject). The stimuli were chosen with

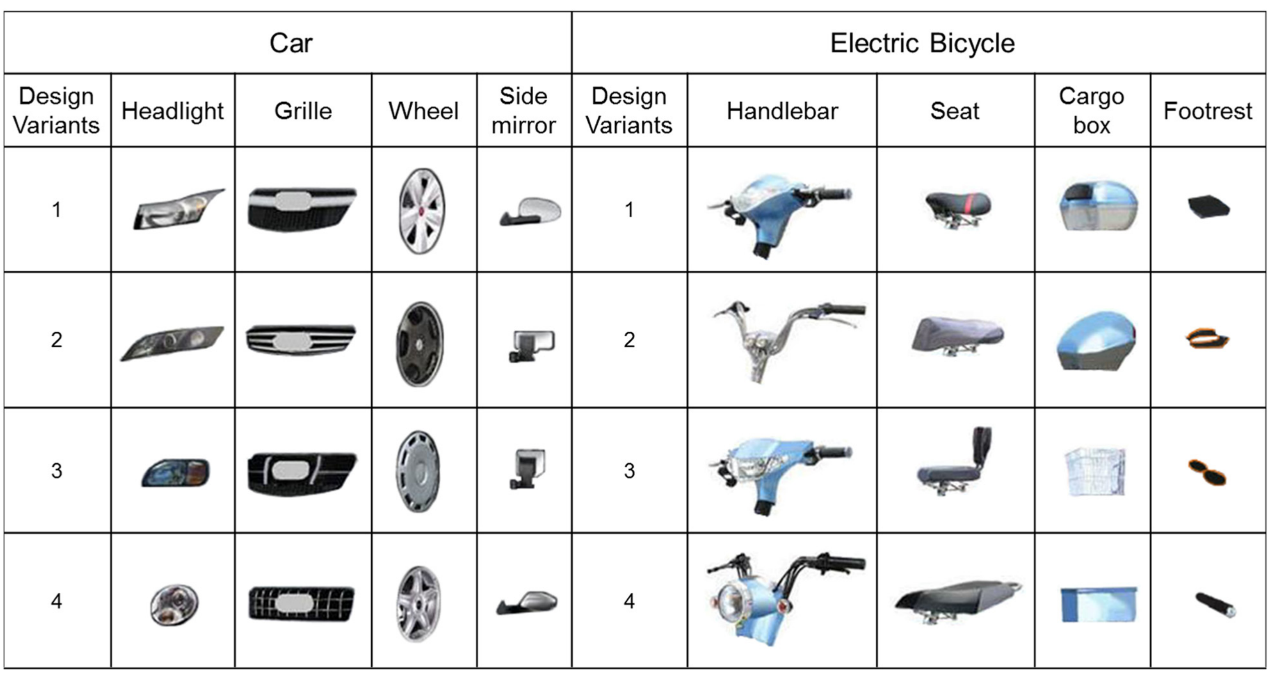

Fig. 3 Design pool for varied features

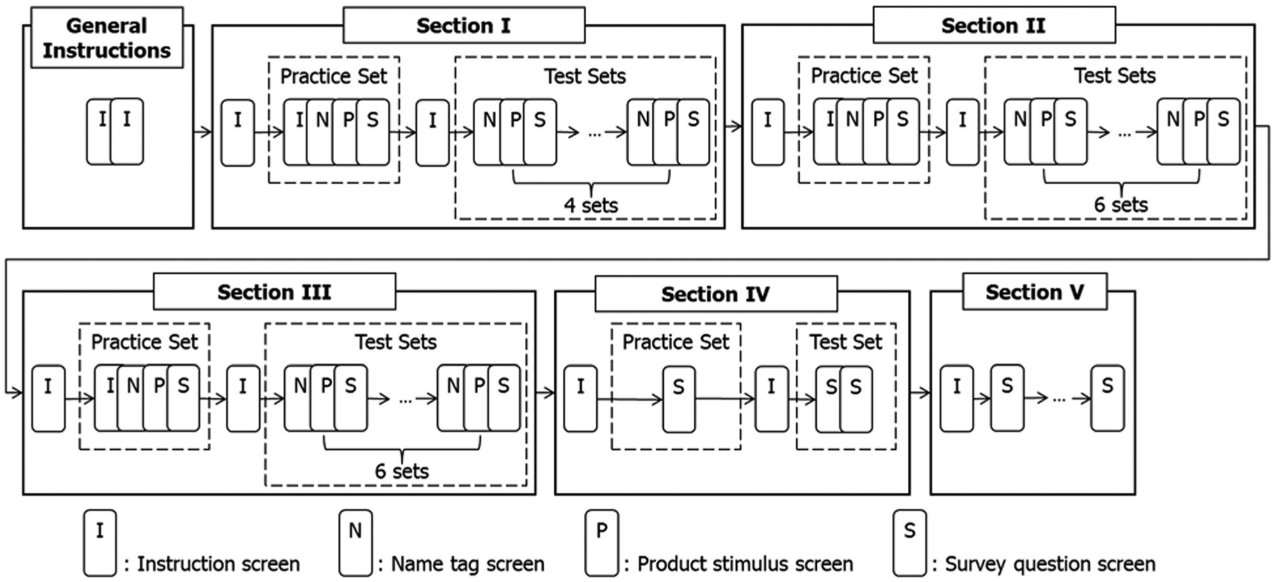

Fig. 4 An illustration of the experiment flow (demonstrated by the SBS condition)

Fig. 5 Screens from experiment section II (images from SBS condition, with enlarged text)

the requirement that in each product category the subject saw three different levels of size variants (enlargements) of different features. As an example, three pairs of cars presented to a subject could be a pair in which the headlights of product B were enlarged 15%; a pair in which the grille of product B was enlarged 25%; and a pair in which the side mirror of product B was enlarged 20%. Each of the 18 pairs of stimuli prepared for this section was seen by 12 subjects in an experimental condition. After each pair of stimuli was presented and evaluated, subjects were asked to indicate their preferences using an eight-level scale as in section I.

