

声道模型による音声ダイナミックスを再現する試み*

○荒井隆行・鈴木良平・△EARP, Chandler・辻慎也（上智大・理工）

1 はじめに

声道模型は人間の声道を模擬した物理模型で、音声生成に関する仕組みを説明するためにも有効であり、音や音声に関わる教育やデモンストレーションにも広く応用されている[1-4]。声道形状を機械制御することで発話ロボットなどに応用されることになるが、物理的な発声機構と組み合わせたりスピーカを介して音源信号を与えることによって、音声のダイナミックスが実現されてきている（例えば [5-8]）。

我々も以前から、梅田・寺西[9]による声道模型を改造し、アクチュエータを接続して音声のダイナミックスに関する実験やデモンストレーションを行ってきた[10,11]。しかし、その詳細については未解明な点も多く、研究を進める意義がある。

そこで本稿では、音声ダイナミックスを実現する3つのPC制御型システムを見ると共に、その一部の性能を検討した。

2 声道模型による3つのシステム

2.1 梅田・寺西式声道模型によるシステム

Fig. 1 に示すように、角棒を抜き差しして声道形状を変化させる梅田・寺西式の声道模型 VTM-UT において、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にしたものである [10,11]。現在は、口腔内において喉頭側の1本は固定である一方、残りの10本がアクチュエータによる制御対象となっている。鼻腔には8本の角棒があるが、それらは手動で変えられるものの、アクチュエータには接続されていない。また、鼻咽腔結合は1基のアクチュエータにより弁の開閉が制御可能となっている。

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ない、機械的なトリガー装置が別途必

要になるなど、工夫を要する状況となっている。

2.2 スライド式声道模型によるシステム

Fig. 2 に示すように、アクリルパイプ (外径 24 mm, 内径 20 mm) の中を、直径が 18 mm の丸棒がスライドする声道模型 VTM-S24 [12,13] において、そのスライド部にリニアアクチュエータを接続し、PC による制御を行うシステムを新たに開発した。喉頭側にはホーンスピーカのドライバユニット (TOA, TU-750) が接続されているが、スライド部の動きの制御と音源再生の両者について、共に Matlab から指令を出すことが可能である。

2.3 ロボットアームによるシステム

その他のシステムとして、Arai [14]の提案に基づきロボットアームを介して VTM-S24 を制御するシステムが実現された。

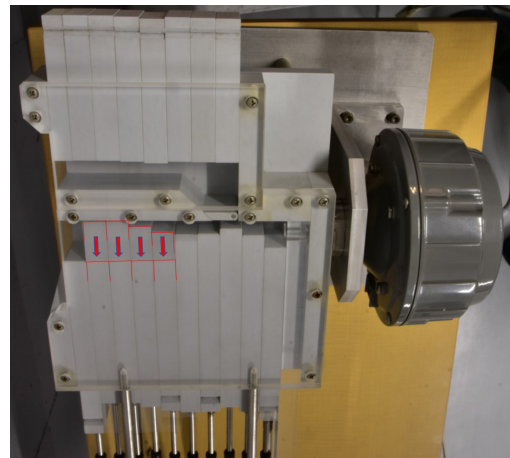


Fig. 1: PC-controlled Umeda-Teranishi vocal-tract model VTM-UT [10,11].

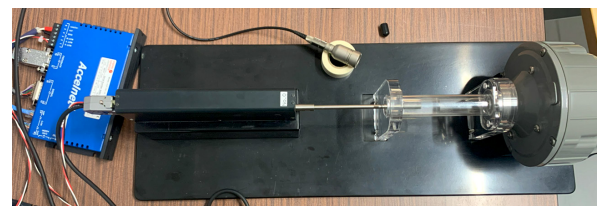


Fig. 2: PC-controlled Sliding vocal-tract model VTM-S24 [12,13].

* PC-controlled vocal-tract models for speech dynamics, by ARAI, Takayuki, SUZUKI, Ryohei, EARP, Chandler, and TSUJI, Shinya (Sophia University).

