

## 子音発声を目的とした人間形発話ロボットの開発

西川 員史, 小河原 隆行, 今井 彰浩 (早稲田大学),  
高信 英明 (工学院大学), 持田 岳美 (NTT), 高西 淳夫 (早稲田大学)

本研究の目的は、発声器官、調音器官を有し、人間の発声動作を模擬した発話ロボットを開発し工学的な見地より発声系のメカニズムを明らかにすることである。本発表では、母音と子音の実現を目指し、WT-1 (Waseda Talker-No.1) を改良した WT-1R (Waseda Talker-No.1 Refined) について述べる。WT-1R は全 15 自由度を有する人間形発話ロボットであり、調音器官 口唇 (4 自由度)、歯 (1 自由度)、舌 (6 自由度)、鼻腔、軟口蓋 (1 自由度)、発声器官 声帯 (2 自由度)、肺 (1 自由度) からなる。WT-1R は日本語の母音 (/a/, /i/, /u/, /e/, /o/)、子音 (/s/, /h/, /m/, /p/) を発声することが可能である。

## Development of a Talking Robot for Consonant Sounds

Kazufumi Nishikawa, \*Takayuki Ogawara, Akihiro Imai (Waseda University),  
Hideaki Takanobu (Kogakuin University), Takemi Mochida (NTT), Atsuo Takanishi (Waseda University)

**Abstract** - The purpose of this research is to clarify a human vocal mechanism from engineering viewpoints by simulating the vocal movement with a robot. Therefore, the authors developed an anthropomorphic talking robot WT-1R (Waseda Talker-No.1 Refined) that improved on WT-1 (Waseda Talker-No.1) for the realization of natural vowels and consonant sounds. It simulates a human vocal movement, and has vocal organs (the 1-DOF lung and 2-DOF vocal cords) and articulators (the 6-DOF tongue, 4-DOF lips, 1-DOF teeth, a nasal cavity and 1-DOF soft palate). Total DOF of the robot is 15. WT-1R could utter Japanese vowels (/a/, /i/, /u/, /e/, /o/) and consonant sounds (/s/, /h/, /m/, /p/).

### 1. はじめに

音声言語の生成つまり発声に関しては、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」 (<http://www.brl.ntt.co.jp/cs/speech/crest-j.html>) として、日本電信電話株式会社 (NTT) を中心に全国 15 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。

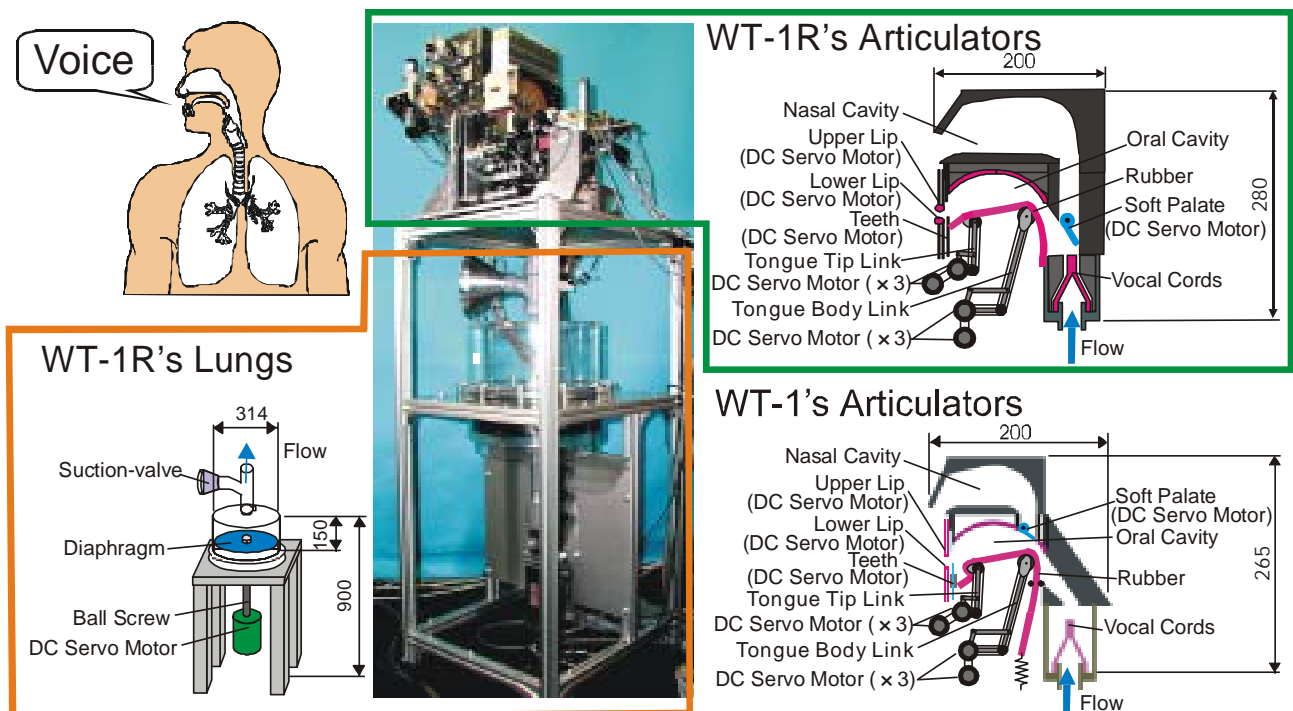


Fig. 1 Mechanical overview of talking robot WT-1R (that improved WT-1) and its articulators

