

## サイズの異なるスライド式3音響管 による母音について\*

☆男澤和果 (立命館慶祥高), 荒井隆行 (上智大), △渡邊儀輝 (立命館慶祥高)

### 1 はじめに

Fant の音響理論[1]に基づいたスライド式3音響管モデル (以後, S3T モデル) [2] は, シンプルな構造で人間らしい母音を生成することができる. 音源フィルタ理論や声道の形状と母音の関係など, 音声生成の基本的な概念を確認することが容易なため, 教育の場面でも応用されている[2,3].

Fig. 1 は, 本研究で用いた S3T モデルの模式図で, 外側の筒 (以後, 外筒) と内側の丸棒 (以後, スライド部) で構成されている. 外筒の中でスライド部の位置を変えて外筒の一端に音源を入力することによって, さまざまな母音が出る. このモデルでは, スライド部によって声道内で舌などが作る狭窄を形成され, スライド部の位置を変えることで母音の音質が変わる. スライド部の位置の小さな変化でも, 母音の質を大きく, そして比較的自由に変化させることができる.

S3T モデルにおいてスライド部の位置以外には, 狭窄の程度や唇の丸みといった 2 つの自由度を追加することで, モデルが作り出す母音の種類をさらに増やすことができる (例えば[2]).

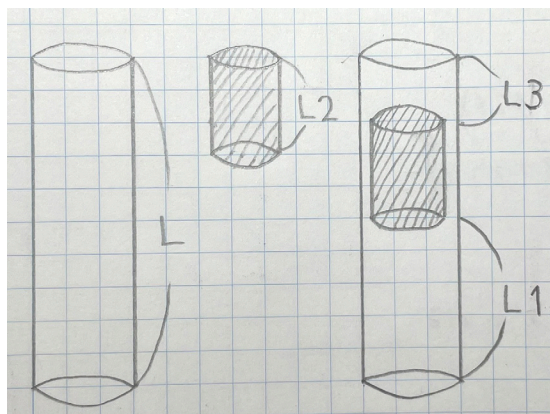


Fig. 1: Schematic designs of the sliding three-tube (S3T) model.

シンプルな構造のため, S3T モデルの第1から第3フォルマント周波数は比較的簡単に推定可能である[4]. これは, モデル全体を以下の3つの部分に分解することで可能になる: 1) 前腔=1/4 波長の共鳴器, 2) 後腔=1/2 波長の共鳴器, 3) 後腔+狭窄部=ヘルムホルツ共鳴器. これにより, 大学の音響学の授業における母音生成の基本的な音響理論を教えるために使用される他, このモデルのシンプルなデザインがゆえに, 子供たちが自分で S3T モデルを工作する科学教室などでも使用されている[2].

これまで, S3T モデルを用いた研究では, 主に成人男性の母音について調べられてきているが, S3T モデルについて子供用と大人用の両方があることは教育的にも学術的にも有用であり, このモデルの筒の長さを短くすることで, 女性や子どもの母音を模擬した実験も行われてきている[5]. しかし, 特に文献[5]では, S3T モデルのサイズが4種類である他, 4種類の音源もスピーカからの再生音を用いていた.

そこで, 本研究では S3T モデルの筒の長さのバリエーションを7種に増やし, さら



Fig. 2: Seven outer tubes of the sliding three-tube (S3T) model in this study.

\* Vowels produced by the sliding three-tube models in different sizes, by OZAWA, Waka (Ritsumeikan Keisho High School), ARAI, Takayuki (Sophia University), and WATANABE, Yoshiteru (Ritsumeikan Keisho High School).

に3種類の基本周波数の異なるリード式音源と組み合わせて、どのような母音が生成されるかを確認した。そして、母音の認識をする手段として、既存の自動音声認識システムを使用した。

## 2 実験

### 2.1 使用した S3T モデル

長さの異なる7つの外筒(透明アクリル)を用意した(Fig. 2)。この筒は内径が20 mm、厚さ2 mmで中が基本的に空洞になっている。それぞれの長さは100 mmから160 mmまで10 mm刻みで7本用意した。そして、外筒の一端には、中央にφ10 mmの穴が開いた厚さ4 mmの円板を接着した。

