

특집: 감각과학의 이해와 산업적 활용

감각 특성의 계량: 척도와 데이터의 이해

홍 재 희

국민대학교 식품영양학과

Measurement of Sensory Responses: Understanding Scaling and Data

Jae-Hee Hong

Department of Foods and Nutrition, Kookmin University, Seoul 02707, Korea

서 론

Institutes of Food Technologists(IFT)의 정의에 따르면 감각검사(sensory evaluation)는 “시각, 후각, 촉각, 미각, 청각에 의해 감지되는 제품에 대한 반응을 불러일으키고 측정·분석하며 해석하는 과학의 한 분야”(“A scientific discipline used to evoke, measure, analyze, and interpret those responses to products that are perceived by the senses of sight, smell, touch, taste and hearing”)이다. 바꾸어 말하면, 감각검사는 인간을 측정 기구로 사용하여 감각으로 감지되는 특성이나 기호 반응, 감정적 반응을 정성적으로, 그리고 정량적으로 측정하는 것이라 할 수 있다.

최근 식품 과학 분야에서 다양한 감각적 자극, 즉 식품의 맛과 향기에 대한 인간의 인지적, 정서적, 기호적 반응에 대한 관심이 증가함에 따라 여러 감각검사 기법을 활용하여 이를 측정하고자 하는 노력이 학계와 산업계 모두에서 활발히 이루어지고 있다. 그러나 최신 검사 기법 및 그에 따른 통계 분석법에 대한 관심은 높은 데 비하여 다양한 감각적 자극을 어떻게 인간이 감지, 인식하여 반응을 산출하는지와 같은 이론적 배경에 대한 관심은 상대적으로 약한 편이다. 특히 이 부분은 국내 감각평가 관련 이론서에서 간략히 다루고 넘어가는 경우가 많으며, 강의 역시 평가 방법 위주로 진행되는 경우가 많아 상대적으로 접할 기회가 적은 부분이다.

따라서 본고에서는 감각평가를 처음 접하거나, 감각 자체의 측정에 대한 이론적 배경을 좀 더 알고 싶은 학생이나 연구원들을 위하여 감각이 어떻게 감지, 인식되어 언어나 수치로 표현되고, 어떠한 형태의 자료(data)로 산출되는지에 대한 개괄적인 정보를 제공하고자 한다.

감각의 측정 과정

인간의 감각 측정 과정을 분석 기기의 분석 과정에 비

유해 보면, 인간의 오감은 분석 기기에서 반응 값을 측정하고 이를 수집하는 입력부(input device)에, 뇌는 입력된 정보를 처리하는 중앙 정보 처리 장치(central processing unit: CPU)에, 평가 결과를 구두로 발표하거나 설문지에 기재하는 것은 출력부(output device)에 해당한다고 할 수 있다(Fig. 1). 분석 기기에서 특정 파장에서 흡광도, 전위차, 또는 전도도 등의 물리량이 중앙 정보 처리 장치에서 피크(peak)나 수치 등 실험자가 시각적으로 이해하고 분석할 수 있는 형태로 변환되듯이, 감각 기관에 의해 감지된 특성들은 뇌에서 일련의 처리 과정을 거쳐 언어로 또는 숫자 값으로 표현된다(Fig. 2).

이 과정은 크게 감각적 처리 과정(sensory process)과 인지적 처리 과정(cognitive process)으로 구분될 수 있다(1). 감각적 처리 과정은 외부의 자극이 인간의 감각 수용체와 반응, 전기 신호로 바뀌어 신경을 거쳐 뇌로 전달, 내적 표상체계(internal representation) 또는 지각 표상(percept)으로 부호화(encoding)되는 과정이다. 인지적 처리 과정에서는 내적 표상체계를 어떤 언어로 표현하고 강도를 어느 정도로 부여할지 판단하고 이를 실제로 수행하게 된다. 이를 커피 향을 평가하는 상황에 적용하면, 감각적 처리 과정은 갓 내린 커피에서 휘발된 방향성

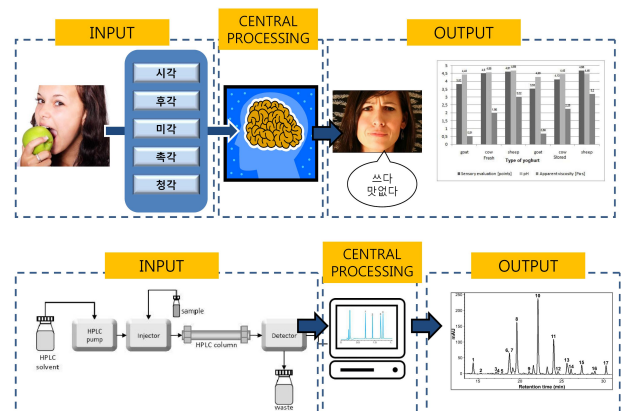


Fig. 1. Sensory information processing as the analogy between human and analytical instrument.

