

信号検出理論の官能検査への応用

畑 江 敬 子*

官能検査とは、味、におい、テクスチャー、色などを人間の感覚で測定する方法である。官能検査の最終的な目標は消費者の嗜好に合った商品を市場に出すことにある。

官能検査に関しては多くの秀れた文献¹⁻⁵⁾があるので、ここでは、どこで何故、官能検査が必要とされるのか等について簡単にふれるに止める。なお、本稿では食物を念頭において話を進める。

信号検出理論は元来ミシガン大学の通信工学の研究者たちが中心になって構築したもので、心理学^{註1)}の分野でも利用されていた。近年、官能検査への応用も試みられている⁶⁻⁹⁾。そこで、信号検出理論について簡単な背景と応用例を紹介したい。

註1) 心理学の方法論には計量心理学と数理心理学の2領域がある。心理現象の解明に統計的方法を用いる計量心理学に対して、数理心理学は数理モデルを用いて現象にせまっていこうとする領域である。信号検出理論はより数理心理学的なものであるとされている¹⁰⁾。

I. どこで官能検査が応用されているか

食物の感覚的な品質特性に対する人間の応答を知りたいときに官能検査が必要となる。それは以下のような場合である。

(1) 消費者の嗜好を把握する

ある商品に対する消費者の嗜好を調べたり、A、B 2社の類似した商品に対する消費者の嗜好に差があるかどうか調べたりするのに、官能検査を行う。

さらに新しい商品を市場に出すにあたっては、はたしてそれが消費者に受け入れられるかどうかを知る必要がある。そこで、予め官能検査を行って消費者の反応を予測する。

消費者がどのような傾向の商品を好むかを調べることもある。

(2) 製品の品質管理

生鮮食品の品質にある程度バラツキがあっても、消費者は止むを得ないと思っている。しかし、加工した商品に対しては、一定の品質を保っていることを要求する。

どの程度の原材料の品質のバラツキ、材料配合の変化、製造条件の変化などが、品質の差として消費者に識別されるか、また、どの程度の貯蔵変化が識別されるかを知るために官能検査を行う。たとえば、これまで使用していた原料の入手が困難になった場合に、安価な原料をどのぐらいまで置き換えても品質の差が消費者に識別されないか、コストを下げるために熟成期間を短縮したいが、どの程度までなら消費者には区別できないか、などを知るために官能検査が行われる。

(3) 人間の感覚と化学特性や物理特性との関係を知るための手段として

人間が好ましいと感じる食物の呈味と、その食物の呈味成分との関係が明らかになれば、呈味成分をコントロールすることによって、好ましい製品を作ることができると。たとえば、熟成によって好ましい呈味成分が形成されるような製品において、熟成により増加あるいは減少する成分と人間の感覚との関係がわかれば、熟成の条件をコントロールして熟成期間を短縮する方法を模索することが可能となる。

(4) 商品の格付けや規格

米、茶、ウィスキー、ワイン、肉などの格付けは熟練者の感覚によって行われている。

(5) その他

化学物質と身体の相互作用の関係をj知る。呈味物質は味蕾の味細胞の表面に存在する受容体(味覚受容タンパク質)と相互作用して味の感覚を誘起するのであるから、化学物質に対する人体の反応を調べることができる。

また、呈味は人間の食物選択に影響を与えているから、何故それを好むかがわかれば食欲のコントロールの手が

* お茶の水女子大学

かりになる。

II. それは何故か

機器の応答と人間の応答は必ずしも同じではないからである。それは以下のような理由による。

(1) 嗜好の判断は人間にしかできない

機器では食物あるいは商品に対する人間の嗜好をはかることはできないので、官能検査に頼らざるを得ない。

(2) 人間の感覚は同時に多くの情報をとり入れる

人間は多数の旨味成分から味の情報を得る。多数の物理的特性からテクスチャーの情報を得る。そしてこれらを総合して、おいしい、まずいと判断している。このように「総合する」ことが人間の感覚の特徴である。しかも総合に要する時間は極めて短時間、瞬時である。このようなことは機器にはできない。

(3) 人間では多くのメカニズムが同時に働く

人間ではある感覚が他の感覚の影響を受けることがある。例えば、バナナの匂いのする液体に赤い色をつけると消費者はイチゴの匂いと間違えることがある。

(4) 含有成分量と人間の感度は平行するとは限らない。

呈味成分の含有量が2倍になったからといって、人間の感覚は必ずしも2倍強くなったと認識するとは限らない。

一方、物理特性が呈味の感覚に影響を与えることも知られている。蔗糖濃度が同じであっても軟らかい食物より固い食物の方を甘味が少ないと感じることが認められている。このような情報は蔗糖量の分析結果からは得られない。

(5) 人間の感度と機器の感度は必ずしも同じではない。

機器測定によっては区別できない微妙な差を人間の感覚が識別することがある。その逆もあり、両者の感度は同じではない。

(6) 人間の感覚の感度は変化する。

人間の感覚の感度は、環境条件や心理的、生理的条件によって変化する。同一パネル員のこのような条件による変動をなるべく少なくしたいときには訓練を行う。この訓練は測定機器の較正と同じと思えばよい。加齢や病気によっても感度や嗜好は変化する。条件によってどのように変化するかを知ることが必要である。例えば加齢によってどのように嗜好が変化するかを知ることが必要になることがある。

III. 官能検査の三種のタイプ⁶⁾

試料をテストするタイプと人間をテストするタイプとがある。

(1) センサリー I

大勢の中から選別され、特定の食物に対する訓練を受けたパネルが微少な差の識別のために官能検査を行う。パネルをあたかも機器のように取り扱う。従って、感度の良いパネル員を感度のよい機器とみなし、ごく少数のパネル員が食物の品質の識別試験を繰り返し行い、有意に識別された場合に両者に差があると結論づける。