次に、スライド部として2種類の丸棒を用意した。この丸棒は直径18 mmで長さが40 mmと50 mmであった。

### 2.2 使用したリード式音源

音源には、3種類のリード式音源を用いた[6]。1つ目はSS-R30で、PET材のリード(厚さ0.2 mm)を、長さ32 mm×幅10 mmで準備し、根元の10 mmはアルミプレートによって固定した(振動するリード長は22 mm)。リードの受けは、Rが30 mmの曲率であった。2つ目はSS-R25で、厚さが0.1 mmのOHP用透明シートを長さ24 mm×幅10 mmで準備し、根元の10 mmをアルミプレートによって固定した(振動するリード長は14 mm)。リードの受けは、Rが25 mmの曲率であった。3つ目はSS-R15で、同じく厚さが0.1 mmのOHP用透明シートを長さ19 mm×幅7 mmで準備し、根元の10 mmをアルミプレートによって固定した(振動するリード長は9 mm)。リードの受けは、Rが15 mmの曲率であった。これらのリード式音源の周りは、内径が34 mm、長さが41 mmの筒(透明アクリル)を含む部品で密閉し、両端に穴を設けた。そして、一端からは足踏みポンプ(Logos, Big Bellow Pump 3000)から空気が流入し、もう一端からは喉頭原音を模擬した音源が出力された。

### 2.3 方法

スライド部を、針金を使ってその位置を移動できるようにした。外筒の下にリード式音源を接続し、ポンプを踏んで音を出した。外筒の先端からスマートフォン(iPhone12)を約20 mm離して配置した。スライド部を2 mmごと動かしながら、スマートフォンの自動音声認識システム(Siri, 日本語モード)による認識結果を記録した。

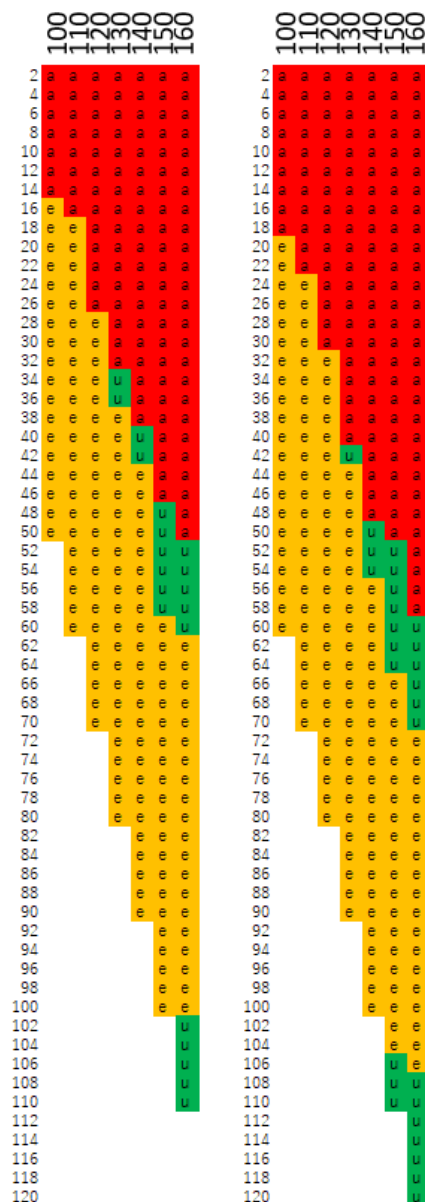


Fig. 3: Experimental results when SS-R30 is used.

