

# S10 RWCP 音声・音響データベースを用いた環境音・効果音合成の検討とオノマトペ拡張データセットの構築

井本 桂右<sup>†</sup>, 岡本 悠希<sup>‡</sup>, 高道 慎之介<sup>‡</sup>, 山下 洋一<sup>‡</sup> (<sup>†</sup>同志社大学, <sup>‡</sup>立命館大学, <sup>‡</sup>東京大学)

## ① 研究背景と目的

### 環境音・効果音合成とは？

- 音声や楽音に限らないあらゆる音を人工的に合成
- 雨の音, 足音, ドアノック音, 鳥の鳴き声, 架空のキャラクターの声 ...

### 環境音・効果音合成技術の活用例

- 映画やゲームコンテンツの制作補助, 自動生成
- コンテンツにふさわしい環境音・効果音を自動付与
- 高度な VR やメタバースサービスの実現
- 仮想空間であらゆる環境音を再現
- 架空のキャラクターになりきって会話
- 動物の行動コントロール
- 畜産動物の行動管理, 獣害の追い払いシステム



### 本研究の目的

- 多様な環境音・効果音を簡単に合成する手法の実現
- 環境音・効果音合成に必要なデータセットの構築

## ② 従来研究の課題と本研究の方針

### 環境音・効果音合成の従来手法

- 個々の環境音に対する物理モデルを構築
- 雨の音を物理的にモデル化して合成 [1]
- 流体力学と空気力学を用いてプロペラの回転音を合成 [2]

⊗ 合成する環境音の数だけモデルが必要になる

### 深層学習を用いた環境音合成技術

- 音声合成手法である WaveNet を用いた環境音合成 [3]
- ⊕ 単一のモデルで複数の環境音を合成可能
- ⊗ 合成する音の種類のみ指定可能
- ➡ 同じ種類の音でも場面によって細かく使い分けたい



### 本研究の方針

- 1つの機械学習モデルで多様な音を合成する手法を検討
- オノマトペを活用して合成音の細かい表現を制御

## ③ オノマトペ拡張データセットの構築

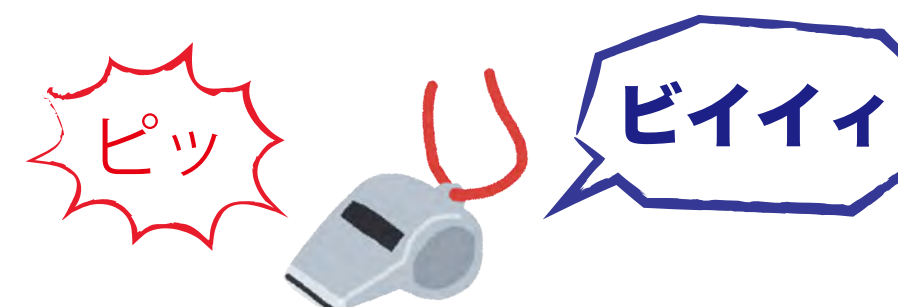
### RWCP 音声・音響データベース (RWCP-SSD) [4] の概要

- 非音声（環境音）のドライソースなどを収録
- 105 種類の環境音, 合計 9,722 サンプル (各音源 100 サンプル程度)
- 各サンプルに対する環境音ラベル
- 音声資源コンソーシアム (NII-SRC) が管理・配布

### RWCP-SSD は環境音・効果音合成に利用可能か？

- 環境音ラベルを入力とした合成手法なら実現可能
- 例) 「笛の音」とだけ指定可能な合成システム

⊗ 音の継続長や音色などは制御できない



※ [https://github.com/Keisukeimoto/RWCPSSD\\_Onomatopoeia](https://github.com/Keisukeimoto/RWCPSSD_Onomatopoeia)