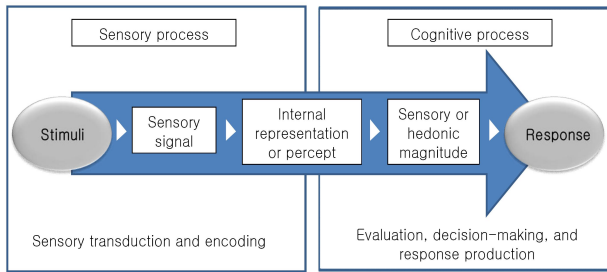


Fig. 2. Sensory stimuli-response process (1).

분자가 코의 후각 수용체에 결합하여 전기적 신호를 생성하고, 이 신호가 후각 신경을 거쳐 대뇌피질로 전달된 후 뇌에서 이에 대한 이미지(내적 표상체계)를 생성하는 단계를 포괄한다. 인지적 처리 과정에서는 이 이미지를 어떠한 단어로 표현할지, 그리고 어느 정도 강도인지 판단한 후, 이를 말로 표현하거나, 강도를 숫자로 적거나, 척도 위에 표시하는 작업을 하게 된다. 따라서 감각평가 시 어떠한 과제를 평가원에게 수행하게 하는가, 즉 어떠한 평가 방법과 척도(scale)를 사용하게 하는가에 따라 다른 형태의 데이터를 산출할 수 있으며, 그에 대한 통계적 분석 방법이 달라진다.

감각검사에서 산출되는 데이터의 종류

명목 자료

통계학에서는 데이터의 형태를 4종류로 구분한다(Table 1)(2). 일반적으로 숫자가 정량적인 의미를 가지는 것이 아닌 부호 또는 표식(label)으로 사용되는 경우 이를 명목 자료(nominal data)라 칭한다. 감각평가 시 검사원의 인구통계학적 정보를 구하기 위하여 성별을 조사할 때 남성을 1, 여성을 2로 표시하는 것이 대표적인 명목 자료라 할 수 있다. 이 경우 1과 2라는 값은 수치로 사용되는 것이 아니고 성별을 구분하기 위한 표식으로 사용되므로, 이들의 평균값 등을 구하는 것은 의미가 없으며 각 표식이 출현한 빈도수를 세어 조사 검사원의 성별 구성비를 조사할 때 더 의미 있는 결과를 얻을 수 있다. 바꾸어 말하면, 명목 데이터(성별)는 관심 대상(검사원)의 특정한 성질의 크기를 비교하거나 측정하는 데 사용되는 것이 아니고 같은 성질을 가진 집단끼리 범주화(categorization)하

기 위해 사용된다.

감각검사에서 차이 식별을 위해 사용하는 삼점 검사(triangle test), 일-이점 검사(duo-trio test) 결과 역시 명목 자료로 산출된다. 즉, 각 시료는 제시된 3개 시료 중 다른 것인지, 아니면 같은 2개 중 하나인지로 구분(범주화)될 수 있다. 최근 감각평가에서 활발히 사용되는 check-all-that-apply(CATA) 방법 역시 명목 자료를 산출하는 검사 기법이다. CATA 방법은 기존의 정량적 묘사분석(quantitative descriptive analysis: QDA) 대신 소비자들로부터 제품 감각 특성의 프로필을 구하는 방법으로(3), 검사원들은 감각 특성의 리스트를 제공받고, 그 리스트에서 제품별로 감지되는 모든 특성을 체크한다. CATA는 제품의 특징적인 향미나 강한 감각 특성은 뚜렷이 감지될 수 있으므로 많은 검사원에 의해 체크되어 빈도수가 높게 나타날 것이고, 약한 특성들은 민감한 소수 검사원에 의해서만 감지되므로 빈도수가 낮게 나타날 것이라는 가정에 기초하고 있다. 검사 결과는 각 감각 특성이 체크된 경우를 1, 체크되지 않았을 경우를 0으로 코딩하여 정리한다(Fig. 3). 따라서 제품별로 감각 특성을 검사원에 의해 선택된 것과 선택되지 않은 것으로 범주화하는 명목 자료가 생성된다. 또한, 기호도 검사에 흔히 동반되는 개방형 질문(open-ended question)이나 단어 연상법(word association task) 등을 통해 정형화되지 않은 텍스트를 얻고, 이러한 텍스트로부터 관련 단어를 분리, 출현 빈도수를 계산하여 단어 간의 관계성이나 제품 묘사에 가장 빈번히 사용되는 묘사 용어를 파악하는 텍스트 마이닝(text mining) 역시 명목형 자료를 산출하는 검사 기법으로 활용되고 있다.