Section III repeated the stimuli of section II for each subject. After each pair was evaluated, the subject was asked to identify and write down the features that were different between the two stimuli in a pair—the experiment gave no direct indication of the

Table 4 Sample survey questions asked in the experiment about cars

Section I	(1) Please compare car B with car A and indicate your preference using the following scale: (sliding scale from 1 = "strongly prefer car A" to 8 = "strongly prefer car B") (2) Please evaluate your decision according to the following requirements: (1) Please think about the car you prefer in this pair and rate your satisfaction with the decision using the following scale: (sliding scale from 1 = "very unsatisfied" to 8 = "very satisfied") (2) Please rate for car A using the following scale: (sliding scale from 1 = "very bad" to 8 = "very good") (3) Please rate for car B using the following scale: (sliding scale from 1 = "very bad" to 8 = "very good")
Section II	Given these two options of cars, which one do you prefer? Please use the following scale to rate your preference for the two cars: (sliding scale from 1 = "strongly prefer car A" to 8 = "strongly prefer car B")
Section III	Please identify the differences between these two cars (just list the names of the parts which are different)
Section IV	We have divided a basic car model into nine components: hood/windshield, grill, headlight, bumper/lower grill, wheel, side door, side mirror, side window, and tail, as shown in the image below. How important are these different components' design in forming your preference for the car? Please rate the importance for the design of these components, respectively, using the following scales: (sliding scale from 1 = "not important at all" to 7 = "very important")

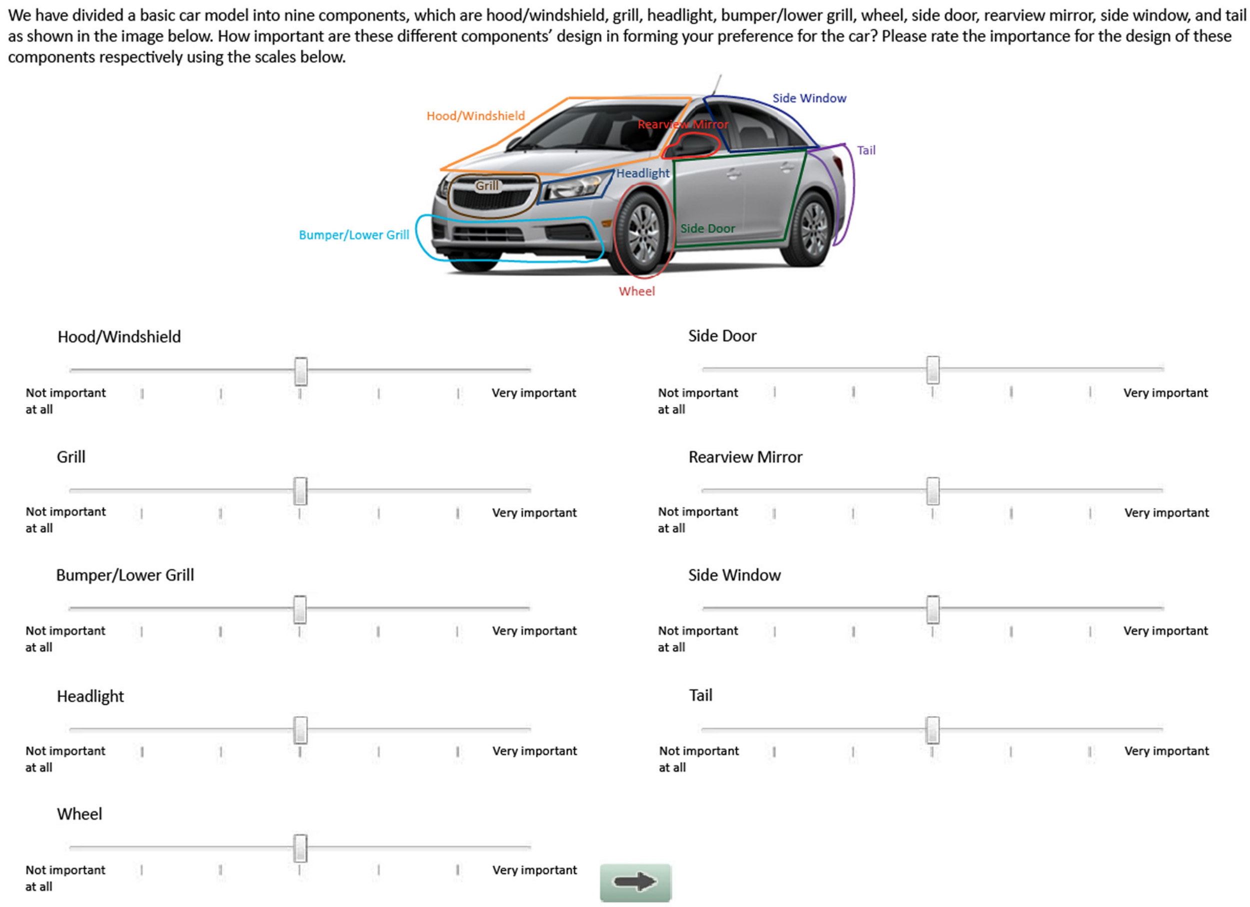

Fig. 6 Car feature-importance-rating survey question screen (rearview mirror referred to as side mirror in this paper)

presence of size variants. This written task allowed for comparison in size-noticing between different types of tasks: implicit and explicit size evaluations, which will be addressed in future work.

Section IV collected importance ratings for all stimuli features. Figure 6 shows the car survey question screen. The rated features included both varied features and unvaried ones, as shown in Table 5, predetermined by the authors. The rating screen showed their names and outlined regions. The section IV rating scales are a typical seven-level Likert scale that made it possible for the subjects to indicate a neutral response. This is different from the preference rating scales in sections I and II (data not analyzed here), which sought to force a preference, as previously discussed in Sec. 4.2. The experiment ended with section V, which collected demographic information.

4.3 Subjects. As stated at the beginning of Sec. 4, the experiment was designed for 72 subjects. However, due to initial computer issues which resulted in unrecorded data for 11 subjects, a total of 83 adults from Iowa State University participated in the experiment, compensated with $5 or extra credit. An online screening survey was conducted to make sure subjects met eye-tracking study requirements. They had normal to corrected vision; did not wear bifocals, trifocals, layered lenses, or regression lenses; did not have difficulty reading a computer screen unassisted; and did not have cataracts, eye implants, glaucoma or permanently dilated pupils [19,41]. Table 6 provides counts of subject gender and ages.

Subjects were randomly assigned to either the Seq or SBS condition. To begin the experiment, subjects were provided with the

Table 5 A collection of features rated in the survey.

