



## 音質の評価

杉本岳大 (NHK)

sugimoto.t-fg@nhk.or.jp

### 1. 音質評価の必要性和その難しさ

音質評価が必要となるのは、新たに考案したアルゴリズムや新たに開発したハードウェアで生成した音声信号の品質を評価したいときである。音声符号化技術にしても音声合成技術にしても、原則として人間が聴取することが目的の技術であり、その音質を見極めるためには人間による評価、すなわち主観評価実験が欠かせない。

音質評価の難しさは、異なる音声を同時に比較できない点にある。画像や映像の場合、ディスプレイを複数設置することで、異なる刺激を並列して同時に提示することが可能である。しかし音声の場合、複数の刺激を同時に再生すると、音声同士が混ざり合ってしまう。そのため、例えばA、B、二つの刺激を比較する場合、AとBを交互に聴取しながら評価せざるを得ない。この構成において評価者に課されるタスクは、先ほど聴取した一方の刺激の聴感印象の記憶と、今現在再生しているもう一方の刺激の聴感印象を比較するタスクに換言されることから、評価者は複雑な手続きで評価していることが分かるだろう。

また、刺激間の微小な音質差を評価することの困難さに加え、刺激間に音質差がない場合に、差がないと正しく判断することにも同程度の困難さが存在する。音質評価に限った話ではないが、評価者の心理状態・生理状態は要因の内的・外的を問わず、常に変動している。再生されている刺激が同じであっても、評価者の心理状態・生理状態が一定である保証はないため、刺激が同一なら評価結果も同じくするとは限らない。そのため、主観評価実験前のトレーニングにおいて、評価者に自身の評価の精度や安定度を把握してもらうことが肝要である。

近年、音響技術者を対象とした、音の聴感印象と音の物理的性質の対応付けスキルを獲得するためのトレーニングである「聴能形成」の活用が、産学問わず盛んになってきている<sup>(1),(2)</sup>。このような流れは、音質を含む音の評価全般の困難さが広く認識されつつあることの証左であろう。

音質の評価手法としては、過去に様々な手法が考案されてきているが<sup>(3),(4)</sup>、特に信頼性の高い主観評価を実施するための手法は、ITU-R (International Telecommunication Union—Radiocommunication Sector) や ITU-T (International Telecommunication Union—Telecommunication Standardization Sector) などの国際標準化機関で規格化されている。

### 2. 劣化の小さい音響システムの主観評価法

本章では、ITU-R 勧告 BS. 1116-3<sup>(5)</sup> として規格化された劣化の小さい音響システムの主観評価法について紹介する。ITU-R 勧告 BS. 1116-3 には、評価手法だけでなく、評価する室の特性や使用する音響機器の性能、評価結果の統計分析法についても詳細に取り決められている。信頼性の高い音質評価を行うための国際的な基準として、MPEG (Moving Picture Experts Group, ISO/IEC SC29/JTC1/WG6) をはじめとする国際標準化団体での評価で活用されてきた。

採用されている評価手法は、「隠れ基準付き三刺激二重盲検法」である。基準となる音声信号（以下、基準音・音声符号化劣化を評価する場合は、非圧縮音になる）と、評価音（音声符号化劣化を評価する場合は、圧縮符号化音になる）の2種類の刺激を用いる。

図1に、隠れ基準付き三刺激二重盲検法用の評価インタフェースを示す。図中、REFの再生ボタンには基準音が割り当てられる。A、Bの再生ボタンの片方には基準音（隠れ基準音と呼ぶ）が、もう一方には評価音が、ランダムに割り当てられる。評価者は、A、B、それぞれの刺激と基準音を比較聴取し、基準音との間の総合的な音質 (Basic Audio Quality) の違いを主観的に評価する。評価にあたっては、5段階の劣化尺度 (表1) に沿って、少数点以下0.1の精度で評点を与える。A、Bのどちらかは基準音と同一の音声信号であることから、A、Bの一方には必ず5.0が与えられるこ

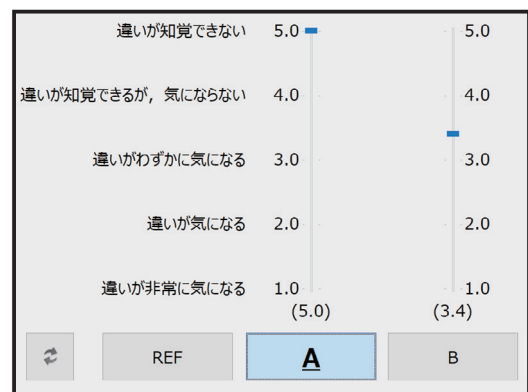


図1 隠れ基準付き三刺激二重盲検法用の評価インタフェース REFの再生ボタンには基準音が割り当てられる。A、Bの再生ボタンの片方には隠れ基準音が、もう一方には評価音が、ランダムに割り当てられる。

本会ハンドブック「知識の森」  
[https://www.ieice-hbkb.org/portal/doc\\_index.html](https://www.ieice-hbkb.org/portal/doc_index.html)

表 1 5段階の劣化尺度

評点	評価語
5.0	違いが知覚できない
4.0	違いが知覚できるが、気にならない
3.0	違いがわずかに気になる
2.0	違いが気になる
1.0	違いが非常に気になる

となる。本主観評価においては、各刺激の音質を絶対評価しているのではなく、あくまで基準音からの音質の違いの割合を評価する。仮に基準音よりも好ましい音質と感ずる刺激があったとしても、基準音と異なる音質であるため、音質の違いの割合に応じて評点を下げる必要がある。

