PAPER

Studies on association characteristics between vowels and visual colors using multiple speakers' speech

Win Thuzar Kyaw* and Yoshinori Sagisaka†

Department of Pure and Applied Mathematics, Graduate School of Fundamental Science and Engineering, Waseda University, 3-14-9 Ookubo, Shinjuku-ku, Tokyo, 169-8555 Japan

(Received 22 September 2020, Accepted for publication 9 March 2021)

Abstract: Aiming at scientific understanding of color association after listening to vowel sounds, cross-modal association characteristics was investigated using parameters directly extracted from speech and color. A perceptual experiment was conducted by employing five Japanese vowel sounds produced by multiple male and female native speakers differing in spectral characteristics. For color description, three attributes (hue, saturation and value) of 153 color tips of Practical Color Coordinate System (PCCS) selected after listening to vowel speech were employed. To enable scientific interpretation, acoustic features of vowel sounds (F0, pitch range, intensity, F1, F2, F3, F4) were used. Regression analyses were carried out between these speech and color parameters. From the results using multiple linear regression and neural networks, significant correlations could have observed between F1, pitch range & F2 and hue, F0 & F2 and value of color. These experimental findings can exactly and clearly show vowel-color association characteristics which have been partially studied in phonetics and cognitive science fields mainly using speech and color categories. These precise parametric correlations between color and speech can provide scientific knowledge for further investigations of cross-modal correlations in multiple research fields which will give the new possibilities for multi-modal information expression.

Keywords: Cross-modal association, Vowel-color association, Cross-modal information expression, Sentiment information

1. INTRODUCTION

When we are asked to show impressions after listening to speech sound, expression by color has been studied in many research fields. In particular, speech-color association experiences have widely studied for synesthetes who can concurrently and consistently see visual perceptions on hearing speech sounds [1–5]. This speech-color cross-modal association has been studied not only for synesthetes but also for ordinary people [6–13]. In these previous studies, most of them have analyzed using vowel categories and color categories to describe the cross-modal impression association.

Concerning the quantitative analysis of vowel-color correspondences, it has been proposed that acoustic factors also influence on the cross-modal correspondences in both synesthetes [4] and non-synesthetes [12]. Recently, the analysis of correlations between acoustic parameters and visual color parameters has been started, which provided

more clear and precise speech-color correlations. Color is composed of three properties: hue, saturation and value. Watanabe *et al.* [10,11] investigated sentiment correlations between speech features and color features. Their results showed high correlations between F0 and value meaning lightness of color (0.85), sound pressure level and saturation representing the strength of the color (0.82) and vowel categories and hue indicating the types of color. For the correlations between vowel categories and color categories, their findings showed similar mappings between vowel categories and color categories as previous studies [1,2]. Anikin and Johansson [5] investigated the association between acoustic parameters and color parameters based on Implicit Associations Test (IAT) experiments using vowel-like synthesized speech sounds. Their results showed implicit associations of pitch with value and saturation, loudness and spectral centroid with saturation. They have not yet observed the association between hue and acoustic features from their IAT experiment. Moreover, these studies have not yet investigated the color selection differences resulting from speech input differences. Our study tried to understand color association characteristics resulting from acoustic differences by using

*e-mail: winthuzarkyaw@akane.waseda.jp

†e-mail: sagisaka@waseda.jp

[doi:10.1250/ast.42.161]

natural speech of multiple speakers including both male and female.

Concerning speech-color association characteristics for multiple speakers, two studies [3,9] could have found. Fernay *et al.* [3] checked out the associations between the properties of vowel sounds (gender of speaker, F0 and phoneme) and color & spatial characteristics in synesthesia. For vowel-color associations, they suggested that the color selection characteristics of the synesthetes varied based on the speaker and the fundamental frequency by examining on four different speakers and four different vowel categories. Miyahara *et al.* [9] confirmed the patterns of correspondence between vowels and colors in the weak synesthesia (non-synesthetes) by carrying out both explicit and implicit measures based on male and female vowel sounds. They reported the mapping between /a/ and red in normal female voice, /o/ and blue in normal male voice and /i/ and yellow in normal-female and high-male voices.