### RWCP-SSD のオノマトペ拡張データセットの構築

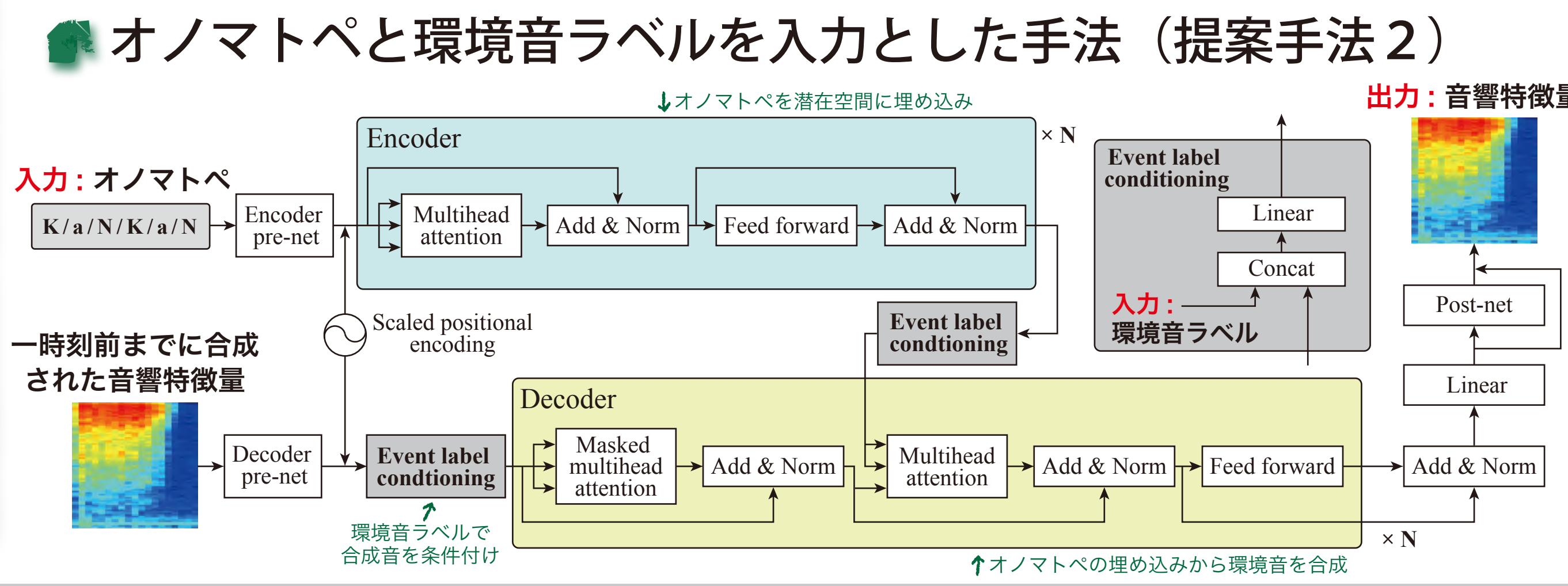
- 拡張データセット (RWCP-SSD-Onomatopoeia) \* の収録内容
- 計 155,568 個のオノマトペ (1つの環境音に対し 15 以上のオノマトペを収集)
- オノマトペに対する 155,568 個の自信度
- オノマトペに対する 548,367 個の許容度 (オノマトペ付与者以外の作業者が付与)
- 作業者のワーカ ID (匿名化済み)
- 取得されたオノマトペの例
- 笛の音 (whistle1) → 「ピーツ」「ピィィィ」「フィー」「ミーツ」など
- ひげ剃りの音 (shaver) → 「ジュー」「ギューン」「ビィィン」「ギウギウ」など
- ➡ 環境音に付与されるオノマトペは大きくばらつく
- ➡ オノマトペ付与者の作業自信度と他の人が感じたオノマトペの許容度も収集

## ④ 環境音・効果音合成の提案手法

### オノマトペを入力として環境音・効果音を合成

- Sequence-to-sequence (seq2seq) や Transformer を利用
- オノマトペを環境音・効果音へ変換する深層学習モデル
- 複数の環境音・効果音を 1つのモデルで合成可能
- 提案手法 1: オノマトペのみを入力として音を合成
- 提案手法 2: オノマトペと環境音ラベルを入力として音を合成

