

# 大規模主観評価と周波数分析による 心地よいインパルス応答測定信号の設計\*

☆高橋 勇希 (長岡高専), 小口 純矢, 高道慎之介 (東大院・情報理工),  
矢野 昌平 (長岡高専), 猿渡 洋 (東大院・情報理工)

## 1 はじめに

音響インパルス応答 (Acoustical Impulse Response, IR) は, 系の基本特性を表す物理量でコンサートホールの設計やヘッドホンの設計など様々な分野で測定される. IR の測定では MLS (Maximal Length Sequence) 信号や TSP (Time Stretched Pulse) 信号などの IR 測定信号を用いた測定法 [1] が一般的である. これらの信号を用いた測定では, 通常, 測定対象の系に人間が含まれているかどうかを考慮していないため, 信号が人間に与える知覚的ストレスが無視されている. 他方, 新たな生体認証として, 人間の外耳道の IR が提案 (耳音響認証として知られる [2]) されており, 個人に応じた音メディア提示などへの応用が期待されている. 耳音響認証では IR 測定対象となる系に人間の聴覚器官が含まれる. そのため, IR 測定音を受聴した場合の受聴者の心理的ストレスが小さいものを利用すべきである. そこで本研究では人間にとって心地よいとされる自然環境音での IR 測定手法, クラウドソーシングを利用した大規模主観評価と周波数分析を用いた評価指標を提案する. 次に, 自然環境音が IR 測定信号として用いる有効性をシミュレーション IR 測定実験により実験的に検討する.

## 2 IR 測定信号評価手法

### 2.1 板倉斎藤距離による信号の周波数分析

IR 測定信号は測定対象となる周波数帯域を欠落なく含んでいる必要がある. そこで, 式 (1) の板倉斎藤擬距離 (Itakura-Saito divergence, ISD) [3] を使った周波数分析を提案する.

$$D(x|y) = \frac{y}{x} - \log \frac{y}{x} - 1 \quad (1)$$

IDS の導出手順について説明する. まず, IDS 導出対象である IR 測定信号に解析時間範囲をシフトさせながら, STFT (Short-Time Fourier Transform) 処理を行い, 周波数振幅特性のスペクトログラム (Spectrogram, SG) を導出する. SG の全時間範囲で周波数振幅特性の平均をとったものを, 式 (1) の  $x$  とする. 次に, 全周波数範囲で  $x$  の振幅特性平均値を求めたものを, 式 (1) の  $y$  とし, 基準値とする. IDS は, IR 測定信号の周波数振幅値と基準値  $y$  との差が全周波数帯域で小さい場合, その ISD は小さな値となる. ISD は基準値を中心に非対称な距離値をとる. IR 測定信号の周波数振幅特性が基準値  $y$  より大きい場合に比べ, 小さい場合は IDS が大きな距離値となる. IR 測定信号の周波数帯域にピークが含まれている場合よりも, デイップつまり周波数帯域の欠落がある場合に ISD は大きくなる.

### 2.2 大規模主観評価による心地よさの数値化

IR 測定信号の心地よさを客観指標を用いて定量化することが求められる. そこで, クラウドソーシングを用いた大規模主観評価により, その心地よさを定量化する. 主観評価として, 5段階 MOS (Mean Opinion Score) 評価を利用する. クラウドソーシングを用いた主観評価実験では, 評価者は Web 上でブラウザから実験に参加する. IR 測定信号を受聴し, その心地よさを, 「1) 全く心地よくない», 「2) 心地よくない», 「3) どちらとも言えない», 「4) 心地よい», 「5) とても心地よい」の5段階で回答する.

## 3 自然環境音を用いた IR 測定

本研究では, 自然環境音を用いた IR 測定に逆フィルタを利用する.