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工学的な視点から明らかにすることを目的としている．また将来，音声から口の動きを逆推定する技術が開発できれば，発声の訓練や外国語の習得の有効な手段を提案できるものと考えられる．

本研究のように，機械的に音声合成を試みた研究は古くからおこなわれており，これまでに1771年のKempelen, W.V. [1]をはじめ，梅田 [2]，川村 [3]，Martin Riches [4]，大須賀 [5]によって機械式発声装置の研究がなされている．しかし，いずれも人間の発声を機械で忠実に再現しているとは言い難い．

また，計算機シミュレーションによる解析では形状モデルや乱流モデルの信頼性など多くの問題があり子音発声の時間変化のシミュレーションは非常に困難である．現在，シミュレーションでは母音と子音の中でも定常的な摩擦音はおこなわれているが，動的な形状変化をシミュレートすることは非常に困難である．

そこで，われわれは，2000年 Fig. 1 に示すような人間と同様の発声器官および調音器官を有する人間形発話ロボット WT-1 (Waseda Talker-No.1) を開発し，母音発声を実現した．しかし，WT-1の母音は不自然で，認識は困難であった．そこで，2001年 WT-1 を改良し W-1R (WasedaTalker-No.1Refined) を開発し，母音に加えて子音発声を実現した．

本稿では，第2章に WT-1 と WT-1R の機構，第3章に WT-1R の制御方法を示し第4章に実験結果，第5章に結論と今後の展望について述べる．

## 2. WT-1 と WT-1R の機構

### 2.1 WT-1の概要

発話ロボット WT-1 は，全14自由度を有する人間形発話ロボットであり，発声器官である声帯（1自由度），肺（1自由度），人間の調音器官である口唇（4自由度），歯（1自由度），舌（6自由度），鼻腔，軟口蓋（1自由度）からなる．高さは約1,600 [mm]，声道長さは約210 [mm]であり，成人男性の約1.2倍である (Fig. 1)．その機構は，肺から空気を送り，声帯を振動させることで音を生成し，さらに舌などの調音器官によって調音され，口唇から音が発せられる．

Table 1 Specifications of Lungs of human and WT-1

	Adult Male	Adult Female	WT-1
Lung Capacity [m <sup>3</sup> ] [6]	3.45×10 <sup>-3</sup>	2.50×10 <sup>-3</sup>	7.74×10 <sup>-3</sup>
Breathing Pressure during Speech [Pa] [7]	600-1000		1100
Breathing Flow during speech [m <sup>3</sup> /s] [8]	1.36×10 <sup>-4</sup>	1.43×10 <sup>-4</sup>	7.74×10 <sup>-4</sup>

### 2.2 WT-1の問題点

WT-1の発声は，不自然なブザー音のような音になってしまい，人間のように明瞭な発声を実現できなかった．また，さまざまな問題から子音を発声することができなかった．そこで，WT-1の問題点を改良しWT-1Rを開発した．2.4節にWT-1とWT-1Rでの共通の機構，2.5節にその改良点を示す．

### 2.3 WT-1Rでの改良点

WT-1Rでの主な改良点は，声帯の駆動機構，口唇部の機構，鼻腔・軟口蓋形状，気密性の向上である．2.5節に改良点を示す．

Table 2 Voice range of human and WT-1

	Adult Male	Adult Female	WT-1
Voice Range at the Maximum [Hz]	69.0-587	131-784	97.0-140
Voice Range during Normal Conversation [Hz]	98.0-123	185-221	

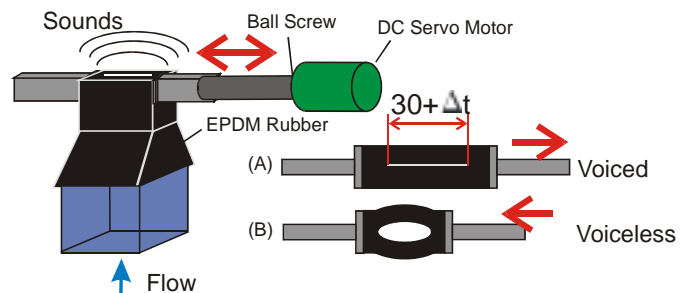
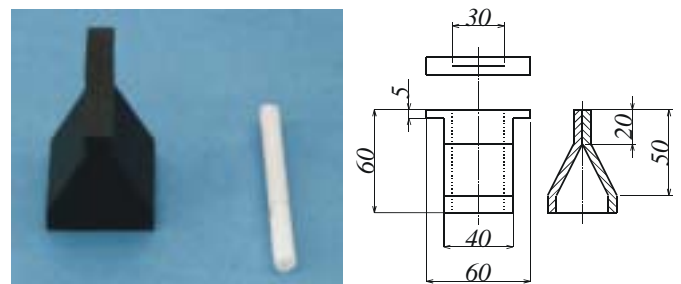