### Transformer による環境音・効果音の合成モデル



## ⑤ 評価実験

### 実験条件

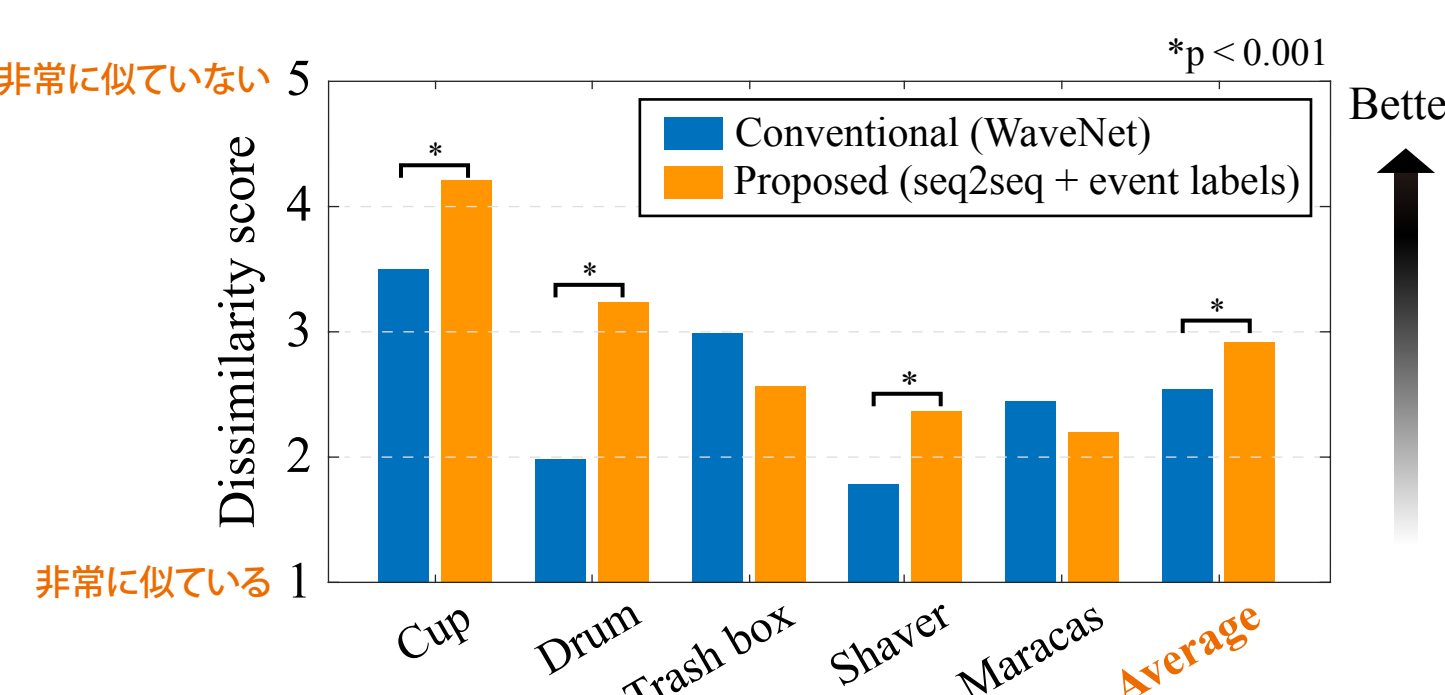
- 利用データ
- RWCP 音声・音響データベースに含まれる 10 種類の環境音 (計 1,000 サンプル)
- RWCP-SSD-Onomatopoeia に含まれる 14,250 個のオノマトペ
- 環境音合成手法
- 従来手法: WaveNet (環境音ラベルを入力して合成)
- 提案手法: seq2seq, Transformer (オノマトペ & 環境音ラベルを入力して合成)

ベルの音, ドラムの音, ひげ剃りの音, カップを叩く音, 目覚ましの音, マラカスの音, コーヒー豆をミルで挽く音, 紙を引き裂く音, ゴミ箱を叩く音, 笛の音

