International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 15 (2017) pp. 4941-4945

A Pattern Recognition of the Volatile Chromatographic Peaks detected in the Korean Distilled Liquors using Principle Component Analysis

Sungmin Myoung[1] and Chang-Hwan Oh,[2]

[1]Associate Professor, Department of Health Administration, Jungwon University, Goesan-gun, Chungbuk 28024, South Korea.

[2]Associate Professor, Department of Oriental Medical Food and Nutrition, Semyung University, Jecheon, Chungbuk 27136, South Korea.

[2]Orcid: 37084513800.

Abstract

The traditional Korean distilled liquor (KDL) has historically been made from rice, which are fermented with the aid of yeast and nuruk, a wheat-based source of the enzyme amylase followed by distillation with "Soju Gori" that is a traditional ceramic distillation device in Korea. In this research, we analyzed the commercial KDLs produced by traditional manufacturing process to find patterns in chromatographic data and investigated the quality control indices for physicochemical properties of KDLs. A total of 21 samples of commercial 7 KDLs were analyzed by GC-FID and GC-MS after sample preparation (dichloromethane partition). Pattern recognition procedures have been applied principle component analysis (PCA). Among the volatile compounds detected in 7 KDLs, phenylethyl alcohol, diethyl butanedioate and methyl-1-butyl acetate are the compounds with high normalized peak area%. In PCA results, Munbaesool, one of 7 KDLs, showed clearly different volatile pattern. Pattern recognition by using statistical methods are employed to extract relevant information from the data of chemometrics. The results of this study will be utilized as the indices for evaluating and improving the quality of individual KDL, and establishing the quality assurance system for Korean traditional spirits.

Keywords: Gas chromatograph-mass spectrometer (GC-MS), Korean distilled liquors, multivariate analysis, principle component analysis (PCA)

INTRODUCTION

Chemometrics is to extract useful information and knowledge from data, and the application of principles of multivariate statistical method and measurement science [1]. In fields of food analysis, one of the fundamental objects of chemometrics is the replacement of the sensory analysis (that is quite objective unless very large number of tests performed with trained personnel) characterizing food and beverages by using various descriptors. Among instrumental analysis, gas chromatograph - mass spectrometer (GC-MS) have been widely used for analyzing volatile/flavor components that affect food quality, and applied to many researches as a factor in chemometrics [2]. But, ion peaks produced by GC-MS is relatively likely lacking precision due to the possible variation caused by the type and condition of analyzer or the ionization conditions of various type of GC-MS. By the way, the alternative GC with flame ionization detector (FID) might be nice tool for chemometrics due to better relative precision of retention time and peak area (or height) caused by simple instrumental mechanism.

Multivariate analysis is a powerful statistical method for analyzing peak area data sets generated in a typical GC or GC-MS analysis of volatile/flavor components distinguishing meaningful knowledge/information from random variation in the data set [3]. In general, this analysis starts with applying the exploratory algorithm which reduces the information to a more comprehensive format in order to recognize hidden patterns of complex data in the data set. Several methods suitable for pattern recognition in multivariate analysis were applied: principle component analysis (PCA), cluster analysis (CA), partial least square (PLS) analysis, linear discriminant analysis (LDA) and partial least square – discriminant analysis (PLS-DA) [4].

Recent advances in pattern recognition for the volatile components (analyzed by 2nd dimensional GC-MS and/or other techniques) were suggested for many applications such as, jet fuel samples for underground spill source identification using PCA and CA [5], classification of premium and regular gasoline using PCA and artificial neural networks [6], classification of different vinegar types using DA and PLS-DA [7], differentiation of Irish whiskeys for higher-alcohol congener analysis using PCA and CA [8], and flavor pattern

of various coffee measured electronic nose using LDA [9].