Fig. 2 Vocal cords that produce voiced and voiceless sounds of WT-1

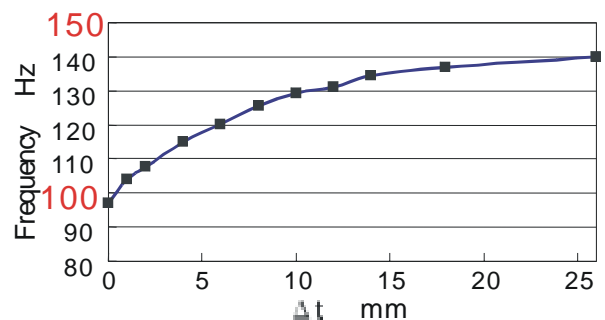


Fig. 3 Pitch of the vocal cords of WT-1

## 2.4 WT-1の機構

WT-1は、人間と同様に発声器官（肺・声帯）と調音器官（舌・唇・歯・鼻腔）を有している。以下に各発声器官・調音器官の具体的構成を述べる。

### 2.4.1 肺

人間の全肺気量は Table 1 に示すように成人男性において平均  $3.45 \times 10^{-3} \text{ [m}^3\text{]}$  [6], 発声時の呼気圧は通常  $6.0 \times 10^2 - 1.0 \times 10^3 \text{ [Pa]}$  [7], そして成人男性の発話時の呼気流量は平均  $1.36 \times 10^{-4} \text{ [m}^3\text{/s]}$  である [8]。Fig.1 に示すように発声器官としての肺は、人間の呼気・吸気系を模擬し、ペロフラムを用いた機構を採用している。WT-1の肺は、Table 1 に示すように人間の仕様を満たしている。

### 2.4.2 声帯

音声の高さは声帯の振動数によって決まり、振動数が増加すれば高い声として、減少すれば低い声として認識される。一般に子供と女性の声は高く、男性の声は低い。成人の声域はおよそ2オクターブであり、Table 2 に示すように成人男性で通常の会話時の周波数は  $98.0 - 123 \text{ [Hz]}$  である [9]。また、音声には、声帯の振動の有無によって有声音および無声音があることがある。

しかし、人間の声帯は非常に複雑であり、機械モデルで再現することは困難である。そこでわれわれは、Fig. 2 に示すように1つの声帯で有聲・無声音が生成可能、かつピッチ変更が可能な声帯を開発した。声門を引っ張り、ゴムの張力を変えることで  $97.0 - 140 \text{ [Hz]}$  のピッチ変更が可能であり(Fig. 2 (A)), 逆に声門部を押し込み開くことで無声音を生成することができる (Fig. 2 (B))。

声帯の素材は、人間の声帯の物性（ヤング率: 約  $10 - 100$  [10]）を持つ素材が最適であり、それに近くかつ強度のある超低硬度ゴム TP010（ショア硬さ: 1 (JIS-A), 引張強さ:  $5.9 \text{ [MPa]}$ , 材質: EPDM (Ethylene Propylene Diene Monomer, 東京ゴム製品工業株式会社製)を用いた。

WT-1の声帯は、Fig. 3 と Table 2 に示すように成人男性の仕様を満たしている。

### 2.4.3 舌

舌は、多くの筋肉からなり、口腔内を自由に動き、調音にとって最も重要な器官の一つである。複雑な舌運動を再現するため、Fig. 4 に示すように舌先・舌体にそれぞれ3自由度パンタグラフ機構を採用し、声帯と同じTP010のゴムを使用し、そのゴムを押し上げることによってさまざまな舌の形状を作ることができる。(A)と(B)のモータによりエンド・エフェクタの位置を、(C)のモータとタイミングベルトによりエンド・エフェクタの姿勢を決定する。舌先には発声の際必要となる舌尖の代わりとなる機構、また

Table 3 Maximum speed of the tongue of human and WT-1

		Human [mm/s]	WT-1 [mm/s]
Tongue Tip	Forward	234	455
	Upward	168	250
Tongue Body	Forward	180	800
	Upward	195	270

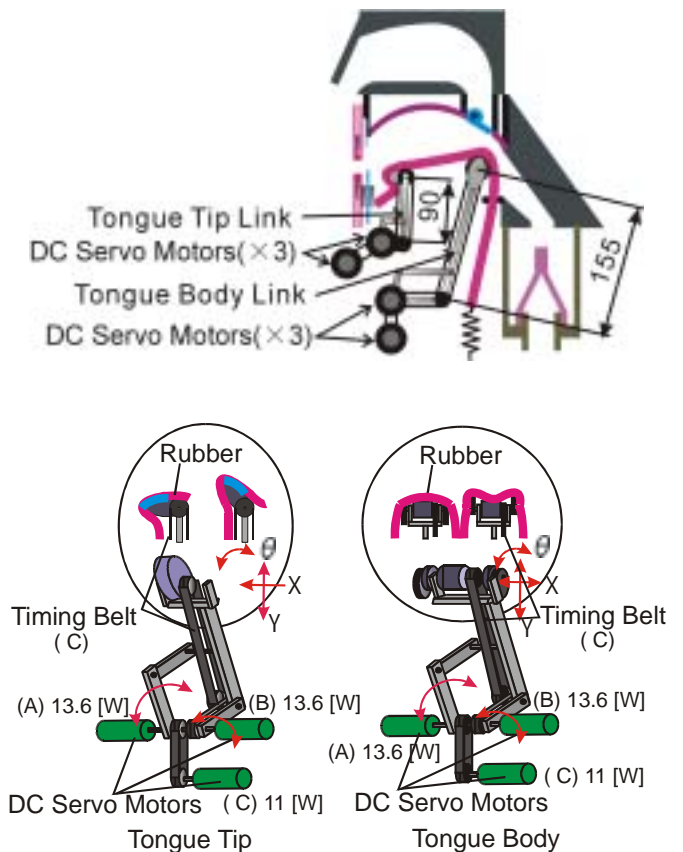


Fig. 4 Tongue using two 3-DOF-pantograph mechanisms

舌体には/s/のように空気ビームの生成に必要な声道の急激な狭めを実現するため舌体の両側を持ち上げる機構を設けている。

WT-1の舌は、Table 3 に示すように人間の速度を上回っており、可動範囲は母音発声時のMRIデータ [11]を参考にした。

### 2.4.4 口唇・歯・鼻腔

声道の終端としての口唇と歯は調音において重要な役割を果たす。口唇は上下方向に対して破裂音などの発声の際高速に運動する必要があるため、Fig. 5 に示すように上下方向にクランク機構、左右方向にワイヤを用い、声帯・舌と同じTP010のゴムを使用し、4自由度の機構とした。歯は下歯にクランク機