### 実験結果

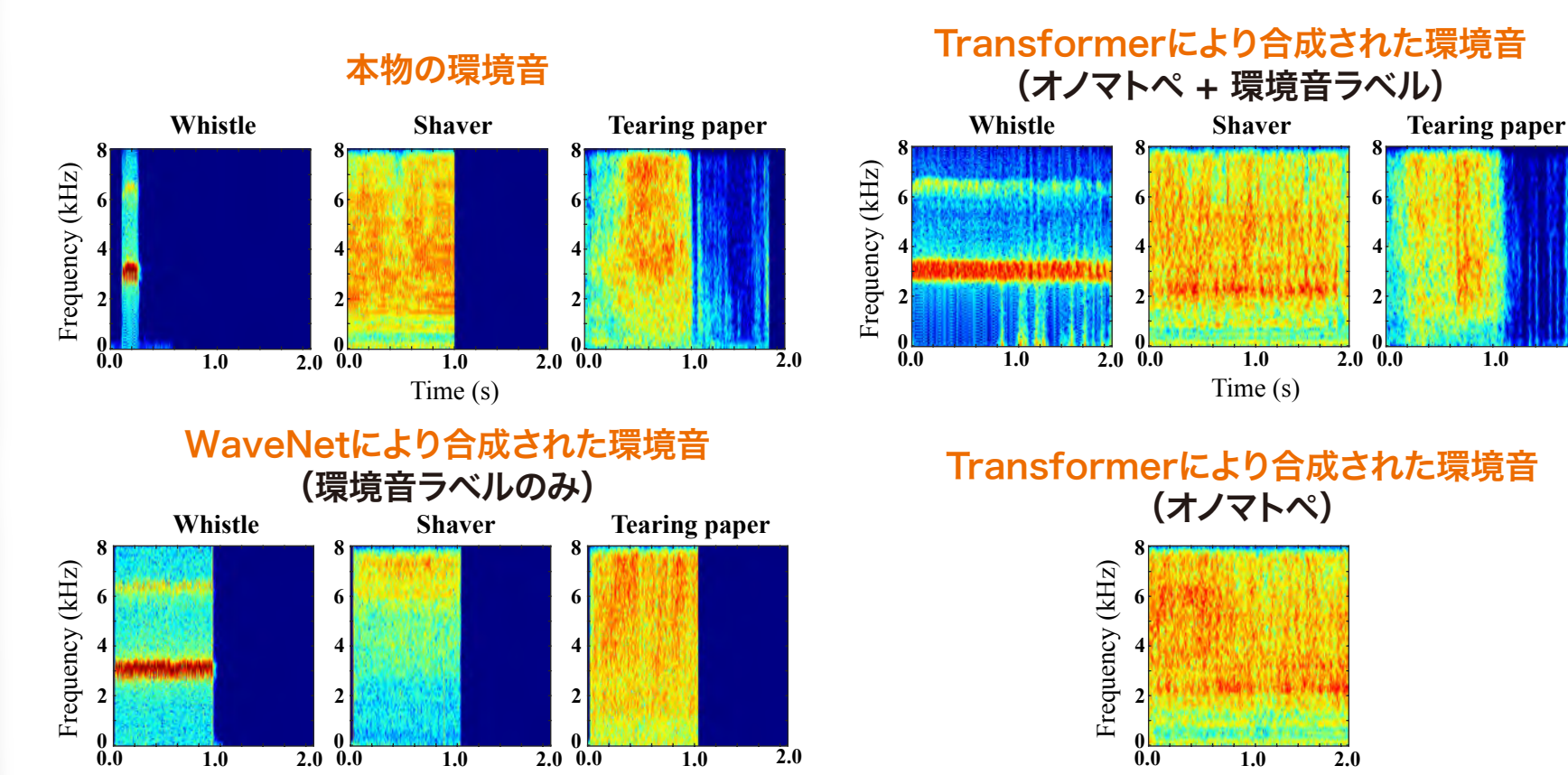
#### 合成された環境音の多様性

- 5 種類 × 5 音の環境音を合成
- 各合成音 30 人 (計 750 人) の評価
- 非類似度の MOS score で評価
- ⊕ 提案手法を用いることで多様な音が合成可能