순서 자료

명목 자료 외 표식으로 사용된 숫자가 논리적인 순서가 있는 경우, 즉 1등, 2등, 3등과 같은 등수, 또는 A, B, C, D와 같은 성적 등은 순위 자료(ordinal data)로 구분된다. 순위 자료는 값들의 크기에 대한 방향성을 보여주지만, 값 간의 간격이 일정하거나 비례 관계가 존재하지 않는다. 1, 2, 3으로 표현된 등수에서 1은 3보다 앞서지만, 3은 1의 3배가 아니다. 감각검사 기법 중 특성 강도가 강한 순서에 따라, 혹은 좋아하는 정도에 따라 순위를 정하는 순위 검사법(ranking test)을 시행할 경우 순위 자료

Table 1. Types of data (2)

Classifications		Examples	Test statistics and measures
Qualitative	Nominal	Male (1), female (2) Soccer (1), basketball (2), baseball (3) ...	Non-parametric; mode, frequencies
	Ordinal	Grades (A,B,C,D,F) Education levels (high school, undergraduate, graduate ...)	Non-parametric; Median, percentiles
Quantitative	Interval	℃, °F	Parametric; Mean, standard deviation
	Ratio	Height (cm), weight (kg), length (m, mile)	Parametric; Log mean, standard deviation

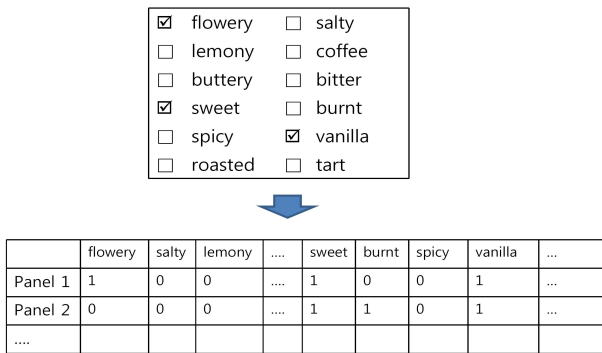


Fig. 3. Check-all-that-apply questionnaire and data coding (3).

를 얻을 수 있다. 명목 자료와 순위 자료는 자료의 값이 아닌 내용이 중요하므로 질적 자료(qualitative data)로 분류되며, 질적 자료는 대부분 해당 값이 얼마나 많이 출현하는지 빈도수를 측정하여 확률을 계산, 통계학적 검정을 하는 비모수적 통계 방법(non-parametric statistics)을 적용하여 분석한다. 감각과학에서 흔히 사용되는 비모수적 통계 방법으로는 차이 식별 검사에서 다른 시료(odd sample)가 선택된 확률이 무작위적으로 한 시료를 선택할 확률보다 유의적으로 큰지 검증하기 위한 binomial test, same-difference test에 사용되는 Chi-square test나 McNemar test, 순위 검사 결과를 분석하기 위한 Wilcoxon signed test, Kruskal-Wallis test, Friedman rank sum test 등이 있다.

등간 자료

질적 자료는 양적인 정보, 즉 특성의 강도나 좋아하는 정도가 얼마만큼 강한지에 대한 정보는 제공하지 않는다. 예를 들어, 순위 검사법으로 제품 특성의 강도를 비교할 경우, 한 제품이 다른 제품보다 강한지 약한지 알 수 있지만, 그 차이가 어느 정도인지는 알 수 없다. 따라서 더욱 정량적인 정보를 수집하기 위하여 간격 척도(interval scale) 또는 비율 척도(ratio scale)를 사용하며, 이들 척도를 이용하여 수집된 자료를 양적 자료(quantitative data)라고 칭한다.

간격 척도 또는 등간 척도(interval scale)에서 얻어진 수치는 일정한 간격을 가지고 있으며, 값 간에 비례적인 관계가 성립한다. 섭씨 또는 화씨온도는 등간 척도의 대표적인 예이다. 우리는 온도를 측정하고 분석할 때 일반적으로 5도에서 10도까지의 간격이 10도에서 15도 사이의 간격과 같으며, 20도는 10도의 2배라는 것을 전제로 받아들인다. 비율 척도로 측정된 자료 역시 등간 자료와 같이 값 간의 간격이 동일하면 비례적 관계가 성립하지만, 등간 척도의 0은 실제 0이 아닌 어떤 임의의 기준점인 반면, 비율 척도에서의 0은 실제의 0, 측정량의 부재를 나타낸다는 점에서 차이를 보인다. 등간 척도의 대표적인 예인 섭씨와 화씨온도 척도에서 섭씨 0도는 물이 어는

