

遠隔教育に向けた音声合成システムの開発

増矢拓郎 緒方公一 園田頼信 (熊本大学)

1. はじめに

現在、教育の多様化、インターネットの普及を背景に、遠隔教育に向けた様々な取り組みが行われている。その中でも Web 技術を応用したシステムの開発は、普段使い慣れているブラウザの操作性、汎用性の観点からユーザフレンドリーなインターフェイスを実現するものとして期待されている。

これまで、本研究室では、Sondhi と Schroeter[1]により提案された声道シミュレータを基に、インタラクティブに声道形状を与え、音声合成を行うスタンドアローン型[2]、Web アプリケーション型[3]のシステムの開発を進めてきた。この音声合成システムの GUI 作成言語には、様々なプラットフォームで動作し、他言語とのリンクが容易な利点から Tcl/Tk を採用している。

しかしながら、Tcl/Tk は一部の Web ブラウザのみに対応し、音声合成システムを使用するユーザが制限されてしまうという問題があった。そこで、本稿ではマルチプラットフォーム言語であり、多くの Web ブラウザで動作する Java を利用して Web 利用に向けた音声合成システムの開発を行った。本稿ではシステムの構成と例用例を中心に紹介する。

2. 音声合成モデル

本システムの音声合成では Sondhi と Schroeter により提案された声道シミュレータを基に調音に基づく音声合成機能を実現している。このシミュレータは、Fig.1 に示すように、1) 声帯、2) 声道、3) 口唇放射の 3 つのモデルから構成されている。本システムでは、声道形状を、20 個の円筒形

の音響管の縦続接続で近似的に表現している。

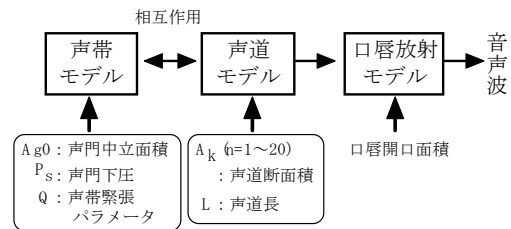


Fig.1 Articulatory speech synthesizer

3. Web 対応システムの構成

スタンドアローン型の音声合成システムでは、これまで、GUI の作成言語として様々なプラットフォームで動作し、他言語とのリンクも容易であることから Tcl/Tk を採用してきた。このシステムでは、FORTRAN や C 言語を用いて記述された音声合成に関する数値計算プログラムを Tcl/Tk のリンク機能を利用することで統合し、GUI を構築している。

Web ベースシステムの開発にあたっては、これまで作成してきた Tcl/Tk で記述されたシステムを活用することで開発コストの面でメリットがある。そこで、開発言語に Tcl/Tk を採用し、Web 対応の音声合成システムを構築した。

しかし、Tcl/Tk プログラムの実行には Web ブラウザプラグインが必要であり、また、一部のブラウザでは Tcl/Tk のプログラムを実行できないという問題がある。そこで、多くの Web ブラウザで動作し、マルチプラットフォーム言語である Java を採用することでネットワーク対応システムの開発を行った。

* Development of a speech synthesis system for distance education

Java はネットワーク処理を念頭に開発された言語であり、現在多くのネットワークシステムの開発言語として採用されている。

Java を使用することにより、プラットフォームからの独立、セキュリティといった問題を解決することができる[4]。

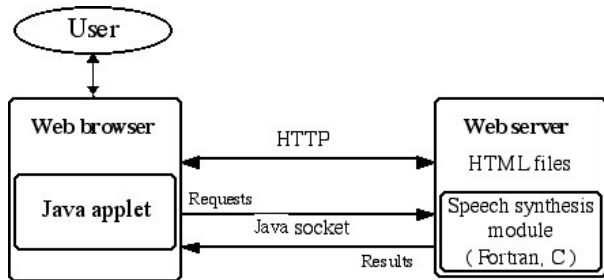


Fig.2 System configuration of the Web-based speech synthesis system

Fig.2 に、開発した Web ベースの音声合成システムの構成図を示す。システムの GUI を記述する Java applet は Web サーバ上に置かれている。ユーザは Web サーバにアクセスし、Java applet をダウンロードすることでシステムの利用が可能となる。また、データ通信には Java のソケット機能を利用し、クライアントとサーバ間のデータの受け渡しを実現している。

本システムではクライアントとサーバ間に 2 本のソケットを作成し、クライアント側から Web サーバ側へ向けた数値計算要求とその計算パラメータの送信、また、その逆方向への数値計算結果送信をそれぞれ実現している。

4. 音声合成システムの利用例

本システムでは、利用方法についての解説のページを設置している。利用者はシステムの利用の前に解説ページを参照することで音声合成システムの使用法を知ることができる。Fig.3 に解説画面の一例を示す。なお、Web ブラウザとしては Internet Explorer6.02 を用いて表示している。このペ

