ORIGINAL RESEARCH

Quantitative evaluation of sensitivity in confidential car exterior design

Takumi Kato*[1], Kazuhiko Tsuda[2]

[1]*Business Analytics Section, Business Development Supervisory Unit, Honda Motor Co.,Ltd. Tokyo, Japan*
[2]*Graduate School of Business Sciences, University of Tsukuba, Tokyo, Japan*

ABSTRACT

In recent years, the manufacturing industry has seen a shift in competition from performance, which can easily be evaluated numerically, to design which much more challenging to express numerically. The rise of companies that focus on design, such as Apple, Samsung and IKEA, is remarkable. However, design presents two challenges for the manufacturing industry. First, the sensory aspect of design is challenging to evaluate quantitatively, and unified evaluation indicators are not yet defined. Second, confidentiality of product design. In many cases, the design is kept in confidence within the companies, so it is often hesitated to investigate large customers. The above two problems increase the influence of the evaluator's experience and cause a situation that it is challenging to create a design desired by the customer. Therefore, the present study aims to enable inexpensive quantitative evaluation of automobile exterior design while maintaining confidentiality. We propose a technique that uses a convolutional neural network to link features extracted from accumulated design images to the sensitivity extracted from the customer's voice. This is then used to quantitatively evaluate an input image.

Key Words: Emotional value, Design engineering, Deep learning, Natural language processing

1. INTRODUCTION

Japanese companies have manufactured world-leading weapons and other high-technology products to high standards of quality and performance. Despite maintaining these high technological standards, strong sales performance has not necessarily followed, and many manufacturing companies are now struggling. In contrast, companies that emphasize design, such as Apple, Samsung, IKEA, and Dyson, are performing well.[1]

As technological development progresses, product performance and quality exceed customer needs and overshoot occurs. Overshoot occurs when excessive competition leads to improvements in function and specification that do not affect the utility to customers. Therefore, in the market for the product, commoditization and price competition occur. In addition, further investment in technological development is ineffective for the manufacturers of such products.[1]

Performance and quality of products can be expressed numerically, and these specifications can be easily published. These are called functional values. To contrast with functional value, emotional value can be evaluated by human sensitivity, such as concept, customer experience, or design. Once overshoot has occurred in functional value, it is necessary to improve emotional value to avoid price competition of goods, to add value to products, and to maintain market competitiveness.

---

*Correspondence: Takumi Kato; Email: Takumi_Kato@hm.honda.co.jp; Address: Business Analytics Section, Business Development Supervisory Unit, Honda Motor Co.,Ltd. 2-1-1 Minami-Aoyama, Minato-ku, Tokyo 107-8556, Japan.

The most important factor in quickly judging the appeal of a product is the design, which customers can assess in as little as in 0.05 s.[2] Companies that emphasize emotional value, especially design, such as IKEA, Apple, Samsung, and Dyson, are emerging in the global market and rapidly taking market share from Japanese companies.

There are three main problems when assessing whether a design is good or bad. The first problem is that it is challenging to quantitatively evaluate design because accounting costs of design are not defined, and a unified evaluation index has not been defined yet. In Kansei engineering, predefined viewpoints are typically evaluated using five- to seven-point Likert scales according to the semantic differential (SD) method developed by Osgood.[3] The second problem is maintaining the confidentiality of a product design. Designs are often highly guarded secrets for companies, and it is challenging to gather customer opinions on a design while maintaining this secrecy. Therefore, internal expert evaluators often assess designs using methods such as the SD method. Therefore, instead of large-scale surveys, neuroscience approach such as functional magnetic resonance imaging (fMRI) and Electroencephalograph (EEG) are getting popular because we can understand the sensitivity precisely with evaluation of few subjects. Li et al. proposed an evaluation method for uniqueness of product appearance by EEG equipment and an eye-tracking device to record a subject's brain activity and eye-gaze data.[4] However, this leads to the third problem: that dedicated facilities are required and subject's load is high, resulting in high costs. Thus, it is still challenging to evaluate many times at the site of design.

Therefore, in the present study, we propose a method to evaluate the sensitivity of new car exterior designs using accumulated designs and the customer's opinions without revealing the design to be evaluated.

2. EVALUATION OF PRODUCT DESIGN

2.1 Understanding the characteristics and impressions of design

Ryoke et al. prepared 30 samples for the survey.[5] These were evaluated using the SD method using 26 pairs of adjectives with a seven-point Likert scale for each pair. The obtained results allowed the authors to quantify the association between the adjectives and the sample. Sato considered the construction of design clusters and their characteristics based on the impression evaluation results for the front grille design of an automobile.[6] Ten front grille designs were investigated with respect to 14 impressions such as "cute" and "fashionable". The results were analyzed using quantification theory class III. A cluster similarity map was constructed by cluster analysis, as shown in Figure 1. As an example of how the map is used, group 4 is the closest to "vintage", and designs close to the impressions "natural" or "adult" have a common feature in that they use wood grain.