온도, 화씨 0도는 얼음, 물, 염화암모늄(NH₄Cl)을 같은 비율로 섞어 열평형에 도달했을 때 온도를 나타내지만, 실제 열량이 부재함을 의미하는 것은 아니다. 그러나 비율 척도의 대표적인 예인 길이, 질량 등에서 0 cm, 0 kg 등은 실제 길이나 질량이 없음을 의미한다. 이와 같은 양적 자료들은 자료가 정규분포를 한다는 가정에 기반을 두는 모수적 통계 방법(parametric statistics)을 적용하여 분석하며, 데이터를 대표하는 값으로는 평균을 구하여 사용한다. 대표적인 모수적 통계 방법으로는 t 검정, 분산분석(analysis of variance, ANOVA) 등이 있다.

감각평가에서는 다양한 척도를 사용하는 평점법(rating test)을 실시하여 양적 자료를 구한다. 가장 널리 사용되고 있는 것은 항목 척도(categorical scale), 선 척도(line scale)이다(Fig. 4). 항목 척도가 양적 데이터, 즉 등간 자료를 산출하기 위해서는 척도상의 강도를 표시하는 용어(‘강한’, ‘약한’, ‘매우 좋은’ 등)가 모든 검사원에 의해 같은 의미로 받아들여져야 하며 각 용어에 대한 심리적 반응이 같은 분산을 가진 정규 분포를 할 수 있어야 한다.

감각 검사에 사용되는 척도 중 가장 많이 알려진 9점 기호도 척도는 1952년 실험심리학자인 Peyram과 Girardot에 의해 도입되었고 도입 즉시 광범위하게 사용되기 시작하였다(1). Jones 등(4)은 9점 척도 점수 간의 심리적 간격을 검증하였는데, 이 연구에서는 51개의 강도를 나타내는 형용사 목록을 900명의 검사원에게 제시하고, -4에서 4까지의 점수(중앙값은 0점)로 나타낸 척도 위에 각 형용사가 어느 점수에 해당하는지 표시하게 하였다. 모든 형용사 출현 빈도수에 대한 형용사별 누적 출현 빈도 비율을 구한 후 이를 z 값으로 변환하였다. 변환된 z 값을 y축, -4~4의 각 점수를 x축으로 하여 회귀선을 구한 후 중앙값(median)에 해당하는 점수를 구해 이 점수가 각 형용사의 대표 점수가 되도록 하였다. 이렇게 하여 “극도로 싫다/좋다(dislike/like extremely)”는 각각 -4.32,

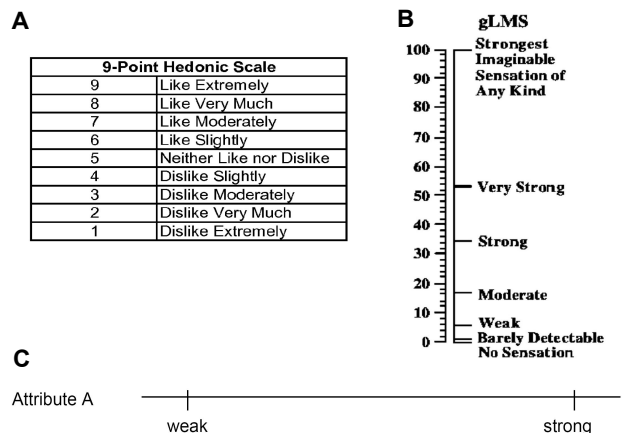


Fig. 4. Interval and ratio scales. A: a 9-point categorical hedonic scale, B: generalized labeled magnitude scale, C: line scale.

4.16, “좋지도 싫지도 않다(neither like nor dislike)”는 0에 해당함을 검증하였다. 그러나 “좋지도 싫지도 않다”에서 “약간 싫다/좋다(dislike/like slightly)”까지의 거리는 “매우 싫다/좋다(dislike/like very much)”에서 “극도로 싫다/좋다(dislike/like extremely)”까지의 거리의 약 1/3 정도인 것으로 나타나, 점수 간의 심리적 거리가 같은 간격이 아니라는 것이 밝혀졌다(4). 따라서 일부 학자들은 항목 척도를 사용해 구해진 데이터가 등간 자료가 아닌 순위 자료로 취급되어야 하며, 관습적으로 사용되는 모수적 통계 방법이 아닌 비모수적 통계 방법으로 분석되어야 한다고 주장하기도 한다. 그러나 산업계나 학계에서는 이를 등간 자료로 취급, t-test나 ANOVA로 결과를 분석하였을 때 그 결과가 현상을 비교적 잘 설명하므로 기존의 연구 방법을 수용하고 있는 편이다.