In our study, for the purpose of mapping from speech parameters to color parameters, F0, pitch range, intensity, F1 F2, F3 & F4 were employed as speech parameters and hue, saturation & value were used as color parameters. Traditional multiple linear regression (MLR) was employed because of the straightforward interpretability of linear regression. However, the mapping between acoustic features and color features may not be as simple as linear relationship. Neural networks (NNs) was also used to achieve much better complex nonlinear mapping and generalization which cannot be provided by multiple linear regression.

This paper is composed of the following sessions. Previous studies on vowel-to-color association are summarized in the following Sect. 2. The perceptual experiment, speech stimuli and their features, color stimuli and their features are explained in Sect. 3. Statistical analyses and the results are described in Sect. 4. Our findings and the current understandings on the association between vowels and colors are summed up and discussed in Sect. 5. Finally, the paper is concluded in Sect. 6.

2. PREVIOUS STUDIES ON VOWEL-TO-COLOR ASSOCIATION

Synesthesia is an extraordinary, automatic and consistent multi-sensory perception reported about a hundred different types. It is a rare phenomenon and the prevalence is roughly 4.4% of the population tested for nine distinct types of perceptions including colors, tastes, smells and so on [14]. Among 4.4%, the most studied grapheme-color synesthesia which is experiencing colors for letters and numbers is more than 1% of the synesthetes [14]. Speech-color synesthesia is less common than grapheme-color synesthesia and can find only around 10% of all forms of

synesthesia [15]. The studies of synesthesia mainly focus on testing whether the concurrent perceptions of cross-modal experiences are stronger and more consistent over time for synesthetes than general population or not. It is still an open area and still growing attention recently [4,16]. Cross-modal associations between speech and color have also studied not only for synesthetes but also for general people [6–13].

Color is composed of three main properties including hue, saturation and value. Many studies found robust associations between pitch and value of colors for both synesthetes and non-synesthetes [2,3,6,9–11,13]. High pitch tended to associate with light colors. Mok *et al.* [13] revealed that the pitch effect is stronger than vowel associations as a high tone induced lighter colors than a low tone in general. Fernay *et al.* [3] checked out how the color changes according to the phoneme, speaker and pitch by using both male and female voices.

For the association between vowels and hue of colors, similar color categories to vowel categories mapping could be understood in multiple languages as well as both in synesthetes and non-synesthetes from previous findings [1,2,7,8,10,11,13]. In synesthesia research, Jakobson [1] proposed that synesthetes tended to select red for /a/, darker colors for /o/ and /u/, brighter colors for /e/ and /i/ concerning vowel-to-color mappings. Marks [2] also reported that /a/ is associated with red and blue, /e/ and /i/ with yellow and white, /o/ with red and black, and /u/ with brown, blue, and black. For the purpose of enhancing L2 pronunciation training, Wremble studied associative colors of Polish vowels [7] and English vowels [8] with non-synesthetes. By comparing the color associations of Polish vowels and English vowels, they observed cross-modal associations of high front vowels with light colors such as yellow and green, back vowels with dark colors like brown, blue and black, central vowels with gray. In recent study for non-synesthetes, Mok *et al.* [13] investigated the cross-modal association between Cantonese vowel sounds and 11 basic colors. They found strong associations between /a/ and red, /i/ and light colors, /u/ and dark colors. Miyahara *et al.* [9] confirmed similar color selection characteristics of non-synesthetes like synesthetes based on implicit and explicit measures. They proposed that the most selected colors are red for /a/, yellow for /i/ and blue for /o/ and /u/. Moreover, they explored gender and pitch effects on vowel-color associations. The most frequent choice of colors is red for /a/ in normal female voice, blue for /o/ in normal male voice and yellow for /i/ in normal-female and high-male voices.