These approaches are suitable for understanding the emotional aspects of design characteristics and impressions. However, evaluations that maintain the confidentiality of new designs have not been investigated.
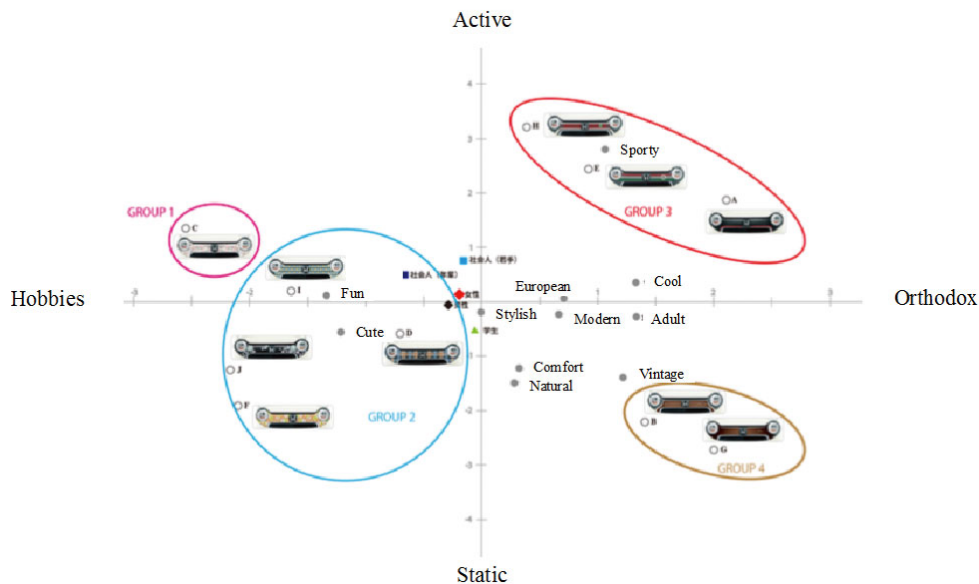

Figure 1. Similarity map of front grille designs

2.2 Causal relationship between morphological elements and images by rough set analysis

Unlike specifications in which each function can be numerically evaluated, design needs to be evaluated as a comprehensive whole with many interrelated attributes. Thus, it is important to understand not only the trend but also factors including combination synergy and offset effect. Therefore, rough set theory can be used to acquire valuable information on emotional design.

Inoue et al. revealed the causal relationship between form elements, attributes and images using rough sets for the front mask design of cars.[7] The form factor and attribute "shape of air intake" was evaluated as a negative influence by correspondence analysis. However, rough set analysis it was shown that a combination of "the internal structure of the light stands out" and "the grill is thin" contributes to a positive influence.

However, the rough set approach requires a huge number of validations when covering all combinations of form elements and attributes, which is unrealistic. In the above study, two to three attributes are prepared for each of the eight elements, each of which has 4,374 combinations, but these were evaluated with only 17 samples. In actual designs, the numbers of form elements and attributes become enormous, so it is difficult to efficiently extract meaningful results. In other words, we should explore methods in which humans do not define feature quantities of design elements. Nevertheless, there are still few proposals for the method of customer's sensitivity in product design.

Li et al. proposed a method to optimize the design of product form based on supporting vector and artificial fish swarm algorithm in order to meet customers' demand for perceptual cognition of product form. However, in this verification targeting refrigerators, 10 critical control points have been set, and it is necessary for humans to designate feature quanti-

ties.[8] Yang et al. validated that multiple regression analysis based on design elements and perceptual image helped designers design car console which could meet users' psychological needs. However, for example, navigation radio panels are three types of square, polygon, and arc line, so that humans have to define the design elements as well.[9]

2.3 Feature value extraction by deep learning

It is unrealistic for humans to set all feature quantities. Therefore, recent studies have used deep learning for extracting features from images. A competitor in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015, an international competition for computer object recognition, reached a human-level accuracy of feature identification, exceeding 5.1%.[10] Schroff extracted features from images of human faces with 22 layers of deep learning and succeeded in identification of individuals with an accuracy of 99.63%.[11] Karpathy et al. extracted features and behaviors from photographs and used this information generated descriptive texts.[12] Mansimov et al. extracted information from sentences, including adjectives describing colors and states, and generated images.[13] However, in these studies, objects, colors and actions that are not blurred by humans are considered objects, and they do not have features that are not clearly defined. Under these circumstances, Shimoda et al realized the perception of texture of things.[14]

In the Japanese language, there are a lot of onomatopoeic words that expresses the state of things. For example, the word *pom-pom* in Japanese is used to describe something that is round and bulging. By extracting vocabulary features using a convolutional neural network (CNN), the images of Figure 2 are obtained. The accuracy of the CNN in identifying images that match onomatopoeic words was 84.6%.

Although humans were released from work that defines the feature quantity by deep learning, there are few proposals for estimating customer's sensitivity in product design.

Figure 2. Image example of flowers obtained by onomatopoeic terms

3. Methods

3.1 Outline of this research

In this research, we proposed a technique to quantitatively evaluate by linking the design features extracted from the

design images with CNN and the sensitivity extracted from the customer's voice. As a result, not only can the evaluation be carried out inexpensively but also the confidentiality is maintained without investigating to the customer. The eval-

uation of sensitivity in this research is to estimate the ratio of the customer's voice to the sensitivity category defined in Section 3.6.

We did not use the SD method for gathering customer sensitivity data for three reasons. First, the purpose of this study is to understand customer opinions without bias, so words should not be prepared in advance and presented to respondents. Second, to maintain confidentiality, the target feeling words or images should not be presented to the respondents. The third reason is that it is possible to understand the customer with less bias, with the method of obtaining answers without presenting options to respondents. Indeed, Kardes et al. confirmed that respondents overestimate when presenting choices.[15]

Table 1. Overview of customer sensitivity survey

Item	Content
Country	Japan
Period	3/2017
Car	Extracted the top 75 cars in descending order of number of surveys
Sample size	Total: 216196 Target 75 cars: 97492
Sample extraction	Random sampling according to population dynamics (age, place of residence) of Japan for those who bought a car at January to December 2016
Survey method	Web survey
Question	− Gender, Age, Prefecture, Occupation, Marriage − Please answer the manufacturer and car name you bought. − What kind of charm do you feel about the design of this car? Please fill in freely. (FA)

3.2 Data used in this research

The customer sensitivity data for exterior designs used in this research were collected by web survey for 216,196 people. These people were sampled at random in accordance with demographic dynamics of Japan from a database of people who bought a car between January and December 2016. As shown in Table 1, the questions covered the demographics of the respondents, the manufacturer and model of the purchased car, and the respondents' opinions and impressions of the car design.

