調音のダイナミックスを考慮した声道模型の動きと出力音声の関係*

○荒井隆行・△中川千沙希・△川端千尋・鈴木良平・辻慎也（上智大・理工）

1　はじめに

人間の声道を物理的に模擬した声道模型において、声道形状の時間的な動きを実現し調音のダイナミックスを実現するアプローチは、以前から複数試みられている[1-4]。声道模型については、音声生成の仕組みを直感的に理解できるため、教育応用やその他研究応用でもその有用性が報告されている（例えば[5]）。一方で、動的な声道模型において、特に子音を再現することについては、比較的時間変化が遅い接近音はその実現が容易である一方、破裂音においてはある程度の工夫が必要である。例えば、我々が以前開発した BMW モデル[6-8]では、/b/の音を実現するため下唇用のブロックが上下する機構を備えており、バネの弾性力も手伝って破裂の解放を実現している。また、同様に BMW-DN モデル[8]では/d/の音を実現するために舌尖が歯茎にて閉鎖を作ることが可能であり、素早くレバーを倒すことによって、やはり破裂の解放が実現される。

そこで、本稿では以前から改良を続けてきた梅田・寺西[9]による声道模型について，いくつかの改良版[10,11]の他、新規モデルも含めて、破裂音をターゲットとした際の調音動作を検討し、一部については速度測定を行った。また、フレーズを出力する試みについても分析した。

2　動的な声道模型

2.1 梅田・寺西式 VTM-UT45-D11

Fig. 1 に、梅田・寺西[9]による声道模型の改良版[10-12]を示す。もともとの梅田・寺西による模型は、手動にて角棒を抜き差しして、声道形状を変えていた。この模型には従来から鼻腔も設けられており、ダイヤルを回すことで鼻咽腔結合の程度がコントロールされる仕組みとなっていた。

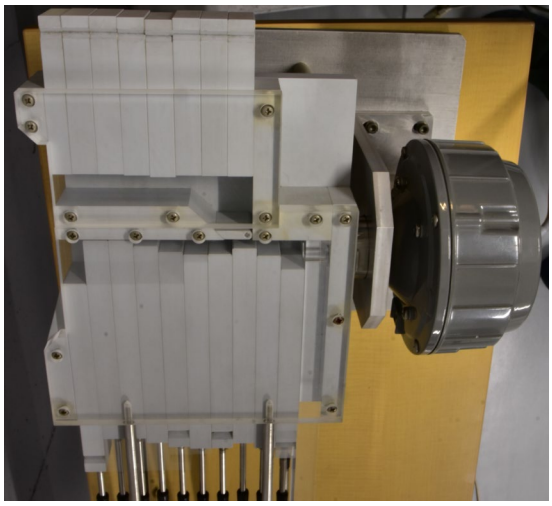

Fig. 1: VTM-UT45-D11 featuring ten bars and a dial for the velopharyngeal port connected with a set of actuators [10-12].

このようなオリジナルの模型に対し、Fig. 1 が示すように、各角棒をリニアアクチュエータによって自動的に動かせるような改良を施した。当初の改良では 11 本すべての角棒に対し、11 基のアクチュエータを接続した[10]。しかし現在の改良版では、喉頭側の 1 本以外の 10 本の角棒に 10 基のアクチュエータが備え付けられており、11 基目のアクチュエータについては鼻咽腔結合部の開閉をコントロールするようにしている。以後、この声道模型を VTM-UT45-D11 と呼ぶこととする。

喉頭側にはホーンスピーカのドライバユニット（TOA, TU-750）が接続されており，任意の音源を再生可能となっている．ただし，制御系と音源再生用のソフトウェアが独立しているため両者の同期を取らなければならず，機械的にトリガーを検知して音源が出力するような工夫を施している．

* Relationship between movements and output speech sounds of vocal-tract models taking articulatory dynamics into consideration, by ARAI, Takayuki, NAKAGAWA, Chisaki, KAWABATA, Chihiro, SUZUKI, Ryohei, and TSUJI, Shinya (Sophia University).

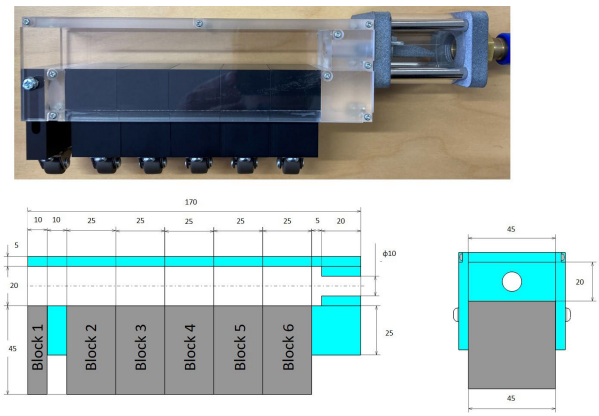

Fig. 2: VTM-UT45-D6 featuring six blocks.

