

聴覚印象を考慮したインパルス応答測定信号設計

高橋 勇希^{4,a)} 小口 純矢^{2,b)} 高道 慎之介^{1,c)} 矢野 昌平^{3,d)} 猿渡 洋^{1,e)}

概要: 音響インパルス応答 (IR) の測定には, Maximal Length Sequence や Time Stretched Pulse などの測定信号の利用が一般的である. 通常, 被測定系に人間が含まれるかどうかは考慮されないため, これらの測定信号が人間に与える聴覚印象が無視されている. 他方, 新たな生体認証として, 人間の外耳道の音響インパルス応答が提案されており, 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. この被測定系は人間の聴覚器官を含むため, 受聴者の心理的ストレスが小さい測定信号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主観評価と周波数分析を用いた評価指標を提案する. 実験的評価結果より, (1) 測定信号としての妥当さと聴覚的な心地良さを両立する自然環境音が存在すること, (2) 板倉斎藤擬距離に基づく測定信号選定は, 雑音環境下の高精度 IR 測定にあまり寄与しないことを明らかにする.

YUKI TAKAHASHI^{4,a)} JUNYA KOGUCHI^{2,b)} SHINNOSUKE TAKAMICHI^{1,c)} SHOHEI YANO^{3,d)}
HIROSHI SARUWATARI^{1,e)}

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. これらの信号は, 全周波数帯域の成分が含まれることを理論的に保証しているため, IR 測定信号として妥当である. しかしながら, これらの信号を用いた測定では, 通常, 被測定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, これらの信号が人間に与える聴覚印象は無視されている.

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, 外耳道入口から IR 測定信号を放射し, 外耳道内の伝達と鼓膜による反射を通して得られる信号を, 再び外耳道入口にて收音する. この枠組みは, 被測定系として人間の聴覚器官を含む. すなわち, 認証対象の人間は IR 測定信号を受聴する. そのため, IR 測定信号として, 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, 聴覚印象を考慮した IR 測定信号の設計法を提案する. 人間にとって心地よいとされる自然環境音から, IR 測定信号としての妥当さと聴覚的な心地良さを両立する IR 測定信号を選択する. IR 測定信号としての妥当さは板倉斎藤擬距離, 聴覚的な心地良さはクラウドソーシングを利用した大規模主観評価により数値化する. 実験的評価では, これらの 2 要素を両立する IR 測定信号が存在しうるかと, 選択した IR 測定信号がシミュレーション IR 測定において有効であるかを検証する. その結果, (1) 測定信号としての妥当さと聴覚的な心地良さを両立する自然環境音が存在すること, (2) 板倉斎藤擬距離に基づく測定信号選定は, 雑音環境下の高精度 IR 測定にあまり寄与しないことを明らかにする.

2. IR 測定信号評価手法

2.1 板倉斎藤擬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

¹ 東京大学 大学院情報理工学系システム情報学専攻, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.

² 明治大学, Meiji University, Nakano, Tokyo 164-8525, Japan.

³ 長岡高専 電気電子システム工学科, 888 Nishikatai, Nagaoka, Niigata 940-8532, Japan.

⁴ 新潟大学 工学部工学科電子情報通信プログラム, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan.

a) t20b914a@mail.cc.niigata-u.ac.jp

b) cs202027@meiji.ac.jp

c) shinnosuke_takamichi@ipc.i.u-tokyo.ac.jp

d) syano@nagaoka-ct.ac.jp

e) hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

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。

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

ISD 値の計算手順について説明する。まず、ISD 値計算対象である 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) を行い、周波数振幅特性のスペクトログラム (Spectrogram, SG) を計算する。SG を全時間周波数で平均したものを基準値 y とし、各時間周波数の SG を x とする。各時間周波数で式 (1) の値を計算し、それを全時間周波数で平均したものを ISD 値とする。ISD 値は、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 値は小さな値となる。ISD 値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 ISD 値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、ディップつまり周波数帯域の欠落がある場合に ISD 値は大きくなる。

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、 δ 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 δ 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) の逆フィルタ $x_i(t)$ を得る。

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

4. 実験的評価

4.1 実験条件

測定信号としての妥当性と聴覚的な心地よさに関する実験を実施する。人間にとって心地よいとされる自然環境音 [4] として森、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音信号を5つずつダウンロードした。音信号の無音区間を除き、372ms 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 値を求め、各ジャンルより ISD 値の小さい上位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] 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