構を採用した。

WT-1の口唇と歯は、Table 4に示すように人間の可動範囲と速度を満たしている。

鼻音発声に必要な鼻腔は人間と同様に一對の共鳴

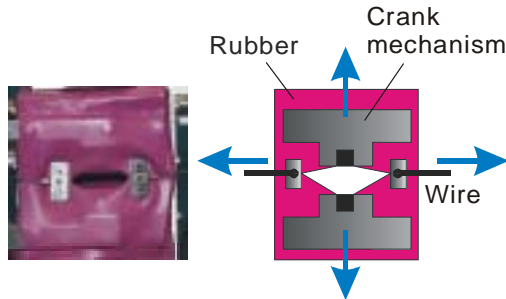


Fig. 5 Lips Mechanism of WT-1

Table 4 Movable Range and Maximum speed of lips and teeth between human and WT-1[12] [14]

		Human	WT-1
Moving Area [mm]	Upper Lip	3.70	20.0
	Lower Lip	10.9	30.0
	Lower Teeth	7.90	20.0
Maximum Speed [mm/s]	Upper Lip	108	223
	Lower Lip	170	294
	Lower Teeth	100	163

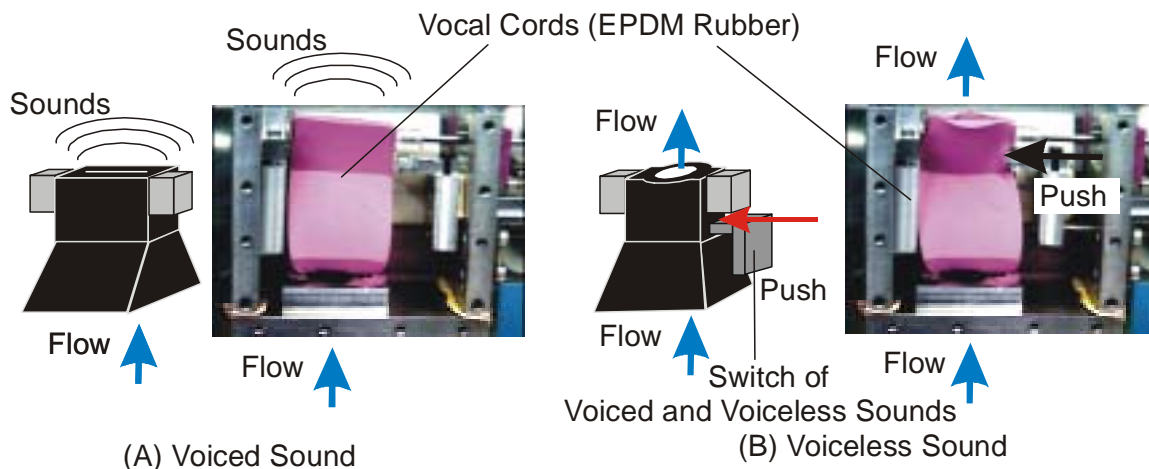
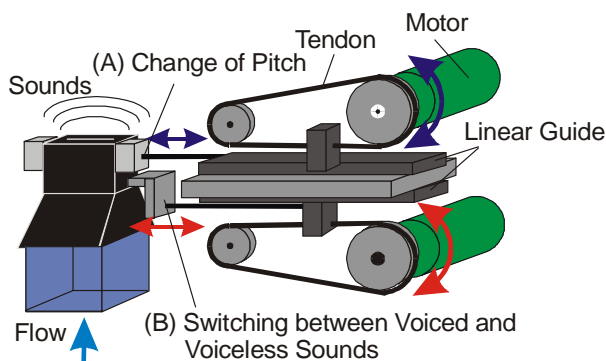


Fig. 6 WT-1R's mechanism of the vocal cords that can produce voiced and voiceless sounds

腔で構成しており、軟口蓋を開くことによって/m/や/nなどの鼻音発声を実現する。

## 2.5 機構部改良

われわれは2001年、WT-1の問題点を改良しFig. 1に示すような人間形発話ロボットWT-1R (Waseda Talker-No.1Refined)を開発した。

また、WT-1Rでは声帯の駆動機構が1自由度増え全15自由度となった。

以下に機構部改良について述べる。

### 2.5.1 声帯駆動機構

人間の音声は、声帯の振動によって「有声音」と「無声音」に区別され、/s/, /h/などの摩擦音の場合、無声音から有声音への遷移時間が40~60 [ms]と報告されている [7]。しかし、WT-1の機構ではボールねじを使用していたため、切り替えに200 - 400 [ms]を要し、無声音から有声音に瞬間的に切り替わる子音の発声が困難であった (Fig. 2)。

そこで、WT-1Rでは声門を引っ張り、ゴムの張力を変えることでWT-1同様97.0 - 140 [Hz]のピッチが変更可能であり (Fig. 6 (A))、逆に声門を押し込み開くことで無声音を生成可能な機構を開発した (Fig. 6 (B))。その結果、独立した駆動機構により、ピッチを保ったまま有聲・無声音の切り替えが可能、かつ有聲・無声音の切り替えが人間同様20~100 [ms]でおこなえるようになり、子音発声の可能性が高まった。