Car									
Varied features				Unvaried features					
Grille	Headlight	Side mirror	Wheel	Bumper/lower grille	Door	Hood/windshield	Tail	Window	
Electric bicycle									
Varied features				Unvaried features					
Cargo box	Footrest	Handlebar	Seat	Front frame	Kick stand	Pedal	Rear frame	Rearview mirror	Tire

Table 6 Counts of subject gender and ages; N = number of subjects

Gender	Male	Female	Age	18–24	25–34	35–44	45–54	55–64	65–74
N	37	35	N	27	22	15	4	3	1

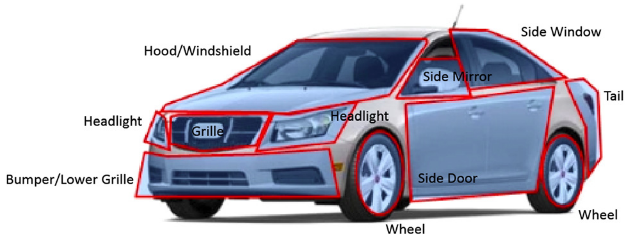

Fig. 7 An example of the AOIs generated for a car

informed consent document. They were then instructed to sit in front of the eye tracker, and adjusted themselves such that their eyes were in the optimal position according to the ATTENTION TOOL's Eye Finder. The subjects were instructed to maintain a consistent posture while completing the experiment. They then performed a calibration exercise, after which the experiment began automatically.

4.4 Data Preparation. The gaze data, together with the survey responses, were collected and managed using the ATTENTION TOOL software. The authors used this software to manually define an area of interest (AOI, see Sec. 2.1.1), for each feature of each stimulus, shown in Fig. 7. The car wheels, the car headlights, and the electric bicycle's rearview mirrors each required the creation of two AOIs for which the gaze data were combined. Gaze data, organized by AOIs, were exported to R, a free statics software platform, for postprocessing, and then analyzed using the statistical software package JMP.

While the ATTENTION TOOL software worked well to identify fixations overall, it failed for two subjects, which had very few fixations identified for almost all product stimuli. An additional eight subjects had no fixations for only a few stimuli. These missing fixations were identified in postprocessing and treated as missing data in the analysis.

5 Proposition 1 Results

The relationships between importance ratings and gaze data from sections I and II were tested separately. The reason for this is that the relevant subject group from section I had only 12 subjects and showed only feature design variants with no size variants, while section II included 36 subjects and showed both design and size variants. The gaze data from section II were used to fit linear regressions to predict the importance of product features. Data from the Seq condition and the SBS condition were analyzed separately, and data from the car and the electric bicycle were analyzed both separately and jointly. The collected survey data from sections I and II are not analyzed. The survey questions served to make the subjects evaluate the stimuli and their related features, but knowing the results (like what product stimuli are preferred and how satisfying the decisions are) is not the interest of this study.

5.1 Experiment Section I Results. Subject-level or individual-level averages were calculated for fixation time, percentage-fixation time, fixation count, and first-located time—throughout Sec. 5, these four types of data will be collectively referred to as *gaze data*. As one subject saw two pairs of stimuli (four stimuli) for each product category, to calculate the subject-level averages of the gaze data, four measurements were averaged for each feature. Example calculations for fixation time are shown in Table 7 as $\bar{T}_{1,1}$ through $\bar{T}_{n,19}$, where $I_{1,1}$ through $I_{n,19}$ represent importance ratings directly collected from experiment section IV.

Next, average gaze data for each feature across all the subjects in an experimental condition were calculated, shown in the last row of Table 7. As a few stimuli for ten subjects were excluded in the analysis due to missing fixations as explained in Sec. 4.4, this process ensured each subject's gaze data contributing equally to the average gaze data. Average importance ratings across the subjects for each feature were also calculated, shown in Table 7 as $\bar{I}_1$ through $\bar{I}_{19}$.

The average gaze data and average importance ratings were used to conduct Pearson correlation tests, shown in Table 8. Conclusions indicated from tests are consistent: there are significant positive correlations between feature importance rating and fixation time, percentage-time and count; and there is a significant negative correlation between feature importance rating and the first-located time on the feature. Hypotheses 1a–1d are strongly supported by these results.

To visualize gaze data across feature importance ratings, observations of how a feature was visually evaluated by each subject were grouped based on the subject's importance rating for the feature, i.e., all features that received a rating of 1 had their gaze data averaged together. Thus, seven averages for each type of gaze data were plotted against the corresponding feature importance ratings. Similar trends were obtained for fixation time, percentage-fixation time, and fixation count, so fixation time is used as a demonstration. Data from the car and the electric bicycle stimuli are combined. Figure 8 shows that for both the ISeq

Table 7 Data used for the correlation analysis, demonstrated with fixation time calculations

	Features				
	Headlight		…	Kick stand	
Subject	Fixation time	Importance rating	…	Fixation time	Importance rating
1	$\bar{T}_{1,1}$	$I_{1,1}$	…	$\bar{T}_{1,19}$	$I_{1,19}$
2	$\bar{T}_{2,1}$	$I_{2,1}$	…	$\bar{T}_{2,19}$	$I_{2,19}$
…	…	…	…	…	…
n	$\bar{T}_{n,1}$	$I_{n,1}$	…	$\bar{T}_{n,19}$	$I_{n,19}$
Averages across subjects (used for the correlation test)	$\bar{T}_1$	$\bar{I}_1$	…	$\bar{T}_{19}$	$\bar{I}_{19}$

Table 8 Correlations between feature's average importance and associated average gaze data for sections I and II ($^{+}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.0001$, one-tailed test)

Fixation metric		Time (ms)	% Time	Count	First-located (ms)
ISeq	Car	0.54^{+}	0.59*	0.56^{+}	−0.51^{+}
	Electric bicycle	0.58*	0.65*	0.60*	−0.51^{+}
	Car and electric bicycle combined	0.51*	0.63**	0.54**	−0.54**
ISBS	Car	0.86**	0.89**	0.81**	−0.70*
	Electric bicycle	0.63*	0.58*	0.66*	−0.58*
	Car and electric bicycle combined	0.70**	0.69**	0.70**	−0.63**
IISeq	Car	0.65*	0.68*	0.68*	−0.72*
	Electric bicycle	0.59*	0.58*	0.57*	−0.51^{+}
	Car and electric bicycle combined	0.61**	0.61**	0.61**	−0.58**
IISBS	Car	0.55^{+}	0.56^{+}	0.50^{+}	−0.50^{+}
	Electric bicycle	0.79**	0.81**	0.78**	−0.70*
	Car and electric bicycle combined	0.73**	0.74**	0.71**	−0.64**