検査に際してはある特定の品質のみに集中して判定できるように、それ以外の品質の影響をできるだけ排除するような条件を設定する。例えば、味の識別のためには、においの影響を除くべく鼻のクリップを使用することもあるし、テクスチャーの影響を除くために細かく刻んで供することもある。色の影響を避けるために赤いライトの下で検査を行うこともある。

パネルは1~10名ぐらいで構成する。1人のパネル員に何回も繰り返し検査を行い、有意に識別できたかどうかを検定する。さらに有意に識別できたパネル員の数が有意かどうかを検定する。試料にバラツキの可能性がある場合や1人のパネルがすべての品質に対して感度がよいとは限らないので複数の品質をテストする場合は、さらにパネル員を増やした方がよい。しかし、単に人数を増やすだけでは、感度のよいガスクロマトグラフィーの結果と感度の悪いガスクロマトグラフィーの結果を平均することになってしまう場合があるので注意を要する。

センサリー I で識別されない程度の微少な差は、それ以上の検査の必要はない。なお、センサリー I のパネルに嗜好試験をするのは意味がない。

(2) センサリー II

センサリー I のパネルが識別できた試料間の差を、消費者が識別できるかどうか知るために、より消費者に近いパネルで官能検査を行う。パネルはセンサリー I のような選別や、特定の商品に対する特別な訓練を必要としないが、官能検査の一般的な手法の訓練や、ある程度の商品に対する知識は必要とされる。会社内で、その商品の開発に携わっていない、一般の社員などでパネルを構成する。

検査は、通常、その試料を消費する条件に出来るだけ近いところで行い、特別な照明や鼻のクリップの使用、味やにおい等の特定の品質だけに集中させるような質問は、あまり適当とはいえない。二点比較法やトライアングルテストなどの方法が適している。

パネルの数は100~200名が望ましい。識別できたパネルの数によって有意であるかどうかを検定を行う。

消費者の嗜好を知るための予備テストとして嗜好テストを行ってもよいが、センサリー II のパネルに嗜好テストをすることは適当ではない。

(3) 消費者テスト

センサリーIIで差があるとされた場合に、全く訓練を受けていない消費者を対象に大規模な官能検査を行って、消費者の嗜好を調べる。質問は、先ず嗜好を尋ねた後に識別できるかどうか尋ねた方がよい。

パネルは全く訓練を受けていないのであるから、二点嗜好試験法、二点識別試験法、ランキング法等の単純な方法が適している。パネルに複雑な質問をすることは結果を混乱させるだけである。

試料(商品)によってターゲットとする層がある場合は、できるだけターゲット層に合わせた場所を選び、スーパーマーケットの入口、大学のコーヒーハウス等で官能検査を行う。アメリカ合衆国のような多民族国家では、特に、民族、地域、職業などの考慮が重要である。

パネルの構成人数は、目的にもよるが100~200名でテストし、その結果によってさらに人数をふやす。

統計処理によって有意差の検定を行う場合センサリーIにおいて n はくり返し数(とパネル員の数)、センサリーIIと消費者テストにおける n はパネル員の数である。テストされるのはセンサリーIでは試料で、センサリーIIおよび消費者テストではパネル員である。

IV. 信号検出理論

(1) パネル員の慎重さの問題

試料間に品質の差が大きければ、パネル員はその差を容易に識別できる。しかし、官能検査が必要とされるのは試料間に差の小さい場合の方が多い。

クッキーNとXを提示されて、XはNより甘いかと尋ねられた場合を想定する。あるパネル員は、XはNより甘いと答えるのに慎重で、Nより甘いと答えるための境界線のおき方が厳しい、つまり図1のIに境界線がある

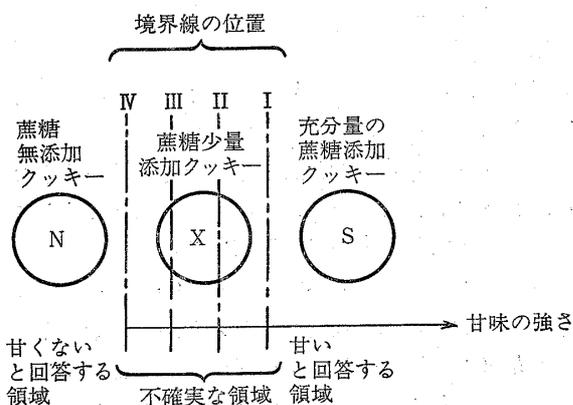


図1. パネル員の境界線の位置と回答

とする。パネル員は境界線より右側ならNより甘いと回答し、境界線より左側ならNより甘いとはいえないと回答するのであるから、この場合のパネル員の回答は、クッキーXはNより甘いとはいえない、というものである。反対に、XはNより甘いと答えるための境界線をIVに置き、Nより甘いと答えるのに大胆になったとしたら、パネル員は、クッキーXはNより甘いと回答する。

同様にNとXを提示されて、これらの甘さは違うかと尋ねられた場合を想定する。境界線を慎重にIまたはIIに置いたパネル員は甘さが違うとは言えないと回答するし、逆に、大胆にIIIあるいはIVに置いたパネル員なら甘さが違うと回答する。

このように境界線の置き方で答えが異なるので、パネル員は境界線をどこに引くべきか悩まねばならない。その上、テストをした人にはパネル員の回答が異なるのはパネル員の感度に差があるのか、慎重さに差があるのか、区別がつかない。