### 2.5.2 口唇

口唇は、声道の終端として調音において大きな役割を担っている。開口面積を変化させ、唇を突き出し、声道長さを変化させることができる。しかし、Fig. 5に示すようにWT-1の機構では、口唇の開口面積のみに注目し、/u/などのように唇を突き出し声



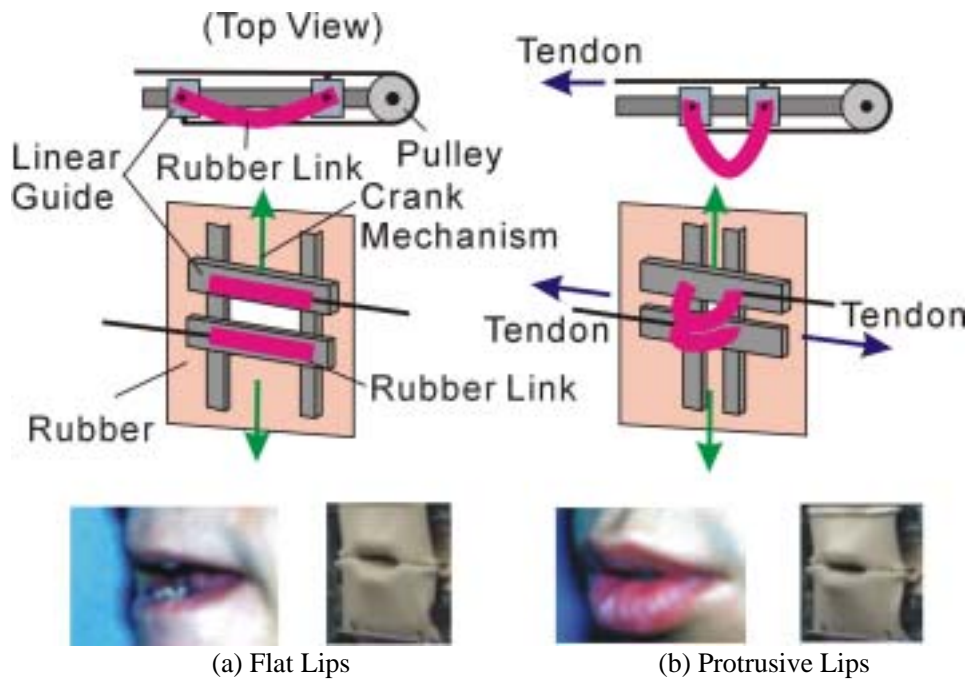


Fig. 7 WT-1R's lips mechanism and it's shape change

道長さを伸ばす必要のある音が再現できない可能性があった。

そこで、Fig. 7に示すように開口面積、声道長さの両方を変化させることが可能な新たな口唇駆動機構を開発した。この機構では、ワイヤとゴムでできたリンクが拮抗し、ワイヤを引っ張ることでリンクが変形し唇を突き出す。また、開口面積はWT-1同様、上下のクランク機構により変化できる。

### 2.5.3 鼻腔・軟口蓋

人間は、軟口蓋を開き鼻腔に空気が流れることで鼻腔内で共鳴し、/m/、/n/などの鼻音が生産される。WT-1は、単音発声させながら軟口蓋を開閉してもあまり音が変化しないという問題点があった。われわれはその原因として

- (1) 軟口蓋の位置
- (2) 鼻腔の大きさを考えた。

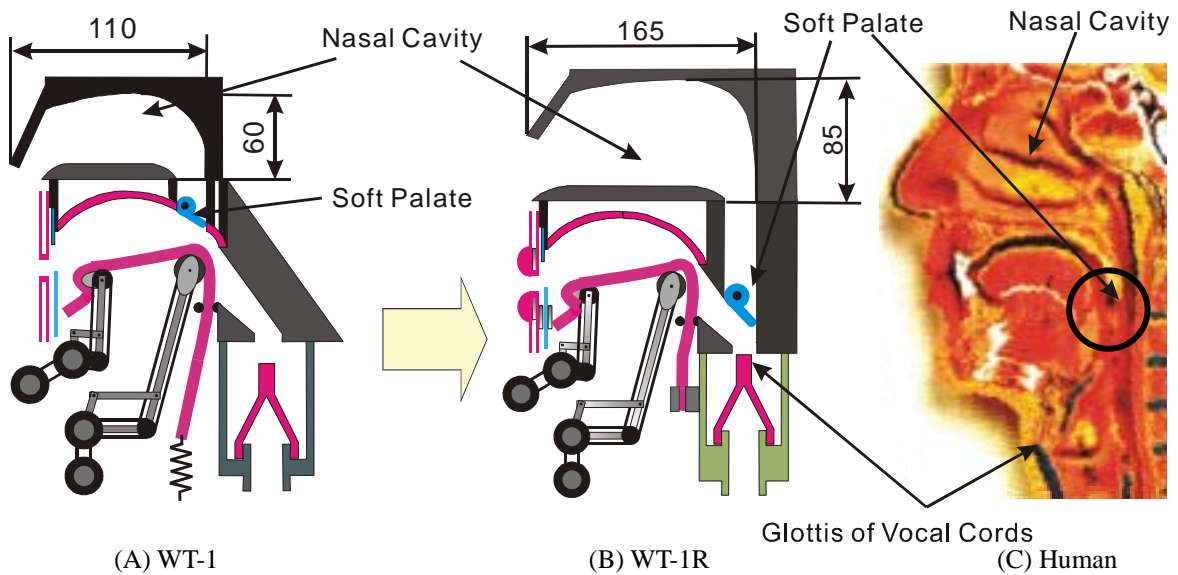


Fig. 8 Mechanism of the nasal cavity and soft palate of WT-1R compared with WT-1 and human