IR 測定信号を  $x(t)$  ( $t$  は時間インデックス) とし, その逆フィルタ  $x_i(t)$  を, 正規方程式を解くことで求める. ここで, 逆フィルタをベクトル表現したものを

$$\mathbf{x}_i \equiv [x_i(1), \dots, x_i(t), \dots, x_i(T)]^T \quad (2)$$

とする.  $T$  はサンプル数である.  $x(t)$  との畳込み演算を行列表現したものを  $\mathbf{A}$  とすると, 逆フィルタの関係より式 (3) が得られる. ここで,  $\delta$  は, 目的信号をベクトル表現したものである. 本稿では目的信号  $\delta$  にバンドパスフィルタの IR 信号を利用した.

$$\mathbf{A}\mathbf{x}_i = \delta \quad (3)$$

式 (3) の両辺に対して  $\mathbf{A}$  の転置行列である  $\mathbf{A}^T$  を両辺の左側から掛けることで Toeplitz 行列  $\mathbf{A}^T\mathbf{A}$  を含む式 (4) が得られる.

$$\mathbf{A}^T\mathbf{A}\mathbf{x}_i = \mathbf{A}^T\delta \quad (4)$$

式 (4) の両辺に  $\mathbf{A}^T\mathbf{A}$  の逆行列を掛け, 式 (5) 逆フィルタ  $x_i(t)$  を得る.

$$\mathbf{x}_i = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\delta \quad (5)$$

## 4 大規模主観評価実験

### 4.1 実験条件

人間にとって心地よいとされる自然環境音 [4] とし森, 川, 波の3ジャンルを選択した. Freesound [5] から各ジャンルに付き楽音信号を5つずつダウンロードした. 楽音信号の無音区間を除き, 372 ms の信号長で切り出し, 森は 148 個, 川は 193 個, 波は 357 個の合計 698 個を用意した. 次に, これらの楽音信号の ISD を求め, 各ジャンルより ISD の小さい上位 100 個, 計 300 個を IR 測定の候補信号とした. 候補信号の心地よさの評価を 2.2 節で述べた MOS 評価を用い, クラウドソーシングサービス「ランサーズ」 [6]

\*Design of comfortable impulse response measurement signal by large-scale subjective evaluation and frequency analysis, by Yuki Takahashi (National Institute of Technology, Nagaoka College), Junya Koguchi, Shinnosuke Takamichi (The University of Tokyo), Shohei Yano (National Institute of Technology, Nagaoka College), and Hiroshi Saruwatari (The University of Tokyo).

上で実施した。MOS 評価において、各評価者は 3 つの各ジャンルからランダム抽出された 6 個、計 18 個の候補信号を受聴し、その心地よさスコアを回答した。評価には 294 名が参加し、合計 5292 個の評価スコアを得た。各候補信号の評価スコアを平均したものを MOS 値とした。

## 4.2 実験結果

大規模主観評価実験結果を Fig. 1 に示す。Fig. 1 において縦軸は ISD を、横軸は MOS 値を示す。波ジャンルの候補信号は、MOS 値が 2.5 から 3.5 の範囲で分布している。これは聴覚的な心地よさが「どちらとも言えない」であることを示す。また、ISD が 20 周辺であり、IR 測定信号としてはある程度適していることを示す。川ジャンルの候補信号は、MOS 値が 2 を中心とするものと 4 を中心とする 2 クラスに分かれている。これは、聴覚的な心地よさが「心地よい」ものと「心地よくない」ものとに二分していることを示す。「心地よい」と回答された候補信号の ISD は 20 を超えており、「心地よくない」と回答された候補信号の ISD は 20 を下回っている。これは、僅かではあるが聴覚的には心地よくないが IR 測定信号としては適していることを示している。森ジャンルの候補信号は、MOS 値が 2 から 4 に分布している。これは、聴覚的な心地よさが「心地よい」ものから「心地よくない」ものまで幅広く分布していることを示す。また、ISD は 0 周辺であり、森ジャンルの候補信号は IR 測定信号として適していることを示している。

## 5 自然環境音用いた IR 測定実験

### 5.1 シミュレーション実験

4.1 節で作成した候補信号の内、ISD が最大と最小のものを用いて、IR 測定実験を行った。実験において系の伝達特性は外耳道音響伝達特性を使用する。候補信号を用いて系の伝達特性を導出するシミュレーション実験を行った。系の伝達特性  $h(t)$  には予め測定しておいた外耳道音響伝達関数の IR を用いた。式 (6) に示すように、候補信号  $x(t)$  を伝達特性  $h(t)$  を持つ系に入力し、その応答  $y(t)$  を得る。

$$x(t) * h(t) = y(t) \quad (6)$$

\* は畳み込み演算を表す。次に式 (7) に示すように、逆フィルタを用いて伝達特性  $h(t)'$  を導出する。

$$\begin{aligned} x_i(t) * x(t) * h(t) &= x_i(t) * y(t) \\ \delta * h(t) &= x_i(t) * y(t) \\ h(t)' &= x_i(t) * y(t) \end{aligned} \quad (7)$$