비율 자료

비율 자료는 크기 추정(magnitude estimation, ME) 방법에 의해 산출된다. 이 방법은 감각 반응을 정량적으로 측정하여 자극(stimuli)과 반응 간의 관계를 수학적으로 모델링하기 위하여 개발되었다. 크기 추정 방법에 의해 산출된 자료는 Stevens의 power law를 따르는 것으로 관찰된 바 있다(Fig. 5). 빛의 밝기를 감지할 때는 광선의 강도가 증가할수록 더 밝게 느끼지만, 자극의 강도가 증가할수록 반응 값이 증가하는 정도가 점점 작아진다. 이와 같은 상관관계는 지수식으로 표현되며, 이때의 지수 값은 1보다 작다. 반면, 전기 자극의 경우는 자극의 강도가 증가할수록 반응 값이 증가하는 정도가 더 커지며, 이때의 지수 값은 1보다 크다(Fig. 5). ME는 감각과학에서 두 가지 감각의 강도를 매칭하는 데에도 활용되고 있다(cross-modality matching). 예를 들어 빛의 강도를 특정 소리의 강도에 대한 비율로 나타낼 수 있는데, 이를 확장하면 맛이나 냄새 등을 물리적 단위를 가진 소리(db)나 길이(cm)와 매칭시켜 감각 특성에 물리적 단위를 부여할 수 있다. 모든 감각 특성을 하나의 공통 척도로 측정, 비교하게 하는 스펙트럼 묘사분석은 이러한 cross-modality matching에 그 이론적 기반을 두고 있다. 따라서

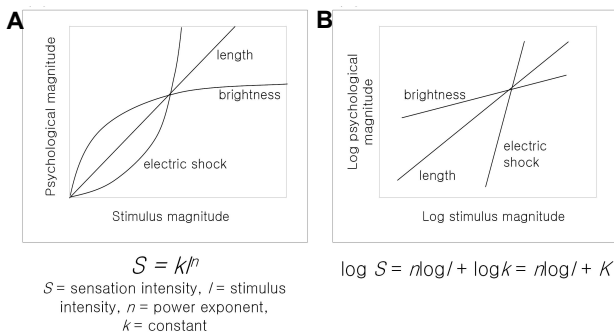


Fig. 5. Psychological functions for three perceptual continua following Steven's power law. A: functions expressed in a linear scale, B: functions expressed in a log scale.

스펙트럼 묘사분석에서는 시료의 단맛이 6점, 짠맛이 3점으로 측정되었다면 단맛이 짠맛의 2배 강도라고 해석한다(5). 나아가 검사원 2명이 크기 추정을 활용하여 짠맛은 같은 점수로, 매운맛에 대해서는 각각 다른 점수로 강도를 표시하였다면, 이는 두 사람의 매운맛에 대한 개인 차이를 반영하는 것이라 해석한다. 따라서 크기 추정을 옹호하는 학자들은 이 방법이 약물 치료군과 대조군이 느끼는 통증을 비교하여 통증 완화 효과를 판정하거나, 노인이나 환자의 미각과 후각이 정상인과 비교하여 어떻게 달라지는지 연구하는 데 활용될 수 있다고 주장한다(6).

ME에서는 하나의 자극을 검사원에게 제시하고 임의의 값을 부여하게 하거나, 사전에 정해진 값(5, 10, 25, 100 등)을 부여한 modulus라고 하는 표준 시료(reference)를 제공한다. 검사원들은 처음에 평가한 자극 또는 modulus에서 감지되는 자극의 수치에 비례하여 다른 자극의 강도를 수치화한다. 즉, 5% 설탕 수용액의 단맛을 10이라고 평가하였을 때 평가할 시료의 단맛이 그의 절반 강도였다면 5, 두 배의 강도라면 20을 부여한다. 그러나 이 방법은 특성 강도를 비율로 표현하므로 전반적으로 이 특성이 강한지 약한지를 알 수 없다는 단점이 있다.

Borg(7)는 육체적 피로도를 측정하기 위하여 비율 척도 위에 강도 표현 용어를 표시한 “항목-비율 척도(category-ratio scale)”를 사용하였다. Green 등(8,9)은 Borg의 연구를 확장하여, 여러 특성 강도 표현 용어를 제시하고 용어에서 감지되는 강도를 magnitude estimate 방법으로 추정하도록 한 후, 산출된 값과 그에 해당하는 용어를 수직선 위에 표시한 labeled magnitude scale(LMS)을 개발하였다(Fig. 4). 등간 척도와 달리 LMS에서는 약한(weak), 보통의(moderate), 강한(strong), 매우 강한(very strong) 간의 간격이 로그(log) 간격과 유사하게 배치되어 있다. 또한, LMS는 척도의 가장 강한 쪽을 “상상할 수 있는 가장 강한(strongest imaginable)”이라고 표시하였다. 이 역시 Borg(7)의 연구 결과에 기반한 것으로, Borg는 연구 참여자들이 최고 피로도에 이를 때까지 자전거를 타게 하면 그때 연구자들이 느끼는 피로 정도는 서로 비슷하리라 추정하였다. 따라서 “상상할 수 있는 가장 강한”이라는 표현 용어를 사용하게 되면 검사원들이 비슷한 수준의 강도를 연상하게 될 것이므로 평가 기준이 동일해질 것이라고 주장한다(6,10). 일반적으로 LMS를 사용할 경우 “상상할 수 있는 가장 강한 입안 감각(strongest imaginable oral sensation)”으로 척도의 가장 높은 부분을 표시하는데, 이는 매우 감각과 같은 통각까지 아우르게 하여 더욱 넓은 척도 범위를 사용하게 하기 위함이다.