測定信号としての妥当性と聴覚的な心地よさに関する実験結果を図1に示す。図1において縦軸は ISD 値を、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 値が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 値は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 値は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。

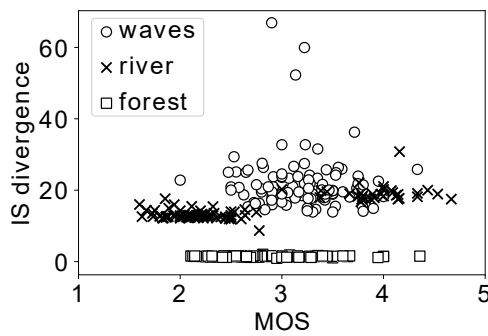


図 1 ISD 値と MOS 値の散布図. waves, river, forest はそれぞれ、波・川・森ジャンルを指す。

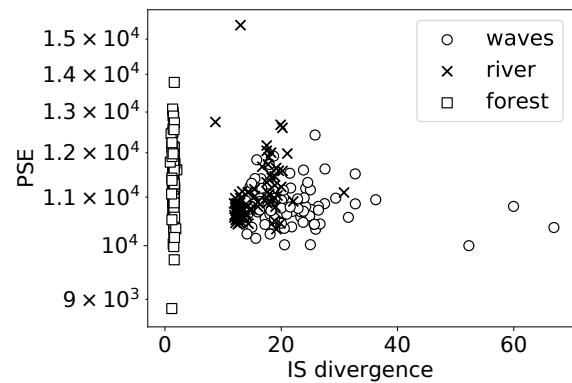


図 2 PSE 値と ISD 値の散布図. waves, river, forest はそれぞれ、波・川・森ジャンルを指す。

5. 自然環境音用いた IR 測定実験

ISD 値では周波数特性にピークが存在するよりもディップが存在する場合に、IR 測定用信号としての評価を下げる指標となっている。そのため、ISD 値の小さい測定信号は雑音の影響を強く受けないと考えられる。雑音下における IR 測定信号の選定における ISD 値との関係を雑音付加 IR 測定シミュレーション実験により検証する。

5.1 雑音付加シミュレーション実験

雑音付加シミュレーション実験では 4 節で作成した候補信号を用いて系の伝達特性を導出する。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。雑音信号 $n(t)$ には DEMAND Corpus [7] の環境音 SCAFE_48k の 1ch 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、雑音信号 $n(t)$ を付加し、その応答 $y(t)$ を得る。

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

* 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 IR 測定精度の評価をした。ここで、 $X(\omega)$ は候補信号のパワースペクトル (ω 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。

$$PSE = \sum_{\omega} \left(20 \log_{10} \frac{X(\omega)}{H(\omega)} \right)^2 \quad (8)$$

5.2 シミュレーション実験結果

シミュレーション結果として、横軸に ISD 値、縦軸に PSE 値を持つ図を図 2 に示す。相関係数は、森ジャンルが

-0.08、川ジャンルが 0.296、波ジャンルが -0.09 となり、いずれのジャンルにおいても、ISD 値と PSE 値に強い相関は見られない。以上より、ISD 値に基づく IR 測定信号選定は、高精度 IR 測定にあまり寄与しないことが分かる。

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。また、板倉斎藤擬距離に基づく測定信号選定は、実環境シミュレーション下の高精度 IR 測定にあまり寄与しないことが分かった。

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」の支援を受けた。

参考文献

- [1] G.-Bart, S., J.-Jacques, E. and Dominique, A.: Comparison of different impulse response measurement techniques, *Journal of the Audio Engineering Society*, Vol. 50, No. 4, pp. 249–262 (2002).
- [2] Arakawa, T., Koshinaka, T., Yano, S., Irisawa, H., Miyahara, R. and Imaoka, H.: Fast and accurate personal authentication using ear acoustics, *Proc. APSIPA ASC*, Jeju, Korea, pp. 1–4 (2016).
- [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, 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>.
- [7] : DEMAND: a collection of multi-channel recordings of acoustic noise in diverse environment, https://zenodo.org/record/1227121#.Xqa_rWj7SHt.