

## 環境音の模倣音声を利用した環境音合成とデータセット構築

岡本 悠希<sup>†,††</sup> 井本 桂右<sup>†††</sup> 高道慎之助<sup>††,††††</sup>

永瀬亮太郎<sup>†</sup> 福森 隆寛<sup>†††††</sup> 山下 洋一<sup>†††††</sup>

† 立命館大学大学院情報理工学研究科 〒567-8570 大阪府茨木市岩倉町 2-150

†† 東京大学大学院情報理工学系研究科 〒113-8656 東京都文京区本郷 7-3-1

††† 同志社大学文化情報学部 〒610-0321 京都府京田辺市多々羅都谷 1-3

†††† 慶應義塾大学理工学部 〒223-8522 神奈川県横浜市港北区日吉 3-14-1

††††† 立命館大学情報理工学部 〒567-8570 大阪府茨木市岩倉町 2-150

E-mail: †y-okamoto@ieee.org

**あらまし** 環境音の特徴を表現する方法の一つに、人間の声によって環境音を模倣した音声（以下、模倣音声）が挙げられる。従来の環境音合成における入力情報（画像、テキストなど）では、環境音の音高やリズムの特徴を表現できない。一方で模倣音声は、環境音の音高やリズムを表現するために有効である。そこで本論文では、環境音に対する模倣音声を収録したデータセットを構築する。また、構築したデータセットを用いた環境音合成手法を提案する。環境音合成モデルの入力に模倣音声と音響イベントラベルを用いることで、合成音の音源の種類、音高、リズムの制御が期待できる。評価実験より、模倣音声の利用が、合成音の音高やリズムの制御に有効であることを示した。

**キーワード** 環境音合成, 環境音変換, 音声模倣, 音響イベントラベル

## Environmental sound synthesis and creation of dataset using vocal imitations

Yuki OKAMOTO<sup>†,††</sup>, Keisuke IMOTO<sup>†††</sup>, Shinnosuke TAKAMICHI<sup>††,††††</sup>,

Ryotaro NAGASE<sup>†</sup>, Takahiro FUKUMORI<sup>†††††</sup>, and Yoichi YAMASHITA<sup>†††††</sup>

† Graduate School of Information Science and Engineering, Ritsumeikan University

†† The University of Tokyo

††† Faculty of Science and Engineering, Doshisha University

†††† Faculty of Science and Technology, Keio University

††††† College of Information Science and Engineering, Ritsumeikan University

E-mail: †y-okamoto@ieee.org

**Abstract** One way to represent the characteristics of environmental sounds is to imitate the environmental sounds by human voice called vocal imitation. The input information (e.g., image, text, etc.) in conventional environmental sound synthesis cannot represent the pitch and rhythm of environmental sounds. On the other hand, vocal imitation is effective in representing the pitch and rhythm of environmental sounds. We thus create a dataset of vocal imitations of environmental sounds. We also propose an environmental sound synthesis method using vocal imitation of environmental sounds. Using vocal imitations and sound event labels as input to the environmental sound synthesis model, we can control the pitch, rhythm, and sound event of the synthesized sounds. Experimental results show that using vocal imitations effectively controls the pitch and rhythm of synthesized sounds.

**Key words** Environmental sound synthesis, environmental sound conversion, vocal imitation, sound event label