

# 概要

本論文では、音声や楽音に限らないあらゆる音である環境音を対象とした音の合成技術を扱う。映画やアニメ、舞台などのコンテンツ制作において、「雨音」や「風の音」などといった環境音は状況の説明や登場人物の心象の象徴など、非常に重要な役割を担う。従来、音声や楽音に対しては、新たな音を生成するための合成・変換技術に関する検討が数多く行われてきた。一方、環境音を対象とした合成・変換技術に関する研究は極めて少ない。また、様々な環境音の特徴をどのようにすれば制御可能であるかについては明らかになっていない。本論文では、多様な環境音の特徴を柔軟に制御するため、様々な入力情報を用いた統計的手法に基づく環境音合成手法を提案する。多様な音を柔軟に制御可能とする環境音合成技術の実現は、映画や動画コンテンツ、ゲームなどの背景音、効果音作成、環境音分析のためのデータ拡張など様々な用途への応用が期待できる。

本研究ではまず、環境音に対してオノマトペを付与した環境音合成のためのデータセット構築を行う。統計的手法によって環境音を合成するにあたり、合成モデルの学習に使用するデータセットの整備は非常に重要な課題である。また、既に公開されているデータセットでは、環境音に対して音響イベントラベルが付与されたデータは多く存在する一方、その他のラベル情報が付与された環境音データセットは極めて少ない。そこで本論文では、環境音の時間的特徴を表現するために有効であるとされるオノマトペを環境音に付与したデータセットを構築する。オノマトペ付与の際には、オノマトペを記述した本人による自信度のスコアと、他者からのオノマトペに対する許容度のスコアを収集する。収集した自信度と許容度のスコアを分析することで、収集されたオノマトペが環境音に対して適切であることを示す。

続いて、環境音合成のための評価方法の整理を行い、合成された環境音をどのように評価するべきかについて検討する。本論文では、環境音合成のための主観評価手法の提案を行う。評価実験にて、環境音合成で使用される主観評価と客観評価手法による結果を比較して、合成された環境音をどのように評価するべきかについて示す。

さらに本論文では、様々な入力情報を用いた環境音合成手法を提案する。まず、「雨音」や「風の音」といった音響イベントラベルから環境音を合成するための手法を提案する。音響イベントラベルでは、生成したい環境音全体の印象を表現す

ることが可能である。そのため、音響イベントラベルを環境音合成の入力情報として使用することで、合成音の音源の種類が期待できる。合成音の品質評価実験により、入力情報として音響イベントラベルを利用することが有効であることを示す。また、半数近くの音響イベントにおいてはデータセットに含まれる自然音と同程度の自然性を獲得できることを示す。

続いて、本論文にて構築したデータセットを用いて、オノマトペからの環境音合成手法を提案する。オノマトペは音の時間的変化を表現するために有効であるため、環境音合成の入力情報として利用することで、合成音の繰り返し回数などといった時間的変化の制御が期待できる。また、オノマトペと音響イベントラベルを同時に環境音合成の入力情報として用いることで、合成音の時間的変化と音源の種類の同時制御を行う。合成音に対する評価実験より、オノマトペを利用することが合成音の時間的変化の制御に有効であることを示す。また、音響イベントラベルとオノマトペを同時に入力することで、音響イベントラベルのみを入力する場合よりも多様な環境音が生成可能であることを示す。

最後に、環境音を模倣した音声を用いた環境音合成手法を提案する。環境音の音高やリズムを表現する方法として、人の声による環境音の模倣が挙げられる。声による環境音の模倣は、直感的に環境音の音高やリズムを表現することができる。そのため、環境音を模倣した音声（模倣音声）を環境音合成に利用することで、合成音の音高やリズムの制御が期待できる。合成音の評価実験より、環境音を模倣した音声は合成音の音高とリズムの制御に有効であることを示す。また、入力に使用する模倣音声の音高やリズムを変化させた場合、それに追従して合成音の音高とリズムが変化することも評価実験にて示す。