As shown in Table 2, the target cars are the top 75 vehicles for which we obtained more than 100 records per car. The "Text Freq" column represents the number of data record obtained, so total data records amount to 97,492. The "body type" column represents the shape of the 8 types shown in Figure 3. Here, "Kei" refers to a category of micro car sold primarily in Japan. Kei cars have an engine displacement of

660 cc or less. "Kei_Tall" refers to kei cars that are taller than standard, a subcategory that grown rapidly in popularity in recent years in Japan. Of the 75 cars included, 60 cars are used as learning objects and 15 cars as validation targets for five-fold cross validation of the proposed method. Table 3 shows the target of each validation. To make the targets of each validation a homogeneous group, body types are allocated equally.
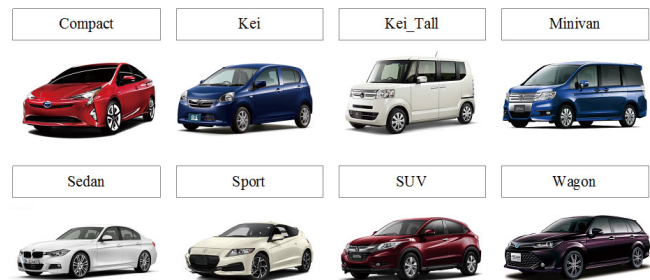

Figure 3. Examples of each body type

We prepared 30 images per car for a total of 2,250 design images to be used in the study. The design images are all front 3/4 views showing the front and side of the car. This viewing angle allows the viewer to see the overall design characteristics in a single view, and is commonly used by automobile manufacturers on websites and in catalogs.

3.3 Procedure of this research

As shown in Figure 4, the procedure that the present study followed includes 4 phases. Details of each phase are explained in the following subsections.
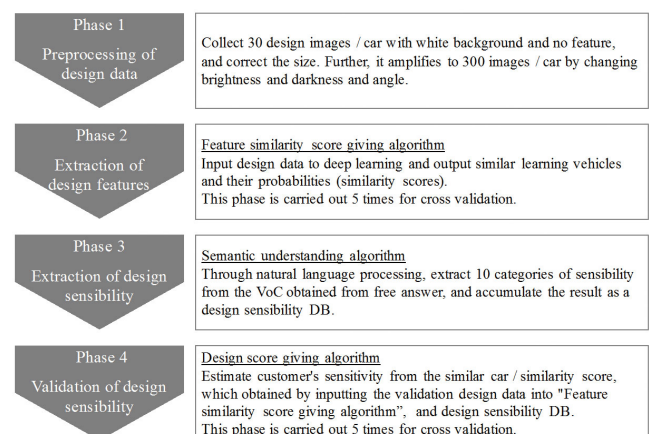

Figure 4. Research procedure

3.3.1 *Phase 1: Preprocessing design data*

Phase 1 involves preprocessing the design data. First, images of the car to be learned are collected for each car. Human

volunteers select 30 images with plain color backgrounds for each car. At that time, it is confirmed that the car in the image is a model that was on sale during the target period between January and December 2016. These images are compressed to 256 × 256 pixels. Using rotations and contrast changes, the number of images is increased ten-fold to 300 images per car.

Table 2. Target cars and volume of customer sensitivity data

No	Validation 1				Validation 2				Validation 3			
	Maker	Car	Body Type	Text Freq	Maker	Car	Body Type	Text Freq	Maker	Car	Body Type	Text Freq
1	MINI	COOPER	Compact	147	BMW	3SERIES	Sedan	641	Mazda	AXELA	Sedan	761
2	Honda	CRZ	Sport	190	Toyota	ALPHARD	Minivan	959	MB	C_CLASS	Sedan	682
3	Nissan	DAYS	Kei_Tall	941	Suzuki	ALTO_LAPIN	Kei	915	Toyota	COROLLA_FIELDER	Wagon	945
4	Mitsubishi	EKWGN	Kei	680	Toyota	AQUA	Compact	5,153	Nissan	ELGRAND	Minivan	564
5	Toyota	ESTIMA	Minivan	1,300	Mazda	ATENZA	Sedan	198	Honda	FIT	Compact	6,200
6	Honda	INSIGHT	Sedan	301	Nissan	CUBE	Kei	913	Honda	FREED	Minivan	3,370
7	Toyota	ISIS	Minivan	798	Mazda	CX5	SUV	924	VW	GOLF	Compact	1,055
8	Suzuki	JIMNY	Kei	429	Mazda	DEMIO	Compact	2,114	Toyota	HARRIER	SUV	582
9	Honda	NWGN_CUSTOM	Kei_Tall	258	Subaru	FORESTER	SUV	880	Suzuki	HUSTLER	Kei	1,120
10	VW	POLO	Compact	523	Daihatsu	MIRA_COCOA	Kei	731	Honda	LIFE	Kei	963
11	Toyota	PORTE	Compact	680	Daihatsu	MOVE	Kei_Tall	2,438	Nissan	MARCH	Compact	780
12	Mazda	PREMACY	Minivan	1,120	Honda	NBOX	Kei_Tall	2,695	Daihatsu	MIRA_ES	Kei	2,028
13	Suzuki	SOLIO	Kei_Tall	840	Honda	ODYSSEY	Minivan	912	Daihatsu	MOVE_CUSTOM	Kei_Tall	1,244
14	Subaru	STELLA	Kei	597	Nissan	SERENA	Minivan	3,707	Toyota	NOAH	Minivan	1,879
15	Honda	STREAM	Wagon	296	Toyota	SPADE	Compact	888	Suzuki	SPACIA	Kei_Tall	1,328
No	Validation 4				Validation 5							
	Maker	Car	Body Type	Text Freq	Maker	Car	Body Type	Text Freq				
1	Lexus	CT200H	Sedan	230	Fiat	Fiat500	Compact	124				
2	Subaru	IMPREZA	Sedan	914	Subaru	LEGACY	Sedan	756				
3	Nissan	JUKE	SUV	485	Toyota	MARKX	Sedan	497				
4	Nissan	MOCO	Kei	1,329	Honda	NONE	Kei	1,036				
5	Suzuki	MRWGN	Kei	385	Toyota	RACTIS	Compact	908				
6	Honda	NBOX_CUSTOM	Kei_Tall	1,548	Nissan	ROOX	Kei_Tall	775				
7	Nissan	NOTE	Compact	2,862	Honda	SHUTTLE	Wagon	705				
8	Toyota	PASSO	Compact	1,241	Honda	STEPWGN	Minivan	1,999				
9	Toyota	PRIUS	Compact	4,832	Suzuki	SWIFT	Kei	876				
10	Toyota	PRIUS_a	Wagon	1,529	Daihatsu	TANTO_CUSTOM	Kei_Tall	1,901				
11	Toyota	SIENTA	Minivan	1,351	Toyota	VELLFIRE	Minivan	1,224				
12	Honda	STEPWGN_SPADA	Minivan	629	Toyota	VITZ	Compact	2,265				
13	Daihatsu	TANTO	Kei_Tall	2,571	Toyota	VOXY	Minivan	2,576				
14	Honda	VEZEL	SUV	825	Suzuki	WGNR_STINGRAY	Kei	1,062				
15	Suzuki	WGNR	Kei	2,302	Nissan	XTRAIL	SUV	1,086				