### 3. その他の主観評価法

ITU-R では、前述した ITU-R 勧告 BS. 1116-3 以外にも、ITU-R 勧告 BS. 1534-3「中程度の品質の音響システムの主観評価法」<sup>(6)</sup> や、ITU-R 勧告 BS. 2132-0「基準音を含まない複数刺激による音響システムの主観評価法」<sup>(7)</sup> が規格化されている。

ITU-R 勧告 BS. 1534-3 で採用されている評価手法は、「隠れ基準および隠れアンカ付き多重刺激二重盲検法(MUSHRA : Multi Stimulus test with Hidden Reference and Anchor)」である。MUSHRA は、ITU-R 勧告 BS. 1116-3 が対象とする評価音に比べて、より基準音との違いが大きい評価音が対象である。図 2 に、MUSHRA の評価インターフェースを示す。基準音と評価音を比較して評価する点では ITU-R 勧告 BS. 1116-3 同様であるが、隠れ基準音以外の評価音が複数ある点、隠れアンカ音が存在する点、20 点間隔の 5 段階の評価語 (Excellent, Good, Fair, Poor, Bad) に対応した 100 点満点で評価する点、に違いがある。MUSHRA における隠れアンカ音は、評点を適切に分布させるためのカウンターバランスの役割を担っている。例えば圧縮符号化音の評価では、非圧縮音に 3.5 kHz の低域フィルタを掛けた音声信号を隠れアンカとして用いることが多いが、圧縮符号化音における最低ラインの音質を想定した設定である。

一方、ITU-R 勧告 BS. 2132-0 は、確定した基準音が存在しない条件での主観評価法である。例えば、Virtual Reality (VR) 空間におけるレンダリング法やマルチチャンネル音響のダウンミックス法などは、音声符号化における非圧縮音のように、基準を一意に定めることが困難である。そこで ITU-R 勧告 BS. 2132-0 では、性能の異なる各アルゴリズムで生成された複数の刺激を相互に比較する手法を規定している。

### 4. 音質の客観評価法

一般に主観評価実験は、評価者への負担が小さくない上に、評価に適した実験環境の構築にも手間が掛かる。技術開発においては、アルゴリズムを調整するたびに調整結果を評価する必要があるが、その評価を主観評価実験のみに依拠することは、前述の理由から非現実的である。そこで、主観評価に頼らずに音質を評価する目的で、音質の客観評価法が開

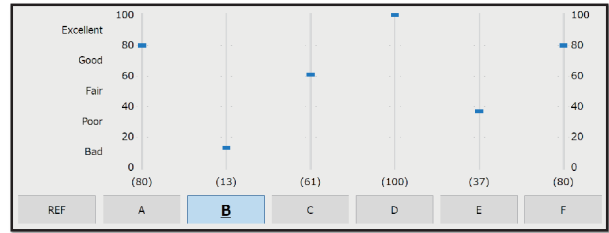


図 2 MUSHRA の評価インターフェース REF の再生ボタンには基準音が割り当てられる。評価音が 4 個の場合、A~F の再生ボタンには、隠れ基準音、隠れアンカ音、4 個の評価音が、ランダムに割り当てられる。

発された。これまでに、楽音を含む音声信号を対象とした ITU-R 勧告 BS. 1387-2 (通称 PEAQ)<sup>(8)</sup>、人の声を対象とした ITU-T 勧告 P. 862 (通称 PESQ)<sup>(9)</sup> と ITU-T 勧告 P. 863 (通称 POLQA)<sup>(10)</sup> などが規格化されており、評価装置への実装実績も多い。本章では、PEAQ の評価アルゴリズムを紹介する。

PEAQ ではまず、基準となる非圧縮音と評価対象の圧縮符号化音を、人間の聴覚を模擬したアルゴリズム (聴覚モデル) に入力する。次に、聴覚モデルから出力された各音声信号の、聴覚的なひずみ量を求める。最後に、前述のひずみ量をニューラルネットワークで構成された認識モデルに入力し、音質評価結果を算出する。なお、PEAQ で用いられるニューラルネットワークは、規格化に際して実施された主観評価実験に基づいて学習を行ったものである。

客観評価は、手軽に何度でも音質を評価できるため、アルゴリズムや装置の開発段階においては重宝する。しかしながら、主観評価値と客観評価値は必ずしも一致するわけではないため、最終的な音質評価のための主観評価実験は欠かせない点を留意されたい。

## 文 献

- (1) 河原一彦, 西村 明, “大学における聴能形成の展開とその展望,” 音響誌, vol. 77, no. 11, pp. 726-733, 2021.
- (2) 大脇達生, “「聴能形成」を取り込んだ社内音響教育—「音感訓練」を通して見えてきたこと—,” 音響誌, vol. 77, no. 11, pp. 734-740, 2021.
- (3) 難波精一郎, 桑野園子, 音の評価のための心理学的測定法, コロナ社, 1998.
- (4) 佐藤 信, 統計的官能検査法, 日科技連出版社, 1985.
- (5) Recommendation ITU-R BS. 1116-3, “Methods for the subjective assessment of small impairments in audio systems,” 2015.
- (6) Recommendation ITU-R BS. 1534-3, “Method for the subjective assessment of intermediate quality level of audio systems,” 2015.
- (7) Recommendation ITU-R BS. 2132-0, “Method for the subjective quality assessment of audible differences of sound systems using multiple stimuli without a given reference,” 2019.
- (8) Recommendation ITU-R BS. 1387-2, “Method for objective measurements of perceived audio quality,” 2023.
- (9) ITU-T Recommendation, “Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” p. 862, 2001.
- (10) ITU-T Recommendation, “Perceptual objective listening quality prediction,” p. 863, 2018.

(2023 年 6 月 25 日受付)