### (1) 軟口蓋の位置

Fig. 8 (C)に示すように人間の軟口蓋は声帯直後の上部付近にあるのに対し、Fig. 8 (B)に示す WT-1 の軟口蓋は声帯からかなり離れた舌上部にあった。そのため、鼻腔に空気が流れにくい構造になっていたと考えられる。そこで WT-1R は、Fig. 8 (B)に示すように軟口蓋の位置を変更した。これにより、軟口蓋を開いた際に、声帯からの空気の流入がスムーズに鼻腔に流れるようになった。

### (2) 鼻腔の大きさ

WT-1 の鼻腔も口腔に比べ小さく、それも原因の1つとして考えられた。そこで、Fig. 8 (B)に示すように人間の MRI データを参考とし、その大きさと形状を変更した。

また、鼻腔の内部には音のエネルギーを適度に吸収する素材として、綿を詰めている。

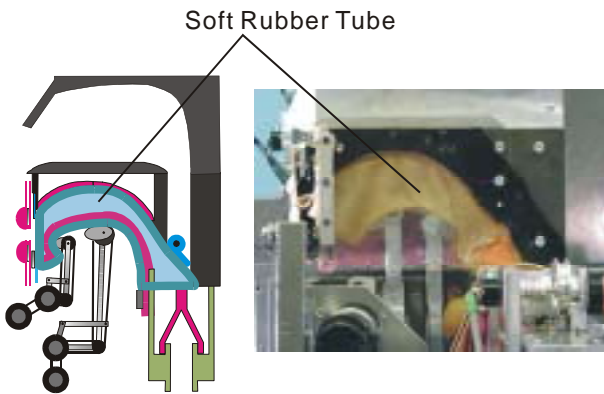


Fig. 9 WT-1R's sealing up for air tightness

### 2.5.4 気密性の向上

声道の空気の密閉は、摩擦音や破裂音などの子音発声の際、非常に重要である。しかし、WT-1 では空気の密閉が十分ではなかったため摩擦音の/s/を発声させようと試みても、口唇以外の声道部分から空気が漏れるため、人間のような「スー」という摩擦音を発声することはできなかった。

そこで、Fig. 9 に示すように声帯上部から口唇までの声道をゴム製の筒でつなぎ、声道からの空気漏れを防ぐことで、摩擦音を実現した。

### 2.5.5 肺

WT-1は、肺から送られた空気によって声帯を振動させて音を生成する。人間が人工声帯を吹いた場合、Fig. 10 (A)のようにステップ状に0.2~0.3 [s]で目標値に達する。この肺圧の立ち上がりは子音発声など瞬間的に空気を送る時に重要になる。WT-1の肺は、90 [W] DCモータ、減速比4.3のアクチュエータで制御していたため、Fig. 10 (B)に示すように声門下圧が目標値に達し安定するまでに2.0-2.8 [s]を要していた。

そこで、250[W]DCモータ、減速比1のアクチュエータを用いて制御をおこなった。その実験結果は4.1節に述べる。

## 3. 子音の制御

子音の音響的特徴は母音よりも複雑である。調音運動と声門の開閉運動を見ても、母音が定常的であるのに対し、子音は過渡的である。さまざまな音響的特徴が複雑に絡み合い、それらの音響現象をロボットで再現することは困難である。