この問題をさけるには2つの方法がある。強制選択法と信号検出理論の応用である。

(2) 強制選択法

パネルは試料間に差のあることを予め知らされ、必ずどれかを選ぶよう指示される。二点比較法、トライアングル法、3-AFC法^{註2)}、1・2点試験法などがそれである。必ずどれかを選ぶのであるからパネル員は2個の試料の間に境界線を引くことになり、境界線をどこに置くべきか悩む必要はない。つまり一方は必ず甘いのであるから迷う必要はないわけである。

註2) 3-AFC法(3-Alternative Forced Choice法)トライアングル法で2個は等しく1個は異なる、3個の試料をパネル員に呈示し、半端な試料を選ばせる。このとき、半端な試料はどのようなものか教えてからテストする。例えば甘さの強いものであるとか、食塩水である等を予め指示する。トライアングル法よりパネル員の識別能は高くなる。特に半端な試料が刺激の弱いものである場合よりも、刺激の強いものである場合の方がパネル員の識別能は高い¹¹⁾。

この方法によれば、試料を1個示してこれは甘いかどうか尋ねたり、試料間に差のあることを知らせずにテストを行う場合よりも識別の精度は高くなる。

(3) 信号検出理論とは

脳の中核は神経を通して伝えられる入力の受け取り手であるとみなし、感覚の神経システムを通信ネットワークシステムのように取り扱い理論である¹²⁾。

あたかも電話の受話器にノイズがあるように、あるいは暗闇でも何か光がチカチカしているかのように、何も刺激のないときでも感覚神経はノイズを脳に送り続けている。感覚の受容器官に刺激として信号が与えられる

信号検出理論の官能検査への応用

と、感覚神経はノイズの上に刺激の信号を重ねて大脳に送る。大脳が受け取るのは「ノイズ+刺激」という入力である。

刺激が大きければ容易にノイズと識別できるが、刺激が小さければノイズとの区別はつきにくい。

ノイズの強さは一定ではなく、これはある値を中心とした正規分布をするものと仮定する。そして受容器官に一定の刺激が与えられた場合、大脳の入力はこの一定の刺激と、正規分布を示すノイズとが相加的に重畳したもので、これは先のノイズの分布中心に刺激を加えた値を中心とし、同じ標準偏差をもつ正規分布に従うものと仮定する(図2)。

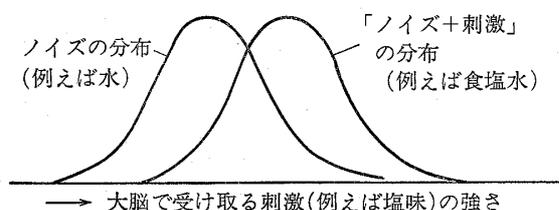


図2. ノイズと刺激の分布

呈味物質を含まない水の味をノイズとすると、食塩水溶液という塩味の刺激に対して、大脳は「ノイズ+塩味」という入力を受け取ったと考える。もし、クッキーで、砂糖を添加していないものと、砂糖を添加したものとを比較するとしたら、砂糖無添加クッキーの味がノイズで、砂糖添加クッキーの味は「ノイズ+甘味」となる。

(4) d' (d' -プライム)

「ノイズ」の分布の中心値と「ノイズ+刺激」の分布の中心値の差、すなわち刺激の大きさそのものを「ノイズ」の分布の標準偏差で割ったものを d' と定義する。 $d'=0$ とは刺激がないことであり、 $d'=3$ とは刺激が標準偏差の3倍であり刺激が強いことを示している註3)。
註3) d' とパネルの予測される正答率との関係は以下の通りである¹³⁾。

d'	2点比較法するとき	トライアングルテストのとき
0	50%	33.3%
0.5	64	
1.0	76	
1.5	86	
2.0	92	60
2.5	96	70
3.0	98	80
3.13		90
4.03		99
6.34		

図3は、パネル員Aが2種のテストを行ったところ、A-1の刺激よりA-2の刺激の方が強かったことを示している。あるいはA-1とA-2の刺激の強さが同じであったのにA-1のテストのときよりもA-2のテストのときにパネルがより敏感であったと考えてもよい。一方、A-2とBを比較すると、パネル員Bはノイズの標準偏差が大きい。A, B両パネル員は同じ刺激を与えられたにも拘らず、A-2では $d'=3.0$ 、Bでは $d'=2.0$ となり、パネル員Aの方が感度のよいことがわかる。このように d' によってパネルの感度を知ることができる。

d' は強制的に選択させる方法と同様に、境界線の位置による変動を排除できる。その理由は次のように考えられる。

図4において、あるパネル員の境界線がIにあると仮定すると、Iより強い刺激に対してパネル員は刺激であると答える。このとき、Ⓐは「ノイズ」を正しく「ノイズ」と答える領域である。Ⓑは「刺激」を正しく「刺激」と答える領域(Hits)である。Ⓒは「ノイズ」を「刺激」と間違えて答え(False Alarms)、Ⓓは「刺激」を「ノイズ」と間違えて答える領域である。1人のパネル員に何回かテストして、その答えがⒶ、Ⓑ、Ⓒ、Ⓓの各領域に何回出現したかを求め、ⒹとⒷの比率、ⒶとⒸの比率を計算し、ⒷとⒸから d' の大きさを推定できる。図4より d' は $(e+f)$ である。Ⓓ-0.5に相当する点(Iの