Liquor, one of representative preferable food, has specific sensory characteristics combined with alcohol. Multivariate analysis is also suitable to differentiate volatile characteristics which is related to quality of liquors. From the earliest record of liquor, the distilled liquor from outside of Korea was named as "Arakaju" that must be originated the name of distilled liquor "Arak or Araq" (it is an Arab alcoholic spirit). The representative traditional Korean distilled liquor (KDL) "Soju" that is named as "Arakju (a name derived from Arakaju)" in some place, has historically been made from rice, which are fermented with the aid of yeast and nuruk, a wheat-based source of the enzyme amylase followed by distillation with "Soju Gori" that is a traditional ceramic distillation device in Korea. There were hundreds of different types of alcoholic beverages (including various Soju) in Joseon dynasty of Korea. But many of them were disappeared due to the proclamation of Liquor Tax Law and the erasure policy for Korean culture by the colonial administration of Imperial Japan in early 20th century. Many modern popular Soju brands in Korea is the mixture of the fermented industrial grade spirit (95% alcohol by volume) with water and food additives such as stevioside and aspartame. Therefore there is no intrinsic chemical difference between the modern popular Soju brands except added food additives. However the top note of KDLs (Such as the different types of traditional Soju) are characteristic due to the different materials, fermentation method and distillation procedure. Especially the distillation procedure influences the occurrence and concentration of volatile flavor compounds in the distillate [10]. That's why these volatile compounds could be the good candidates to examine the quality of the spirit products. Many studies have been conducted for KDLs into investigate physicochemical and sensory characterization [11, 12]. However, no research of volatile component of KDLs have yet been attempted for the development of quality control indices yet.

In this study, we analyzed KDLs to find patterns in chromatographic data and investigated the quality control indices for physicochemical properties of KDLs. The volatile indicator peaks characterizing the seven commercial KDLs were analyzed by GC-FID after liquid-liquid extraction (LLE) with dichloromethane. The volatile peaks in KDL samples were analyzed by PCA to figure out the potent characteristic indicators. The results of this study will be utilized as indices for evaluating and improving the quality of individual KDL, and contribute establishing of quality assurance system for Korean traditional spirits.

Table 1. Characteristics of the Korean distilled liquor samples

Name	Characteristics	EtOH (%)	Symbol
Leegangju	Pear, ginger, tumaric, cinnamon and honey added rice liquor	19	IK
Oakzen	Party blended distillate (rice fermented soju) aged in oak barrels for ten years	25	OK
Ilpoom Andong Soju1	Distillate (rice fermented soju)	21	IA1
Ilpoom Andong Soju2	Distillate (rice fermented soju)	40	IA2
Ilpoom Jinro	Distillate (rice fermented soju) aged in oak barrels for 10 years	25	IJ
Hwayo	Distillate (rice fermented soju) aged in pottery for 3 months	41	HW
Moonbaesool	Distillate with rice, foxtail millet, glutenous sorghum fermented rice	23	MB

MATERIAL AND METHODS

A. Network components

A total of 21 samples of commercial 7 KDLs, such as Munbaesool (MB, 40% of EtOH; Moonbaesool brewery co., Gimpo-si, Gyeonggi-do, Korea), Leegangju (IK, 25% of EtOH; Chonju Leegangju, Jeonju-si, Jeollabuk-do, Korea), Ilpoom-Andong Sojus (IA1, 21% and IA2, 40% of EtOH; Ilpoon Andongsoju Inc., Andong-si, Gyeongsangbuk-do, Korea), Ilpoom-Jinro (IJ, 25% of EtOH; Hitejinro, Hongchoeon-gun, Kangwon-do), Hwayo (HW, Hwayo, Yeoju, Gyeonggi-do, Korea) and Oakzen (OK, 25% of EtOH; Kumbokju Co. Ltd., Dalseo-gu, Daegu, Korea), were purchased in local mart or department store in Seoul, Korea.

Ethyl alcohol concentration of sample liquor was adjusted to 20% by HPLC grade water before LLE using dichloromethane (Burdick and Jackson, Honeywell corp., Morristown, NJ, U.S.A.). The sample (10 mL, 20% ethanol concentration base) was extracted with 1mL of dichloromethane after salting-out (with 3g of anhydrous sodium chloride). The extract was filtered with 0.45μm PTFE syringe filter (Acrodisc, Pall Corporation, Port Washington, NY, USA) followed by removing moisture with anhydrous sodium sulfate (Samchun chemical, Pyungtaek, Gyoungi-do,

South Korea) before GC-FID or GC-MS analysis. All reagents and solvents were analyzed by GC-FID for the assurance of their purity. Table 1 presents symbols, name, and characteristics of samples.