Recent quantitative analyses of vowel-to-color relationships could be seen for synesthetes [4] and non-synesthetes [10–12]. In psychological area, Moos *et al.* [4] proposed that acoustic factors also influence on color correspond-

ences with vowels like grapheme factors based on their experiment using 16 synthesized vowels and 16 focal colors. They quantitatively analyzed on associations between vowel formants and hues. Kim *et al.* [12] studied the effects of the tongue body's position on association of vowels with colors employing synthesized vowel sounds by manipulating the tongue body articulations. In speech processing, Watanabe *et al.* [10,11] carried out sentiment correlation analyses between speech attributes and color attributes. From their studies, high correlations were found between F0 and value, sound pressure level and saturation and vowel categories and color categories. For the relationship between vowel categories and color categories, their results were accordant with the previous findings. /a/ was found to associate with red or orange, /i/ with yellow or green, /u/ with blue, green or purple, /e/ with green or orange and /o/ with blue, green or purple. Anikin and Johansson [5] presented implicit associations between acoustic parameters and color parameters. Although they found associations of pitch with value and saturation, loudness and spectral centroid with saturation, they did not observe the associations between hue and acoustic features.

In our study, the vowel-to-color associations were tried to quantitatively characterize by carefully looking at the color responses based on vowel sounds differing in acoustic characteristics which have never been investigated in previous studies by employing many natural male and female vowel sounds. In particular, the feature-based understanding of the mappings between hue and acoustic features was intensively examined with multiple speakers. To understand cross-modal correlations, not only traditional statistical method which was mostly used in previous studies, multiple linear regression (MLR), but also artificial neural networks (ANNs) were employed to achieve much better mapping between vowel parameters and color parameters.

3. EXPERIMENT

To study the color selection differences based on speech input differences, color association experiment was carried out using natural vowel sounds uttered by multiple speakers including both male and female.

3.1. Experimental Setup

Color selection experiment was conducted using five Japanese vowel sounds by participating 34 Japanese listeners (22 males and 12 females) ranging from 18 to 24 years old. Each participant was asked to listen to individual speech stimulus played in random order by taking a seat in front of a computer in a quiet room. After listening to each speech sample, the subject was asked to choose the most suitable color among 153 colors at a time based on his/her perceptual impression.

3.2. Speech Stimuli and Features

For the speech stimuli to be used in our experiment, Japanese vowel sounds of many male and female speakers were collected. Five Japanese single vowels (/a/, /i/, /u/, /e/, and /o/) were produced by seven native speakers including four males and three females. Total thirty-five speech stimuli (5 Japanese vowels $\times$ (4 male speakers + 3 female speakers)) were recorded with sampling frequency of 44.1 kHz for our experiment.

For the associations between vowels and colors, acoustic features representing the characteristics of vowel sounds were employed including formant frequencies, F0, pitch range and intensity of vowel sounds. Since vowel features can be roughly characterized by the first two peaks of sound spectrum called first and second formants (F1 and F2), the previous studies evaluated vowel formants and color correlations by manipulating F1 and F2. The higher formants (F3 and F4) are also important features for gender differences. For our analyses, not only F1 and F2 but also F3 and F4 were taken into our consideration as the inputs to regression tasks. Mean fundamental frequency (F0), variation of F0 in a speaker (pitch range) and intensity of speech sounds were also measured as input features.

All acoustic features were calculated using Praat [17]. The frequencies of the formants (F1, F2, F3 and F4) in Hertz were measured from Linear Predictive Coding (LPC) [18] and Fast Fourier Transform (FFT) using the Burg algorithm.

3.3. Color Stimuli and Features

Most of former studies [4,5,13] mainly used approximated color space such as light-dark or red-green and 11 basic colors [19]. By providing a larger set of colors, we can expect to explore cross-modal correspondences more precisely. In our study, Practical Color Co-ordinate System (PCCS) was employed for the color stimuli to be used in the experiment. PCCS is a discrete color space emphasizing on human perception announced by Japan Color Research Institute (JCRI) in 1964.

Color is usually represented by three attributes consisting of hue, saturation and value in cognitive science studies. PCCS is a two-dimensional hue-tone system in which tone dimension is a combination of saturation and value. Hue refers to color quality such as "reddish" or "bluish." Tone shows differences in tints and shades of color like "pale red," "deep blue" or "soft purple." For the color choices, PCCS is supposed to be the most suitable color system since it consists of almost all colors at same psychological intervals fitting to human visual perception. In addition, two-dimensional color display is easier for the participants to evaluate than other three-dimensional color systems like Munsell color system.