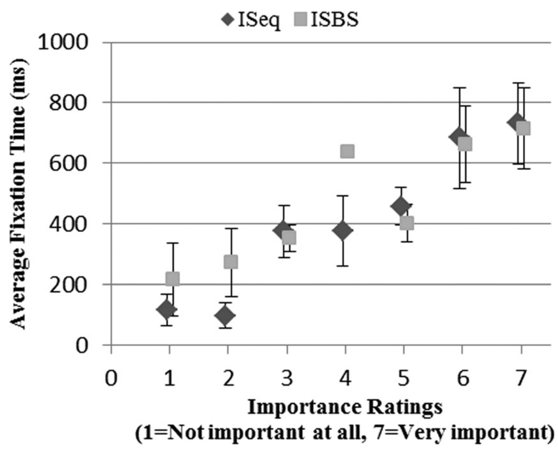

Fig. 8 Average fixation time spent on a feature increases with its importance rating (section I); error bars indicate ±1 standard errors (the two series of data are nudged along the horizontal axis to avoid overlapping of the error bars)

condition and the ISBS condition, there is a clear trend showing that a longer average fixation time is spent for a higher importance rating. These trends are consistent with the significant positive correlations found for hypothesis 1a. The average first-located time was plotted with the importance rating, shown in Fig. 9. The first-located time for a feature decreases with its importance rating in ISeq; while in the ISBS condition, the pattern is less clear. This may be due to the limited data in that condition (12 subjects).

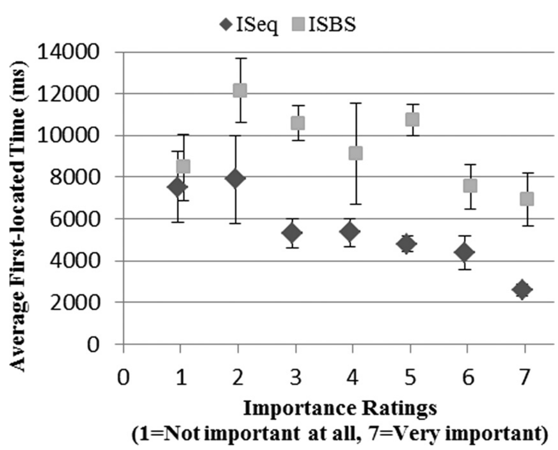

Fig. 9 Average first-located time on a feature varies with importance ratings (section I); error bars indicate ±1 standard errors

5.2 Experiment Section II Results. The method described in Sec. 5.1 was used with the data in section II as well. Average importance ratings for each feature are provided in Table 9. Results of the Pearson correlations are shown in Table 8. In all situations, there are significantly positive correlations between feature importance rating and the fixation time, percentage-time, and count. There is a significantly negative correlation between feature importance rating and the first-located time. Hypotheses 1a–1d are strongly supported by these results. Graphs of trends are shown in Figs. 10 and 11. In both conditions, there is a clear trend showing that the average fixation time spent on a feature increases and the average first-located time on the feature decreases as its importance rating goes up. This indicates that features considered as more important are examined earlier and for longer. When the car and electric bicycle were examined separately, similar trends were observed, as shown in Fig. 12.

Linear regressions were applied to predict feature importance using the gaze data. The regressions were based on the average gaze data and average importance rating for each feature, which was the same data set as that used to obtain the correlation values shown in Table 8. To avoid any potential multicollinearity problems [42] caused by involving correlated variables, only one type of gaze data was chosen as the independent variable. For each situation considered here, the type of gaze data that had the largest correlation with the importance rating, as indicated in Table 8, was chosen. The summary of the fit is shown in Table 10. In all situations, except for the car in the SBS condition, the chosen types of gaze data are significant predictors of the feature importance.

6 Proposition 2 Results

Based on the write-in responses from experiment section III (indicating what size changes are noticed), data from the AOIs of features that had size variants were classified into two sets: noticeable size changes and unnoticeable size changes. These two sets were compared using fixation time/count and Δ fixation time/count. The results from the IIISeq and IIISBS conditions were analyzed separately and tested by one-way ANOVA.

6.1 Results from Experiment Section III, Sequential (IIISeq). As detailed in Table 11, noticeable size changes in product B have significantly larger values of average fixation time and count than unnoticeable ones. For feature pairs, for example, the two car grilles in a stimuli pair, noticeable-size-change feature pairs have a significantly larger average fixation time and count than unnoticeable ones. Average Δ fixation time for the noticeable size changes is significantly different from the unnoticeable ones, with the former value above zero and the latter one below zero. Average Δ fixation count shows the similar results. These results strongly support hypotheses 2a, 2b and 4a, 4b.

A logistic regression used gaze data to predict saliency (noticed versus unnoticed) of a size change. This regression is suitable

Table 9 Average importance ratings for all features

	Car								
	Varied features				Unvaried features				
	Grille	Headlight	Side mirror	Wheel	Bumper/lower grille	Door	Hood/windshield	Tail	Window
Average importance rating	4.73	5.57	4.56	5.14	4.37	4.29	5.16	4.20	3.94
Standard error	0.16	0.13	0.16	0.16	0.16	0.15	0.17	0.18	0.15

	Electric bicycle									
	Varied features				Unvaried features					
	Cargo box	Footrest	Handlebar	Seat	Front frame	Kick stand	Pedal	Rear frame	Rearview mirror	Tire
Average importance rating	5.13	3.57	5.60	6.22	4.54	3.38	4.00	4.00	4.59	4.65
Standard error	0.17	0.17	0.12	0.11	0.15	0.18	0.18	0.16	0.17	0.19

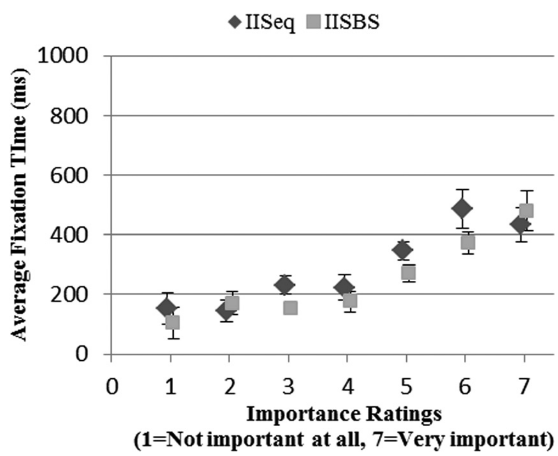

Fig. 10 Average fixation time spent on a feature increases with its importance rating (section II); error bars indicate ±1 standard errors