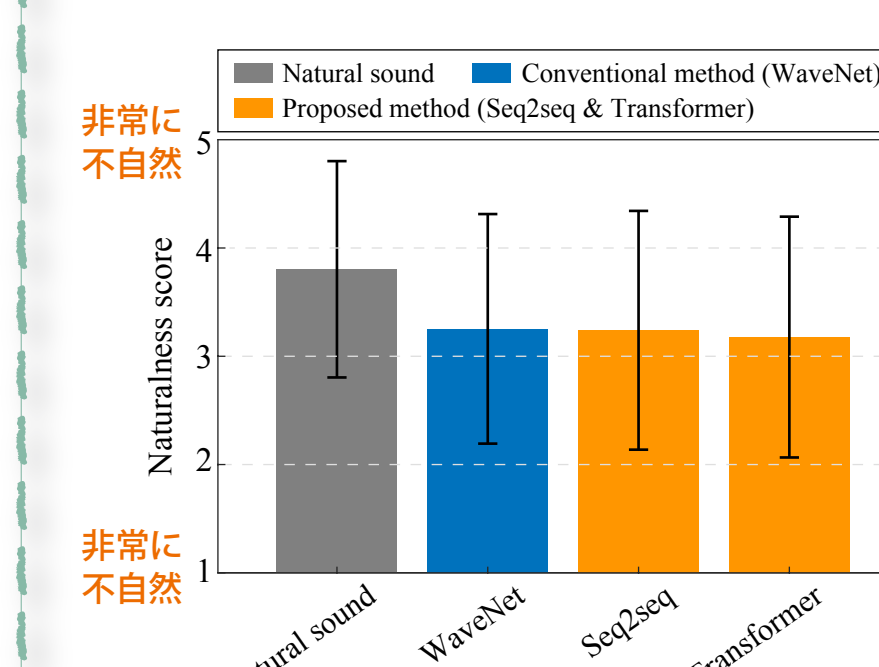
#### 合成された環境音の可視化と合成音の自然性

- 合成音のスペクトログラム
- ⊕ オノマトペからそれらしい環境音が合成可能



#### 自然性の主観評価結果

- 10 種類 × 10 音の環境音
- 各合成音 30 人 (計 3,000 人)
- 自然性の Mos score で評価
- ⊕ オノマトペから環境音を合成しても自然性は低下しない



#### 合成音のデモページ

- [https://y-okamoto1221.github.io/Onoma\\_to\\_wave\\_Demonstration/](https://y-okamoto1221.github.io/Onoma_to_wave_Demonstration/)
- [https://y-okamoto1221.github.io/Transformer\\_onoma\\_to\\_wave/](https://y-okamoto1221.github.io/Transformer_onoma_to_wave/)

## ⑥ 今後の計画

- オノマトペに対する自信度, 許容度を活用した環境音合成
- 映像なども入力としたマルチモーダル環境音・効果音合成

### 参考文献

- [1] S. Liu, H. Cheng, et al., "Physically-based Statistical Simulation of Rain Sound," ACM Transactions on Graphics, vol. 38, no. 4, pp. 1–14, 2019.
- [2] R. Selfridge, D. Moffat, et al., "Physically Derived Sound Synthesis Model of a Propeller," Proc. AM, pp. 1–8, 2017.
- [3] Y. Okamoto, K. Imoto, et al., "Overview of Tasks and Investigation of Subjective Evaluation Methods in Environmental Sound Synthesis and Conversion," arXiv, arXiv:1908.10055, 2019.
- [4] S. Nakamura, K. Hiyané, et al., "Sound Scene Data Collection in Real Acoustical Environments," Journal of ASJ, vol. 20, no. 3, 225–231, 1999.

### 謝辞

本研究の一部は, 2020 年度ならびに 2021 年度 国立情報学研究所 公募型共同研究 (20S0401, 21S0502) の助成を受けたものである。ここに記して謝意を表す。