Table 3. Assignment of body type in cross validation

Type	Total	Validation 1	Validation 2	Validation 3	Validation 4	Validation 5
Compact	15	3	3	3	3	3
Kei	15	3	3	3	3	3
Kei_Tall	11	3	2	2	2	2
Minivan	14	3	3	3	2	3
Sedan	9	1	2	2	2	2
Sport	1	1	0	0	0	0
SUV	6	0	2	1	2	1
Wagon	4	1	0	1	1	1
Total	75	15	15	15	15	15

3.3.2 *Phase 2: Extraction of design features*

The design data are input into the CNN, and the 60 cars to be learned are categorized according to class and assigned feature similarity scores. The algorithms learned by the CNN during this process are called the feature similarity scoring algorithms. AlexNet, which identified 1,000 classes of cars in ILSVRC 2012, is composed of eight layers.[16] Because there are 60 types of vehicles to be learned per each validation in this study, fewer layers are necessary. As shown in

Figure 5, the feature similarity scoring algorithm, constructed with MXNet, consists of an input layer, three convolutional layers, a fully connected layer, and a soft max layer. In each convolutional layer, introduction of dropout, in addition to convolution or activation, improves generalization performance so as not to over fit. The unit selection probability of dropout is set to $p = .5$, at which level the regularization effect is maximized. The activation function adopts rectified linear units, which have a gradient that propagates without attenuation for positive units and converges quickly. Finally, the similarity scores for the 60 vehicles learned by the soft max layer are output. This phase is carried out 5 times for cross validation.
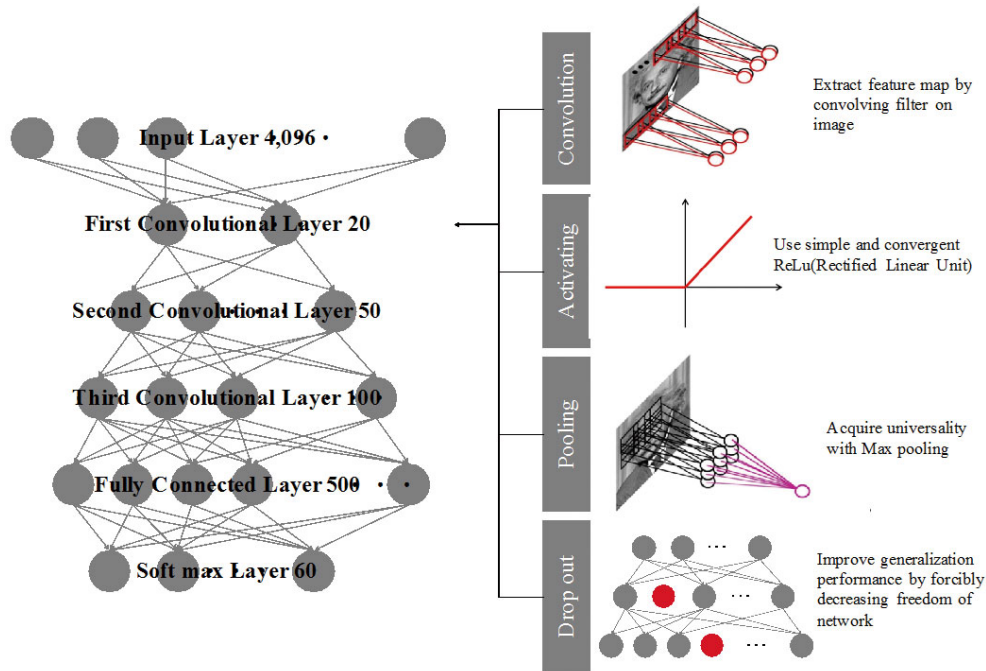

Figure 5. Structure of the feature similarity scoring algorithm

3.3.3 Phase 3: Extraction of design sensitivity

Phase 3 extracts the sensitivity category for the design of the learning object. The customer opinion obtained by the free answer is processed using by natural language processing and the sensitivity category is input into the design sensitivity database. To achieve this, a sensitivity dictionary is compiled and a semantic understanding algorithm constructed.

