

7M-7 NeXT Computer を用いた音声合成実験環境

友田 一郎 山下 泰樹 Kris Sheahan 今井 徹 竹林 洋一 斎藤 光男
(株)東芝 総合研究所 情報システム研究所

1.はじめに

近年デジタル信号処理機能を持ったワークステーションが開発されるようになり、音声や音響信号処理を行うための環境が整いつつある。その中でも特に NeXT Computer [1] はデジタル信号処理プロセッサ (DSP) を搭載し、また、NextStep と呼ばれるウインドウシステムを基礎とした優れたユーザーインタフェースと、オブジェクト指向に基づく豊富なソフトウェア開発環境を備え、音声信号処理の研究開発に対して優れた環境を提供している。

一方我々は以前から Klatt のフォルマント合成器 [1] を用いて音声合成の研究を行なっているが、その従来の環境では、今後求められるマルチメディアや知識処理を取り入れた、優れた音声合成の研究開発に十分であるとは言えなかった。そこで我々は NeXT Computer の持つ環境を利用して、音声合成の研究開発が容易に行え、音声信号処理についてのより深い理解が得られるような環境の構築に取り組んでいる。その1つとして今回、Klatt の合成器を NeXT Computer 上に移植し、パラメータの入

力をグラフィカルに行なえるようなウインドウベースのユーザーインタフェースを開発した。

2.システムの構成

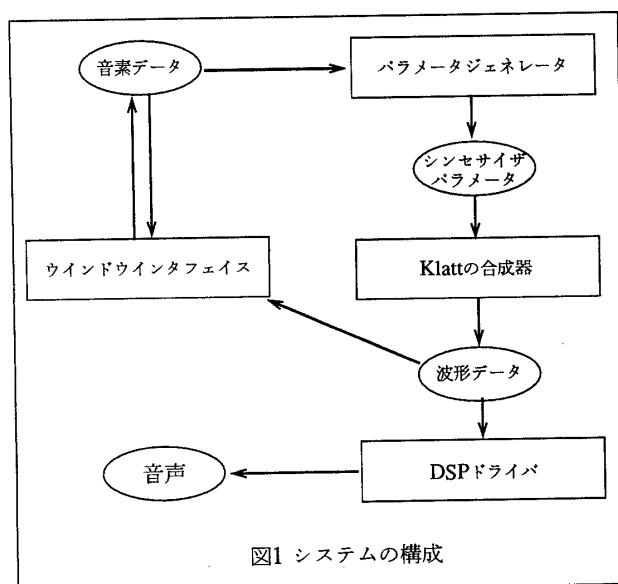
本システムのおおまかな構成を図1に示す。

Klatt のフォルマント合成器は原理的には有声音源(インパルスジェネレータ)、無声音源(ノイズジェネレータ)およびフォルマント(共振特性)を表現するデジタルフィルタにより構成されている。これらを制御するために合成器は40個のパラメータを持ち、それらを5msの周期で更新しながら音声合成する。本システムでは音声の韻律的特徴をより簡単に表現するために、ユーザーは音素を単位として入力を行う形になっている。すなわち合成音はいくつかの音素の時系列であると考え、ユーザーはそれぞれの音素ごとに 1.音素の種類を表すコード 2.基本周波数 3.強さ 4.時刻 の4つのパラメータを入力する。これらの音素パラメータはパラメータジェネレータ部により合成器パラメータに変換される。

ウインドウインタフェース部はユーザーへのインタフェースを提供するとともにシステムの全体的な制御を行う。DSPドライバ部はサンプリング周波数の変換(シンセサイザの出力データ:22.05kHz → NeXT Computer 内蔵D/Aコンバータ:44.1kHz)を行いD/Aコンバータへ出力する。合成音の最終出力は NeXT Computer のヘッドホン端子、ライン出力端子から出力される他、内蔵のスピーカから聞くこともできる。

3.使い方

本システムは図2に示すユーザーインタフェースを持つ。使い方としてはまず始めに、左上部の Input String と書かれたフィールドに音素文字列を入力すると、その下の PitchView と AmplitudeView の2つのサブビューに各音素のデフォルト値が表示される。PitchView は縦方向に周波数、横方



A Speech Synthesis Tool on the NeXT Computer
Ichiro TOMODA, Yasuki YAMASHITA, Kris SHEAHAN,
Toru IMAI, Youichi TAKEBAYASHI, Mitsuo SAITO
Research and Development Center, Toshiba Corp.