묘사분석이나 기호도 검사에서 흔히 사용되는 항목 척도나 선 척도에 비해 LMS는 비율 자료를 산출할 수 있으며, 또한 매우 강렬한 감각 경험을 척도의 최상위 부분의 정박점(anchor)으로 사용함으로써 항목 척도에서 흔히

발견되는 천정 효과(ceiling effect: 감각의 강도가 척도의 최고점 이상으로 감지되어도 응답자가 주어진 척도를 활용하여 평가해야 하므로 최고점 이상의 점수를 줄 수 없게 되는 현상)를 피할 수 있는 장점이 있다고 주장된다(6,10). 그러나 실제로 ME나 LMS가 비율 자료를 산출하는지에 대해서는 학자들 간에도 논란이 있다. 길이와 같은 명확한 시각적, 물리적 자극이 아닌 맛과 향기와 같은 화학적 자극을 감지하고, 이를 비율 자료로 정량화하는 것은 인간에게 있어서 쉬운 작업이 아니다. 특히, 2배나 1/2배와 같은 명료한 비율이 아니고 시료의 단맛이 modulus 또는 처음 맛본 시료의 단맛의 1.8배로 산출되었다면, 이는 검사 요원들이 단맛의 강도가 1.8배임을 실제로 감지하고 산출한 수일까, 아니면 단순히 “2배보다는 약간 낮은”의 개념으로 산출한 수일까? 이에 대하여 실험을 통한 완전한 검증 역시 아직 이루어지지 않은 상태이다(5).

척도 사용 시 고려 사항

산업계와 학계에서 감각 특성을 측정하고 계량하기 위해서는 등간 척도 또는 비율 척도를 흔히 사용한다. 이는 등간 척도나 비율 척도를 통해 얻은 양적 자료가 질적 자료보다 더 많은 정보를 가지고 있기 때문이다. 차이 식별 검사나 순위 검사를 실시하면 시료 간의 차이 유무나 어떤 시료가 단맛이 더 강한지를 알 수 있으나 그 차이의 정도가 얼마만큼 인지하는 알 수 없다. 반면 척도를 사용하여 측정하면 단맛은 실제로 몇 점이고, 또 다른 시료에 비해 얼마나 더 강한지를 알 수 있다. 또한, 실무적으로도 차이 식별 검사의 프로토콜이 두 시료 간의 차이를 식별하는 데 초점이 맞추어져 확립되어 있으므로, 여러 시료의 차이를 파악하기 위해서는 여러 번의 차이 식별 검사를 수행하여야 한다. 따라서 검사의 목적이나 예민도에 대한 고려보다는 비용이나 시간 등 현실적 문제점을 우선으로 고려하여 척도를 사용한 검사를 수행하는 경우가 빈번하다.

도출된 자료를 분석하여 결과를 얻고 그 결과를 올바르게 해석하여 의미 있는 결론을 얻기 위해서는 척도에 대한 이해가 필요하다. 척도는 인간이 감지하는 특성의 강도는 하나의 연속선상에서 증감한다는 이론적 모형(unitary psychophysical continuum)에 근거하여 개발되었다. 실제로 빛, 휘발성 성분, 맛 성분, 음파 등 자극의 물리적 강도나 농도가 증가할 경우 우리가 감지하는 밝기나 냄새, 맛, 소리 역시 증가한다. 자극의 강도와 감지된 감각의 강도 간의 관계를 구한 여러 수학적 모형이 개발된 바 있으며(11), 이러한 모형들이 감각 특성 계량의 이론적 타당성을 제공하고 있다.

그러나 Fig. 2와 같이 감각 특성을 계량하기 위해서는 감각적 처리 과정과 인지적 처리 과정을 거쳐야 한다. 이때 각 처리 과정은 여러 요인에 의해 영향을 받을 수 있다.