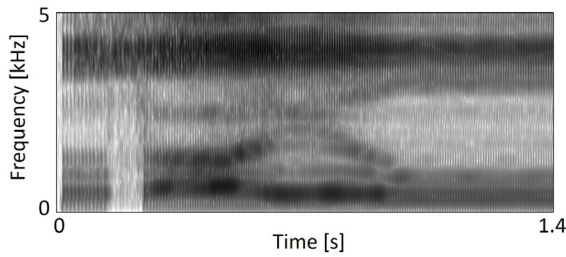


Fig. 3: A Japanese utterance of “Ohayo:” produced by the VTM-UT system.

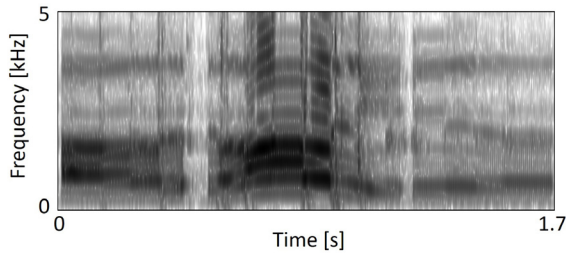


Fig. 4: An English utterance of “How are you?” produced by the VTM-S24 system.

3 音声ダイナミックスの検証

3.1 VTM-UT システム

VTM-UT システムを用いて、「おはよう」というフレーズを出力させる試みを実施した。男性の原音声から、Praat [15]を用いて音源（phonation）を求めた。出力音声に対するスペクトログラムを Fig. 3 に示す。

3.2 VTM-S24 システム

VTM-S24 システムを用いて，“How are you?” というフレーズを出力させる試みを実施した。女性の原音声から、Praat [15]を用いて音源（phonation）を求めた。出力音声に対するスペクトログラムを Fig. 4 に示す。

4 考察

本稿では、声道模型を用いた PC 制御型システムに対し、音声のダイナミックスを実現する試みとして、フレーズの生成を中心に検討した。その結果、VTM-UT と VTM-S24 の両方において、ある程度は明瞭な出力音声を得たが、まだ改良の余地も同時に確認された。単音では明瞭度が低い場合も、フレーズにすることで声道形状がダイナミックに変化すると、明瞭度が上がることも確認された。特に第 2 フォルマント (F2) などに注目し、その軌跡を追うこと

は重要なポイントとなることが予想される。今回は、イントネーションなどの韻律情報が音源信号の中にオリジナルに近い形で保たれており、そのことも明瞭度向上に寄与していると考えられる。

5 おわりに

今回の調査では、Fig. 1 に示す通り、口唇側（左側）から 2 本目～5 本目の角棒を、赤い矢印のように瞬時に引くことで、歯茎音（例えば/d/のような音や日本語ラ行子音など）を近似できることも確認された。口唇側から 1 本目が両唇音を近似することも分かっていたため（例えば[16]），これらを組み合わせることで、実現可能なフレーズの質の向上ならびにバリエーションを増やしていきたい。また、音源に関しても物理模型の応用も進めている。さらに、解剖模型風[11]の改良も検討していく予定である。

謝辞

内容の一部は、JSPS 科研費 21K02889 の助成を得た。

参考文献

- [1] Arai, *AST*, 28(3), 190–201, 2007.
- [2] Arai, *JASA*, 131(3), 2444–2454, 2012.
- [3] Arai, *AST*, 37(4), 148–156, 2016.
- [4] Arai, *JASA*, 152(5), 2746–2757, 2022.
- [5] Mochida *et al.*, *INTERSPEECH*, 1533–1536, 2002.
- [6] Hofe and Moore, *Connection Science*, 20(4), 319–336, 2008.
- [7] Fukui *et al.*, *INTERSPEECH*, 1021–1024, 2010.
- [8] Sawada and Hashimoto, *JSME International Journal*, Series C, 43(3), 645–652, 2000.
- [9] Umeda and Teranishi, *JASJ*, 22(4), 195–120, 1966.
- [10] Arai, *INTERSPEECH*, 1025–1028, 2010.
- [11] 荒井, 音講論(春), 781–784, 2022.
- [12] 男澤他, 音講論(春), 809–812, 2022.
- [13] 荒井, 音講論(春), 609–612, 2023.
- [14] Arai, *ICA*, ABS-0792, 2022.
- [15] Boersma, *Glott International*, 5:9/10, 341–345, 2001.
- [16] Arai, *INTERSPEECH*, 1099–1103, 2016.