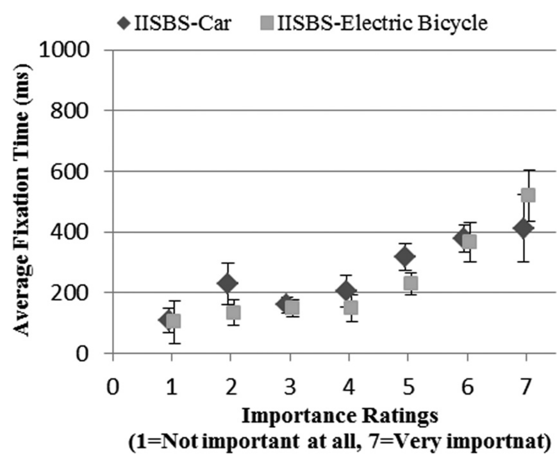

Fig. 12 Trends of average fixation time spent on a feature as its importance varies are similar for the car and the electric bicycle (section II—SBS condition); error bars indicate ±1 standard errors

such a binary dependent variable [43,44]. It models the odds of a size change to be noticed by a subject and estimates "the effects of independent variables on these odds [44]." The following equation represents the standard regression model:

$$P(y=1) = \frac{1}{1+\mathrm{e}^{-(\mathrm{Intercept}+\beta x)}} \tag{1}$$

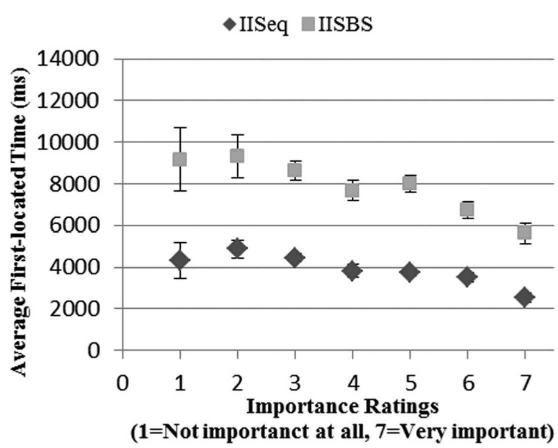

Fig. 11 Average first-located time on a feature decreases with its importance rating (section II); error bars indicate ±1 standard errors

In the model, y is a dummy variable, indicating whether a size change is noticed (1) or not (0); $P(\ldots)$ stands for the probability of an event; β is the coefficient for the independent variable x; intercept is the constant term in the model. The model is fit following the maximum likelihood principle. All gaze data measurements produce significant results when used as the independent variable (separately). Fixation time for a size-changed feature in product B is shown as an example; a summary of the fit is provided in Table 12. The small value of "Prob > ChiSq" demonstrates that the current model with gaze data as an independent variable is significantly better than a model with intercepts alone. The significant nonzero coefficient for the independent variable validates that fixation time for the size-changed feature in product B has a significant effect on differentiating noticeable and unnoticeable size changes. Similar results were obtained for data combined from pairs of size-changed features.

6.2 Results from Experiment Section III, Side-by-Side (IIISBS). The analysis in Sec. 6.1 was performed for data from the IIISBS condition. Detailed results are demonstrated in Table 11. Noticeable size changes in product B have significantly larger values of average fixation time and count than the unnoticeable ones. When considering the feature pair, noticeable-size-change feature pairs have significantly larger values of average fixation time and count than unnoticeable ones. Average Δ fixation time and count show no differences between the two sets of features that are compared. Hypotheses 3a and 3b are strongly supported here.

Table 10 Linear regressions show gaze data predict feature importance; the intercept is the constant term ($^{+}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.0001$)

			Intercept			Coefficient for x		
		Independent variable (x)	Estimate	Standard error	t	Estimate	Standard error	t
IISeq	Car	First-located time	5.90	0.48	12.30***	-3.50×10^{-4}	1.29×10^{-4}	−2.72*
	Electric bicycle	Fixation time	3.93	0.36	10.85***	1.83×10^{-3}	8.85×10^{-4}	2.07^{+}
	Car and electric bicycle combined	Fixation time	3.98	0.23	17.29***	1.80×10^{-3}	5.68×10^{-4}	3.16**
IISBS	Car	Percentage-fixation time	4.16	0.33	12.64***	0.21	0.12	1.79
	Electric bicycle	Percentage-fixation time	3.91	0.27	14.70***	0.34	0.09	3.86**
	Car and electric bicycle combined	Percentage-fixation time	3.96	0.19	20.50***	0.30	0.07	4.56**

Table 11 Proposition 2 is supported by results ($^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.0001$)

		Condition	
Associated AOI	Fixation metric	Seq Noticeable versus unnoticeable	SBS Noticeable versus unnoticeable
Size-changed feature in Product B	Time	1080 ms versus 400 ms***	906 ms versus 483 ms**
	Count	4.13 versus 1.63***	3.66 versus 2.38**
Feature pair with size change	Time	1843 ms versus 1110 ms*	1697 ms versus 919 ms***
	Count	7.22 versus 4.33**	7.25 versus 4.43**
Feature pair with size change	Δ time	319 ms versus −303 ms**	115 ms versus 47 ms
	Δ count	1.03 versus −1.06**	0.07 versus 0.32

Table 12 Logistic models show fixation time (and other gaze data not shown here) predict saliency of size changes (with the unnoticeable size change as reference level) ($^{**}p<0.01$, $^{***}p<0.0001$)

Seq			SBS		
Prob > ChiSq	<0.0001		Prob > ChiSq	<0.001	
Term	Estimate	Standard error	Term	Estimate	Standard error
Intercept	2.23***	0.27	Intercept	1.49***	0.22
β for "fixation time for size-changed feature in product B"	-8.97×10^{-4}**	2.41×10^{-4}	β for "fixation time for size-changed feature in product B"	-7.38×10^{-4}**	2.23×10^{-4}

The logistic regression in Sec. 6.1 was applied. All available measurements, except the Δ fixation time/count for a feature pair with size change, produce significant results when used as the independent variable separately. Fixation time for a size-changed feature in product B is shown as an example, in Table 12.

7 Discussion

The experiment results support both research propositions: (1) feature importance is correlated with gaze data in preference choices between two products and (2) saliency of size changes can be predicted by gaze data.