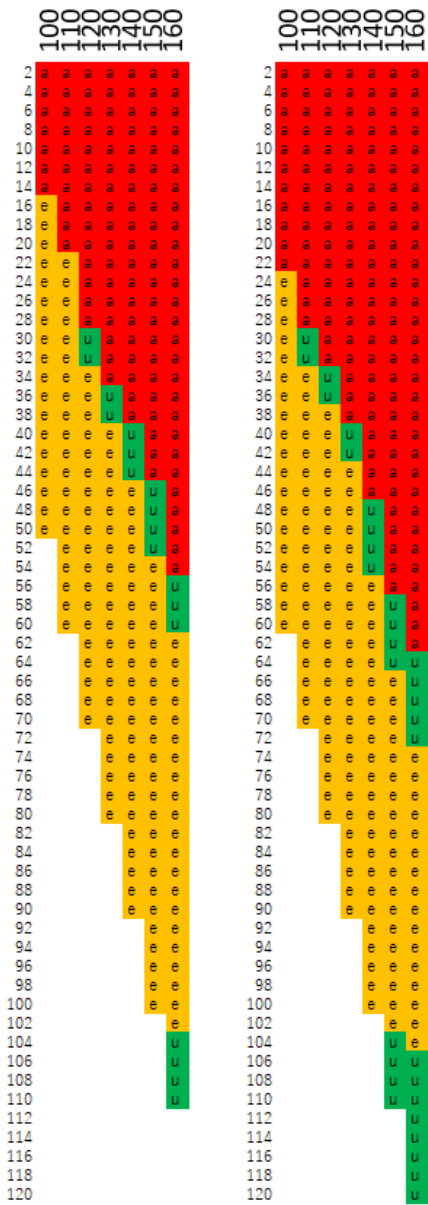


Fig. 4: Experimental results when SS-R25 is used.

### 3 結果

外筒とスライド部の長さをそれぞれ  $L$  と  $L_2$  とし, さらに後方および前方の空洞の長さをそれぞれ  $L_1$  および  $L_3$  とした. スライド部が外筒の中をスライドするとき,  $L = L_1 + L_2 + L_3$  の条件の下で,  $L_1$  と  $L_3$  は 0 から  $L - L_2$  の間で変化する. そこで, 結果の図は縦軸を  $L_1$ , 横軸を外筒の長さ (左の 100 mm から右の 160 mm まで) とし, 認識された母音ごとに次のように色分けした: 赤が/a/, 緑が/u/, 黄色が/e/. Fig. 3 から Fig. 5に, それぞれリード式音源 SS-R30, SS-R25, SS-R15 に対する結果を示す (各図は左半分が  $L_2 = 50$  mm, 右半分が  $L_2 = 40$  mm).

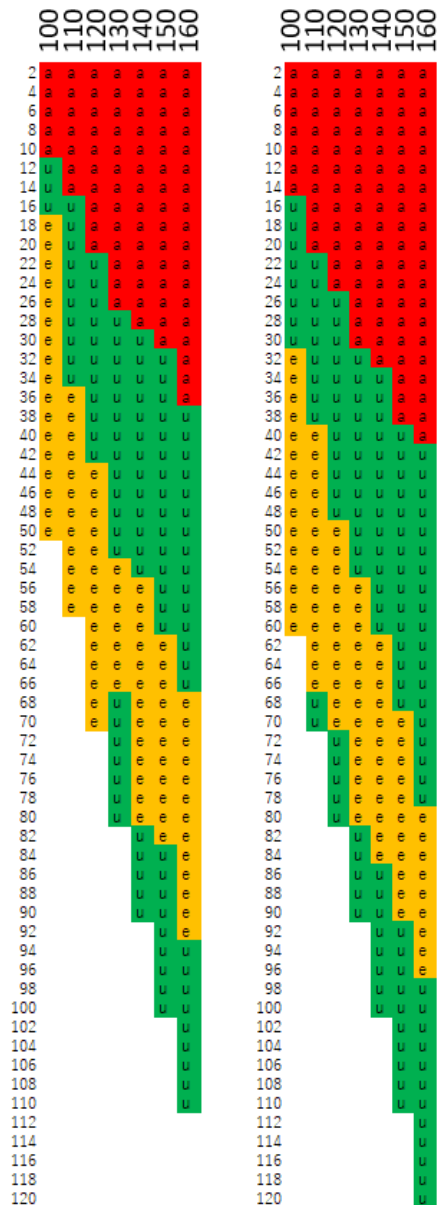


Fig. 5: Experimental results when SS-R15 is used.

### 4 考察

まず, 結果の図を見るといずれの場合においても, 声道模型から作られる音はどれも自動音声認識システムによって母音として認識されていることが確認された. 文献 [2] などにおいては, 外筒の直径が 34 mm などと実際の成人男性のサイズに近いものであった. しかし, 今回使用した外筒の内径は 20 mm と細い. しかし, 母音の生成にはその程度の細さでも問題なく, 聴感上もはっきりと母音として聞き取ることが出来た. そして, 今回の条件下で観測された母音は, /a/, /e/, /u/ の 3 つであった.

