発話ロボットの弾性体柔軟舌形状のモデル化 および音響モデルとの統合*

西川 員史^{*1,*6}, 小河原 隆行^{*2}, 高信 英明^{*3}
持田 岳美^{*4}, 菅田 雅彰^{*5}, 高西 淳夫^{*1}

Modeling and Analysis of Elastic Tongue Mechanism of Talking Robot for Acoustic Simulation

Kazufumi NISHIKAWA^{*7}, Takayuki OGAWARA, Hideaki TAKANOBU,
Takemi MOCHIDA, Masaaki HONDA and Atsuo TAKANISHI

^{*7} Department of Mechanical Engineering, Humanoid Robotics Institute, Waseda University,
3-4-1 Ookubo, Shinjuku-ku, Tokyo, 169-8555 Japan

This paper describes the modeling and analysis of an elastic tongue mechanism of a talking robot WT-3 (Waseda Talker-No. 3) and the development of an acoustic simulator for the construction of the computational model of human speech production. WT-3 consists of lungs, vocal cords and articulators. Because they were composed of an elastic rubber to make various shapes and to seal the air and sound, it was difficult to estimate and analyze the mechanisms. Therefore, we constructed a modeling method to simulate the shape of the elastic mechanisms using rigid links. This method is applicable to all the elastic body. We developed an acoustic simulator for WT-3 based on this method.

Key Words: Humanoid Robot, Speech Production, Mechanical Model, Elastic Mechanism, Speech Synthesis

1. 緒言

音声言語の生成に関しては、多くの研究がなされているが、未だ脳における発声の運動計画処理機構から運動器官における音声生成の運動までを包括的に研究された例はなく、また人の発声運動は十分に解明されていないのが現状である。1998年より、科学技術振興事業団(JST) 戦略的基礎研究推進事業(CREST)「脳を創る」研究プロジェクト「発声力学に基づくタスクプランニング機構の構築」として、日本電信電話株式会社(NTT)を中心に全国10の医学、音響学および工学の研究機関で人間の発声に関する共同研究が開始された。その中でわれわれは、人間の発声運動を再現する人間形発話ロボットの開発を担当している。

本研究は、発声器官(肺、声帯)および調音器官(舌、

唇、歯、鼻腔)を有し、人間の発声動作を模擬した発話ロボットを開発し、これを用いて人間と同様の発声を実現することにより、計算機シミュレーションのみでは解明困難な発声系のメカニズムをロボット工学的な視点から明らかにすることを目的としている。

われわれは2000年に人間形発話ロボットWT-1 (Waseda Talker-No. 1)⁽¹⁾, 2001年にWT-1R (Waseda

Fig. 1 Talking robot WT-3 to mechanically reproduce human speech

* 原稿受付 2004年4月13日.

^{*1} 正員, 早稲田大学理工学部ヒューマノイド研究所(☎ 169-8555 東京都新宿区大久保 3-4-1).

^{*2} 早稲田大学大学院.

^{*3} 正員, 工学院大学工学部(☎ 163-8677 東京都新宿区西新宿 1-24-2).

^{*4} 日本電信電話(株)コミュニケーション科学基礎研究所(☎ 243-0198 厚木市森の里若宮 3-1).

^{*5} 早稲田大学スポーツ科学部(☎ 359-1192 所沢市三ヶ島 2-579-15).

^{*6} 日本学術振興会特別研究員.

E-mail: nishik@ruri.waseda.jp

Talker-No. 1 Refined)⁽²⁾, そして2002年に人間の発声器官を機能的側面から模擬した発話ロボットWT-2 (Waseda Talker-No. 2)を開発し, 明瞭性の高い母音および破裂性子音・摩擦性子音・鼻子音の生成を実現し, 日本語の五十音すべての発声が可能となった⁽³⁾. しかし, その音声は人間と異なっており, 十分発声可能なシステムを実現しているとは言い難く, より人間に近い音声生成を実現するため, 図1に示すように新型機構を有する発話ロボットWT-3 (Waseda Talker-No. 3)を開発した.

また, 発話ロボットの発話器官(舌・口唇・声帯等)は弾性体であるゴムで構成されており, それらの形状を正確にモデル化し, 解析することは非常に困難であった. そこで, 弾性体で構成された舌形状をモデル化し解析する手法を考案した. この手法は, ロボットの舌形状のみならず, その他さまざまな大規模変形を伴う弾性体に応用可能であると考ええる. その結果, 正確な声道形状が推定可能となり, さらに音響モデルと統合することで, 声道形状からロボットの発声可能な音声までを一貫して推定計算する音響シミュレータを開発した.

本論文では, WT-3 の弾性体で構成された舌形状をモデル解析する手法, そして音響シミュレータについて述べる.