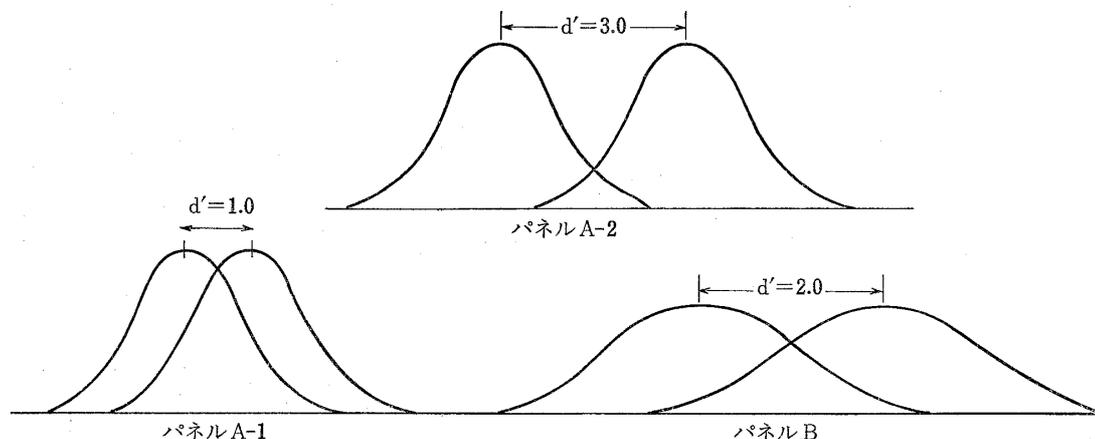


図3. 刺激の強さとパネルの感度

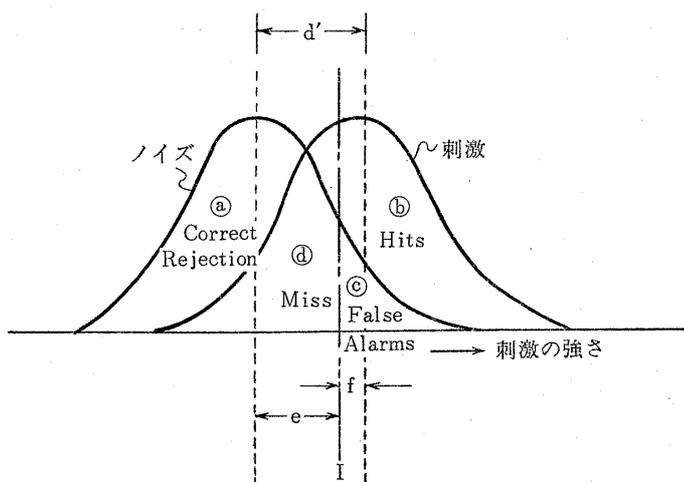


図 4. 境界線がIにあるときのパネル員の応答

位置)を正規分布から読みとれば f が求められる。同様に $0.5 - \text{㉑}$ の面積に相当する点を正規分布表から読みとれば, e が求められる。これより $(e+f)=d'$ を得る。I の位置が変わっても $d'=(e+f)$ であるから, d' は I の位置に影響されない。

V. ROC カーブ

d' の計算には ROC (Receiver Operating Characteristic) カーブを用いるのが容易である。

(1) ROC カーブの求め方

図5のように同一パネル員に同じ刺激を何回か与えて ㉑の領域 (Hits) と, ㉒の領域 (False Alarms) との交点を先ず一点プロットする。もし境界線の置き方が厳しければ Hits も False Alarms も少ないので, 点は図5の左下方に打たれる。何らかの方法で境界線の位置を変えるようパネル員を誘導して註4), テストを行い先の点とは異なった False Alarms と Hits の交点を得る。これを繰り返していくつかの点を求め, それらを結んで得ら

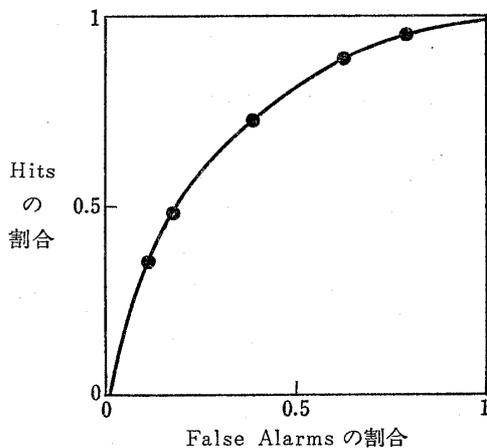


図 5. ROC カーブ

れた曲線が ROC カーブである。ROC カーブから d' が計算される。

註4) 例えば, Guns and Money 法。「刺激をあてたら褒美をあげる」と言われると, パネル員は刺激かノイズか迷った場合には刺激と答えるから, 境界線は図1のIVの位置にくる。逆に「ノイズを誤って刺激と答えたら罰を与える」と言われると, パネル員は境界線を図1のIに置く。

たとえ同じ試料を別の日にテストして, パネル員の境界線の位置が変わったとしても, Hits と False Alarms の交点は同じ ROC カーブの別の位置にくるだけである。1つ1つの点を問題にするのではなくて全体としてのカーブを問題にしているのであるから, ROC カーブは境界線の位置による差を排除していることになる。

ただし, d' を求めるためには, 「ノイズ」, 「ノイズ+刺激」の2つの分布はいずれも正規分布であり, 両者の標準偏差は等しいという統計的前提条件が満たされなければならない註5)。しかし, 刺激が与えられたときにはノイズの変動は, 刺激のないときとは異なるかも知れない。そうだとすると両者の標準偏差は異なり, ROC カーブはゆがみを生じるし, d' の計算は複雑になる。正規分布であるかどうか疑わしい。