Table 4. Sensitivity categories and word examples

Sensitivity category	Word example 1	Word example 2	Word example 3	Sensitivity category	Word example 1	Word example 2	Word example 3
D01_Cool	Cool	Handsome	Fearless	D06_Individual	Individual	Unique	Stimulating
D02_Charming	Charming	Cute	Lovely	D07_Simple	Simple	Monotone	Basic
D03_Sophistication	Sophistication	Stylish	Beautiful	D09_Masculine	Masculine	Wild	Solidly
D04_Highclass	Highclass	Luxury	Gorgeous	D09_Family	Family	Casual	Familiarity
D05_Sporty	Sporty	Dynamic	Energetic	D10_Traditional	Traditional	Classic	Retro

First, 97,492 records of text data related to the attractiveness of design obtained by free answer are morphologically analyzed by the open source engine MeCab. The number of emotion categories increases tremendously if words that occur only once or twice are subject to dictionary registration. Therefore, only adjectives or adverbs that occur three or more times are selected, leaving 547 items. Words used to describe products differ depending on the product, and such terms for car exterior design have not been systematized. Therefore, referring to previous research,[6] categories are defined which are considered important for car design. In this case, shape expressions such as "round" or "large", functional expressions such as "convenience" or "safety", price expressions such as "expensive" or "cheap", direct expressions such as

"good" or "wonderful", and other irrelevant items are excluded. Also, since the question asked for positive responses, negative expressions are excluded. From among the 547 total items, 169 words describing emotion related to the design are extracted and registered in the dictionary.

The majority of the free answers were single sentences or phrases, and at most two or three sentences. Furthermore, because the descriptions tend to contain duplicate words and phrases, the number of registered words is not very large. As shown in Figure 6, just 33 words account for 80% of the total appearances. The 169 extracted emotion words were arranged into 10 sensitivity categories, as shown in Table 4.
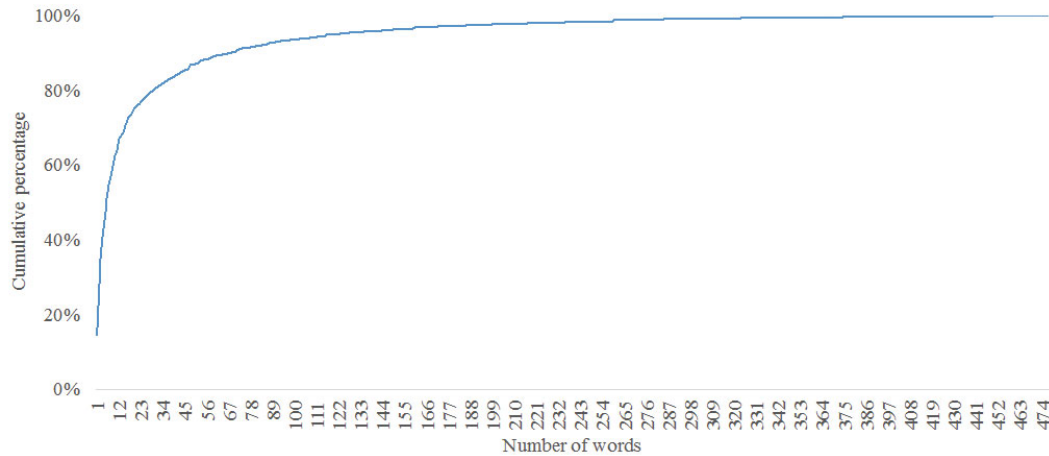

Figure 6. Number of words and cumulative appearance ratio

The semantic understanding algorithm is used to parse the customer opinions. Techniques to understand spoken words can be divided broadly into rule-based methods and statistical methods. Tsuchiya et al. judged emotions according to the rule of 8,024.[17] Harada et al. used statistical semantic matching of, for example, LDA to understand spoken words.[18] Statistical processing is difficult when there are few input words, such as with our free answers. In addition, we are focused on expressions that describe designs, so we do not need a large database. Therefore, we adopt the rule-based approach.

We constructed four rule functions for the semantic understanding algorithm, as shown in Table 5 along with examples.

Table 5. Examples of semantic understanding result

No	Algorithm	Example	Target car	Sensitivity category	Score
1	Dependency	It is not stylish, but I feel familiarity.	GOLF	D03_Sophistication	0
				D09_Family	1
2	Multiple negation	The surroundings say that the Prius is not cool, but I can not agree at all.	PRIUS	D01_Cool	1
3	Affirmative doubt	Which is a luxury car, Toyota Alphard or Benz Viano? Is it alphard?	ALPHARD	D04_Highclass	1
4	Comparison	Mercedes' C-class is cooler than the BMW 3 Series at all.	3 SERIES	D01_Cool	0

The first function, "dependency", is focused on the affirmation or negation of the term stored in the dictionary, and when a term is negated, no score is given. For example, "stylish" and "familiarity" are detected. Because "stylish" appears with "not" it is negated no score is given. In contrast, "familiarity" is given a score.

In the second function, "multiple negation", when multiple negative words are detected, if it is judged positive overall, a score is given. Sixty-five negative words were detected, including "disgusting", "disappointed", "bad", and "terrible". For example, the term "cool" is negated twice; this is judged to be positive, so a score is given.

The third function, "affirmative doubt", gives scores to question sentences that are seeking confirmation. Four words are registered. For example, a score is given for "luxury" because the question is asking for agreement that the target

car (ALPHARD) is a luxury model.

The fourth function, "comparison", understands the dependency around comparison words, and does not give a score in the context where another car or an older model prevails. For example, no score is given for "cool" because a different car (C-class) is noted as cooler than the target car (3 SERIES).

Natural language processing consisting of the above dic-

tionary, morphological analysis, syntax analysis, and rules is called semantic understanding algorithm which is implemented in Python. The composition ratio of the sensitivity categories accumulated for each car for each sensitivity category by this algorithm is called the sensitivity database. Table 6 shows examples of four cars. By car, we can see that the distribution of sensitivity voices changes greatly.