Fig. 2 WT-3's vocal cords mechanism

2. 発話ロボットWT-3の機構

発話ロボットWT-3は, 図1に示すように肺 (1自由度), 声帯 (3自由度), 調音器官である口唇 (5自由度), 歯 (1自由度), 舌 (7自由度), 鼻腔, 軟口蓋 (1自由度) の全18自由度を有し, 声帯から口唇までの声道長さは175[mm]であり成人男性と同程度の大きさを持つ. その発話器官は, 動きを確保しつつ大規模に変形し, かつ空気・音の密閉性を確保しなければならない, そのため弾性体の超低硬度ゴムTP010⁽⁴⁾で構成されている. その中心的な機構として舌機構がある.

WT-3 の声帯は 2 枚の超低硬度ゴム TP010⁽⁴⁾を張り合わせることでスリットを形成し, そこに肺からの呼気を通し振動させ, 音源を生成する. 図2に示すように左右のゴムを独立した機構で均等に巻き取ることで, ピッチ (声帯振動の基本周波数)の変更を可能にした. 安定した声帯振動を実現し, ピッチが 100-260[Hz]と広域になり, より抑揚のある発声を期待できるようになった. さらに人間の声帯と同様に, 基本周波数を保ったまま有声無声音の切り替えを可能にした.

また, WT-3 の舌機構は, 図3に示すようにワイヤによる 2 次元の舌形状変更機構であり, 舌は超低硬度ゴム TP010⁽⁴⁾で形成されている. 舌表面のゴムの厚みは 8 [mm], そして舌の幅は 30 [mm]と一定である. その舌表面のゴムは 5 [mm]と 3 [mm]のゴムの板を重ね合わせて作られており, その間に 7 個の金属部材を舌の幅と同じ長さで挿入・接着している. その金属部材の中心点からループ状にしたワイヤを口蓋と下顎の間に通し, ループ状にしたワイヤでその中心点を駆動させることで, 舌の形状を変化させた. この機構は舌の幅が一定の 2 次元の駆動機構であり, 金属薄板の中心点である 7 個の駆動点で舌形状を形成している. この機構により人間に近い多様な構えの舌形状が生成可能となった. しかし, その各駆動点間はゴムで形成されて

Fig. 3 WT-3's flexible tongue mechanism using looped tendons

いたため、舌表面の形状をモデル解析することは非常に困難であった。そこで、第3章に述べる柔軟舌形状のモデル化手法を考案した。

3. 柔軟舌形状のモデル化

図4に示すように剛体リンクを用い、弾性体で形成される舌の形状をモデル化し、推定する手法を考案した。この手法においては定常的な母音発声を推定することを目指し、舌が静止しているとしてモデル化された。また、舌駆動機構は2次元機構であることから、舌表面の3次元的な変形はないものと仮定しモデル化された。そのモデル化手法を下記に述べる。

駆動点間のゴムの形状を複数の剛体リンクの連結であるとモデル化し、その形状を推定した。各リンク間に仮想的なばね（ばね定数 k ）を、また連結されたリンクの始点・終点には、注目している駆動点の前後の駆動点の位置の影響を考慮するため、仮想的なリンクを考える。この時各リンクの終端の点は、

第1リンク:

$$\begin{cases} x_1 = l_1 \cos q_1 \\ y_1 = l_1 \sin q_1 \end{cases} \quad (1)$$

第 n リンク:

$$\begin{cases} x_n = x_{n-1} + l_n \cos(\sum_{i=1}^n q_i) \\ y_n = y_{n-1} + l_n \sin(\sum_{i=1}^n q_i) \end{cases} \quad (2)$$

ただし、 $q_n = \frac{\alpha}{180} - (\sum_{i=1}^{n-1} q_i)$ 、 α は任意に与えられる角度[deg]とする。

Fig. 4 Modeling of elastic material using rigid links

さらに、力 F により駆動点に (δ_x, δ_y) の仮想変位が生じたとき、

$$\begin{cases} \delta_x = x_n - P_x \\ \delta_y = y_n - P_y \end{cases} \quad (3)$$

とおく。ただし、 (P_x, P_y) は目標点である。

ポテンシャルエネルギーは、仮想仕事の原理より

$$U = \frac{1}{2} k_1 \left(\frac{\pi}{2} - q_1 \right)^2 + \sum_{i=2}^{n+1} \frac{1}{2} k_i q_i^2 + F_x \delta_x + F_y \delta_y \quad (4)$$

となる。静止していると考えれば、運動エネルギー

Fig. 5 Acoustical simulator using rigid links

$T=0$ となり、ラグランジアンは $L=T-U$ となり、ラグランジュの運動方程式は以下ようになる。

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = - \frac{\partial U}{\partial q_i} = \frac{\partial U}{\partial q_i} = 0 \quad (i=1, 2, \dots, n) \quad (5)$$