註5) これらの前提条件が満たされ, パネル員の判定の変化が感度の変化には依らず, 慎重さの変化のみになるとき, これらの点は1つの d' につき一定の曲線に乗ることを示すことができる。 d' の異なる ROC カーブは両端を除いて交点を持たないので, この平面上1つの測定点を与えられれば対応する d' は一意的に定まることになる。

(2) P(A) と ROC カーブ

そこで, このような前提条件が満たされなくても, 刺激の強さを表わす方法として, d' の代わりに ROC カーブそのものを用いることが考えられた。刺激が強ければ ROC カーブの曲線は上方に凸となり, ROC カーブの下方の面積を $P(A)$ とすると, $P(A)$ は大となり, 刺激の強さを表わすことができるからである (図6, 7)。

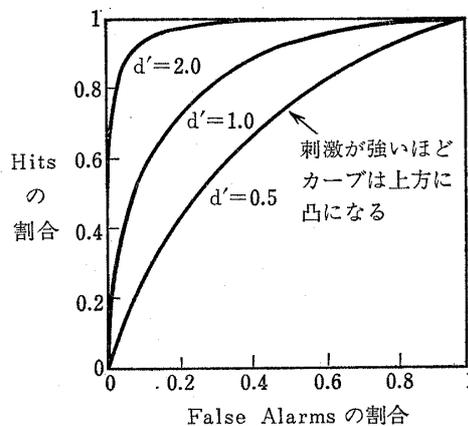


図 6. 刺激の強さと ROC カーブ

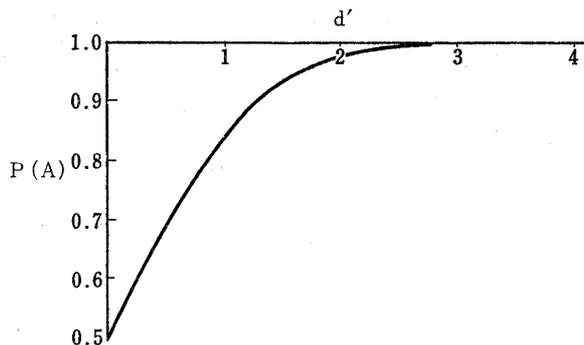


図 7. d' とそれに対応する ROC カーブから求められた面積 $P(A)$ との関係。注 3) も参照のこと

(3) ROC カーブの簡単な求め方

ROC カーブをもっと短時間に求めるために、パネル員に「sureness rating」をつけて答えさせる方法がある¹²⁾。これは、刺激かノイズかときかれたときに、パネル員は自分の答えに「充分確信が持てるかどうか」をつけ加えて答える方法である。パネル員の答えは「刺激—確信がある (sure)」、または「刺激と思うが確信がない (not sure)」、 「ノイズ—確信がある (sure)」、 「ノイズ—確信がない (not sure)」の 4 通りとなる。さらにもう 2 段階「わからないが多分刺激だと思う (don't know but I guess)」および、「わからないが多分ノイズだと思う (don't know but I guess)」を加え 6 通りの中からパネル員に答えを選ばせてもよい。

パネル員はこの方法によって評点をつけているのではなく、単に分類しているだけであることがポイントである。

実際にはパネル員は 100 個の刺激 (例えばうすい食塩水) と 100 個のノイズ (例えば水) をランダムに 1 個ずつ与えられ、それぞれについて上記 4 通りあるいは 6 通りの答えの中からどれかを選ぶことになる。パネル員は境界線の位置が毎回変わっても、同時にいくつかの境界線を持つことになるから、この方法によって ROC カーブを描くことができる。

d' の計算には先に述べたように前提条件が満たされる必要があるが、前提条件が満たされなくても $P(A)$ を求めることはできる。

VI. R-index

R-index は J. Brown¹⁴⁾ によって認識とメカニズムに関する研究において開発されたものであるが^{註 6}, $P(A)$ と同様に信号検出理論の簡単な計算方法であることがわかった。

註 6) 彼の論文の例は、パネル員に先ずよく似た写真を 10~20 枚見せたのち、それらを隠して次に再びよく似た写真を 20~40 枚見せる。そのうち何枚かは最初に見せた写真

と同じものである それぞれについて、その写真を前に見たかどうか尋ねる。このときパネル員の答えとして「確かに前に見た」、「見たと思うが確信がない」、「確かに見ていない」、「見ていないと思うが確信はない」と sureness rating をつけさせた。

(1) R-index の求め方

R-index を理解するには計算方法を見るのが早道である。今ここに 2 個の試料がある。1 つは刺激 (例えばごく稀い蔗糖液) で、もう 1 つはノイズ (例えば水) である。この 2 つの試料間の僅かな味の差をパネルは識別できるかどうか知りたい。

そこで 10 個のごく稀い蔗糖液と 10 個の水をランダムに 1 つずつ 1 人のパネル員に提示する。パネル員は「蔗糖液—確信がある (S)」、「蔗糖液—確信がない (S?)」、「水—確信がある (N)」、「水—確信がない (N?)」のいずれかで答える。このとき、パネル員は予めごく稀い蔗糖液と水を与えられて、それらの呈味の差がわかるまで訓練あるいはウォーミングアップを行うと結果の精度が高まる。

試料の数は 10 個と限らずいくつでもよい。パネル員の答えが、仮に蔗糖液に対して S と答えた個数を 8, S? と答えた個数を 1, N? と答えた個数を 1 とする。さらに水に対して S? と答えた個数が 1, N? が 2, N が 7 とする。これらをまとめると表 1 のようになる。