Fig. 1 Color palette used in the experiment of color selection by listening to the vowel sound. (R = red, rO = red-orange, yO = yellow-orange, Y = yellow, YG = yellow-green, G = green, BG = blue-green, gB = green-blue, B = blue, V = violet, P = purple, RP = red-purple and N = achromatic).

The color palette for our experiment consists of 153 color tips (12 hues $\times$ 12 tones + 9 achromatic colors) as shown in Fig. 1. All 153 color tips were selected in psychologically same interval and ordered in a plane. Horizontal axis represents hue. Vertical axis corresponds to tone. Hue consists of four primary colors (red, blue, green, and yellow) and the complementary colors of them, supplemented by additional four hues to arrange them in the psychologically identical intervals. Tone includes 12 tones like “vivid,” “bright,” and “dark” based on the impressions they impart in terms of vividness.

In recent studies on speech-color associations [4,5,12], CIE-Luv and CIE-Lab color systems were used for quantifying colors. Both CIE-Luv and CIE-Lab color systems are composed of one channel for lightness and two channels for color. For our study, HSV color model was used to calculate three color attributes individually. HSV represents hue, saturation and value attributes of color. PCCS color tips were converted into HSV parameters.

In HSV color space, hue shows types of color which is measured by an angle with red starting at 0 degrees, green at 120 degrees and blue at 240 degrees. Complementary colors are in-between: yellow is at 60 degrees, cyan is at 180 degrees, and magenta is at 300 degrees. Saturation parameter represents the strength of the color that is how grey or pure the color will be. Value refers to the lightness.

Because hue is a circular quantity, Fernay *et al.* [3] suggested to treat hue values in a different way for the quantitative analysis of association between hue and acoustic features. The standard way of computing the average cannot be used for hue in calculating the average values of HSV parameters. The average hue values were calculated using the mean of circular quantities by employing the following equations.

$$H = \{h_1, h_2, \dots\}, \quad h_i \in [0, 359] \quad (1)$$

where H is a set of N hues.

Hue angle values were converted to radians by Eq. (2).

$$h_i^r = \frac{\pi h_i}{180} \quad (2)$$

The average vector of all the hue angles were determined by Eq. (3).

$$h^{-r} = \arctan(\text{avg}_i \sinh_i^r, \text{avg}_i \cosh_i^r) \quad (3)$$

The averaged values were converted back to degrees according to the following equation.

$$\frac{180}{\pi} h^{-r} \pmod{360} \quad (4)$$

For further processing such as training hue angles in neural network, hue angles were encoded using trigonometric encoding. Each averaged hue angle was transformed from a single variable into two continuous variables. Before calculating x and y coordinates of hue value, the angle was converted to the radian by Eq. (5).

$$r = \frac{\pi}{180} h \quad (5)$$

Two continuous variables (x and y coordinates) were calculated by taking the sine and cosine of the radian by the following equations.

$$x = \cos(r) \quad (6)$$

$$y = \sin(r) \quad (7)$$

To understand the actual colors, the angle value of hue can be reconstructed from x and y coordinates using the following decoding function.

$$h = \frac{180}{\pi} \arctan\left(\frac{y}{x}\right) \pmod{360} \quad (8)$$

4. ANALYSIS AND RESULTS OF ASSOCIATIONS BETWEEN SPEECH AND COLOR FEATURES

For the cross-modal association between acoustic parameters and color parameters, multiple linear regression to enable the interpretation of the correlations and neural networks (NN) to achieve precise mapping were employed.

4.1. Data Preparation

To be free from the variations in color selections between listeners, the data was averaged across all participants (34 participants) to obtain a representative sample for each speech sample. To measure the central tendency, the mean value which includes every value in the data were calculated. Although the mean is sensitive to

Table 1 Summary of the speech and color parameters.