仮想変位=0より、

$$\begin{cases} \delta_x = x_n - P_x = 0 \\ \delta_y = y_n - P_y = 0 \end{cases} \quad (6)$$

式(5), (6)より、舌の駆動点の位置である目標点 (P_x, P_y) を与えれば、ニュートン・ラフソン法⁽⁵⁾による

近似解からリンク角度 $(q_1, q_2, \dots, q_n)$ を導出でき、この

剛体リンクを連結することにより柔軟な舌形状を推定できる。さらに、このシミュレータにおいては、微小時間 Δt 前のリンク角度を初期値として与えており、動的に変化する舌の動きに対しても形状推定が可能である。この結果を第5章に述べる。また、第4章に述べるように正確な声道形状を推定することが可能となり、その声道形状からロボットの発声可能な音声を推定可能となった。

4. 音響モデルとの統合

機構部の改良指針を与え、さらに母音・子音発声の効率的なパラメータ探索を行うため、図5に示すような音響シミュレータを開発した。

このシミュレータは図6に示す音響理論の「音源-フィルタ」モデル⁽⁶⁾と音響管モデル⁽⁷⁾を用いており、その「音源-フィルタ」モデルは、母音発声時における声帯音源と声門から口唇までの声道部分を分離して考えるモデルである。そして、音響管モデルは声道が音響管をつなぎ合わせたもので表せるという仮定に基づき、 Δ の小さい短い管を多数使って表したモデルである。このモデルでは連続した管の断面積がそれぞれ異なり、音響管内を音波が伝搬した際の体積速度と音圧から声道の共振周波数であるフォルマント（低い順に第1フォルマントF1・第2フォルマントF2・第3フォルマントF2…と呼び、母音の識別においてはF1とF2が重要である⁽⁸⁾）が導き出される。そして、声道伝達関数はフォルマントを極としたものであり、声帯の音源スペクトルが掛け合わされることで音声スペクトルとなる（図6(a)下図）。そして、人間の声帯音源スペクトルはピッチの倍音構造であり、約12[dB/oct]でほぼ滑らかに減衰するため、その音声スペクトルは右下がりとなります。

以上、このように短い音響管が多数連結されたものをWT-3の声道のモデルとして仮定することで、そこから発声される音声を推定することが可能であり、本研究のシミュレータとしても十分に適応できると考えられる。

よって、このシミュレータでは調音器官の舌（7自由度）・歯（1自由度）・口唇（5自由度）の各自由度への位置制御指令である制御パラメータを入力することにより声道形状を推定し、上記の音響理論を利用し、その声道形状からWT-3が生成可能な音声を推測できる。代表的なものとして以下の機能を持つ。

- (1) 第3章に述べたモデル化手法を用い、柔軟材で構成されたWT-3の舌形状を複数の剛体リンクの連結として仮想し、制御パラメータを入力することによってその舌形状を推定できる（図5(a)）。
- (2) (1)により求められた舌形状から声道形状を推定でき、声道中心線、およびそれに垂直な声道断面積（各音響管の断面積）が導出される（図5(b)）。
- (3) (2)によって得られたWT-3の各音響管の断面積から母音の音響的特徴であるフォルマントを推定できる（図5(c)）。

これらの結果を第5章に述べる。

Fig. 6 Acoustic model

5. 実験

5.1 柔軟舌形状のモデル化

第3章に述べた手法により、弾性体の形状推定を行った。まず、Maplesoft によって開発されたシミュレーションソフトウェア Maple7 を使い、図7に示すように弾性体の推定を行った。図7(a)は、リンク数が3・4・8本のときに駆動点の位置 (x, y) を $(0, 27)$ の位置から間隔 1[mm] で $-y$ 軸方向に移動させたときの結果である。これよりリンクの本数が多いほど形状は丸みを帯び、実際の形状に近くなることを確認した。

そして、リンクの前後には注目している駆動点の前後の形状による影響を考慮するため、仮想的なリンクを考えており、その方向は形状を決定する上で重要な要素であると考えられる。つまり、2個の駆動点間のみに注目して、形状を推定するのではなく、その前後の駆動点の位置を考慮することで、形状の推定を正確に行うことが可能になると考える。上記について、図7(b)は、仮想リンクの角度 α を可変し、 $(5, 27)$ の位置から -1 ずつ変化させたときの結果を示している。この図より、仮想リンクの角度が 60° 、 90° 、 120° と変化するにつれて、他の条件がすべて同じであっても、リンクの形状が異なることが確認できる。これより2個の駆動点のみに着目するのではなく、その前後の駆動点の位置を考慮に入れ、仮想リンクを想定することは十分に意味があることを確認した。