NeXT は NeXT Inc. の商標です。
NextStep は NeXT Inc. の登録商標です

向に時刻をとって基本周波数の変化を、AmplitudeViewは縦方向に強さをとってその変化をグラフィカルに示すものである。ユーザーはこのウインドウ上でマウスを使って各音素の位置を動かすことができ、各音素の基本周波数、強さ、時刻をグラフィカルに編集することができる。また現在選択されている音素の各パラメータの値は右中部のSelected Phonem と書かれたフィールドに表示され、ここから数値で入力することもできる。

編集終了後は右上部の Compute ボタンをマウスでクリックすることにより、パラメータジェネレータとフォルマントシンセサイザが起動されて合成音声計算され、その結果の波形が SoundView に表示される。合成音声は Play ボタンをマウスでクリックすることにより内蔵スピーカおよびライン出力、ヘッドホンに出力される。出力中の音声は、Pause ボタンで一時停止、また Stop ボタンで停止することができる。また発声中は右上部の SoundMeter にピークレベルが表示される。

4. 結果

本システムでは特にグラフィカルなユーザーインタフェイスにより音声を視覚的にとらえることがで

きるようになり、実際の研究 ([3]等) において音声合成の実験に役立っている。

一方、実際にこのシステムを使ってみると合成器の計算が非常に遅く感じられる(合成音1秒につき計算に約20秒かかる)点が問題となっていたが、これに対しては、現在CPU上で実行されている合成器のプログラムをDSP上で実行させることにより、より速い計算ができるものと期待されている。

今後の発展としては、このシステムによって得られる実験結果を蓄積・管理するシステムを開発し、音声合成のための知識ベース構築を支援する環境の実現について考えている。

参考文献

- [1] Bruce F. Webster "The NeXT Book" Addison-Wesley, 1989
- [2] Dennis H. Klatt "Software for a cascade / parallel formant synthesizer" Journal of the Acoustical Society of America, 67(3), 1980
- [3] Kris Sheahan, 山下 泰樹, 竹林 洋一 "Synthesis of Nonverbal Expressions for Human-Computer Interaction" 日本音響学会講演論文集 2-4-6, 平成 2年

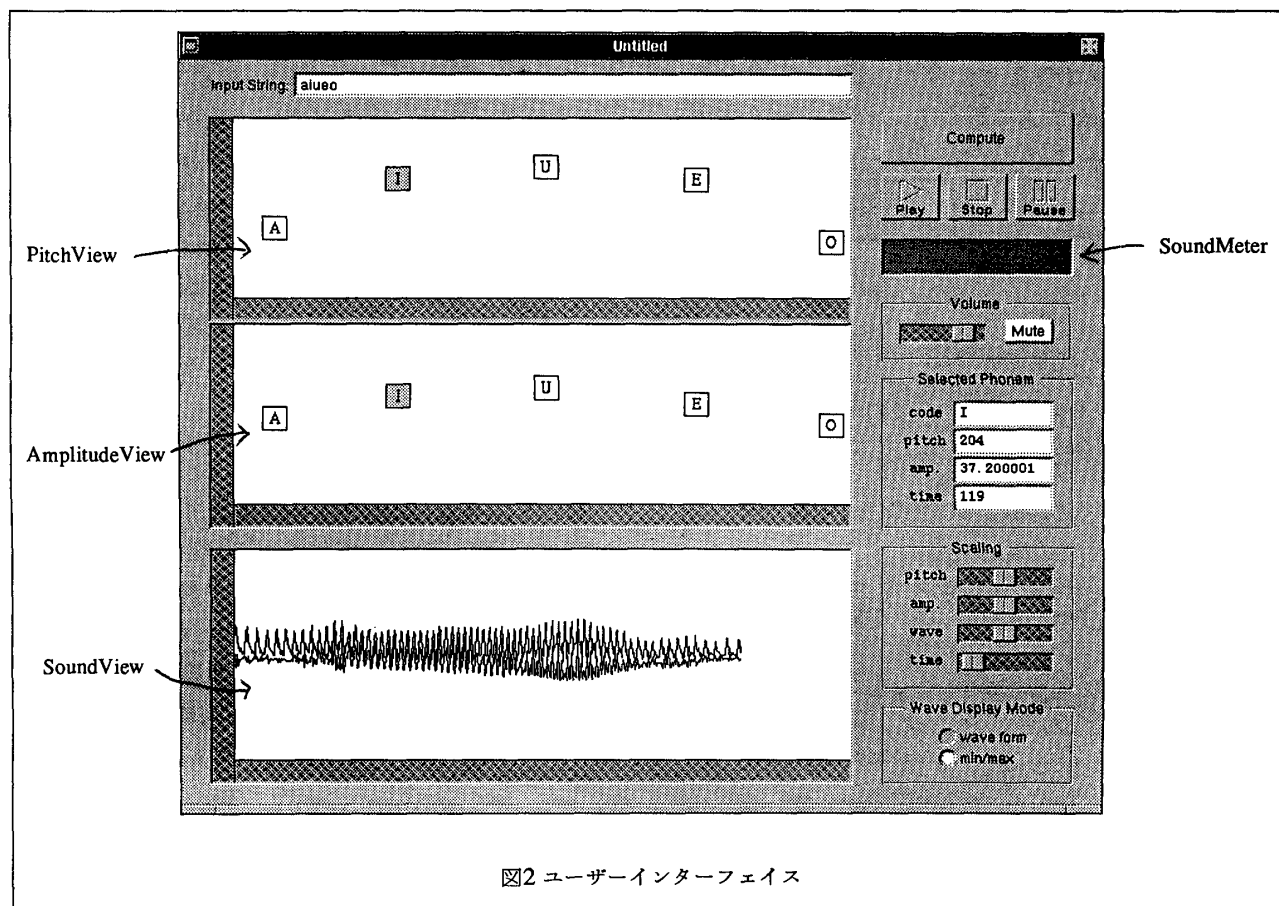


図2 ユーザーインタフェイス