Hypotheses 1a–1d hold true in all situations tested here. During the processes of making preference decisions, there are positive correlations between feature importance and three types of gaze data: the fixation time, the percentage-time, and the count. Each shows a clear trend with increasing feature importance. These findings are consistent with the conclusions of Bettman et al.: People pay more attention to the information that has a larger weight in achieving the decision goal [23]. There is a negative correlation between feature importance and the feature's first-located time. As the feature importance rises, its first-located time decreases. The results from the SBS condition in section I did not show a clear decreasing trend. As mentioned in Sec. 5.1, this may be due to the limited data in that condition. In section I, each condition has only 12 subjects, while in section II, each condition has 36 subjects. In section I, each subject only evaluated two pairs for a product category, while in section II each subject evaluated three pairs for a product category. Evaluating more pairs is more likely to indicate the true measurement of how the subject evaluates a feature. The negative correlation found between feature importance and the feature's first-located time may suffer if people are evaluating a new product, an extremely unfamiliar product, or a product with new features, because in these cases people may not have clear ideas ready in mind about the feature importance and would look around the product stimulus in a less systematic manner to learn the product. But this concern is mitigated by the significant results found for the electric bicycle, which is less familiar (compared with the car) to the subjects. It may be that even for an unfamiliar product, after a one or two evaluation "burn-in," the negative correlation builds.

Hypotheses 1a–1d were analyzed both separately and together for the car and the electric bicycle stimuli. When analyzed separately, the absolute values of the correlations for the car range from 0.50 to 0.89 and those for the electric bicycle range from 0.51 to 0.81; when analyzed together, the absolute values of the correlations range from 0.51 to 0.74. Similar results are obtained in both cases (separate and together), indicating the potential robustness of the tested correlations across different levels of product familiarity. However, more products would need to be tested to confirm this robustness.

For hypotheses 1a–1d, type I error (falsely accepting the research hypothesis) could occur for a number of reasons. One potential cause for type I error is that an important product feature could attract more gaze attention for reasons other than its importance in preference decisions. For example, some features varied and other features did not. One might expect that the varied features attracted more eye attention because they were changing, and also because they were changing they "primed" subjects to exaggerate their importance ratings (as compared with unvaried features). However, the average importance of the features listed in Table 9 indicate that unvaried features remained important in decisions, see especially the car hood, although for the electric bicycle, varied features are on average more important in decisions than unvaried ones. Furthermore, the average importance ratings in Table 9 suggest that no one feature received a rating of "6" or "7" from all subjects, lending a useful variability to the data (although the bicycle seat is understandably rated as very important to many subjects).

Other reasons for increased gaze time could include a unique design that requires further mental processing; a feature that occupies a larger area of the screen; or a feature that stands out because it is unrealistic or unharmonious with the rest of the design. Arguments against these sources of error include the inclusion of familiar and unfamiliar stimuli, multiple feature variations, more and less harmonious features, and small and large features. Another potential cause of the type I error could be the participation of engineering students, who may have importance-oriented viewing strategies as compared with normal consumers. This error is mitigated by the fact that engineering students comprise only 21% of the sample population.

Type I error could also occur if the important features are in easy-to-locate positions, which could increase the correlation between first-located time and feature importance. This source of error is mitigated by the strong support of the hypotheses from three other types of gaze data. The use of product photographs as compared with real products could increase the correlation between first-located time and feature importance because real products contain rich information that can distract people and delay the located time for some important features. The differences in gaze data between photographic and real evaluations of products are likely substantial, as is true of many types of product preference data.

Type II error (falsely rejecting the research hypotheses) does not apply in this study because the results for hypotheses 1a–1d are strongly accepted. But when researchers apply the conclusions of this study to other experiments, type II error is a possibility. One factor to consider is the amount of time provided for stimulus evaluation; providing unlimited time can allow people to "invest" gaze in features that are not important to them. Another potential cause is placing special design efforts (like flashy designs, attractive colors, etc.) on some features, which could enable these features to attract unbalanced gaze attention.

This study proved that there is correlation between the importance of product features and associated gaze data, but they are not perfectly correlated. This suggests that including other variables in the regression would improve prediction of feature importance. Such factors could be (1) determining customers' use of particular features to extrapolate information missing from the decision, like price, comfort, brand, and safety information, and (2) recording customers' willingness-to-pay for designs.

Results from section III of the experiment show that the saliency of a size change can be predicted with gaze data. In both the IIISeq and IIISBS conditions, noticeable and unnoticeable size changes can be differentiated with gaze data. Hypotheses 2a, 2b, 4a, and 4b are strongly supported by the data of IIISeq. A noticeable size change in product B has significantly larger values of fixation time, count, Δ fixation time, and Δ count than an unnoticeable one. The noticeable size change attracts extra attention; while the unnoticeable one is ignored. Therefore, when a feature with an unnoticeable size change appears in two stimuli shown sequentially, its latter appearance (in product B) is considered as a repetition of the former one and attracts less attention. These findings are consistent with Refs. [1,40]; when a pair of stimuli is evaluated for preference decisions, their shared information is likely to be ignored in its second appearance. Even though not originally hypothesized, results show that a noticeable-size-change feature pair has significantly longer total fixation time and higher count.

In IIISeq, the break between the two stimuli weakens the memory of the first stimulus. It is possible that only abstract representations of the first stimulus remain while its details are overwritten by the second stimulus, according to the "overwriting" explanation for "change blindness" [45]. Even in this situation, there are some size changes that trump overwriting. This is worth further investigation, as exposure to minor product variations is a common situation for customers, for example, when viewing products on websites, such as Amazon.com.

As hypotheses 3a and 3b predict, in the IIISBS condition, a noticeable-size-change feature pair has significantly larger fixation time and count than an unnoticeable one. Even though not hypothesized, in the IIISBS condition, a noticeable size change in product B has significantly larger values of fixation time and count than an unnoticeable one. This may have resulted from the fact that gaze typically moves from left to right, and product B is on the right, thus mimicking sequential behavior even though the two stimuli are shown simultaneously. The Δ fixation time and count do not support a hypothesis for the IIISBS condition. This suggests that different viewing strategies are adopted in the Seq and SBS conditions. Presenting stimuli side-by-side enables pairwise comparisons between options, so it is less likely that the saliency of a size change will be identified with Δ fixation time or count.