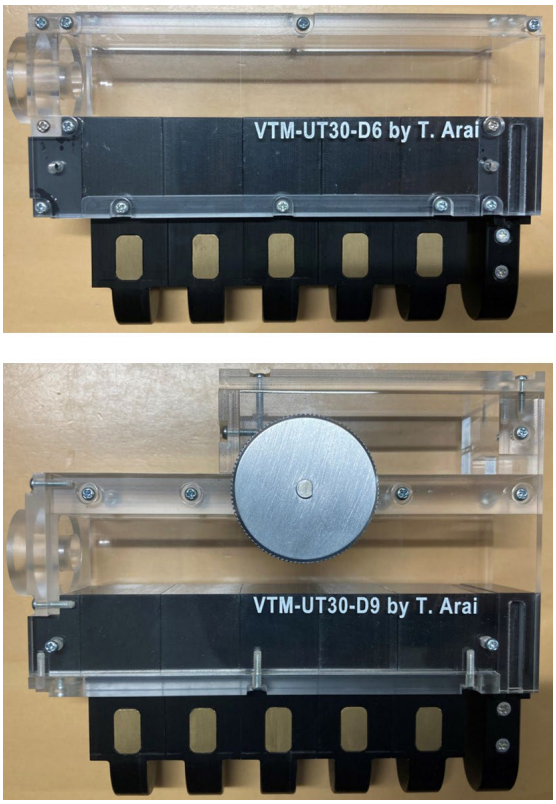

Fig. 3: VTM-UT30-D6 (top panel) and VTM-UT30-D9 (down panel).

2.2 梅田・寺西式 VTM-UT45-D6

VTM-UT45-D11 と同じように声道の側面からブロックが抜き差しされるような模型の変種として、VTM-UT45-D6 が開発された（Fig. 2）。違いとして、VTM-UT45-D6 には鼻腔は付いておらず、抜き差しされるブロックの数も 6 個である。さらに、アクチュエータによる操作ではなく、手動あるいはカム機構でブロックを上方に押し上げ声道内に閉鎖や狭めを作り、手や機構が外れることによってブロックが自由落下することで声道が再び元の開いた状態に戻る。

Fig. 4: Lip model with an actuator connected to the lower lip.

2.3 梅田・寺西式 VTM-UT30-D6/D9

VTM-UT45-D6 と同じように声道の側面から 6 個のブロックが抜き差しされるような模型であるが、横幅が 45 mm から 30 mm に、全長が 170 mm から 135 mm に変更されたのが VTM-UT30-D6 である。さらに、VTM-UT30-D6 に鼻腔を付けたモデルである VTM-UT30-D9 も製作した。D9 では D6 と違って、透明アクリルの側板を固定するネジの方向が改良されたため、ブロックの動きを鈍らせる締め付けが改良されている。

2.4 口唇モデルのシステム

現在、文献[13]で開発された口唇モデルの下唇をアクチュエータによって操作可能とし、口唇の開閉を PC によってコントロールするシステムを開発中である。このシステムでは、口唇の裏にスピーカ（ELECOM, MS-P08UBK）が隠されており、そこから音声信号が出力される。

3 測定

3.1 VTM-UT45-D11

CV あるいは VCV（C は子音、V は母音）の短いフレーズに対し、特に C が破裂音[b], [d], [g]をターゲットとした（V は [a]に固定）。[b]については、口唇側の角棒だけを動かした。一方、[d]と[g]の調音器官の動きは、実時間 MRI 動画[14]を参考にした。Fig. 5 に、測定に用いた 12 パターンを示す。