ージでは、単母音の合成の手順について示されており、使用環境の設定に関する注意とともにブラウザでの表示イメージが示されている。

Fig.4 はシステムの利用時の音響管表示ウィンドウを示している。声道形状は、本来は曲がりを持つ音響管であるが、本システムでは簡単のため直円筒型の形状としている。声道は、20 個の円筒形の音響管の縦続接続で近似表現されており、各音響管の断面積はファイル入力による設定のほかに、キーボード入力等による形状変更などが可能となっている。

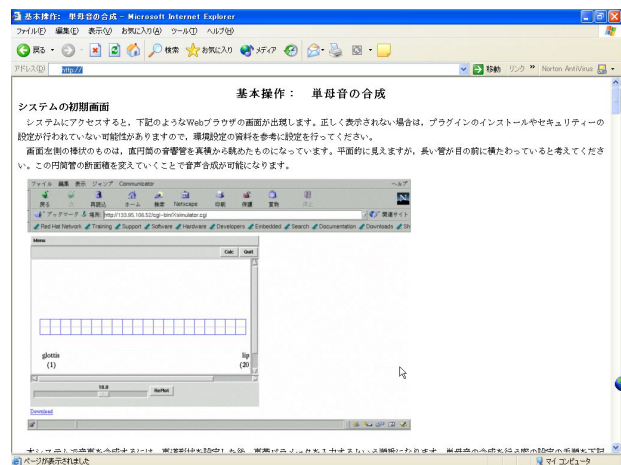


Fig.3 Example of manual pages.

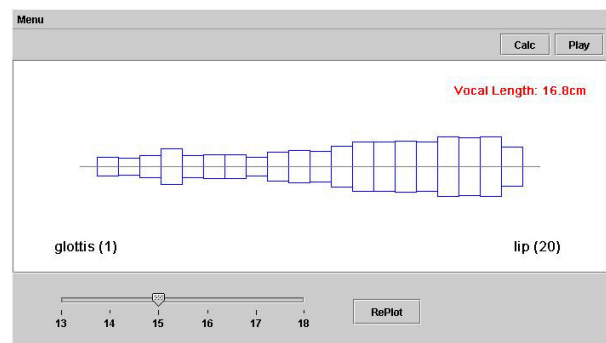


Fig.4 Vocal tract display window.

また、ウィンドウ下部のスライダーで声道長を設定することが可能であり、マウスを利用してその長さを調整することで、長短の声道長を模擬することができる。図形の描画は Java の paint メソッドを利用してお

り、Java appletがWebブラウザ上で描画を行っている。また、Java appletではこれまで開発してきたTcl/Tkを用いたスタンドアロン型、ネットワーク対応型のシステムとほぼ同等のイベント処理が実現されている。

これまでTcl/Tkを利用したネットワーク対応のシステムでは、(1)単母音の合成、(2)任意形状設定による合成、(3)連続母音合成が実現されている。Java appletを用いたシステムでは現在、(1)のみが実現されており、以下本稿では、(1)の機能の利用例について説明する。

5. 単母音の合成機能

ここでは本システムを利用した単母音の合成例について示す。

5.1 声道形状の設定

Fig.5は単母音合成時のウィンドウ表示例である。ユーザは、Menu → Synthesis → Vocal tract → Image と選択することで、母音の声道断面積、及び声道長のデータ設定を行うことができる。ここでの声道形状のデータはMRIデータを画像処理することで得られた断面積、及び声道長を用いている。

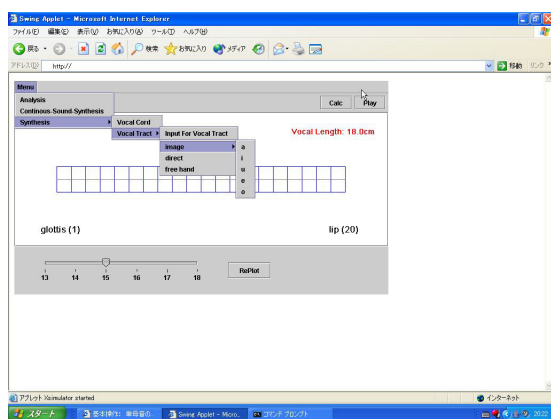


Fig.5 Selecting menus to display a vocal tract shape.

断面積の設定後は、音響管表示ウィンドウ上部の“Calc”ボタンを押すことにより、Fig.6に示すようなグラフが表示され、声道

伝達特性を確認することができる。また、グラフには第4までのホルマント周波数の値も表示されるため、様々な声道形状に対応するホルマント周波数の変化を確認するツールとしても利用可能である。

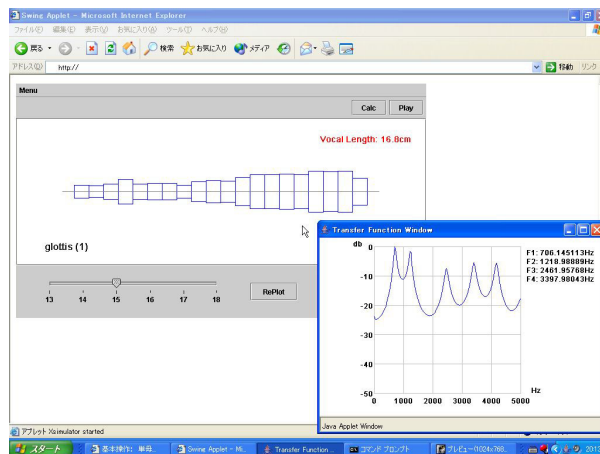


Fig.6 Vocal tract and its transfer function.