B. Instrumental Analysis

Quantitative analysis was performed by Agilent 7890 GC-FID with oven temperature started from 40°C (holding for 2 min), prompted to 250°C (by ramping speed, 3°C/min) and held for 5 min. Splitless injection mode was used with purge delay time, 1 min. The oven condition of GC and GC-MS was same. Mass spectral database searching was applied for qualitative analysis of GC-MS peaks. GC-MS peaks were matched to GC-FID peaks by comparing of retention index (RI) calculated by the below calculus with n-paraffin mixture (C_{10}-C_{26}) [13]. The blank solvent (HPLC grade dichloromethane) was injected to GC-FID after each sample analysis for assuring no carry-over.

$$RI = 100\times\left[z+\left(\frac{RT_u-RT_x}{RT_{x+1}-RT_x}\right)\right] \quad (1)$$

where

z : number of n-paraffin

RT_u : retention time of unknown compound

RT_x : retention time of n-paraffin eluted before the unknown compound

RT_{x+1} : retention time of n-paraffin eluted after the unknown compound

The n-paraffin mixture was analyzed every day for the accurate precise calculation of RI of the detected peaks.

C. Statistical Analysis

Statistical analyses of GC-FID/GC-MS data were performed using R 3.3.1 and R-studio 1.0.44. Each KDL sample was considered as an assembly of 12 components (variables) which were the selected volatile peaks for data analysis. All the results expressed as percent of the normalized peak area.

Pattern recognition procedures have been applied PCA [14]. The aim of PCA is for reducing dimension, and it defines new variables, termed principle components (PCs), consisting of linear combination of original variables (volatile components) without much loss of information. PCA result consists of two values such as PC-score and PC-loading. PC-score means similar behavior between samples, PC-loading means similar behavior between variables. By comparing the score and loading, the relationships can be identified between samples and variables. The selection of number of principle component is decided with eigenvalues of more than 1. According to this rule, two PCs were selected in this study. Statistical significance was p-value<0.05.

Table 2. Normalized peak area% of the major identified volatile peaks in KDLs

Compound Name	Abbre-viation	RI*	Normalized Peak Area%			
			MIN	MAX	MEAN	%RSD**
2-Methyl-1-butyl-acetate	P1	878	1.10	25.52	9.94	79.42
Ethyl sorbate	P2	919	0.00	1.95	0.42	172.04
1-Octen-3-ol	P3	980	0.00	2.38	0.52	158.48
Ethyl hexanoate	P4	1000	0.96	12.27	3.26	114.83
Benzeneacetaldehyde	P5	1047	0.00	3.17	0.45	251.00
Phenylethyl alcohol	P6	1118	24.15	90.49	66.60	29.30
Diethyl butanedioate	P7	1180	1.05	29.20	11.19	94.57
Ethyl octanoate	P8	1196	0.00	7.80	3.01	85.63
2-Phenylethyl acetate	P9	1256	0.56	7.50	2.96	83.75
4-Ethyl-2-methoxyphenol	P10	1277	0.00	5.58	0.79	251.00
Methyl decanoate	P11	1393	0.00	2.40	0.71	142.14
3-Methylbutyl octanoate	P12	1441	0.00	1.09	0.16	250.99

*RI: Retention index, **%RSD: Relative standard deviation

RESULTS

The normalized peak areas of volatile components are given in Table 2. A total of 12 volatile components was detected among total 21 KDLs using GC-FID. Among components of 7 KDLs, the component with the highest average normalized peak area% (66.60%) was P6 (phenylethyl alcochol) followed by P7 (diethyl butanedioate, 11.19%) and P1 (2-methyl-1-butyl acetate, 9.94%). Other components occupied under 5% of normalized peak area%.