表 1. 稀い蔗糖液と水に対するパネル員の回答

提示した試料	応答のカテゴリー				繰り返し数
	S	S?	N?	N	
蔗糖液	8	1	1		10
水		1	2	7	10

10 個の蔗糖液と 10 個の水をもし二点比較法でテストすると、 $10 \times 10 = 100$ のペアができる。パネル員は 100 回の二点識別試験を行わなければならないことになる。100 のペアのうち偶然正答となる可能性は 50% であり、75% 識別できれば正答率は 75%、すべて識別できれば正答率は 100% である。この予測される正答率、いいかえればペアの中から誤りなく刺激を選択する確率の予測が R-index である。

パネル員が蔗糖液を正しく S と答えた数は 8 個であるが、この 8 試料はもし水を S?, N?, および N と答えた試料とペアにして提示し、どちらが蔗糖液かと尋ねられたら、パネル員は蔗糖液の方を S であると正しく選択すると考えてよい。つまり、パネル員は $8 \times (1 + 2 + 7) = 80$ のペアに対して正しい答えをすると考えられる。

次に、蔗糖液を S? と答えた 1 試料は、水を N?, および N と答えた試料とペアにして、どちらが蔗糖液かと

尋ねられたら、パネル員はここでも蔗糖液の方を正しく選択すると考えてよい。1×(2+7)=90のペアに対してもパネル員は正しい回答をする筈である。

同様に蔗糖液をN?と誤って答えた試料でも水をNと答えた試料とペアにしてどちらが蔗糖液かと尋ねられたら、パネル員はN?の方を蔗糖液であると答えると考えられるから1×7=7のペアに対しても正しい選択をする。

ところで残りのペア、蔗糖液をS?, 水S?をと答えたペア(1+1)と蔗糖液をN?, 水をN?と答えたペア(1+2)に対しては正答の確率は1/2である。

これらを合計するとこの20個の試料に対するパネル員の回答から予測される正答率は、

$$\begin{array}{r} 8 \times (1+2+7) = 80 \\ 1 \times (2+7) = 9 \\ 1 \times 7 = 7 \\ (1+1) \times \frac{1}{2} = 1 \\ (1+2) \times \frac{1}{2} = 1.5 \\ \hline \text{合計} \quad 97.5 \end{array}$$

となり、100のうち97.5つまり97.5%が予測される正答率: R-indexである。

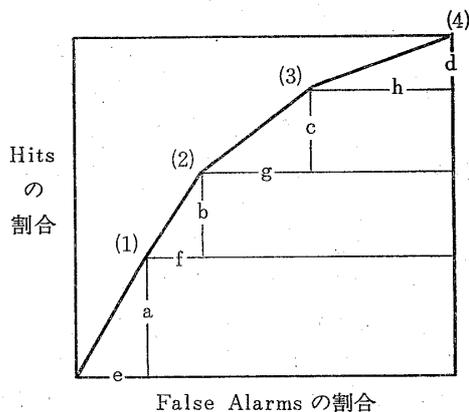
(2) R-index と P(A) の関係

R-indexの求め方は書き方を変えれば、次のようになる。

	S	S?	N?	N	
S	a	b	c	d	N ₁
N	e	f	g	h	N ₂

$N_1 = a + b + c + d$

$N_2 = e + f + g + h$



	S	S?	N?	N	
S	a	b	c	d	N ₁
N	e	f	g	h	N ₂

図 8. (1)~(4) 点より求めた ROC カーブ

$$R = \frac{a(f+g+h) + b(g+h) + ch + \frac{1}{2}(ae+bf+cg+dh)}{N_1 N_2}$$

この計算により得られた R-index は P(A) と同じであることは以下のように示される。

図 8 の ROC カーブにおいて 4 種の異なった境界線における Hits と False Alarms の値はそれぞれ

- (1): (e, a)
- (2): (e+f, a+b)
- (3): (e+f+g, a+b+c)
- (4): (e+f+g+h, a+b+c+d)

P(A) は ROC カーブより下方の面積であるから

$$P(A) = \frac{a(f+g+h) + b(g+h) + ch + \frac{1}{2}(ae+bf+cg+dh)}{N_1, N_2}$$

となり、R-index を求める式と同じである。

(3) 2 試料の識別に対する R-index の応用

試料 A (例えば従来のクッキー) と試料 B (配合割合を変えたクッキー) があり、配合割合を変えたことがパネルに識別できるかどうか知りたいとする。試料 A, B を何こかずつ用意し、パネル員にランダムに 1 つずつ与える。パネル員はそれぞれに、「試料 A—確信がある (S)」、「試料 A—確信はない (S?)」、「試料 B—確信がある (N)」、「試料 B—確信はない (N?)」のいずれかで答える。この答の割合から R-index を求める。

別の例を挙げると、標準試料を与えて、A または B どちらの試料が標準試料と同じであるかパネルに尋ねる場合がある。パネル員の答えは「標準試料と同じ—確信がある (S)」、「標準試料と同じ—確信がない (S?)」、「標準試料とは異なる—確信がある (N)」、「標準試料とは異なる—確信がない (N?)」のいずれかである。

R-index の利点はパネル員の評点をつけさせることなく差の程度を示すことができるという点にある。

R-index の差の検定は Mann-Whitney rank sum test によって行う。

(4) 3 試料の識別に対する R-index の応用

試料 A に何らかの処理を施した 3 種の試料 B₁, B₂, B₃ がある。パネル員に A, B₁, B₂, B₃ をそれぞれ 10 個ずつ合計 40 個をランダムに与え、それぞれが A であるかそれとも B であるか答えさせる。パネル員の答えはこれまでと同様に sureness rating をつけて、「A—確信がある (A)」、「A—確信がない (A?)」、「B—確信がある (B)」、「B—確信がない (B?)」のいずれかとする。