Parameter	Min	Max	Mean	Std. deviation
Inputs				
F0	120.240	274.227	191.212	45.306
Pitch range	3.762	409.451	142.641	137.673
Intensity	71.594	84.503	78.360	3.187
F1	249.415	1,052.209	539.142	207.755
F2	641.721	3,003.208	1,668.597	640.647
F3	2,237.010	3,716.691	2,940.168	312.683
F4	3,228.677	4,420.472	3,839.458	343.000
Outputs				
Hue (x)	−0.999	0.996	0.078	0.820
Hue (y)	−0.989	0.989	0.357	0.467
Saturation	59.382	86.941	73.253	5.992
Value	51.324	91.206	71.633	11.322

F0: fundamental frequency; F1: first formant; F2: second formant; F3: third formant; F4: fourth formant; Hue (x): x coordinate of hue value; Hue (y): y coordinate of hue value.

extreme values, an outlier in our data may be due to the variability inherent in the observed data. For this reason, all data values in the data set were used without dropping any outlier. In total, 35 observations (5 Japanese vowels $\times$ (4 male speakers + 3 female speakers)) were prepared for our analyses. Speech features were considered as independent variables and color features were used as dependent variables to look for their relationships. Statistical measures for speech inputs and color outputs are as shown in Table 1.

High correlations amongst independent variables may result in the issue of multicollinearity and effect on the performance of regression models. Before carrying out regression tasks, multicollinearity among the independent variables was tested by the variance inflation factor (VIF). VIF values greater than 10 infer that there is a serious multicollinearity problem in the dataset [20]. From VIF plot and the eigenvalues plot in Fig. 2, there is no multicollinearity problem in our data.

Before the regression analyses, the data were standardized by subtracting the mean and dividing by the standard deviation of the data according to the following equation.

$$Z = \frac{x - \mu}{\sigma} \quad (9)$$

where Z = standard score, x = observed value, μ = mean of the sample, σ = standard deviation of the sample.

For standardization, preProcess option of train function in caret package [21] was used.

4.2. Multiple Linear Regression

Simple linear regression is one of the widely used

Fig. 2 Plots of multicollinearity test.

statistical techniques to understand the relationships between continuous variables. Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable by fitting a linear equation to observed data [22].

To understand the relationships between acoustic features and HSV color features, multiple linear regression was carried out using caret package [21] of R [23]. The acoustic features including fundamental frequency (F0), pitch range, intensity and frequencies of first four formants (F1, F2, F3 and F4) were employed as independent (input) variables. The color attributes representing x coordinate of hue, y coordinate of hue, saturation and value (HSV) were used as dependent (output) variables. To avoid overfitting, multiple linear regression (MLR) was performed using Leave-one-out cross-validation (LOOCV) method. The models with the lowest Root Mean Square Error (RMSE) were selected to understand the relationship between speech and color features. The significant acoustic features for each color feature of the MLR models are as shown in Table 2.

4.3. Artificial Neural Network

Expecting the better mapping, a feed-forward neural network with a single hidden layer was employed. Three fundamental components in the neural network are an input layer, a hidden layer and an output layer [24]. The nodes of

Table 2 Significant speech features for each color feature produced by multiple linear regression analysis.

Response variable	Independent variables	t	p
Hue (x)	F1	3.070	<0.01
	Pitch range	−2.336	<0.05
Hue (y)	F2	4.595	<0.001
Saturation	Intensity	2.439	<0.1
Value	F0	3.209	<0.01

our input layer are the speech features and the output layer consists of each of color features. These input layer and output layers are linked by a hidden layer.

For the neural network training, the number of nodes in the hidden layer of neural networks was set as a free parameter that we had to select a value through a process of model tuning. Other free parameter called weight decay together with number of neurons in the hidden layer was applied to deal with over-fitting [24]. The modeling process began by assigning random weights to the connections between nodes, which were then iteratively updated as predictions were checked against reality and error was back-propagated. The modeling process was iterated through the number of hidden layer nodes: 1, 2, 3, 4, 5, to 20 and through decay rates of: 0.1, 0.2, 0.3, 0.4, and 0.5. The results of the grid search for hyper-parameters of neural network models were as shown in Fig. 3.

To study the effects of various independent variables on the output and provide insight into the helpfulness of individual variables, variable importance analysis was conducted using `varImp()` function in `caret` package of R. `VarImp` is a generic method that uses combinations of the absolute values of the weights based on Gevrey *et al.* [25] in neural networks. The important variables for each neural network model were depicted in Fig. 4.

