

## 音響信号の直接入力による概念からの音合成

吉田 誠一<sup>†</sup> 小坂 直敏<sup>‡</sup>  
 東京電機大学<sup>†</sup> 東京電機大学<sup>‡</sup>

## 1 はじめに

音楽、映画、ゲームなど、アミューズメントから先端芸術までのさまざまなデジタルコンテンツの制作において、音合成システムは音制作者の重要なツールである。

音合成にあたり、ユーザは必要な音を明確に指定できないため、サンプル音、あるいは合成音から基本となる音を設定し、エフェクトを付加し、試行錯誤により加工しながら目的とする音に辿りつく。この合成法は最終的に意図する合成は得られるかもしれないが、操作に多大な時間を要し、一定品質の合成音を効率的に作る方法としては好ましくない。ユーザの表現意図を明確に物理的に機械に入力できれば、この問題は解決する。

このため本稿では、入力として目的の音に対するイメージやアイデアを音声や一般の音などの音響信号から直接入力し、その振幅とピッチなどの韻律情報を基に合成音を作成する手法を提案する。

## 2 表現意図の入力方法

音の音色、高さ、音量などのユーザの表現意図をここでは概念と呼ぶ。概念を何らかのテキストで表現しこれを用いて伝える場合と、アナログ情報で直接伝える場合が考えられる。本研究では後者の方法を選び、目的の音に対するイメージやアイデアを音響信号で表現し、これをマイクから直接入力する。

本研究の特徴を以下に示す。

## (1) 音声などの音響信号を入力する

入力した声の振幅とピッチの変化を基に合成音を作成する。ピッチ算出手法はケプストラム法を用いる。

## (2) 自然な音声の高速かつ高品質な合成

「乾杯」という掛け声を例に挙げると、現状では「か・ん・ぱ・い」というテキストを朗読調で合成し、感情や自然性の感じられないものが合成されてしまう。

これは、音声コーパスに朗読データを前提としたものが一般的だからである。

本研究では、ユーザが入力した音声の韻律を合成音に転写することで感情や自然性のある合成音を得られる。

## (3) 新しい音表現

クラリネットに歌声のビブラートを付加する、または喋るバイオリンといった混成音(いくつもの音の特徴を組み合わせる一つの音とする合成音; sound hybrid)を合成し、アミューズメントから先端芸術までのさまざまな分野に用いることができる。

## 3 システムの全体像

図1のように、マイクから入力された音響信号の振幅とピッチの変化を分析し、それを基に音合成エンジン STK(Synthesis Tool Kit)[1][2]を用いて合成音を作成し出力する。

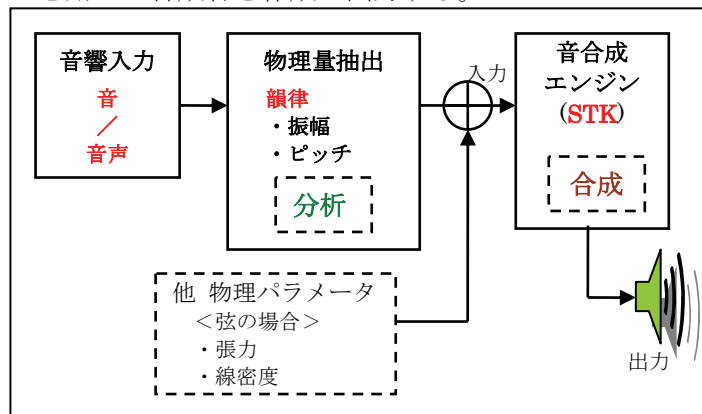


図1 システムの概略図

## 4 音合成エンジン (STK)

音合成エンジンとは、与えられたデータを基にシンセサイザのように様々な音を作り出すものであり、本研究で用いる STK はこれらの機能を提供するツールキットである。

STK は Cook らによって制作され、様々な楽器の物理モデルが C++ 言語で記述されている。物理モデルによる音合成とは、自然楽器の音響振動の発生機構を数学的・物理的にモデル化することにより音を合成する方法である。

「Sound synthesis from concept by direct input of audible signal」

<sup>†</sup> 「Seiichi Yoshida・Tokyo Denki University」

<sup>‡</sup> 「Naotoshi Osaka・Tokyo Denki University」

## 5 合成音の事例

ここで歌声を入力に用いた場合の例を示す。ソプラノ歌手の歌声(ビブラート付き)の振幅とピッチの変化を基に音を合成する。

### 5.1 入力に用いた歌声

入力に用いる歌声の波形を図2に示す。これを分析し、得られた振幅とピッチの時間変化がそれぞれ図3、図4である。分析には短時間フーリエ変換を用いた。

### 5.2 混成音合成の例

5.1節で分析したソプラノの時変の振幅とピッチ情報を付与した擦弦モデル(バイオリンなどの弦を擦って演奏する楽器を対象としたモデル)の例を図5に示す。比較のため、同じ擦弦モデルで振幅とピッチ(100[Hz])を一定にした場合の例を図6に、図5の波形のピッチの時間変化を図7に示す。図5と図6を比較すると振幅の変化が見られ、図4と図7の比較によりソプラノのピッチ変化が擦弦モデルに付与できていることがわかる。このように、入力に用いた音響信号により異なる表現を持った合成音を得られる。