そこで、WT-1Rでの音声合成は人間の調音パラメータの時間的変動を正確に求めることが必要である。

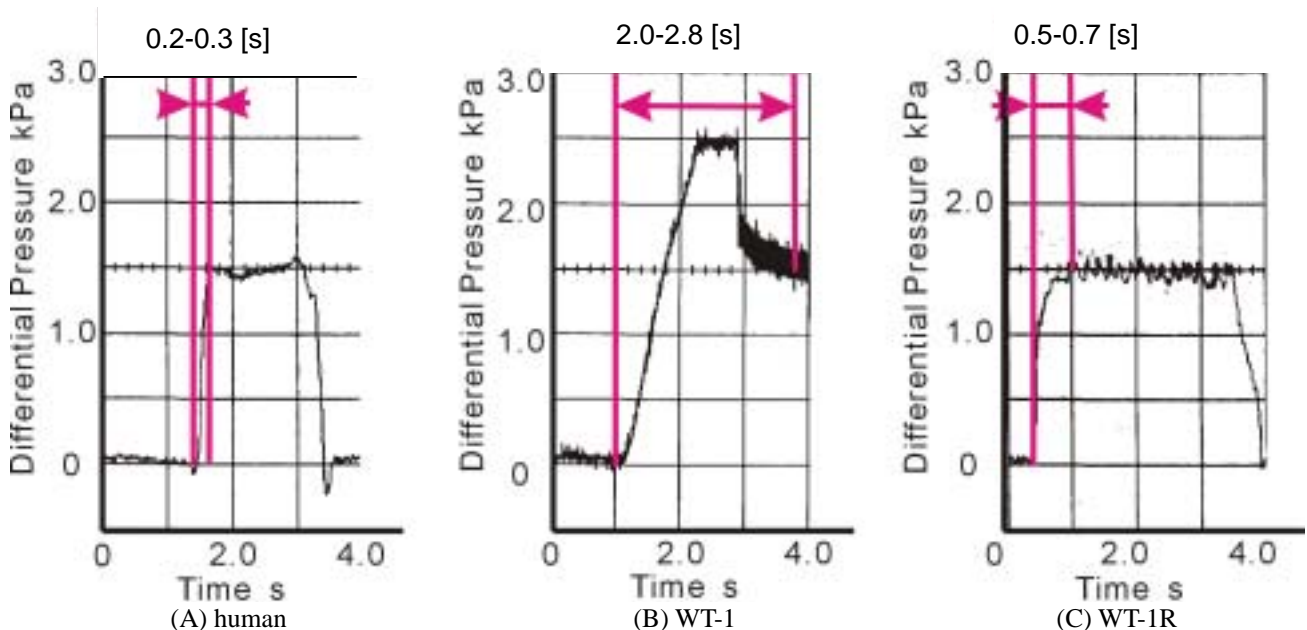


Fig. 10 Comparison of the lung pressure between WT-1, WT-1R and human

これは NTT において計測された調音運動と声門の運動データを利用し、15 自由度のパラメータを時間軸上で協調的に制御し、子音発声を試みた。具体的には、Fig. 11 に示すように子音発声時の各パラメータの位置、母音発声時の各パラメータの位置を設定し、さらに、各パラメータに対しては子音発声時の持続時間 ( $0-t_1$ )、子音から母音への遷移時間 ( $t_1-t_2$ )、母音の持続時間 ( $t_2-t_3$ )を設定する。子音 + 母音の 2 音素の組として、位置と時間のパラメータを設定し、これを目標軌道として与え、モータを制御する。Fig.11 に示すのはある発音をする時の、各時間でのアクチュエータの動きである。実際は WT-1R の全 15 自由度すべてのパラメータを設定して制御する。

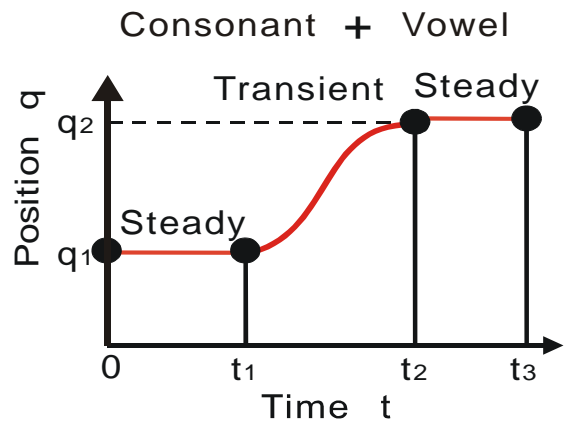


Fig. 11 Motion planning for consonant sounds

#### 4. 実験

##### 4.1 肺

肺のアクチュエータにステップ状の位置指令を送り、そのときの肺圧を圧力センサで計測した。Fig. 10 (B) に示すように、WT-1Rの肺圧の立ち上がりに要する時間は0.5 ~ 0.7 [s] とWT-1に比べ人間に近い肺圧の立ち上がりを確認した。これにより、発声の瞬間からはっきりと発声することが可能になり、より明瞭な母音、子音発声の可能性を見いだした。

##### 4.2 摩擦音 (/s/, /h/)

WT-1Rでは、子音発声のために気密性の向上をおこなった。その結果、WT-1では困難であった /s/, /h/ の区別が可能となった。/s/の発声は下歯と上歯の隙間を小さくし、舌により狭められた声道から肺のペロフラムの駆動速度を上げることで空気ビームを生成し発声させた。

また、/h/の発声は/s/の発声で閉じていた歯を開き狭めを作らず、ある程度の断面面積を保ったまま発声させた。その結果、人間の耳で聞いたところ/s/, /h/ の発声の区別をすることができた。また、Fig. 12 に示すように/s/の特徴である広範囲での強い雑音を観察することができた。