특히, 인지적 처리 과정에서 인간은 감각을 통해 수집된 정보 자체, 즉 sensory input만을 평가하는 데 그치지 않고 감각 정보를 처리하는 데 주변 환경, 사전에 형성된 경험이나 지식 등을 적극적으로 활용한다(top-down processing)(12,13). 예를 들어, 0.5% 소금 농도를 가진 국의 짠맛 강도를 평가할 때 평소 짜게 먹어온 평가원과 싱겁게 먹어온 평가원의 응답은 다를 수 있다. 이는 각 평가원이 국의 짠맛에 대해 사전에 형성된 경험 범위에 기반을 두어 짠맛의 강도를 판단하기 때문이다. 또한, Lawless(14)의 연구에 따르면 0.25% 소금 농도를 가진 국의 짠맛을 0.12~0.25% 범위의 소금 농도를 가진 국 시료들과 같이 평가했을 때, 0.25~0.5% 범위의 소금 농도를 가진 국 시료들과 같이 평가하였을 때보다 점수가 유의적으로 더 높아졌다. Lawless(14)는 이와 같은 결과를 평가원이 제공된 시료의 짠맛 범위라는 “맥락(context)”을 고려하는 context effect에 의한 것으로 해석하였다.

또한, 자극 자체만을 평가하기보다는 맥락을 고려하는 경향에 의해 실험 조건이나 환경이 응답에 영향을 주어 원래의 응답과 다른 편차(bias)를 유발하는 여러 종류의 response bias가 발생할 수 있다(15,16). 그중 하나인 centering bias는 평가원들이 제시된 시료 중 중간 강도를 가진 시료를 척도의 가운데에 위치시키는 경향이다(17). Centering bias는 특히 적합 척도(just-about-right scale)를 사용하는 경우 문제가 될 수 있는데, 평가원들이 실제 적합한 강도 범위와는 무관하게 제시된 시료 세트 중 중간 강도를 가진 시료가 척도의 가운데, 즉 “적당한” 강도를 가진다고 평가하는 경향을 보이기 때문이다. Dumping effect는 흔히 발생하는 response bias의 또 다른 예이다. 시료를 평가할 때 평가원이 감지한 감각 특성이 설문지에 기재되어 있지 않으면 해당 특성에 대한 반응을 다른 특성을 평가하는 부분에 “투척(dump)”하게 되어 평가 점수가 인위적으로 증가하게 되는 경향을 의미한다. 실제로 평가원들에게 과일 향기와 단맛을 평가하는 항목이 있는 설문지를 제공하고 단맛을 평가하게 하였을 때보다 단맛 항목만을 포함한 설문지를 제공하였을 때 평가원들이 과일 향기에 대한 응답을 단맛 항목에 투척함으로써 단맛의 평가 점수가 높아지는 현상이 관찰되었다(18).

학계나 산업계 모두 어떠한 환경에서도 동일 시료에 대해서는 항상 동일한 측정치(absolute value)를 산출하는 검사 기법에 대해 고민하여 왔다. 절댓값을 측정할 수 있다면 과거와 현재의 감각 특성 또는 기호도의 변화를 추적하여 품질 관리에 활용하거나, 각 개인 간의 감각 인지의 차이를 조사하는 데 유용하게 사용될 수 있다. 그러나 절댓값을 측정하기 위해서는 인지된 자극만을 환경이나 맥락과 분리하여 단독으로 평가할 수 있어야 하지만, 이는 맥락을 고려하는 인간의 감각 인지 기전(sensory perception mechanism)에서는 가능하지 않다. 절댓값을 산출한다고 주장했던 스펙트럼 묘사분석이나, 스펙트

럼 묘사분석의 이론적 타당성을 제공한 ME나 LMS 역시 context effect나 response bias로부터 자유롭지 않음이 증명된 바 있다(19).

따라서 척도를 이용하여 감각을 측정할 때에는 이러한 인간 고유의 정보 처리 기전에 대한 이해를 바탕으로 context effect와 bias를 최소화하는 여러 전략에 대해 고민할 필요가 있다. 또한, 결과를 해석할 때에도 다른 환경이나 실험 조건에서 측정된 자료를 분석할 때에는 검사 수행의 맥락이 같은지 검토하여야 한다. 만약 작년 3월 미국에서 5개의 비교 시료와 함께 측정된 초콜릿의 단맛 점수가 7점이고, 올해 9월 한국에서 다른 7개의 비교 시료와 함께 측정된 값이 5점이었다면, 초콜릿의 단맛이 감소하였으니 당 함량을 올려야 한다고 해석하는 것이 타당할 것인가? 앞서도 강조하였듯이 검사의 맥락과 환경, 그리고 감각 특성이 어떻게 수치화되는지에 대한 이해가 없다면 부정확한 결과를 도출할 수 있다.