The performance of multiple linear regression and neural network models were as described in Table 3. Based on the results of the model performance through statistical measures consisting of Root Mean Squared Error and R -squared (RMSE and R^2), all NN models outperformed the multiple linear regression models. As expected, neural nets can provide better mapping than linear regressions.

From the statistically significant results of multiple linear regression (MLR) and the most contributed predictors derived from neural network (NN), parametric associations were found between speech features and color features. From the common findings of both MLR and NN models, the x coordinate of hue value was strongly associated with the frequency of the first formant (F1) of speech. On the other hand, y coordinate of hue was highly correlated with the frequency of the second formant (F2). Moreover, value of color was significantly associated with the fundamental frequency (F0) and F2. For the association between saturation attribute of color and speech features, very low correlation scores were observed in both MLR model and NN model and we may need further investigation for this association.

5. DISCUSSION

In previous studies, cross-modal mappings between speech and color parameters have been assessed based on controlling synthesized and natural speech samples. In this study, vowel-color association characteristics was precisely

(a) X coordinate of hue

(b) Y coordinate of hue

(c) Saturation

(d) Value

Fig. 3 Grid search for hyperparameters of neural networks model. The hyper parameters included the number of hidden units and the weight decay.

analyzed by employing speech and color parameters directly extracted from speech samples uttered by multiple male and female speakers to investigate color responses to

Fig. 4 The important speech features for each color feature produced by neural networks.

speech input differences. A perceptual experiment was conducted by employing five Japanese vowels uttered by 4 male and 3 female native speakers and 153 colors of

Table 3 The performance of multiple linear regression and neural networks.

Response variable	Method	RMSE	R^2
Hue (x)	MLR	0.685	0.335
	NN	0.484	0.652
Hue (y)	MLR	0.407	0.264
	NN	0.378	0.329
Saturation	MLR	6.220	0.094
	NN	6.200	0.062
Value	MLR	8.414	0.454
	NN	8.511	0.494

PCCS color system. Using fundamental frequency (F0), pitch range, intensity and first four formant frequencies (F1, F2, F3 and F4) as speech features and hue, saturation and value (HSV) as color features, their mappings were tried to find out by implementing on multiple linear regression and neural networks.

Through the analyses, the relationships between hue and the first and the second formants (F1, F2) could have observed. By transforming hue angular values into two continuous x and y coordinate values, it could confirm that the correlation of x coordinate of hue with F1 and y coordinate with F2. In addition, it was found that pitch range also associated with x coordinate of hue. Although rough analyses have been carried out on the associations between vowel categories to color categories in the previous studies, there were no quantitative analysis. The investigation using speech and color parameters, we could have observed the exact correlations between formants and hue parameters. In psychological area, Moos *et al.* [4] reported that the acoustic influences of F1 and F2 on u^* (the red-green axis) but not on v^* (the blue-yellow axis) of color.

For cross-modal correspondences between lightness and acoustic characteristics, our findings were in accordant with the previous reports. The significant associations were observed between value (lightness) and F0 and F2 features of vowel sounds. Many studies proposed the association of higher pitch with lighter color. For the direct mapping, the high correlation between lightness and F0 has been proposed [5,10,11]. Although there were no statistical measures, Jakobson [1] and Marks [2] pointed out the relationship of F2 with light-dark dimension of colors.

For the association of saturation with spectral features, Anikin and Johansson [5] found that loudness or markedness of speech were correlated with contrast or saturation of colors. Watanabe *et al.* [11] described high correlations between sound pressure level (SPL) and saturation. Although our correlation score was very low, there was a tendency to associate saturation with intensity and F1.

These correlations using modality driven parameters showed the possibility of better description and modeling in speech variations using colors. To treat undescribed information embedded in speech, it has been proposing to use its sentiment information as its descriptor and show the possibilities of their use in speech synthesis. For their description, language information expressing their perceptual impressions has been successfully applying. By using Multi-Dimensional Scaling (MDS), it could have found that communicative prosody information can be reduced to three dimensional expressions (doubtful-confident, unacceptable-allowable, positive-negative) nicely corresponding to its prosody characteristics [26–28].