観測された 3 母音のうち, 特に母音/a/と

/e/は共に安定的に作られる様子がかがえた。比較的喉頭側にスライド部があるとき母音/a/が、比較的口唇側にあるとき母音/e/が生成された。なお、外筒と狭窄部における隙間との断面積比については、文献[2]では一番大きいもので約 12:1 であったのに対し、今回は約 5.3:1 であった。つまり、今回の実験では狭窄の程度が少し弱かったことから、特にスライド部が口腔の前のほうに位置した際に母音が/i/にならず/e/になったものと考えられる。スライド部の直径を 18 mm ではなく、19 mm にした場合については今後、検討予定である。

一方、母音/u/についてはスライド部が外筒の中央付近にある場合、あるいは口唇側にある場合の2か所で観測された。文献[2]においても、S3Tモデルにおける母音/u/が生成される調音位置は、スライド部が声道中央付近であった。また、母音/u/は口唇の狭めも伴う（ただし必ずしも口唇が丸いとは限らないようである）。つまり、口唇側にスライド部がある場合も、母音/u/になることが報告されている[2]。本研究においても、その2か所で母音/u/が観測されたことから、その様子が忠実に再現されていることが分かる。ただし、/u/が観測されるか否かは、外筒やリードのサイズによって限定的であった。全般的な傾向として、外筒が長いほど、そしてリードが短いほど母音/u/が観測された。母音/u/の重要な調音位置は声道中央付近であることから、本研究においてスライド部が外筒の中央付近にある時のほうが母音/u/が出やすい事実と一致した。

スライド部の長さについては、40 mm と 50 mm にはそれほどの違いは観測されなかった。しかし、例えば子どもの場合は声道の長さが短く声帯のサイズも小さいので、Fig. 5 のそれぞれのブロックの左側（外筒が短い側）に該当する。その条件で 40 mm と 50 mm を比較すると、右ブロックの 40 mm のほうが、外筒中央付近で母音/u/が出る箇所が多く、より安定していることが分かる。外筒が 100 mm である場合には、スライド部は 50 mm よりも 40 mm のほうが3母音の安定した生成には適していると言える。

なお、今回使用した3つの音源にわたっ

て、Fig. 4 と Fig. 5 の結果が似ていることの原因を探るため、3つの音源の基本周波数を測定した。その結果、SS-R30 は約 130-135 Hz, SS-R25 は約 145-160 Hz, SS-R15 は約 290-370 Hz の範囲であった。SS-R25 の基本周波数が当初の狙いよりも少し低かったことから、今後は 200-250 Hz 程度のものも試してみたい。

## 5 おわりに

本研究では、異なるサイズのスライド式3音響管とリード式音源を用いて母音を生成し、母音認識の実験を行った。声帯を模擬したリードの長さを短くして基本周波数を上げると同時に、声道を模擬した音響管の長さを短くしても、良好な母音の質を維持することができた。一方、短いリードと長い声道では音の出方が不安定になった。また、長いリードと短い声道では母音/u/が出づらかった。今後は、さらにバリエーションを変えながら、母音の範囲の限界や音源とフィルタの干渉なども深く探りたい。

## 謝辞

内容の一部は、JSPS 科研費 21K02889 ならびに上智大学重点領域研究の助成を得た。

## 参考文献

- [1] Fant, G., *Acoustic Theory of Speech Production*, Mouton, 1960.
- [2] Arai, T., "Education in acoustics and speech science using vocal-tract models," *J. Acoust. Soc. Am.*, 131(3), Pt. 2, 2444–2454, 2012.
- [3] Arai, T., "Vocal-tract models and their applications in education for intuitive understanding of speech production," *Acoust. Sci. & Tech.*, 37(4), 148–156, 2016.
- [4] Stevens, K. N., *Acoustic Phonetics*, MIT Press, 1998.
- [5] Arai, T., "Vowels produced by sliding three-tube model with different lengths," *Proc. INTERSPEECH*, 2190–2193, 2012.
- [6] Arai, T., "Sound sources used speech production research with physical models of the human vocal tract," *Proc. Int'l Workshop on the History of Speech Communication Research*, 2019.