得られた  $h(t)'$  と  $h(t)$  との誤差を式 (8) で示すパワースペクトル誤差 (Power Spectral Error, PSE) を導入し評価した。ここで、 $X_{ISg}(\omega)$ 、 $X_{ISb}(\omega)$  は ISD が最小及び最大の候補信号のパワースペクトル ( $\omega$  は周波数)、 $H(\omega)$  は系の伝達特性  $h(t)$  のパワースペクトルである。

$$PSE = \sum_{\omega} \left( 20 \log 10 \frac{X_{ISg, ISb}(\omega)}{H(\omega)} \right)^2 \quad (8)$$

### 5.2 実験結果

シミュレーション実験結果は ISD が最小の候補信号  $X_{ISg}$  では、ISD = 0.918 となり、PSE = 23430 となった。また、ISD が最大の候補信号  $X_{ISb}$  では、

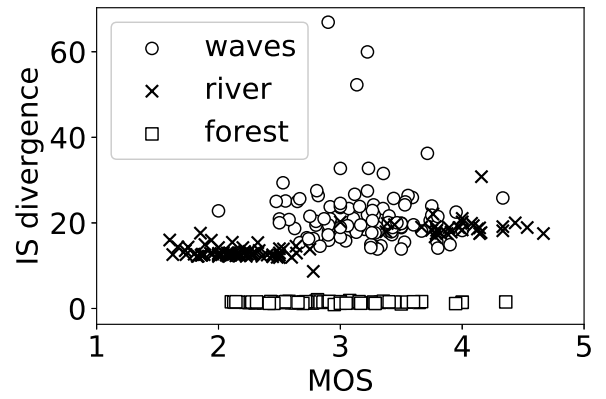


Fig. 1 ISD 値と MOS 値の散布図。waves, river, forest はそれぞれ、波・川・森ジャンルを指す。

ISD = 66.9 となり、PSE = 40024 となった。これは、ISD が IR 測定信号の評価指標として有効であることを示している。

## 6 おわりに

本研究では人間にとって心地よいとされる自然環境音での IR 測定手法、クラウドソーシングを利用した大規模主観評価と周波数分析を用いた評価指標を提案し、IR 測定信号に適した自然環境音をシミュレーション実験より評価した。大規模主観評価実験の結果、森、川、波の 3 ジャンルに共通して、距離値と MOS 値に強い相関が見られないことが分かった。このことは、IR 測定信号の有効性と人間の知覚する心地よさを両立する IR 測定信号を設計しうる可能性を示唆していると考えられる。自然環境音用いた IR 測定実験の結果、ISD が IR 測定信号の評価指標として有効であり、自然環境音を用いた IR 導出手法の有効性を実験的に検証した。

謝辞：本研究は、東京大学 GAP ファンドプログラム「音声合成技術の研究開発・商用利用を加速させる音声コーパスの設計・構築」の支援を受けて実施した。

## 参考文献

- [1] S. G.-Bart, E. J.-Jacques, and A. Dominique, "Comparison of different impulse response measurement techniques," *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 249–262, 2002.
- [2] T. Arakawa, T. Koshinaka, S. Yano, H. Irisawa, R. Miyahara, and H. Imaoka, "Fast and accurate personal authentication using ear acoustics," in *Proc. APSIPA ASC*, Jeju, Korea, Dec. 2016, pp. 1–4.
- [3] 板倉 文忠, "統計的手法による音声分析合成系に関する研究," 名古屋大学博士学位論文, 1972.
- [4] 島井 哲志 and 田中 正敏, "環境音の快-不快評価と音圧の関係," *日本音響学会誌*, vol. 49, no. 4, pp. 243–252, 1993.
- [5] "freesound," <https://freesound.org>.
- [6] "Lancers," <https://www.lancers.jp>.