Though discrete language expressions can be effectively employed to specify communicative speech, it has not yet succeeded to describe their differences systematically. In color science, the color impressions about hue, saturation and value of PCCS color system were tried to express in linguistic terms. Wakata and Saito [29] reported that the impression of “bright-dark” and “light-heavy” corresponded to value of colors. These facts indicated that categorized characterization identified by language can specify only its very vague features but cannot specify them clearly.

To be free from the constraints naturally imposed by language use, sentiment correlation analysis between speech and other media was started by replacing language medium to image medium to understand its scientific background. The results show the possibilities of speech information processing and mutual information transformation among language, speech and image. The systematic correspondence between the speech variation and the color system could be applicable not only in communicative prosody generation and perception, but also in more general purposes such as speech visualization and cross-modal information mapping where many useful applications can be considered.

6. CONCLUSIONS

Aiming at finding exact cross-modal correspondences among features, the association characteristics between vowels and colors was investigated by using multiple male and female speech stimuli. By utilizing Japanese vowel sounds with gender differences showing distinct resonance characteristics, the mappings between acoustic and color features were tried to find out. From our results, the significant correlations were found between F1 & pitch range and *x*-coordinate of hue, F2 and *y*-coordinate of hue which have not been reported in preceding studies. Moreover, the remarkable associations between F0 & F2 and value could be confirmed. These results showed the existence of sentiment correlations between vowels and colors more exactly than the conventional studies.

The current study gives the promising result to pursue media transform which are expected in many application fields. This kind of cross-modal association between speech and other media suggests a possibility for description of speech variations using image related information such as color parameters. As an extension of this work, further correlations between speech source parameters and texture image parameters are started to be investigated.

REFERENCES

- [1] R. Jakobson, *Selected Writings: Phonological Studies* (Mouton, The Hague, 1962).
- [2] L. E. Marks, “On colored-hearing synesthesia: Cross-modal translations of sensory dimensions,” *Psychol. Bull.*, **82**, 303–331 (1975). doi:10.1037/0033-2909.82.3.303
- [3] L. Fernay, D. Reby and J. Ward, “Visualized voices: A case study of audio-visual synesthesia,” *Neurocase*, **18**, 50–56 (2012). doi:10.1080/13554794.2010.547863
- [4] A. Moos, R. Smith, S. R. Miller and D. R. Simmons, “Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder,” *i-Perception*, **5**, 132–142 (2014). doi:10.1068/i0626
- [5] A. Anikin and N. Johansson, “Implicit associations between individual properties of color and sound,” *Atten. Percept. Psychophys.*, **81**, 764–777 (2019). doi:10.3758/s13414-018-01639-7
- [6] J. Ward, B. Huckstep and E. Tsakanikos, “Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?,” *Cortex*, **42**, 264–280 (2006). [https://doi.org/10.1016/S0010-9452\(08\)70352-6](https://doi.org/10.1016/S0010-9452(08)70352-6)
- [7] M. Wrembel, “Still sounds like a rainbow — A proposal for a coloured vowel chart,” *Proc. Phonetics Teaching and Learning Conf. PTLT 2007, London*, pp. 1–4 (2007).
- [8] M. Wrembel, “On hearing colours — Cross-modal associations in vowel perception in a non-synaesthetic population,” *Poznan Stud. Contemp. Linguistics*, **45**, 581–598 (2009). doi:10.2478/v10010-009-0028-0
- [9] T. Miyahara, A. Koda, R. Sekiguchi and T. Amemiya, “A psychological experiment on the correspondence between colors and voiced vowels in non-synesthetes,” *Kansei Eng. Int. J.*, **11**, 27–34 (2012). doi:11.10.5057/kei.11.27.
- [10] K. Watanabe, Y. Greenberg and Y. Sagisaka, “Sentiment analysis of color attributes derived from vowel sound impression for multimodal expression,” *Proc. IEEE Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conf. (APSIPA)*, pp. 1–5 (2014). doi:10.1109/APSIPA.2014.7041586
- [11] K. Watanabe, Y. Greenberg and Y. Sagisaka, “Cross-modal description of sentiment information embedded in speech,” *Proc. ICPhS 2015*, A-117 (CDROM). doi:10.1109/APSIPA.2014.7041586
- [12] H. W. Kim, H. Nam and C. Y. Kim, “[i] is lighter and more greenish than [o]: Intrinsic association between vowel sounds and colors,” *Multisens. Res.*, **31**, 419–437 (2018). <https://doi.org/10.1163/22134808-00002581>
- [13] P. P. K. Mok, G. Li, J. J. Li, H. T. Y. Ng and H. Cheung, “Cross-modal association between vowels and colours: A cross-linguistic perspective,” *J. Acoust. Soc. Am.*, **45**, 2265–2276 (2019). <https://doi.org/10.1121/1.5096632>
- [14] J. Simner, C. Mulvenna, N. Sagiv, E. Tsakanikos, S. A. Witherby, C. Fraser, K. Scott and J. Ward, “Synaesthesia: The prevalence of atypical cross-modal experiences,” *Perception*, **35**, 1024–1033 (2006). <https://doi.org/10.1068/p5469>