Table 6. Examples from the design sensitivity database

Sensitivity category	JUKE		MIRA_COCOA		VELLFIRE		VOXY	
	Freq	Rate	Freq	Rate	Freq	Rate	Freq	Rate
D01_Cool	22	18.33%	2	0.53%	69	33.66%	127	64.47%
D02_Charming	23	19.17%	327	87.20%	6	2.93%	14	7.11%
D03_Sophistication	8	6.67%	16	4.27%	4	1.95%	0	0.00%
D04_Highclass	4	3.33%	2	0.53%	103	50.24%	8	4.06%
D05_Sporty	2	1.67%	0	0.00%	3	1.46%	6	3.05%
D06_Individual	53	44.17%	8	2.13%	12	5.85%	9	4.57%
D07_Simple	1	0.83%	12	3.20%	1	0.49%	6	3.05%
D08_Masculine	7	5.83%	1	0.27%	6	2.93%	10	5.08%
D09_Family	0	0.00%	0	0.00%	1	0.49%	17	8.63%
D10_Traditional	0	0.00%	7	1.87%	0	0.00%	0	0.00%
Total	120	100.00%	375	100.00%	205	100.00%	197	100.00%

Because the semantic understanding algorithm in Phase 3 was commonly used for each cross validation, implementation was done only once. The results of the semantic understanding algorithm were input into the database and extracted for each validation.

3.3.4 *Phase 4: Evaluation and validation of design sensitivity*

In Phase 4, the composition ratio of the sensitivity category of the vehicles to be validated are estimated and the accuracy is validated. An estimated value of the customer sensitivity of the validation target car is calculated from the similarity score obtained in Phase 2 and the sensitivity database obtained in Phase 3.

The design scoring algorithm is shown in equation (1):

$$Score_{i,j} = \frac{\sum_{k=1}^{60}(w_{i,k} * Score_{k,j})}{\sum_{j=1}^{10} Score_{i,j}} \tag{1}$$

where, i is the target car to be validated, j is the sensitivity category, k is the target car to be learned, and w is the similarity score. The sensitivity estimated by the above formula is compared with the customer sensitivity, and its accuracy is evaluated by mean absolute percentage error (MAPE).

4. RESULTS AND DISCUSSION

4.1 Algorithm accuracy

In this study, the feature similarity scoring algorithm for extracting design features (Phase 2) and the semantic understanding algorithm for customer design sensitivity (Phase 3) are centralized. We carried out accuracy validation for both algorithms.

As shown in Figure 7, all five repetitions of the validation of Phase 2 converge at about 95% accuracy. Although the learning data included 60 models, the accuracy of 95% was achieved after only 30 models were used as learning data. As shown in Table 7, we confirmed an accuracy level of over 93% in all models. Thus, we confirmed the accuracy of the feature similarity scoring algorithm.

Table 7. Validation accuracy of the convolutional neural network

	Validation 1	Validation 2	Validation 3	Validation 4	Validation 5
Correct	59	59	57	56	58
Miss	1	1	3	4	2
Correct Rate	98.33%	98.33%	95.00%	93.33%	96.67%

The semantic understanding algorithm was validated using the indicators of precision and recall. Both of these indicators were verified with data of 10 records in each sensitivity

category, 100 records in total. As shown in Table 8, we confirmed a precision accuracy of 86.00% and a recall accuracy of 94.00%. The process of extracting customer sensitivity by

the semantic understanding algorithm was conducted only once, regardless of cross validation; therefore, Precision and Recall were also validated only once.
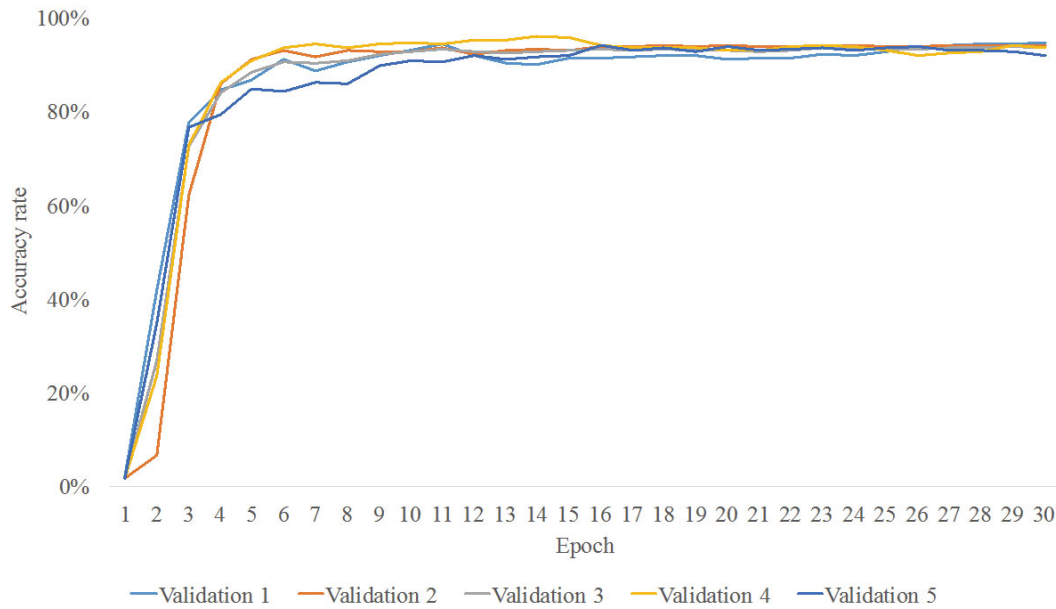

Figure 7. Accuracy rate of convolutional neural network learning

Table 8. Precision and recall

Sensitivity category	Precision		Recall	
	Total	Correct	Total	Correct
D01_Cool	10	10	10	10
D02_Charming	10	10	10	8
D03_Sophistication	10	9	10	9
D04_Highclass	10	8	10	10
D05_Sporty	10	8	10	10
D06_Individual	10	7	10	10
D07_Simple	10	8	10	9
D08_Masculine	10	9	10	8
D09_Family	10	9	10	10
D10_Traditional	10	8	10	10
Total	100	86	100	94

Example sentences that failed semantic comprehension are shown in Table 9. There are two examples of where the precision declined. In the first example, the text "feel the status" is not related to design. In the second example, "rare" is not referring to the car but to the respondent, giving a false positive score. There are also two examples in which the recall rate decreased. In the first example, "lovely" is negated by "although" so no score is given. However, "although" is not necessarily negative and this should have received a score. In the second example, "too cute" is excessive, considered as a negation and not given a score, but it does have a positive

meaning. To correct such failures, it is necessary to continue improving dictionaries and rules while taking side effects into account.