결론

감각과학의 여러 검사는 기기 분석을 활용하는 측정 방법보다 단순히 보이기 때문에 단기간의 훈련 과정을 거쳐 바로 수행하는 경우가 빈번하다. 그러나 원리나 수행 방법에 대한 깊은 이해가 수반되지 않는다면, 척도를 이용하여 산출된 결과가 실질적으로 인간이 감지한 특성의 강도를 반영하지 못할 수 있다. 대표적인 척도의 오용 사례로는 각 항목의 성질을 나타내는 표식(label)을 연구자의 편의에 맞게 변형시킨 유사 항목 척도(pseudo-numerical scale)가 있다. 예를 들어, 제과 제빵 제품 평가 시 10점은 “완벽한 조직감(perfect texture)”, 8점은 “약간 딱딱한 정도(slight dryness)”, 6점은 “검과 같은 느낌(gumminess)”을 주게 할 경우, 단일한 특성의 강도 변화(unitary psychophysical continuum)라는 척도의 기본 심리학적 모형을 위반하게 되므로 엄연히 말해서는 척도가 불리기 어려우며(5) 그 결과를 등간 자료로 취급해서는 안 된다. 따라서 현상에 더욱 접근한 자료를 산출하여 바르게 해석하기 위해서는 단순히 각 검사의 프로토콜뿐 아니라 그러한 프로토콜이 개발된 배경 지식과 이론을 이해하려는 노력이 선행되어야 할 것이다.

참고문헌

- Lim J. 2011. Hedonic scaling: A review of methods and theory. *Food Qual Prefer* 22: 733-747.
- Zar JH. 1999. *Biostatistical Analysis*. 4th ed. Prentice Hall, Upper Saddle River, NJ, USA. p 1-5.
- Valentin D, Chollet S, Lelièvre M, Abdi H. 2012. Quick and dirty but still pretty good: a review of new descriptive methods in food science. *Int J Food Sci Technol* 47: 1563-1578.
- Jones LV, Peryam DR, Thurstone LL. 1955. Development of a scale for measuring soldiers' food preferences. *J Food Sci* 20: 512-520.
- Lawless HT, Heymann H. 2010. *Sensory Evaluation of Food: Principles and Practices*. 2nd ed. Springer, New York, NY, USA.
- Bartoshuk LM, Fast K, Snyder DJ. 2005. Differences in our sensory worlds: invalid comparisons with labeled scales. *Curr Dir Psychol Sci* 14: 122-125.
- Borg G. 1982. A category scale with ratio properties for intermodal and interindividual comparisons. In *Psychophysical Judgment and the Process of Perception*. Geissler HG, Petzold P, eds. North-Holland Publishing Company, New York, NY, USA. p 25-34.
- Green BG, Shaffer GS, Gilmore MM. 1993. Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chem Senses* 18: 683-702.
- Green BG, Dalton P, Cowart B, Shaffer G, Rankin K, Higgins J. 1996. Evaluating the 'Labeled Magnitude Scale' for measuring sensations of taste and smell. *Chem Senses* 21: 323-334.
- Bartoshuk LM, Duffy VB, Green BG, Hoffman HJ, Ko CW, Lucchina LA, Marks LE, Snyder DJ, Weiffenbach JM. 2004. Valid across-group comparisons with labeled scales: The gLMS versus magnitude matching. *Physiol Behav* 82: 109-114.
- Luce RD, Kruschke AL. 1988. Measurement, scaling, and psychophysics. In *Steven's Handbook of Experimental Psychology*. Atkinson RC, Herrnstein RJ, Lindzey G, Luce RD, eds. John Wiley & Sons, Inc., New York, NY, USA. p 3-73.
- Orne MT. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Am Psychol* 17: 776-783.
- Smith EE, Kosslyn SM. 2007. *Cognitive Psychology: Mind and Brain*. Pearson/Prentice Hall, Upper Saddle River, NJ, USA.
- Lawless HT. 1983. Contextual effect in category ratings. *J Test Eval* 11: 346-349.
- Furnham A. 1986. Response bias, social desirability and dissimulation. *Pers Individ Dif* 7: 385-400.
- Meilgaard MC, Civille GV, Carr BT. 2007. *Sensory Evaluation Techniques*. 4th ed. CRC Press, Boca Raton, FL, USA.
- Poulton EC. 1989. *Bias in Quantifying Judgments*. Lawrence Erlbaum Associates, Hove and London, UK.
- Frank RA, van der Klaauw NJ, Schifferstein HN. 1993. Both perceptual and conceptual factors influence taste-odor and taste-taste interactions. *Percept Psychophys* 54: 343-354.
- Diamond J, Lawless HT. 2001. Context effects and reference standards with magnitude estimation and the labeled magnitude scale. *J Sens Stud* 16: 1-10.