- [15] J. Ward and R. Cytowic, “Synesthesia and language,” in *Encyclopedia of Language & Linguistics* (Elsevier Science, Oxford, 2006), pp. 371–376. doi:10.1016/B0-08-044854-2/04196-1
- [16] C. Cuskley, M. Dingemanse and S. Kirby, “Cross-modal associations and synesthesia: Categorical perception and structure in vowel-color mappings in a large on-line sample,” *Behav. Res. Methods*, **51**, 1651–1675 (2019). <https://doi.org/10.3758/s13428-019-01203-7>
- [17] P. Boersma and D. Weenink, Praat: Doing phonetics by computer [Computer program]. (2011). Version 6.0.39, <http://www.praat.org/>
- [18] N. Anderson, “On the calculation of filter coefficients for maximum entropy spectral analysis,” *Geophysics*, **39**, 69–72 (1974). <https://doi.org/10.1190/1.1440413>
- [19] B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution* (University of California Press, Berkeley, 1969).
- [20] D. C. Montgomery and E. A. Peck, *Introduction to Linear Regression Analysis* (Wiley, New York, 1992).
- [21] M. Kuhn, “Building predictive models in R using the caret package,” *J. Stat. Softw.*, **28** (2008). doi:10.18637/jss.v028.i05
- [22] R. B. Darlington and A. F. Hayes, *Regression Analysis and Linear Models: Concepts, Applications, and Implementation* (The Guilford Press, New York, London, 2017).
- [23] R Core Team, “R: A Language and Environment for Statistical Computing,” R Foundation for Statistical Computing (2017). <http://www.R-project.org/>
- [24] M. Kuhn and K. Johnson, *Applied Predictive Modeling* (Springer, New York, 2013).
- [25] M. Gevrey, Y. Dimopoulos and S. Lek, “Review and comparison of methods to study the contribution of variables in Artificial Neural Networks models,” *Ecol. Modell.*, **160**, 249–264 (2003). doi:10.1016/S0304-3800(02)00257-0
- [26] Y. Greenberg, N. Shibuya, M. Tsuzaki, H. Kato and Y. Sagisaka, “Analysis on paralinguistic prosody control in perceptual impression space using multiple dimensional scaling,” *Speech Commun.*, **51**, 585–593 (2009). <https://doi.org/10.1016/j.specom.2007.10.006>
- [27] L. Shao, Y. Greenberg and Y. Sagisaka, “Global f0 control parameter prediction based on impressions for communicative prosody generation,” *Proc. IEEE Int. Conf. Oriental COCOSA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSA/CASLRE 2013)*, pp. 1–4 (2013). doi:10.1109/ICSDA.2013.6709871
- [28] Y. Sagisaka and Y. Greenberg, “Communicative speech synthesis as pan-linguistic prosody control,” in *Speech Prosody in Speech Synthesis: Modeling and Generation of Prosody for High Quality and Flexible Speech Synthesis, Prosody, Phonology and Phonetics*, K. Hirose and J. Tao, Eds., pp. 73–82 (2015). doi:10.1007/978-3-662-45258-5_5
- [29] T. Wakata and M. Saito, “A study of color impression about tone in PCCS color system,” *J. Color Sci. Assoc. Jpn.*, **36**, 224–225 (2012). <https://ci.nii.ac.jp/naid/110009445589/en>