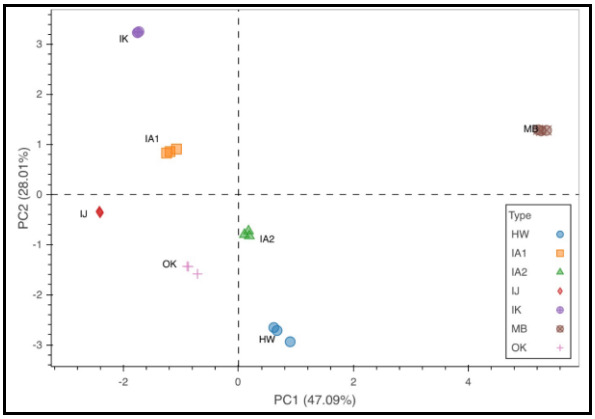

Figure 1. Score plot of PC1 and PC2 for 21 samples of KDLs

The highest RSD% of normalized peak area% was found in P5 (benzeneacetaldehyde, 251%) and P10 (4-ethyl-2-methoxyphenol, 251%), those components showed the very different area proportion while P6 (phenylethyl alcohol, 29.3%) showed the most similar area proportion by tested KDLs.

A PCA was carried out for describing the feature of KDL samples. From the PCA results, the first two principle components, PC1 and PC2 were selected with eigenvalues of more than 1, and were able to explain 75.1% of the total variance.

The biplot of PC1 and PC2 from the PCA of 21 KDLs is shown in Fig.1. In the score plot of PC2 vs PC1, MB was clearly revealed as a different pattern from any other KDLs, and characterized by positive score on PC1. Two KDLs with 20% ethanol content, LK and IA1 appeared in a similar pattern. IJ and OK were characterized as same ethanol content (25%) and aging in oak barrels for 10-years. Two KDLs with 40% ethanol content, IA2 and HW showed similar pattern. The MB clearly showed a different pattern than the rest of the KDLs.

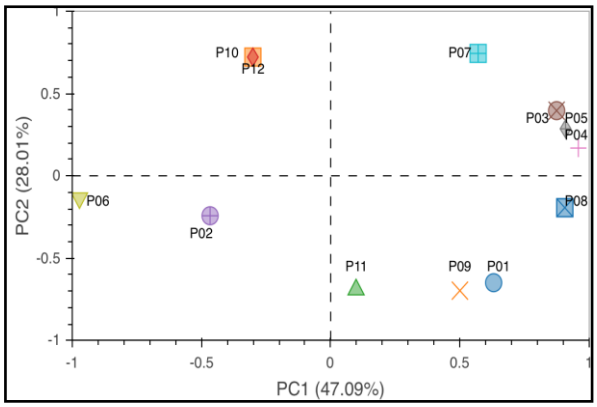

Figure 2. Loading plot of PC1 and PC2 from PCA

According to PC loading plot (in Fig.2.), associated with PC1 and PC2, 1-octen-3-ol (P3), ethyl hexanoate (P4), P5, and diethylbutabedioate (P7) were the most contributive principles characterizing MB from other KDLs. P10 and 3-methylbutyl octanoate (P12) for LK, and P1, 2-phenylethyl acetate (P9), and methyl decanoate (P11) for HW, were considered to be the distinctive components for characterizing for each KDL in the PCA.

As shown in Figure 1 and 2, the components such as P3, ethyl P4, P5, and P7 were the major volatiles distinguishing MB and other KDLs. In order to illustrate that the four selected volatile components may segregate MB from other KDLs, the biplots of variable vs variable, as intuitive method, were shown in Figure 3. MB was clearly distinguished from other KDLs as shown in Fig.3. (A) ~ (C). Therefore, these compounds (P3, P4, P5 and P7) could be the most characteristic volatile compounds in MB.

CONCLUSIONS

Pattern recognition by using statistical methods are employed to extract relevant information from the data of chemometrics. Many studies have been conducted for KDLs into investigate physicochemical and sensory characterization by using GC-MS [13, 14]. However, no research of volatile component of KDLs have yet been attempted.

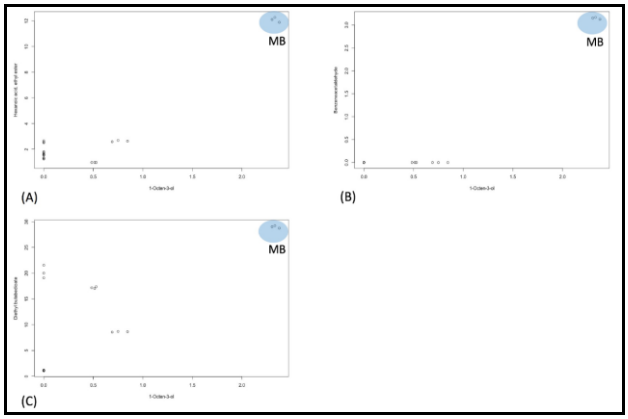

Figure 3. Scatter plot of studied KDLs using a coordinate variables the contents in (A) 1-octen-3-ol and ethyl hexanoate, (B) benzenacetaldehyde, and (C) diethyl butanedioate