また、図7(c)は駆動点の位置 (x, y) を x 軸方向、 y 軸方向ともに移動させたときの結果を示している。これらの図のシミュレーション結果より連続性を確認し、実際の形状に近くなることを確認した。

次に、われわれが開発した音響シミュレータの形状シミュレータを用いて、WT-3の舌形状の推定を行った。代表として母音の結果を図8に示す。左図がそのシミュレータによって得られた声道形状を表しており、特に $/o/$ の舌体に見られるような大きな舌形状の変形を十分推定可能であることが確認できる。

5.2 音声のフォルマント推定

5.1 節の形状シミュレータから声道断面積関数・声道伝達関数を導出し、音声フォルマントの推定を行った。その結果を図9に示す。図ではシミュレータが声道伝達関数を、ロボットが音声スペクトルを示している。これより $F1 \cdot F2$ ともに WT-3 に近い発声を推定できていることが確認できる。

Fig. 7 Simulation of elastic model (shown in Fig. 5) using simulation software Maple 7

しかしながら、人間の声帯音源は 12 [dB/oct] でほぼ滑らかに減衰する⁽⁸⁾のに対し、WT-3 の声帯音源は図9(f)に示すように 500 [Hz] 付近にスペクトルのピークを持つなど人間と異なり、調音器官の共鳴による $F1$ のエネルギーと比較し音源そのものが持つ 500 [Hz] 付近の

エネルギーが大きいので、WT-3 の発声音の F1 は 400-600 [Hz]の間に集中していた。本シミュレータでは音源の影響を考慮しておらず、今後それを考慮したシミュレータの開発を目指す。

6. 結論と展望

発話ロボットの発話器官は弾性体であるゴムで構成されており、その柔軟舌形状を剛体リンクとばねの連結としてモデル化し、推定する手法を考案した。この手法は、ロボットの舌形状のみならず、その他さまざまな大規模変形を伴う弾性体に適用可能であると考えられる。その結果、正確な声道形状が推定可能となり、さらに音響モデルと統合することで、声道形状からロボットの発声可能な音声までを一貫して推定計算する音響シミュレータを開発した。この音響シミュレータは、WT-3 の母音発声を推定可能であり、機構部の改良および制御パラメータの効率的な探索に利用可能である。

今後は WT-3 の声帯音源の改良を行い、人間に近い音源スペクトルの生成を目指す。さらに声帯音源のスペクトルを考慮したシミュレータを開発し、ロボットに近い音声推定を目指す。

謝辞

本研究は、科学技術振興機構(JST) 戦略的創造研究推進事業(CREST)の援助を受け研究された。研究に協力して頂いた共同研究プロジェクトの研究者各位、また機構部製作に協力して頂いたオキノ工業株式会社の沖野晃久氏、3D-CAD ソフトウェアを提供して頂いたソリッドワークス・ジャパン株式会社、テフロン被覆ワイヤを提供して頂いた中興化成株式会社に感謝致します。

参考文献

- 1) 西川員史, 浅間弘一郎, 林宏樹, 高信英明, 高西淳夫: “人間形発話ロボットによる母音発声の実現”, 日本機械学会論文集, 第 68 巻, 第 667 号, 第 C 編, pp.188-193, 2002
- 2) 西川員史, 小河原隆行, 今井彰浩, 高信英明, 持田岳美, 高西淳夫, “人間形発話ロボットによる子音および連続発声の実現”, 日本機械学会論文集 (C 編) 68 巻 675 号, pp.3361-3366, 2002
- 3) 西川員史, 林宏樹, 高信英明, 持田岳美, 菅田雅彰, 高西淳夫, “人間形発話ロボットによる母音および子音発声の実現—第 3 報, より人間に近い発声を目指した新型発話ロボットの開発”, 日本機械学会論文集 (C 編) 69 巻 683 号, pp.139-144, 2003
- 4) 東京ゴム株式会社製, ショア硬さ: 1 (JIS-A), 引張強さ: 5.9 [MPa], 材質: EPDM (Ethylene Propylene Diene Monomer)

Fig. 8 Simulation of elastic tongue shape using the Newton-Raphson method

- 5) W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery: Numerical Recipes in C, Cambridge University Press, 1992
- 6) B. H. Story, “母音の音源に関する生理と物理及びモデル化”, 日本音響学会誌, pp.426-437, 2002

Fig. 9 Formants of the vowel /i/ using the acoustic simulator

7) 勝木保次, 村田計一, 吉田登美男, 亀田和夫, 越川常次, 三浦種敏, 山口善司, 中田和男, “新版聴覚と音声”, pp.478-570, 電子情報通信学会, 1994

8) Ray D. Kent, Charles Read, “The Acoustic Analysis of Speech”, Singular Publishing, 1992