For hypotheses 2a–4b (all accepted), there are a number of sources of type I error. One is the unlimited exposure time of the stimulus, which allows the subjects to carefully check for size changes—with a time limit, gaze patterns may change. Another is the use of digital photographs rather than real products. Subjects can stare at the almost identical images to identify size changes, which could enlarge the gaze attention difference between the noticeable and unnoticeable size changes as compared with reviewing real products. Real products allow for physical interactions, and thus other ways to identify size changes, such as holding small products against each other or measuring precisely. In this experimental setting (section III), only three features have design variants while size is also changing. This could be a source of type I error: As the number of varied features with design variants increases, the gaze attention spent on noticeable and unnoticeable size changes may become more similar.

Type II error does not apply for the testing of hypotheses 2a–4b in this study as these hypotheses are all accepted. But when researchers apply the conclusions of this study to other experiments, type II error is a possibility. Providing stimuli with size changes that are extremely obvious and require little gaze effort to notice would significantly decrease the gaze differences between the noticeable and unnoticeable size changes. Another potential cause could be including clear and constant reference-of-scale, such as a ruler, close to feature size variants as it would decrease the gaze efforts needed to notice the size change.

All hypotheses are tested under two general conditions, showing stimuli sequentially and side-by-side. Based on the results, the relationships found between feature importance and the gaze data are almost the same in the two conditions. This suggests that researchers studying the importance of product attributes using eye-tracking can present two stimuli at a time (such as in choice decisions) and reliably draw conclusions about attribute importance. It is not necessary to show product stimuli individually. But the two conditions have different results when the Δ fixation time and count are used to differentiate noticeable and unnoticeable size changes: The Δ fixation time and count are useful only in the Seq condition.

8 Conclusion

Results from this study indicate that product feature importance is correlated with a variety of gaze data (fixation time, percentage-fixation time, fixation count, and first-located time). The importance rating of a feature can be predicted by the gaze data using linear regression. These findings can help designers by providing a new approach to identify the importance of product features. They suggest that feature importance can be identified at the individual subject level in only three questions, without directly asking about feature importance. This could (a) significantly reduce the subject's mental burden associated with current methods, such as discrete choice analysis and complex rating schemes and (b) remove context effects caused by drawing attention to the purpose of the experiment (ascertaining feature importance), and instead let subjects evaluate products naturally. These directions will be pursued in future research. The study can also be furthered by investigating the effect of product viewing perspective, sizes, etc., and setting time constraints for viewing the stimuli on the results.

This study also demonstrates that gaze data can be used to identify whether or not someone notices a change in the size of a product feature. This can be used in a variety of ways, such as determining when manufacturing imperfections in the form of geometrical variations are noticeable. A considerable amount of time and money has been spent on manufacturing processes to ensure the quality appearance of products [46,47]. If one can predict how likely it is that an imperfection will be noticed, optimization analysis can be performed to reduce the manufacturing costs while maintaining the targeted quality appearance of products. This study could be furthered by developing a method to determine the just-noticeable threshold for size changes, which would be immediately useful to practicing designers.

The study's conclusions have some potential sources of type I error, as noted in Sec. 7. One area that should be noted, in particular, is the difference in the predictive power of gaze data in the evaluation of digital photographs or renderings versus real products. Studies involving real products are considerably more complex, with difficult-to-create stimuli, expensive eye-tracking equipment, and difficult-to-decipher gaze data in three dimensions. These challenges all suggest that for the time being, the usefulness of gaze data in understanding product evaluations is most readily applied to computer screen experiments.

Acknowledgment

The authors would like to thank the iMotions company for providing the ATTENTION TOOL software and the associated technical support; Dr. Frederick Lorenz, Dr. Heike Hofmann, and Kent Kroeger at Iowa State University for statistical advice; and colleagues in the IRIS Design Lab in Iowa State University for their valuable input.

