

Q2: 主観評価がわかりません

日本音響学会2013年春季研究発表会
ビギナーズセミナー
2013.3.13 東京工科大学

とりあえず...

初学者(ビギナー)にお薦めの一冊

音響テクノロジーシリーズ4

音の評価のための心理学的測定法

難波精一郎・桑野園子 共著

コロナ社 (1998)

を読みましょう。

主観評価の目的は？

1. 聴覚特性・心理特性を調べるため

⇒ 生理学的測定代わりに主観評価を行う

被験対象 : 人 (被験者)

制御刺激 : 音 (実験者が制御)

※ 音の違いをどれくらいの精度、どれくらいの範囲で調べるのか

2. 自分の考えた方式の優位性を示すため

⇒ 音の物理測定(客観評価)の代わりに主観評価を行う

被験対象 : 音 (自分が考えた方式と従来手法)

※ 音の違いが分かりやすいか、どれくらいの量を調べるのか

測定装置 : 人 (評価者)

主観評価の目的に合わせて、適切な評価法を選択

主観評価のやり方には

精神物理学的測定法（音と聴覚特性の関係を数量的に測定）

調べられるもの：刺激域（聞こえるか）、刺激頂（これ以上だと分からない）、
弁別域（差が分かるか）、主観的等価点（同じくらいなのは）

調整法
（被験者が自由に調整）

極限法
（段階的にある特徴が変化）

恒常法
（刺激をランダムに提示）

尺度構成法（印象や好みを測定）

調べられるもの：音の大きさ、音の好み、
音の印象（明るい など）、気分（楽しい など）

直接法

ME法（マグニチュード推定法）

どれくらいの大きさ？

間接法

評定尺度法

一対比較法

SD法

どちらが大きい？

順位法

MDS（多次元尺度構成法）

どんな印象？

測定の水準

比率尺度
(比例尺度)

(絶対温度, 質量)
大小, 間隔, 比率

間隔尺度
(距離尺度)

(温度(摂氏))
大小, 間隔

序数尺度
(順位尺度)

(順位)
大小

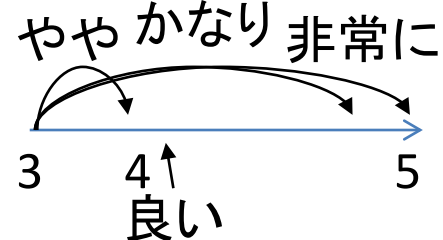
名義尺度
(分類尺度)

(クラス番号)

間隔が等しい
⇒ 4点と5点の間は
4.5点 (平均 ○)

順番が決まる(等距離ではないかも)

5. 非常に良い
4. 良い
3. どちらでもない
2. 悪い
1. 非常に悪い



⇒ 4点と5点の間は4.5点とは
限らない
(平均値 ×, 中央値 ○)

絶対の0がある
音の大きさ

(ME法: 厳密には...)

2倍 ↷ 200 ? dB
100 ↔ 120dB
1/2 ↶ 50 ? dB

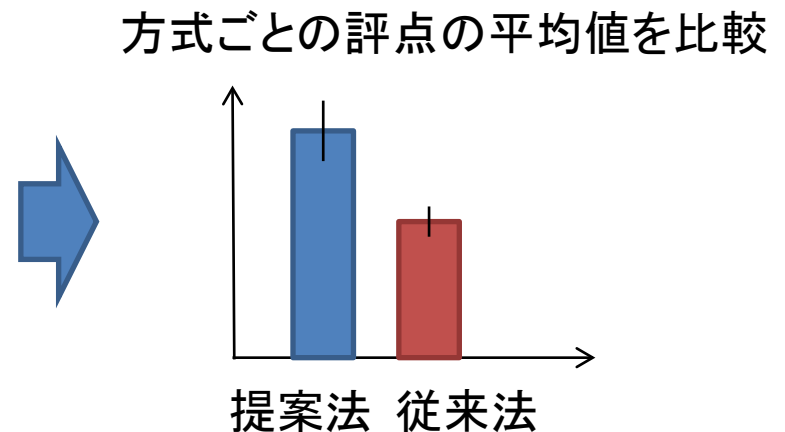
等間隔とみなす
or
間隔尺度に変換
(範疇判断の法則)

とりあえず、提案方式を評価したい

MOS: Mean Opinion Score 平均オピニオン評点
ITU-T P.800(音声品質), ITU-R BS.1284(オーディオ品質)
聴取者の満足度を評価する尺度

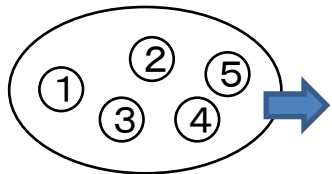


カテゴリ	評点
非常に良い	5
良い	4
普通	3
悪い	2
非常に悪い	1



MOSでは有意差が出ない

一対比較法 (評価音源の差が小さいこと, 時間長が短いこと)
順位(優劣)を付ける ⇒ サーストンの一対比較(間隔尺度)
⇒ ブラッドレイの一対比較(比例尺度)
差の程度を評点で示す ⇒ シェツフェの一対比較法(間隔尺度)



総当たりで2個を選出
 $n(n-1)$ (順序効果を考慮)



どちらが良いか(サーストン)

どの程度良いか(シェツフェ)

- +2. 1が2より非常に良い
- +1. 1が2より幾分良い
- 0. 1と2が同じ良さ
- 1. 1が2より幾分悪い
- 2. 1が2より非常に悪い

選ばれた確率で尺度を構成



より多く好まれた

正解がある場合（劣化尺度）

- DMOS: Degradation Mean Opinion Score
ITU-T P.800 (音声品質),
- ITU-R BS.1116 隠れ基準付き3刺激二重盲検法
- ITU-R BS.1534 MUSHRA法