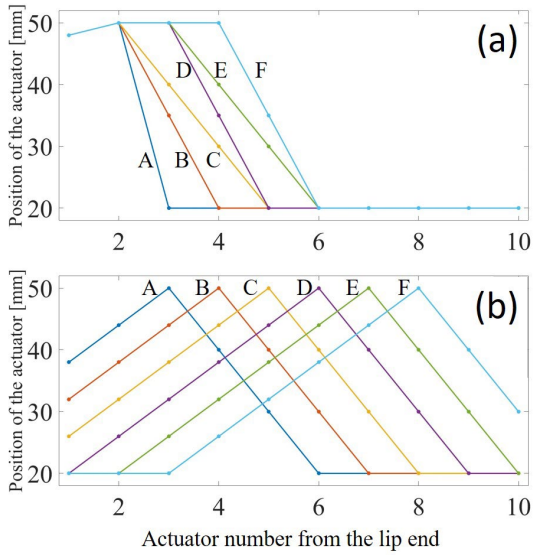

Fig. 5: Twelve patterns of positions of the consonant from Actuators (Bars) 1 through 10 for VTM-UT45-D11: (a) Set 1 and (b) Set 2.

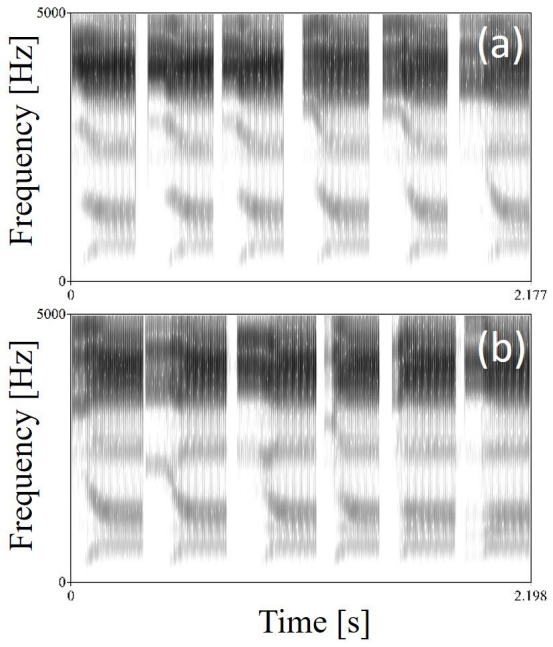

Fig. 6: Spectrograms of 12 patterns using VTM-UT45-D11: (a) Patterns A–F in Set 1 and (b) Patterns A–F in Set 2.

Table 1: Estimated average speeds in m/s (VTM-UT45-D11).

	Inserted steps			
	none	5 steps	10 steps	15 steps
[b]	1.36	0.91	0.89	0.74
[d]	1.80	1.07	0.86	0.78

Table 2: Estimated average speeds in m/s (VTM-UT30-D6/D9).

	Block 1 [b]	Block 2 [d]	Block 3 [g]
D6	0.32	0.24	0.27
D9	0.32	0.30	0.30

Fig. 6 は、最速で動かした際の出力音声のスペクトログラムである。この場合、PC からの命令は 1 time-step の約 10 ms であった。その際の調音速度を Table 1 の “none” の列に示す（[d]は Set 1 の Pattern C）。この際、ビデオカメラ（Sony, FDR-AX700）の高速度モード（960 fps）で測定した。一方、time-step を複数挿入することで速度が遅くなる様子も確認できた。

3.2 VTM-UT30-D6/D9

これらのシステムでは、口唇側から数えて 1 つ目のブロックが[b]、2 つ目が[d]、3 つ目が[g]にほぼ対応する。したがって、これらの破裂音を発する際に、各ブロックがどのくらいの速度で動くかについて測定した。測定は 3.1 節同様に、ビデオカメラの高速度モード（960 fps）にて 10 回の撮影を実施し、ブロックが自由落下するのに要したコマ数から平均速度を推定した。その結果を Table 2 に示す。

さらに、VTM-UT30-D6 を使って、英語の “How are you?” というフレーズを想定した発話を試みた。その際、Fig. 7 のように回転カムの機構を VTM-UT30-D6 の下に設置し、手動でカムを回すことによって上記フレーズを生成し録音した。なお、この声道模型への入力には、リード式音源を用いた。Fig. 8 に、出力された音声に対するスペクトログラムを示す。

3.3 口唇モデル

このシステムは口唇しか動かないため、破裂音[b]を想定して口唇を開く際に、下唇がどのくらいの速度で動くかについて測定した。測定は 3.1 節同様に、ビデオカメラの高速度モード（960 fps）にて 10 回の撮影を実施し、下唇が開くまでに要したコマ数から平均速度を推定した。その結果、約 0.49 m/s であった。