Table 9. Example sentences that failed semantic comprehension

Indicator	Text	Sensitivity category	Score
Precision	I feel the <u>status</u> on the ride.	D04_Highclass	1
	It's <u>rare</u> for me that I thought it cute.	D06_Individual	1
Recall	<u>Although</u> it is <u>lovely</u> looking from far, there is a heavy feeling when looking at nearby.	D02_Charming	0
	The exterior is too <u>cute</u>!	D02_Charming	0

4.2 Sensitivity estimation and accuracy validation

We evaluated the difference between the composition ratio (estimated value) of the sensitivity category obtained by equation (1) shown in Phase 4 and the composition ratio (estimated value) of the customer sensitivity category shown in Phase 3. In the example shown in Table 10, the MAPE of the Mini Cooper is 4.87%. This verification was conducted with five-fold cross validations.

Table 11 shows the MAPE for each sensitivity category obtained by cross validation, and the overall MAPE for each of the 10 categories. The mean MAPE among all categories is 5.26%. Evaluations for all sensitivity categories are within 10%. The MAPE for the 'Charming' category is 9.39%,

which is the most inaccurate. The sensitivity 'Charming' is thought to be difficult to estimate because it has various shapes such as round, square and small. The largest differ-ence between the maximum value and the minimum value is 4.20%, for the 'High class' category. The reason for this is that 'High class' is not specific to a particular car body type.

Table 10. An example of mean absolute percentage error for the Mini Cooper

| | D01 | D02 | D03 | D04 | D05 | D06 | D07 | D08 | D09 | D10 | Total |
	Cool	Charming	Sophistication	Highclass	Sporty	Individual	Simple	Masculine	Family	Traditional	
True value	27.90%	17.36%	9.81%	16.47%	6.50%	9.70%	6.02%	3.48%	2.06%	0.68%	######
Estimated value	27.59%	24.14%	0.00%	10.34%	10.34%	17.24%	0.00%	6.90%	0.00%	3.45%	######
MAPE	0.32%	6.77%	9.81%	6.13%	3.85%	7.54%	6.02%	3.41%	2.06%	2.77%	**4.87%**

Table 11. Mean absolute percentage error by each sensitivity category obtained by five-fold cross validation

| Validation | D01 | D02 | D03 | D04 | D05 | D06 | D07 | D08 | D09 | D10 | Average |
	Cool	Charming	Sophistication	Highclass	Sporty	Individual	Simple	Masculine	Family	Traditional	
1	8.70%	10.98%	6.07%	9.24%	5.89%	4.52%	4.17%	1.79%	3.63%	3.63%	5.86%
2	7.36%	8.96%	4.64%	7.59%	3.16%	5.91%	3.41%	2.19%	2.32%	1.05%	4.66%
3	9.88%	10.21%	2.94%	11.44%	3.84%	4.51%	3.96%	3.09%	3.09%	1.01%	5.85%
4	8.18%	8.14%	3.31%	8.07%	2.11%	7.19%	4.30%	1.86%	1.64%	1.99%	4.76%
5	11.11%	8.65%	4.32%	7.24%	2.59%	4.52%	3.56%	2.03%	2.94%	1.10%	5.18%
Average	9.05%	9.39%	4.26%	8.72%	3.52%	5.33%	3.88%	2.19%	2.73%	1.76%	5.26%
max - min	3.74%	2.85%	3.13%	4.20%	3.77%	2.68%	0.89%	1.30%	1.99%	2.62%	−

Because the number of records varies as shown in Table 3, MAPE for each body type shows the results on total average instead of every cross validation. As shown in Figure 8, the body type with the largest MAPE is Sedan's 6.62%, then 5.90% for Minivan. This factor is thought to be attributed to brand image, which does not appear only in exterior design. For example, VELLFIRE and VOXY shown in Table 6 are made by the same company (Toyota) and their shapes are similar, but the customer's sensitivity is concentrated on high-class for the former and cool for the latter. As in the example, even though the exterior designs are similar, Sedans and Minivans tend to be split into Cool and High-class depending on the brand image.
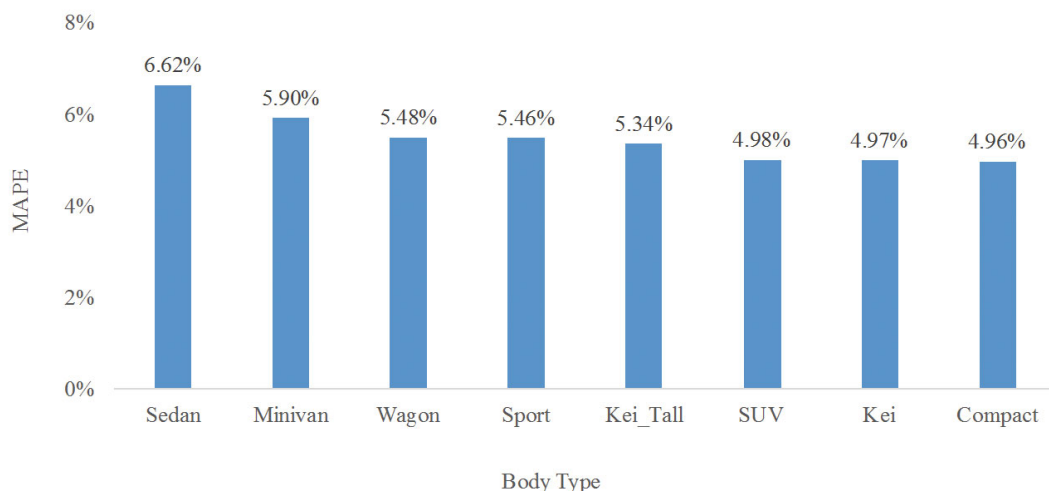