パネル員の答えを仮に表 2 のようであったとすると、A をノイズ、B₁~B₃ を刺激と考えて、B₁ と A, B₂ と A, B₃ と A の組み合わせに対応する R-index が求められる。これら R-index の値は A と B₁, B₂, B₃ との差の程

信号検出理論の官能検査への応用

表 2. 3 試料に対応する R-index

	パネル員の回答数				くり返し数	R-index
	B	B?	A?	A		
B ₁	10				N=10	R=100%
B ₂	8	1	1		N=10	R=98%
B ₃		6	2	2	N=10	R=82%
A		1	1	8	N=10	

度を表わしており、B₁とAとの差が最も大きく、B₃とAの差は最も小さいことがわかる。この方法によれば3試料のR-indexが1度で求められる。

複数のパネル員からR-indexが得られれば、B₁, B₂, B₃それぞれについてR-indexの平均値と標準偏差が求められるから、分散分析によって3試料間に有意の差があるかどうか検定すればよい。

(5) ランキングへの R-index の応用

4種の試料、A, B, C, Dについて、10名のパネル員が、あるいは1名のパネル員が10回、識別(例えば甘味の強い順)のランキング、あるいは嗜好(例えば好ましい順)のランキングを行なう場合についてもR-indexの応用が可能である。

表3に例を示すと、この場合、最も順位の低いものを

表 3. 4 試料のランキング

くり返し No.	パネルのランキング			
	1位	2位	3位	4位
1	A	B	C	D
2	B	A	C	D
3	A	B	D	C
4	B	A	D	C
5	A	B	C	D
6	C	A	B	D
7	A	C	D	B
8	C	A	B	D
9	A	B	C	D
10	A	C	B	D

ノイズと考える。Aの方がDよりも先に選ばれた回数は10回中10回であるからA/D=10/10=100%、同様にB, CについてはB/D=9/10=90%、C/D=8/10=80%となる。

これを通常のR-index計算法で求めてもよい。値がいくらか異なっているのはノイズと刺激が1個ずつではないためである。R-indexはペアの中から刺激を選ぶ確率であることを考えると、表2がそれであり、表3はP(A)の計算である。刺激とノイズが1個ずつであればP(A)=R-indexとなる。

順位合計ではA: 14, B: 23, C: 26, D: 37となる。

R-indexはノイズと刺激のペアの中から刺激を選ぶ確率であるから、パネル員に2個のノイズと2個の刺激

表 4. 4 試料のランキングによる R-index

試料	パネルのランキング回数				くり返し数	R-index
	1位	2位	3位	4位		
A	6	4			10	A/D=100%
B	2	4	3	1	10	B/D=89%
C	2	2	4	2	10	C/D=81%
D			3	7	10	

(刺激の強さは同じ)、合計4個を提示し、刺激の強い順にランキングさせることによってもR-indexを求めることができる。

例えばパネル員が4個の試料を第1位から順にS(刺激), N(ノイズ), S, Nと答えたとすると、彼がSを選ぶ確率はNよりもSの方を強いと答えた数をかぞえればよい。第1位のSは2個のNよりも刺激が強く、第3位のSは1個のNよりも刺激が強いと答えたのであるからペアの中からSを選ぶ確率は4回のうち3回つまり75%(R=75%)ということになる。

これをさらに拡大して、このテストを5回くり返したとする。1名のパネル員に5回でも5名のパネル員に1回ずつでもよい。その結果は例えば次のようになる。

表 5. 2 試料のランキングによる R-index

くり返し No.	パネルのランキング			
	1位	2位	3位	4位
1	S	S	N	N
2	S	N	N	S
3	S	N	S	N
4	S	S	N	N
5	N	S	S	N

SのあとにNが何回来ているか数える。No.1では2+2回、No.2: 2+0, No.3: 2+1, No.4: 2+2, No.5: 1+1, 合計すると20回のうち15回すなわち75%となる。

これを先と同様に

表 6. 2 試料のランキングによる R-index

	パネルのランキング回数				くり返し数	R-index
	1	2	3	4		
S	4	3	2	1	10	75%
N	1	2	3	4	10	

$$R = \frac{4(2+3+4) + 3(3+4) + 8 + \frac{1}{2}(4+6+6+4)}{10 \times 10} = 0.75$$

としても同じである。

(6) 多数の試料のランキングと評点への R-index の応用

複数の呈味物質を混合したときの呈味は、もとの呈味

表 7. R-indices denoting probabilities of rating and ranking stimuli as more mixed than quinine hydrochloride⁸⁾

Stimuli	Rating								
	QHCl	fructose	citric acid	MSG	NaCl	KCl	NaCl	fructose	fructose
Mean R-Indices	54.8	63.4	69.1	80.5	82.2	88.6	QHCl	NaCl	NaCl
S.D. (S)	(3.5)	(19.1)	(13.2)	(14.2)	(15.8)	(12.1)	(10.2)	(5.5)	(3.4)
Stimuli	Ranking								
	QHCl	citric acid	fructose	NaCl	MSG	KCl	NaCl	fructose	fructose
Mean R-indices	54.4	62.8	62.9	74.8	78.0	79.5	QHCl	NaCl	NaCl
S.D. (S)	(3.1)	(16.0)	(16.3)	(15.1)	(11.0)	(9.9)	(10.5)	(5.5)	(6.8)

Horizontal bars link means that are not significantly different (A.O.V., Fisher's L.S.D., $p < 0.05$).