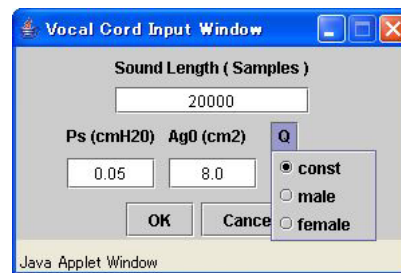


Fig.7 Determining vocal cord parameters.

5.2 声帯モデルの制御パラメータの設定

前述したように、声帯モデルの制御パラメータは、(1)声の高さに関する声門下圧 P_s 、(2)有声無声の区別に関する声門中立面積 Ag_0 、(3)声の高さに関する声帯緊張パラメータ Q の3つである。Fig.7に声帯モデルの制御パラメータの入力画面を示す。画面では、これら3つのパラメータと合成音の時間長を設定する様子が示されている。システムの起動直後は、これらの標準値が初期値として与えられている。これらの3つのパラメータのうち、声の高さに関する声帯緊張パラメータ Q に関しては、その値が一定のもの(const)と、(male, female)とを用意し、選択できるようにしている。

const を選択した場合は、ウィンドウに入力した数値が音声合成に利用され、合成の間一定値として取り扱われる。標準値は 1 に設定されており、この値を大きくすることで高い声を、小さくすることで低い声を実現することができる。male あるいは female を選択した場合にはピッチ変動を与えた合成音の生成が可能となる。

実音声のピッチ周波数には揺らぎが存在し、その時間変動が肉声らしさに深く関わっている。このため、声帯緊張パラメータの値を一定にして合成を行った場合には、無機質な音声となり、肉声とは異なった自然性の損なわれた音声となる。本システムでは、男声、女声の実音声からそれぞれピッチ周波数を抽出し、これを声帯緊張パラメータに反映させたデータファイルを用意することで、揺らぎの効果を確認できるように配慮している。

声帯パラメータ設定後は“OK”ボタンを押すことで音声合成が行われ、Web サーバ上に音声ファイルとして波形データが保存される。その後、Java applet 上の“Play”ボタンをクリックすると、Web ブラウザ上で合成音が再生され、作成した合成音の確認が可能である。

6. まとめ

これまで Tcl/Tk を用いてスタンドアローン型、ネットワーク型の音声合成システムの作成を行ってきた。しかし、ネットワーク型のシステムを利用する際に必要な Tcl/Tk のプラグインは一部のブラウザにしか対応しておらず、システムを利用可能なユーザに限られてしまう問題がある。そのため、本稿では、より多くのブラウザで利用可能な Java を用いた Web ベースの音声合成システムの開発を行った。

システムでは声道形状の描画やイベント処理、さらにソケットによるクライアント・サーバ通信など Java の機能を有効利用

し、ブラウザ上でインタラクティブに声道形状を変化させることで単母音の音声合成を行うことができた。また、Java は数多くのブラウザで動作するため多くのユーザのシステム利用が可能である。以上のことから、本システムはその容易な操作性と汎用性から音響教材としても有効なシステムである。

現在、システムの利用マニュアルを Web ドキュメントとして用意し、ユーザへの便宜を図っているが、利用事例等のコンテンツを充実させ、初心者にもより使いやすい環境の実現を目指す予定である。また、現在の Java を用いた音声合成システムの機能は単母音合成のみとなっているため、今後は連続母音合成、音声分析といった機能を付加し、システムの高機能化を目指す。

参考文献

- [1] M. M. Sondhi and J. Shroeter, "A hybrid time-frequency domain articulatory speech synthesis system," *IEEE Trans. Acoust., Speech & Signal Process.*, vol. ASSP-35, no.7, pp.955-967, July 1987.
- [2] 緒方公一, 加藤央, 中島康博, 園田頼信, "Tcl/Tk を用いた GUI ベース音声分析点合成システムの開発", 熊本大学工学部研究報告, vol.49, no.2, pp.147-155, Dec. 2000.
- [3] 緒方公一, 園田頼信, 藤井心平, "音声生成過程に基づく音声合成システム-Web ベースシステムの開発-", 信学技報, SP2003-74, pp.1-6 Aug.2003.
- [4] Elliotte Rusty Harold 著, 戸松豊和 監訳, 田和勝 訳: Java ネットワークプログラミング 第2版, オライリー・ジャパン, 2003.