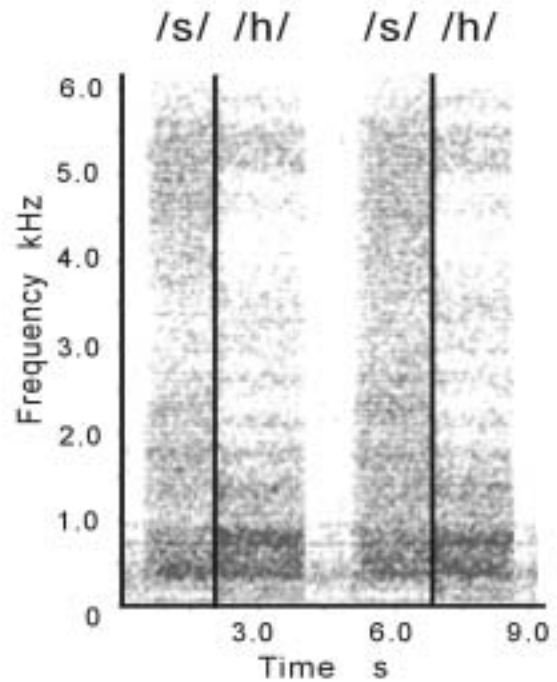


Fig.12 Spectrums of /s/ and /h/ sounds of WT-1R

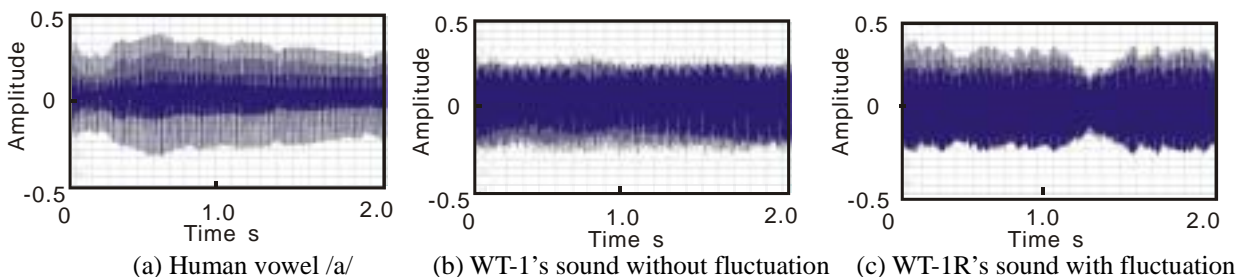


Fig.13 Fluctuation of amplitude



#### 4.3 ゆらぎ

人間の音声には振幅にゆらぎがあり、このゆらぎがあることで、自然な発声になっている[13]。しかし、Fig. 13(b) に示すように、WT-1 の発声は、ほとんど振幅にゆらぎが無く、機械的なブザー音のような発声となっている。

そこで、WT-1R の発声にゆらぎを加え自然な発声にするために肺のペロフラムの速度の制御信号に、 $1/f$  ゆらぎを入れた。

$1/f$  ゆらぎの式は(1) (2) に示す[13]。

Fig. 13(c) に示すように、ゆらぎを加えることにより WT-1R の発声は WT-1 に比べると人間の発声に似た自然な発声になった。

$$x(t) = \sum_{n=1}^{\infty} C_n \cos(2\pi n t / T - \theta_n) \quad (1)$$

$$C_n = \sqrt{1/n} \quad (2)$$

$C_n$  : amplitude of fluctuation

$T$  : period of vocal cords' vibration

$\theta_n$  : random number (0- $2\pi$ )

#### 4.4 子音合成実験

第 4 章 で述べた方法により、子音発声の実験をおこなった。その結果、摩擦音 /s/、/h/、鼻音 /m/、閉鎖音 /p/ の発声を実現した。

/s/、/h/ については 4.2 節で述べたように発声を実現できた。/m/ については、人間同様軟口蓋を開閉することで実現した。/p/ は WT-1R で声道の密閉性と肺圧の立ち上がりを向上させたことにより破裂音を実現した。

### 5 . 結論と今後の展望

#### 5.1 結論

WT-1 の問題点を改良した人間形発話ロボット WT-1R を開発した。その結果、摩擦音 /s/、/h/、鼻音 /m/、閉鎖音 /p/ の発声を実現した。また、肺のペロフラムの速度制御信号にゆらぎを加えることで、WT-1 より人間に近い発声を実現した。

#### 5.2 今後の展望

今後はより効果的な共鳴腔を作り出せる機構の開発が必要であろう。声帯は発声に関しては重要な役割を占めており、今後もさらに周波数域の広い声帯の開発、機構の検討をするべきであろう。

人間の発声の仕組みは未だに解明が困難な部分があるため、今後も試行錯誤の連続であると思うが、最終的に連続発声が可能になれば、発声メカニズムの解明に多いに役に立つことは間違いない。そのためにも、現在、WT-1R の大きさは人間の約 1.2 倍となっているが、人間と同等の大きさの新しいメカニズムの開発し人間の口腔の中で起きている音

響現象を忠実に模擬する必要もあるだろう。

また運動の制御対象とされにくい骨格が無く柔軟な組織から構成されている舌や声帯を制御・機械的に模擬することで、人間形ロボットへの新たなアプローチが見つかるであろう

#### 謝辞

本研究は、科学技術振興事業団 (JST) 戦略的基礎研究推進事業 (CREST) の援助を受けている。

本研究の成果の一部は理工総研プロジェクト研究「発声力学に基づくタスクプランニング機構の構築における機械モデルの開発」の一環として行われたものである。

NTT の菅田雅彰氏をはじめ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、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に感謝致します。

#### 参考文献

- [1] J. L. Flanagan: "Speech Analysis Synthesis and Perception 2nd ed.", Springer, 1972
- [2] 梅田, 寺西: "声の韻質と声質 音響的声道模型による音声の合成", 日本音響学会誌, 第 22 巻, 第 4 号, pp.195-203, 1965
- [3] 井澤, 服部, 松岡, 川村: "機械系制御による音声合成", 第 3 回ロボットシンポジウム予稿集, pp.273-278, 1993
- [4] Martin Riches: "MOTORMOUTH A Speaking Machine", Journal of Experimental Musical Instrument, pp.20-23, 1998
- [5] 澤田, 大須賀, 小野: "機械式音声合成装置の実現に向けて - 第 2 報: 構音のための舌のモデルの提案", 日本ロボット学会誌, 第 17 巻, 第 7 号, PP.1001-1008, 1999
- [6] 大島, "人体の構造と機能", 新思潮社, pp.176-178, 1980
- [7] R. D. Kent, C. Read: "The Acoustic Analysis of Speech", Kaibundo, 1992
- [8] 岩田他: "空気力学的に見た発声機序 声の強さの調節", 音声言語医学会誌, 第 36 巻, 第 1 号, pp.14-21, 1995
- [9] 毛利, "歯科医のための耳鼻咽喉科学", 医歯薬出版, pp.95-99, 1996
- [10] Y. Kakita, M. Hirano and K. Ohmaru: "Physical properties of the vocal fold tissue: Measurement on excised larynges", Vocal Fold Physiology, University of Tokyo Press, pp.377-396, 1981
- [11] Kaburagi, Kyushu Institute of Design
- [12] 沢木他: "みみ・はな・のど 基礎から臨床へ", 南江堂, pp.227, 1977
- [13] T.Wada: "Analysis of Fluctuation and Rhythm in Human Body", Akodansha, 1999