物質とは全く別の新しい単一の味となるのか (synthetic^{註7)}であるか), それとももとの複数の味から成っている (analyticであるか) と感知されるのかについては種々の議論のあるところである。(たとえば McBurney, D. H.¹⁵⁾, J. T. Kuznicki & N. Ashbaugh¹⁶⁾, また, 一種類の呈味物質を単一の呈味と感ずるかどうかにしても定説はない。

註7) synthetic は色の感覚のように, 2つの色を混合すると全く新しい1つの色となる, このような感覚を synthetic とする。これに対して analytic は音のように2つの音を同時に聞いてもそれぞれの音を聞きわけることができるような感覚をさす。(上手な指揮者の耳は analytic である。上手なオーケストラのハーモニーは synthetic に聞こえる)

そこで最も単一である呈味として塩酸キニーネを対照としてパネルに与え, それを比べて9種の試料は単一の呈味であるか, 複数の混合した呈味であるかをランキングと評点によって R-index を求め, 単一さの程度を検討した⁸⁾。

予め synthetic と analytic に関する訓練を受けた12名のパネル員がそれぞれ10回ずつテストを行った。9種の試料は単一の呈味物質あるいは混合物である。パネル員は一回に3種の試料を与えられ, 評点づけではそれぞれの試料について, 「単一である—確信がある」, 「単一である—確信がない」, 「単一でない—確信がある」, 「単一でない—確信がない」のいずれかで答えた。R-index の計算には塩酸キニーネをノイズとした。ランキングでは同時に供された3種の試料に単一であるものから順に順位をつけた。

9種の試料とランキングおよび評点づけから求めた R-index の値を表7に示した。塩酸キニーネが表に含ま

れているのは, 最も単一である呈味として与えた塩酸キニーネを, パネルが確かに単一であると識別していることを確認するために, 試料の一つとして加えたことによる。

R-index が50ということは基準と区別がつかないということで, 単一の呈味であることを示し, R-index が100なら完全に単一ではない, 複数の味の混合した呈味であることを示す。R-index によって一種類の呈味物質でも, 複数の味の混合したものに近い呈味である, と感知されているものがあることがわかる。例えば NaCl や KCl は fructose や citric acid に比べ複数の呈味に近いといつてよい。

VII. おわりに

以上信号検出理論とそれを応用して P(A), R-index を求める方法についてのべた。R-index を求めるには二点比較法やトライアングルテスト等に比べ時間がかかる。センサリー I のテストなら少数のパネル員を機器のように扱うのであるから, 時間がかかることは問題とはならない。センサリー II の場合は目的によって, 二点比較法, トライアングルテスト, 3-AFC, R-index 等, いずれを採用するか考えればよい。

なお官能検査の3種のタイプについては別の考えもあると思う。また, ここでは Descriptive Analysis については取り上げなかったが, センサリー I に分類されるべき性質のもので, 10~20人から成るパネルを長期間訓練の後テストを行う。

本原稿はカリフォルニア大学デービス校の1991年秋学期における Dr. M. O'Mahony の講義その他をもとに

信号検出理論の官能検査への応用

まとめたものであり、ここに感謝の意を表する。

文 献

- 1) 山口静子: 調理科学, 19, 24, 86, 249 (1986), 20, 26 (1987)
- 2) 福場博保, 宮川金二郎編: 調理科学実験ハンドブック, p.291~409, 建帛社, 東京 (1986)
- 3) 日科技連官能検査委員会: 新版官能検査ハンドブック, 日科技連, 東京 (1973)
- 4) 佐藤信: 官能検査入門, 日科技連 (1978)
- 5) 佐藤信: 統計的官能検査法, 日科技連 (1985)
- 6) O'Mahony, M.: Moskowitz, H. ed. "Applied Sensory Analysis of Foods" Vol. 1, p.145~177, CRC Press, Boca Raton, Florida (1988)
- 7) O'Mahony, M.: J. Sensory Studies 7, 1 (1992)
- 8) O'Mahony, M., Atassi-Sheldon, S., Rothman, L. and Murphy-Ellison, T.: Physiology & Behavior, 31, 749 (1983)
- 9) Vie, A., Gulli, D. and O'Mahony, M.: J. Food Sci., 56, 1 (1991)
- 10) 増山英太郎, 小林茂雄: センソリーエバリュエーション——官能検査へのいざない——, 垣内出版, 東京 (19)
- 11) O'Mahony, M. and Odbert, N.: J. Food Sci., 50 1055 (1985)
- 12) Green, D.M. and Swets, J. A.: "Signal Detection Theory and Psychophysics" John Wiley & Sons, New York (1966).
- 13) Elliott, P.B.: "Table of d' , in signal detection, an recognition by human observers" Swets, J. A. ed. "Contemporary Readings" p.651, John Wiley & Sons, New York (1964)
- 14) Brown, J.: British J. Psychol. 65, 13 (1974).
- 15) McBurney, D.H.: Chem. Senses Flav. 1, 17 (1974)
- 16) Kuznicki, J. T. and Ashbaugh, N.: Sens. Processes, 3, 157 (1979)

第19回伝統食品に関する講演会

日 時: 平成5年4月6日(火) 13:00—16:20

場 所: お茶の水女子大学 生活科学部 本館233号室
東京都文京区大塚 2-1-1 (電話 03-3943-3151)

- 演 題: 1. 江戸のくらしと食
石川 寛子氏 (武蔵野女子大学短大)
2. 伝統漬物と現代漬物の間
前田 安彦氏 (宇都宮大学農学部)
3. 香辛料と薬味の効用
森 一雄氏 (兵庫女子短期大学)

会場整理費: 500円 (当日会場にて)

主 催: 日本伝統食品研究会

問合せ先: 〒112 東京都文京区大塚 2-1-1
お茶の水女子大学 生活科学部 畑江敬子
(電話 03-3943-3151 内線 633)