Figure 8. Mean absolute percentage error for each body type

In the present study we considered 10 sensitivity categories; however, additional categories can be evaluated using the proposed method and algorithms. Although not limited to this method, a weak point of machine learning is that it can-not adapt to tendencies that have not been learned. In this research, we used designs of existing vehicles as learning data. However, parts are becoming modularized and new companies are developing electric vehicles. Therefore, it

is difficult to evaluate designs using only past trends of the automobile industry. Manufacturers should collect customer sensitivity data to innovative designs and develop original sensitivity evaluation algorithms.

5. CONCLUSION

In this research, we proposed a technique to quantitatively evaluate design sensitivity by linking design features extracted from the design images and the sensitivity extracted from the customer's voice. The evaluation of sensitivity in this research is to estimate the ratio of the customer's voice to 10 sensitivity category. The estimation accuracy for 10 sensitivity categories was 5.26% MAPE as a result of five-fold cross validation.

This method has three merits. Firstly, it is to estimate sensitivity without disclosing designs which are confidential for the companies. Secondly, it is unnecessary for humans to define design features which is a huge amount work, and it is also possible to evaluate a design as an integrated body rather than elements. As the designs should be evaluated as a comprehensive body, it is not certain whether the perception of the customer is fully grasped regardless of how humans precisely define the features. Thirdly, it can be used affordably at the design site. In recent years, in order to ensure confidentiality and to evaluate design as a comprehensive body, neuroscience attracts attention. However, neuroscience approach not only requires dedicated machines but also has a heavy burden on the examinee, so the cost burden is large.

In recent years, the competitive domain of the manufacturing industry has shifted from functional value to emotional value. Among the emotional value design is the interface closest to the customer, but for companies it is often difficult to evaluate design from the viewpoint of confidentiality. We hope this research will contribute to trial and error at the design site. In future studies, we intend to extract features of brands image from promotional images.

REFERENCES

[1] Nobeoka K. Value Creation and Value Capture at Manufacturing Firms: Importance of Nonfunctional Value. Institute of Innovation Research, Hitotsubashi University Working Paper. 2010 10–01.

[2] Tuch A, Presslaber E, StöCklin M, et al. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. International Journal of Human Computer Studies. 2012; 70: 794-811. https://doi.org/10.1016/j.ijhcs.2012.06.003

[3] Osgood C, Suci G, Tannenbaum P. The measurement of meaning. University of Illinois Press. 1957.

[4] Li B, Wang Y, Wang K. A novel method for the evaluation of fashion product design based on data mining. Advances in Manufacturing. 2017; 5(4): 370-376. https://doi.org/10.1007/s40436-017-0201-x

[5] Ryoke M, Nagase K, Nakamori Y. Kansei evaluation and analysis of traditional craft. Japan Society of Kansei Engineering. 2008; 7: 633-640. https://doi.org/10.5057/jjske2001.7.633

[6] Sato H. Kansei Evaluation for the Front Grill Design of Automobile. Japan Society of Kansei Engineering. 2017; 16: 51-60. https://doi.org/10.5057/jjske.TJSKE-D-16-00040

[7] Inoue T, Harada T. Analysis and Planning of Cars's Front View Design Using Rough Set. Japan Society for Fuzzy Theory and Systems Fuzzy System Symposium Proceedings. 2001; 17: 647-650.

[8] Li Y. Research of Product Design Optimization Based on Supporting Vector and Artificial Fish Swarm Algorithm. International Journal of Hybrid Information Technology. 2016; 9(7): 1-8. https://doi.org/10.14257/ijhit.2016.9.7.01

[9] Yang C, Cheng J, Zhang Z, et al. Evaluation and Research on Interior Decoration Design of Automobile Cabins Based on Intention Recognition–Taking Control Panel Design for an Example. In International Conference on Cross-Cultural Design. 2016: 299-304.

[10] Sergey I, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning. 2015: 448-456.

[11] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 815-823.

[12] Karpathy A, Li F. Automated Image Captioning with ConvNets and Recurrent Nets. Available from: https://cs.stanford.edu/people/karpathy/sfmltalk.pdf (last accessed 2017/07/17).

[13] Mansimov E, Parisotto E, Ba J, et al. Generating Images from Captions with Attention. International Conference on Learning Representations. 2016.

[14] Shimoda W, Yanai K. Gathering and Analyzing Material Images on the Web with DCNN features. IEICE Technical Report. 2015; 114: 67-72.

[15] Kardes F, Sanbonmatsu D, Cronley M, et al. Consideration set overvaluation: When impossibly favorable ratings of a set of brands are observed. Journal of Consumer Psychology. 2002; 12: 353-361. https://doi.org/10.1016/S1057-7408(16)30086-9

[16] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[17] Tsuchiya S, Suzuki M, Imono M, et al. Emotion Judgement Method Based on Knowledge Base and Association Mechanism for Colloquial Expression. The Japanese Society for Artificial Intelligence. 2014; 29: 11-20. https://doi.org/10.1527/tjsai.29.11

[18] Harada T, Suzuki N, Fujita Y, et al. The Estimate Method of the Omission of Japanese Inquiry Texts using an LDA Algorithm. International Journal of Computer Applications in Technology. 2015; 52: 186-195. https://doi.org/10.1504/IJCAT.2015.071980