## 6 終わりに

本稿では、目的の音に対する概念(イメージやアイデア)を音声などの音響信号で表現し、マイクから直接入力して音合成を行う手法の提案を行った。このようなアナログ情報を用いて直接音合成を行うことで、ユーザは音に対する概念を容易かつ効果的に機械に伝達可能となる。

今後は、ピッチ分析の際に、実際のピッチの半分の周波数である半ピッチが検出される問題の検討を行う。

### 参考文献

- [1] Cook, P. R. (1996), "Synthesis ToolKit in C++ Version1.0," In SIGGRAPH 1996, Course #17 & 18, Creating and Manipulating Sound to Enhance Computer Graphics, Available from ACM SIGGRAPH.
- [2] Cook, P. R. and Scavone, G. P. (1999), "The Synthesis ToolKit (STK)," In Proc. 1999 Int. Computer Music Conf, pp.164-166, Beijing, China. Computer Music Association.
- [3] Julius O. Smith III, "Physical Modeling Using Digital Waveguides," Computer Music Journal vol.16, No.4, pp.74-91, the MIT Press, 1992.
- [4] Rodet, p. Depalle, Synthesis by Rule: LPC Diphones and Calculation of Formant Trajectories, IEEE-ICASSP, Tampa, FL., March 85..

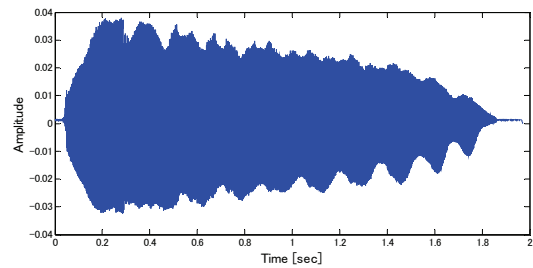


図2 歌声の波形(ソプラノ)

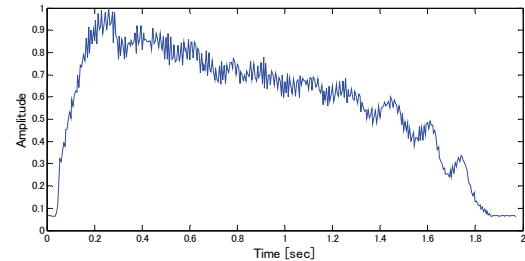


図3 振幅の時間変化

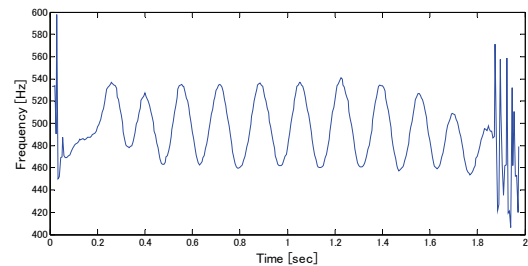


図4 ピッチの時間変化

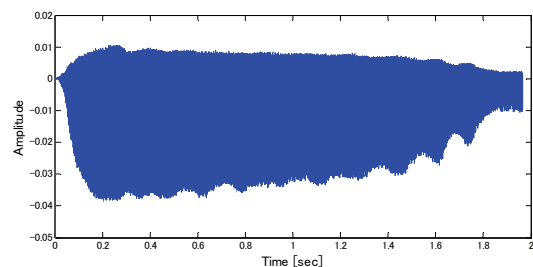


図5 擦弦モデル+ソプラノの韻律(振幅,ピッチ)

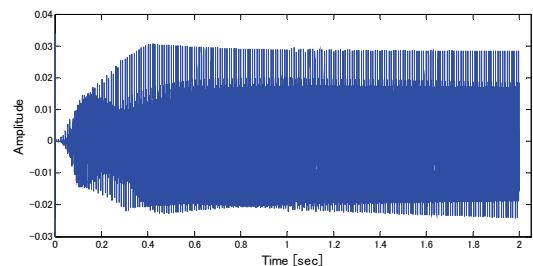


図6 擦弦モデル(振幅・ピッチ(100[Hz])一定)

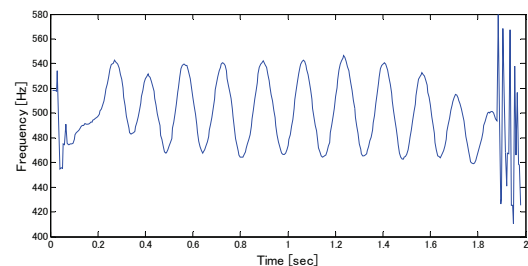


図7 擦弦モデル+ソプラノのピッチ変化