DMOS

リファレンス 評価対象 評価

BS.1116

リファレンス 評価対象1 評価対象2 評価

(評価対象のどちらかは
リファレンスと同じ音)

カテゴリ	評点
差が分からない	5
差が分かるが気にならない	4
差がわずかに気になる	3
差が気になる	2
差が非常に気になる	1

どんな印象なのか

・SD法 (Semantic Differential) 多次元で音の印象を評価
 両極尺度で印象を評価 ⇒ 因子分析



あてはま
る

あてはま
る

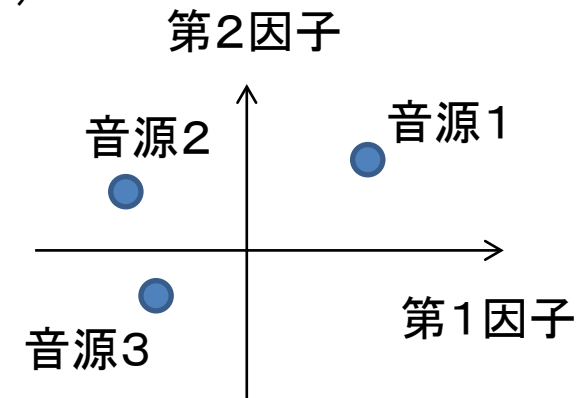
どちらで
もない

明るい	1-2-3-4-5	暗い
かたい	1-2-3-4-5	やわらかい
澄んだ	1-2-3-4-5	濁った
迫力がある	1-2-3-4-5	ものたりない
静かな	1-2-3-4-5	騒々しい

因子分析
(どれが仲間か)

第1因子

第2因子

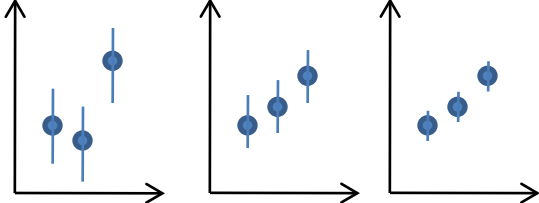


・単極尺度で印象を評価 ⇒ 因子分析

明るい: 1. あてはまらない-5. よくあてはまる

主観評価を成功させるために

評価者はどういう人を選ぶのか

- 10人で大体の傾向がつかめる。⇒ **20人**～30人で有意差
※注 有意差がある ≠ 差が大きい
- 
- 評価の目的に合った人を選ぶ
 - アクセントやイントネーション ⇒ 12歳まで東京に在住
 - 音質の劣化 ⇒ 専門家(高品質), 経験者(中品質)
(十分に練習をして、正解率が悪い人は除外)
 - 先入観のない人を選ぶ
実験の趣旨を理解しないで評価／聴き所が限定される
(専門家は極端な反応を示すことも)
- 実験中に情報交換させない(評価基準を安定させる)
 - やる気を起こさせる演出(お金を支払ったを書かせる論文誌も)

主観評価を成功させるために

どこで実施するのか(どんな装置を使うのか)

- ・ 使える部屋でやるしかない
 - 無響室: 聴覚特性など
 - 評価室: 音響システムの評価など (残響時間などが調整されている)
ITU-R BS.1116-1など、国際規格あり
床面積 40m², スピーカまでの距離 2m, 残響時間 0.25s (500Hz),
音騒音 NR15は超えないこと など
 - 防音室: アクセントやイントネーションなど (外部の音が聞こえない)
- ・ ある装置でやるしかない
 - スピーカ(スモールスピーカ、ITU-R BS.1116-1など、国際規格あり):
周波数応答: 8dB±2dB (250Hzー16kHz)
0° と比較して水平±30° 偏差 4dB (40Hzー16kHz)
ダイナミックレンジ: 112dB以上

やれる範囲でしかできないけれど...

場所, 装置を正確に記載すること・結果に影響がないか考察を

主観評価を成功させるために

音源や評価語を提示する順番(順序効果)

- 音源の提示順、評価語の並び など変えられるものは
評価者ごとに変える
(逆順もやったからOKでなく、ランダムイズされているのかを確認)

評価音源の選び方

- 時間長 : 比較実験では、なるべく短く(BS.1534では10秒)
官能評価(気分など)では、3分後に差が表れることも
- 品質の幅: 一対比較法では差が大きすぎないように
アンカー刺激、同時に聴取する音源によって、尺度の間隔が変化
(パフェ, ケーキ, チョコ) ⇔ (パフェ, ケーキ, レモン)
非常に 甘い やや 非常に 全く

主観評価を成功させるために

教示

- ・ 直感で回答させる。“出来るだけ早く、しかし正確に、戻ることなく”
- ・ 評価語を正しく説明（分野によって意味が異なることも）
 - ⇒ 普段思っている意味と異なると実験中に基準が変わることも・・・
（説明しない方がよいという説も。）

評価語の選び方

- ・ 身近で誤解を招かない言葉を用いる
- ・ SD法（両極尺度）の場合、反意語によって評価軸が異なる
 - 迫力がある — ものたりない（快—不快？ 大きい—小さい？）
 - ⇒ 単極尺度：迫力がある（あてはまらない：迫力がない or 別の印象がある）
- ・ 自分で尺度を作る
 - 音を聞かせて、音を想像させて、形容語を記述させる

ここでご紹介した内容は、主観評価法のごく一部です。
また、小難しい検定などの統計処理は取り上げていません。
よく分かった、まだまだここが分からないなど、感想を教えてください。

音響学入門ペディア運営委員会まで