In this study, 12 volatile peaks of KDL 21 samples generated by GC-FID were analyzed by PCA to find patterns in chromatographic data. The GC-FID volatile peaks of KDLs were successfully applied for the statistical characterization of KDLs. For example, four selected volatile components (P3, P4, P5 and P7) can segregate MB, and two volatile components (P2 and P6) characterized the distilled rice liquors aged in oak barrels for 10 years, such as IJ and OK.

Furthermore, a pattern recognition analysis through PCA provided the possibility facilitating classification of the volatile patterns of various KDL samples.

This pattern recognition analysis would provide a practical strategy for evaluating the quality assurance system of Korean traditional spirits. Research correlating the physicochemical and sensory characteristics of KDLs with the volatile indices established in this study will be needed to improve the quality of KDLs in the future.

ACKNOWLEDGEMENT

This research was supported by a grant (14162MFDS971) from the Korean Ministry of Food and Drug Safety in 2014.

REFERENCES

[1] Siebert, K. J., 2001, "Chemometrics in brewing – a review" Journal of the American Society of Brewing Chemists, 59(4), pp. 147-156.

[2] Ortiz, M., Saez, J., and Palacios, J., 1993, "Typication of alcoholic distillates by multivariate techniques using data from chromatographic analyses", Analyst, 118(7), pp. 801-805.

[3] Marsili, R., 2013, Sensoty-directed flavor analysis, CRC Press, NY, Chap. 2.

[4] Lee, D., Lee, D., and Lee, J., 2014, "Statistical methods for classification of medicinal plants", Journal of chemistry and chemical engineering, 8(7), pp. 698-706.

[5] Johnson, K., and Synovec, R., 2002, "Pattern recognition of jet fuels: comprehensive GCxGC with ANOVA-based feature selection and principle component analysis", Chemonetrics and intelligent laboratory systems, 60(1), pp. 225-237.

[6] Doble, P., Scndercock, E., Petocz, P., Roux, C., and Dowson, M., 2003, "Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principle component analysis and artificial neural networks", Forensic Sci. Int., 123, pp. 26-39.

[7] Pizarro, C., Esteban-Díez, I., Sáenz-Gonzáles, C., and Sáenz-Gonzáles, J., 2008, "Vinegar classification based on feature extraction and selection from headspace solid-phase microextraction/gas chromatography volatile analyses: A feasibility study", Anal. Chim. Acta., 608, pp. 38-47.

[8] Gonzáles-Arjona, D., Gonzáles-Gallero, V., Pablos, F., and Gonzáles, A., 1999, "Authentication and differentiation of Irish whiskeys by higher-alcohol congener analysis", Anal. Chim. Acta., 381, pp. 257-264.

[9] Kim, K., Kim, A., Lee, J., Chun, M., and Noh, B., 2014, "Analysis of flavor pattern of various coffee beans using electronic nose", Korean J. Food Sci. Technol., 46, pp. 1-16.

[10] Korean folk sojues and drinking culture. http://www.arumjigi.org/. Date accessed 26/06/2017.

[11] Kim, H., Jo, S., Lee, S., and Ahn, B., 2008, "Physicochemical and sensory characterization of a Korean traditional rice wine prepared from different ingredients", Korean J. Food Sci. Technol., 40(5), pp. 551-557.

[12] Lee, S., Kwon, Y., Kim, H., and Ahn, B., 2007, "Chemical and sensory characterization of Korean commercial rice wine (Yakju)", Food Sci. Biotechnol., 16(3), pp. 374-380.

[13] Van Den Dool, H., and Kratz, P., 1963, "A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography", Journal of Chromatography A, 11, pp. 463-471.

[14] Wehrens, R., 2011, Chemometrics with R: multivariate data analysis in the natural sciences and life sciences, Spriner Science & Business Media, Philadelphia, Chap. 4.