References

[1] Sütterlin, B., Brunner, T. A., and Opwis, K., 2008, "Eye-Tracking the Cancellation and Focus Model for Preference Judgments," J. Exp. Soc. Psychol., **44**(3), pp. 904–911.
[2] Just, M. A., and Carpenter, P. A., 1976, "Eye Fixations and Cognitive Processes," Cognit. Psychol., **8**(4), pp. 441–480.
[3] Nielsen, J., and Pernice, K., 2010, *Eyetracking Web Usability*, New Riders, Berkeley, CA.
[4] Schiessl, M., Duda, S., Thölke, A., and Fischer, R., 2003, "Eye Tracking and Its Application in Usability and Media Research," MMI-Interact. J., **1**(6), pp. 41–50.
[5] Rayner, K., 1998, "Eye Movements in Reading and Information Processing: 20 Years of Research," Psychol. Bull., **124**(3), pp. 372–422.
[6] Findlay, J. M., and Gilchrist, I. D., 1998, "Eye Guidance and Visual Search," *Eye Guidance in Reading and Scene Perception*, G. Underwood, eds., Elsevier, Oxford.
[7] Glaholt, M. G., and Reingold, E. M., 2011, "Eye Movement Monitoring as a Process Tracing Methodology in Decision Making Research," J. Neurosci. Psychol. Econ., **4**(2), pp. 125–146.
[8] Rosbergen, E., Wedel, M., and Pieters, R., 1997, "Analyzing Visual Attention to Repeated Print Advertising Using Scanpath Theory," Research Institute SOM, University of Groningen, Groningen, Netherlands, Report No. 97B32.
[9] Pieters, R., and Warlop, L., 1999, "Visual Attention During Brand Choice: The Impact of Time Pressure and Task Motivation," Int. J. Res. Mark., **16**(1), pp. 1–16.
[10] Pieters, R., and Wedel, M., 2004, "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-size Effects," J. Mark., **68**(2), pp. 36–50.
[11] Jacob, R. J. K., and Karn, K. S., 2003, "Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises," *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyönä, R. Radach, and H. Deubel, eds., Elsevier, Amsterdam, pp. 573–605.
[12] Poole, A., and Ball, L. J., 2005, "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects," *Encyclopedia of Human Computer Interaction*, C. Ghaoui, eds., Idea Group Reference, PA.
[13] Konstantopoulos, P., 2009, "Investigating Drivers' Visual Search Strategies: Towards an Efficient Training Intervention," Ph.D. thesis, University of Nottingham, Nottingham, UK.
[14] Lohse, G. L., and Johnson, E. J., 1996, "A Comparison of Two Process Tracing Methods for Choice Tasks," Org. Behav. Human Decis. Process., **68**(1), pp. 28–43.
[15] Payne, J. W., Bettman, J. R., and Johnson, E. J., 1988, "Adaptive Strategy Selection in Decision Making," J. Exp. Psychol.: Learning, Memory, and Cognition, **14**(3), pp. 534–552.
[16] Dhar, R., Nowlis, S. M., and Sherman, S. J., 1999, "Comparison Effects on Preference Construction," J. Consum. Res., **26**(3), pp. 293–306.
[17] Gofman, A., Moskowitz, H. R., Fyrbjork, J., Moskowitz, D., and Mets, T., 2009, "Extending Rule Developing Experimentation to Perception of Food Packages With Eye Tracking," Open Food Sci. J., **3**, pp. 66–78.
[18] Koivunen, K., Kukkonen, S., Lahtinen, S., Rantala, H., and Sharmin, S., 2004, "Towards Deeper Understanding of How People Perceive Design in Products," *CADE2004 Web Proceedings of Computers in Art and Design Education Conference*, M. A. Eriksen., L. Malmborg, and J. Nielsen, eds., Sweden.
[19] Reid, T., MacDonald, E., and Du, P., 2013, "Impact of Product Design Representation on Customer Judgment," ASME J. Mech. Des., **135**(9), p. 091008.
[20] iMotions, 2013, retrieved on Dec. 20, 2013, http://imotionsglobal.com/
[21] Qualtrics, 2013, retrieved on Dec. 20, 2013, http://qualtrics.com/
[22] Duchowski, A. T., 2007, *Eye Tracking Methodology: Theory and Practice*, Springer, London.
[23] Bettman, J. R., Luce, M. F., and Payne, J. W., 1998, "Constructive Consumer Choice Processes," J. Consum. Res., **25**(3), pp. 187–217.
[24] Myers, J. H., and Alpert, M. I., 1968, "Determinant Buying Attitudes: Meaning and Measurement," J. Mark., **32**(4), pp. 13–20.
[25] Banks, S., 1950, "The Relationship Between Preference and Purchase of Brands," J. Mark., **15**(2), pp. 145–157.
[26] Orsborn, S., Cagan, J., and Boatwright, P., 2009, "Quantifying Aesthetic Form Preference in a Utility Function," ASME J. Mech. Des., **131**(6), p. 061001.
[27] MacDonald, E. F., Gonzalez, R., and Papalambros, P., 2009, "The Construction of Preferences for Crux and Sentinel Product Attributes," J. Eng. Des., **20**(6), pp. 609–626.
[28] Jaccard, J., Brinberg, D., and Ackerman, L., 1986, "Assessing Attribute Importance: A Comparison of Six Methods," J. Consum. Res., **12**(4), pp. 463–468.
[29] Reisen, N., Hoffrage, U., and Mast, F. W., 2008, "Identifying Decision Strategies in a Consumer Choice Situation," Judgm. Decis. Making, **3**(8), pp. 641–658.
[30] Warell, A., and Nåbo, M., 2002, "Handling Product Identity and Form Development Issues in Design Management Using Design Format Modeling," Proceedings of DMI 2002, the 11th International Forum on Design Management Research and Education Strategies, Resources & Tools for Design management Leadership, Northeastern University, Boston.
[31] Jaccard, J., and King, G. W., 1977, "The Relation Between Behavioral Intentions and Beliefs: A Probabilistic Model," Human Commun. Res., **3**(4), pp. 326–334.
[32] Schkade, D., and Johnson, E., 1989, "Cognitive Processes in Preference Reversals," Org. Behav. Human Decis. Process., **44**(2), pp. 203–231.
[33] Michalek, J. J., Ebbes, P., Adiguzel, F., Feinberg, F. M., and Papalambros, P. Y., 2011, "Enhancing Marketing With Engineering: Optimal Heterogeneous Markets," Int. J. Res. Mark., **28**(1), pp. 1–12.
[34] Coelho do Vale, R., Pieters, R., and Zeelenberg, M., 2008, "Flying Under the Radar: Perverse Package Size Effects on Consumption Self-Regulation," J. Consum. Res., **35**(3), pp. 380–390.
[35] Chandon, P., and Ordabayeva, N., 2008, "Downsize in 3D, Supersize in 1D: Effects of the Dimensionality of Package and Portion Size Changes on Size Estimations, Consumption, and Quantity Discount Expectations," INSEAD Working Papers Collection, Paper No. 2008/46/MKT.
[36] Yang, S., and Raghubir, P., 2005, "Can Bottles Speak Volumes? The Effect of Package Shape on How Much to Buy," J. Retailing, **81**(4), pp. 86–96.

[37] Krider, R. E., Raghubir, P., and Krishna, A., 2001, "Pizzas: π or Square? Psychophysical Biases in Area Comparisons," Mark. Sci., **20**(4), pp. 405–425.
[38] Chevrolet, 2013, retrieved on Dec. 20, 2013, http://www.chevrolet.com/flash.html
[39] Global-Tradekey, 2013, retrieved on Apr. 21, 2013, http://www.global-tradekey.com/company/G02189/C64201.htm
[40] Houston, D. A., and Sherman, S. J., 1995, "Cancellation and Focus: The Role of Shared and Unique Features in the Choice Process," J. Exp. Soc. Psychol., **31**(4), pp. 357–378.
[41] Pernice, K., and Nielsen, J., 2009, "Eyetracking Methodology: How to Conduct and Evaluate Usability Studies Using Eyetracking," Nielsen Norman Group, Fremont, CA.
[42] Paul, R. K., "Multicollinearity: Causes, Effects and Remedies," IASRI, New Delhi.
[43] Agresti, A., 1996, *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Inc., New York.
[44] O'Connell, A. A., 2006, *Logistic Regression Models for Ordinal Response Variables*, SAGE Publications, Inc, Thousand Oaks, CA.
[45] Simons, D. J., 2000, "Current Approaches to Change Blindness," Visual Cognit., **7**(1–3), pp. 1–15.
[46] Forslund, K., Kero, T., and Söderberg, R., 2009, "Appearance FMEA: A Method for Appearance Quality Evaluation of Early Design Concepts," Proceedings of the ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, San Diego, CA, pp. 217–225.
[47] Söderberg, R., Wickman, C., and Lindkvist, L., 2008, "Improving Decision Making by Simulating and Visualizing Geometrical Variation in Non-Rigid Assemblies," CIRP Ann., **57**(1), pp. 175–178.