

論文の内容の要旨

論文題目 Interactive Methods
for Computational Sound Design, Control, and Calibration
(コンピュータによるサウンドデザイン,
コントロール, キャリブレーションのための対話的手法)

氏 名 山本 和彦

Computational sound plays an important role in digital computer entertainment. Traditionally, fixed set of sound data is prepared beforehand and the system simply plays these sounds at runtime. However, as the demand for high quality virtual experience increases, there is an increase need for synthesizing appropriate sound interactively responding to user control at runtime. Unfortunately, technology development in sound effects is lagging behind compared to those for visual effects and there remain various difficulties to bring interactivity in sound processing. For this reason, according to the increases of its interactivity, it arises a problem that the efforts of the artist for designing a lot of sounds one by one would be increased. On the other hand, the user would be dissatisfied with a control device which has only a limited degree of freedom due to the difficulty for manipulating the sound. Furthermore, although it becomes important to render the sounds with appropriately calibrating to the individual specifications or tastes of each user, there is a problem that it is difficult for the designer to gather such the information beforehand. To address these problems, this thesis presents methods to bring higher interactivity to computational sound by reducing computational cost and user's operation cost.

First, we discuss interactivity in the design of physically based sound. Although a sound designer needs to design appropriate material properties for physically based sound, exploration of the design space of physical parameters directly is unintuitive and difficult. This thesis addresses this problem by an example-based user interface that optimizes inversely the material from a few example inputs of sound clips by the user. Second, we discuss interactivity in the control of singing voice synthesizer. A user needs to continuously input lyric and melody to use singing voice synthesizer in improvisational performance.

However, inputting them both simultaneously is difficult with standard input devices. This thesis addresses this problem by predicting latent lyrics desired by the user in realtime from the input of a standard control device. Finally, we discuss interactively in calibration of 3d spatial sound. Although it is necessary to calibrate a spatial audio system to adopt them for a specific user, traditional calibration procedure is expensive and time consuming. This thesis addresses this problem by adaptation of the system for a specific user using an user interface that requires simple pairwise comparison tasks.

To achieve these goals, this thesis builds a computational model behind each task in precomputation, and exploits the model to reduce computational cost or user's operation cost at runtime. Specifically, in the design of physically based sound, we present dramatically fast vibrational analysis using precomputed mesh simplification algorithm using machine learning and hierarchical component mode synthesis, which allows material optimization at interactive rate. In the control of singing voice synthesizer, we present a method to estimate latent lyrics as higher DoF parameters from the input of lower DoF control device using machine learning of lyrics dataset at precomputation phase. Finally, in the calibration of 3d spatial audio, we propose a machine learning model that allows adaptation of the system to a specific user using individual and non-individual factors of dataset which are extracted at precomputation. This thesis describes the details of each method and presents the results of numerical evaluations and empirical studies with end users. These methods, in which the ideas are based on the improvements the interactivity of runtime applications by user interfaces using precomputation, have generality and can be widely applied to similar problems in other domains.

インタラクティブなデジタルコンテンツにおいて音の利用は重要な役割を担う。従来は、あらかじめ用意された限られた数のサウンドファイルを実行時にユーザのインタラクションに対して鳴らす、というのが一般的な音の利用方法であった。しかし、デジタルコンテンツへのインタラクティブ性への要求が高まるにつれて、ユーザのあらゆるインタラクションに対して適切な音で反応することが求められてきているが、それを実現するための音の利用技術は映像技術と比較して非常に遅れている。現状、コンテンツのインタラクティブ性が高まるほど、アーティストにとっては多くの音をデザインする労力が増し、一方ユーザにとっては、限られた自由度の操作手段で多くの要素を制御することが困難であること、に対する不満が高まっていく。また、個人差のあるユーザそれぞれの知覚や好みに合わせてキャリブレーションした上で音をレンダリングすることの重要性が高まるが、そうした情報をデザイナーがあらかじめ収集しておくことは困難である、という問題もある。これらを解決するために、本論文では、コンピュータの計算コストとユーザの操作コスト両方を大きく軽減することによって、デジタルコンテンツにおける音の利用のインタラクティブ性を高める手法を提案する。

第一の手法では、物理ベースのサウンドデザインについて論じる。物理ベースのサウンドをデザインするためには、デザイナーは適切な材質という直感的ではない物理的なパラメータを設定する必要がある、これは非常に困難であるという問題がある。そこで本論文では、例示ベースのユーザインタフェースでデザイナーの所望する音を入力することで逆に材質パラメータを最適化し、この問題を解決する。第二の手法では、歌声合成のリアルタイムコントロールを例に挙げる。ユーザは歌詞とメロディ両方を同時に入力する必要があるが、これは従来の入力デバイスでは非常に困難である。本論文では、従来通りの入力デバイスからの入力から、ユーザの想定した歌詞を、リアルタイム予測をすることで、この問題を解決する。第三の手法では、3次元音響のユーザへのキャリブレーションについて挙げる。3次元音響を再生するためには、ユーザに対して適切なキャリブレーションをおこなう必要があるが、従来これは特別な機材や膨大な時間が必要であり困難であった。そこで本論文では簡単な比較タスクのみによるユーザインタフェースでシステムの個人適応をおこない、この問題を解決する。

これらを実現するために、本論文では、事前計算によってこれらのタスクを表現する計算モデルを構築することで、実行時のコンピュータの計算コストやユーザの操作コストを大幅に軽減するアルゴリズムを提案する。具体的には、物理ベースサウンドのデザインをするために、事前の機械学習によるメッシュ簡略化と階層的モダルフ分解を利用した高速な振動解析を提案し、インタラクティブな速度での材質最適化を可能にする。また、歌声合成のリアルタイム演奏を可能にするために、歌詞データを事

前に機械学習しておくことによって、低自由度の入力からの高次情報のリアルタイム予測をおこなう。さらに3次元音響のユーザへのキャリブレーションをおこなうために、事前にデータセットから個人性と非個人性を分離して個人適応に利用することのできる機械学習モデルを提案する。本論文では、それぞれの手法について詳しく述べるとともに、数値実験やユーザスタディによってその有効性を実証する。本論文で提案している、コンピュータによる事前計算を利用したユーザインタフェースによって、アプリケーション実行時のインタラクティブ性を高める手法は、広く他の分野にも応用することが可能である。