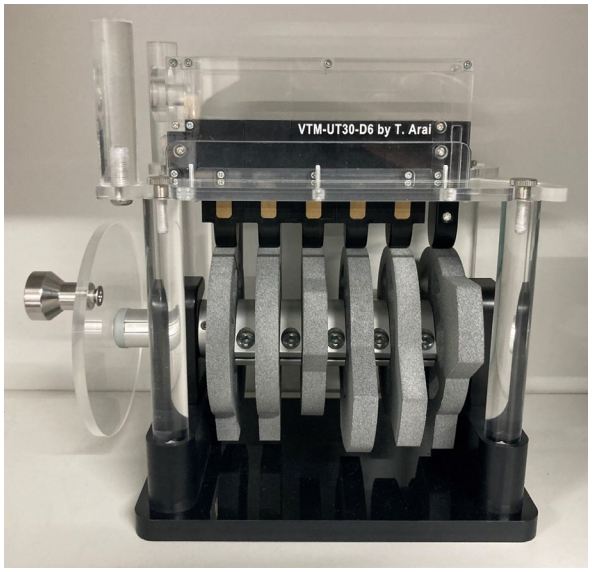

Fig. 7: VTM-UT30-D6 with the rotating cam mechanism for “How are you?”.

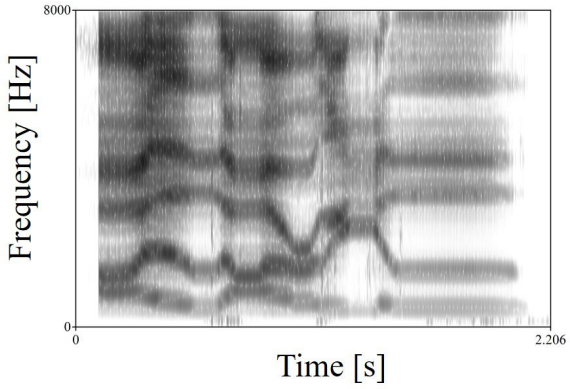

Fig. 8: Spectrogram of “How are you?” produced by VTM-UT30-D6.

4　考察・まとめ

本稿では，調音のダイナミックスを考慮した声道模型のいくつかを対象に、破裂音をターゲットとした際の調音動作に関する速度測定やフレーズの生成などを試みた。従来の声道模型は特定の母音を生成するものから、調音動作を再現するものまでいくつかの試みが展開されている。その中でも、我々は梅田・寺西式の模型を中心に、オリジナルモデルに対する改良版や新しく開発中のものについても動作を確認した。

VTM-UT45-D11 については、アクチュエータの動作速度をなるべく速く動くような調整を過去にしていたため[12]、Fig. 6 を見てもその速度は破裂音を再現するのに十分であることが確認された。

VTM-UT30-D6/D9 については、ブロックが挙上する際の速度は例えば回転カムなどによって制御可能である一方、今回の機構ではブロック自身の自重で自由落下によって閉鎖の解放などの調音動作が実現されるようになっている。この場合、Table 2 からブロックの速度はやや遅くなっている。しかし、Fig. 8 に見られるような接近音を多く含むフレーズを生成する場合には十分であることが確認された。

一方、口唇モデルについてはアクチュエータを使っていることから、その速度もある程度、担保されていることも確認された。

今後は、これらの声道模型について調音動作の調整や音源の工夫などを追加し、さらに発話できる音のバリエーションを増やしたり、その他の改良を進める予定である。

謝辞

内容の一部は，JSPS 科研費 24K06423 の助成を得た．

参考文献

[1] Mochida *et al.*, *INTERSPEECH*, 1533–1536, 2002.
[2] Hofe & Moore, *Connection Science*, 20(4), 319–336, 2008.
[3] Fukui *et al.*, *INTERSPEECH*, 1021–1024, 2010.
[4] Sawada & Hashimoto, *JSME Int’l Journal*, Series C, 43(3), 645–652, 2000.
[5] Arai, *JASA*, 152(5), 2746–2757, 2022.
[6] Arai, *INTERSPEECH*, 1099–1103, 2016.
[7] Arai, *INTERSPEECH*, 979–983, 2017.
[8] Arai *et al.*, *INTERSPEECH*, 2018–2019, 2023.
[9] Umeda & Teranishi, *JASJ*, 22(4), 195–120, 1966.
[10] Arai, *INTERSPEECH*, 1025–1028, 2010.
[11] Arai *et al.*, *INTERSPEECH*, 987–988, 2024.
[12] 荒井他，音講論(春), 663–664, 2024.
[13] Arai, *INTERSPEECH*, 3171–3175, 2021.
[14] Lim *et al.*, *Sci. Data*, 8(1), 187, 2021.