

# 音響教育における音声の可視化・可聴化\*

荒井隆行（上智大・理工）

## 1 はじめに

音声に関する音響学では従来から視聴覚に基づく研究・教育が行われてきたが、昨今の科学技術の発展によって、それがますます実現しやすい時代になった。中でもコンピュータによる可視化・可聴化の技術は、多角的に応用されるようになってきている。

例えば、音声波形やそのスペクトルを表示し、スペクトログラムをリアルタイムで確認することも可能であるし、声道形状を目で確認しながらその母音の韻質を耳で聴き、またそのフォルマント周波数から母音図上の位置を確認することも可能である。その際、その調音の様子を見るために、本来は見えない口腔内の舌の位置や動きを動画で、あるいはアニメーション化して示すことも可能である。

このように、コンピュータ上でマルチメディア化されることの恩恵は非常に大きい。その一方でバーチャルな世界で「音声」を扱うことが多くなり、実際の物理現象を見たり聴いたりすることを疎かにしがちになる、という弊害にも注意しなければならない。特に、実世界では何が起こり得て、何は起こり得ないのか、そこにはどのような物理的な制約があるのか等、実際の現象を積極的に意識することが必要になってくる。コンピュータ上のモデルに加え、実物を観察することは、「より

直感的」であると言える。つまり、「実物」と「コンピュータ・モデル」とをバランスよく共存させることが重要であると考えられる[1,2]。

そのような背景を鑑みながら、音声に関する音響教育における「音声の可視化・可聴化」についてまとめたものが Fig. 1 である[3]。以下では、それぞれについて詳細を述べる。

## 2 音声生成

### 2.1 声道模型による可視化と可聴化

声道の形状を物理的に音響管で実現した声道模型を用いると、音声を視覚的に捉えることができると同時に、それを音として確認することが可能となる。Arai (2001) [4]は、千葉・梶山[5]による声道形状の計測結果から作られる声道模型を復元した (Fig. 2)。この声道模型は、声道形状を単純化することで個人性等の情報は取り除かれている。

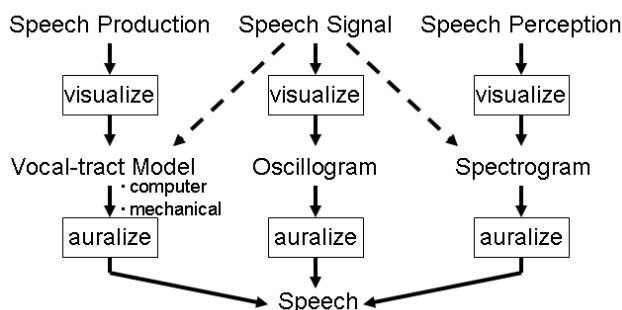
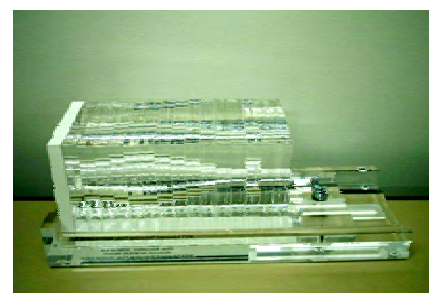


Fig. 1 Visualization and auralization of speech in education in acoustics (after [1])



(a)



(b)

Fig. 2 Arai's models of human vocal tract: (a) cylinder-type models (from /i/, /e/, /a/, /o/ and /u/); and (b) plate-type model

\* Visualization and auralization of speech in education in acoustics, by ARAI, Takayuki (Sophia University).

## Education System in Acoustics of Speech Production Using the Physical Models

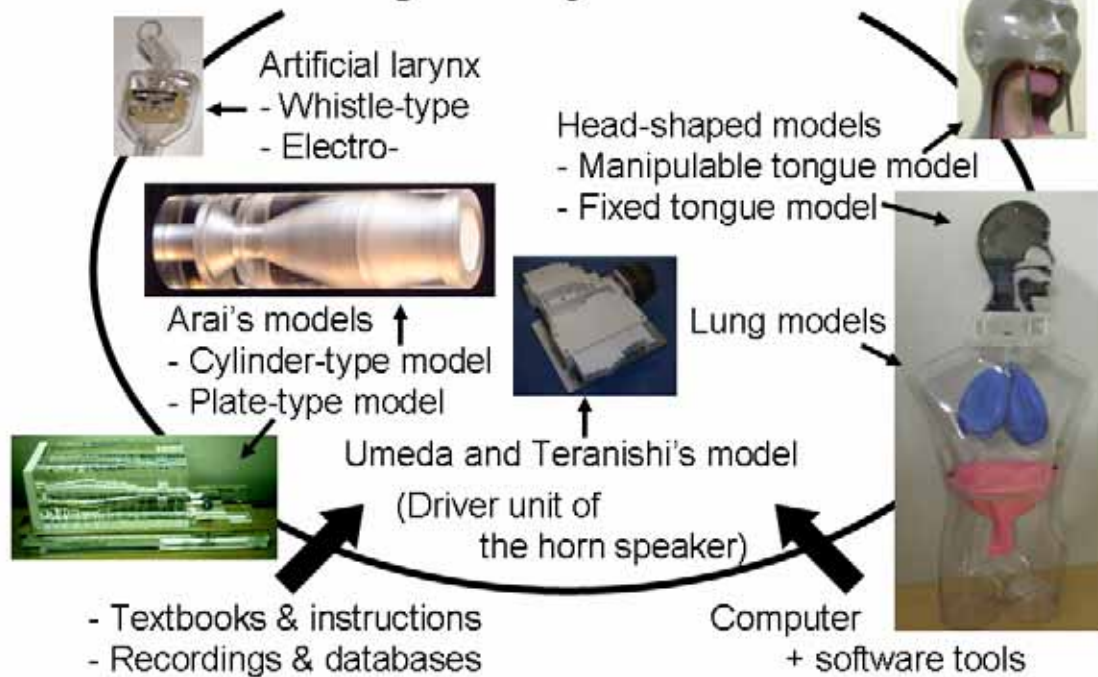


Fig. 3 Education system in acoustics of speech production using physical models of human vocal tract (from [6])

その後，Arai の声道模型は肺の模型を用いた呼吸から頭部形状模型による母音発声まで，音響教育システム[6]（Fig. 3）として発展している（その有効性については[6]を参照）。

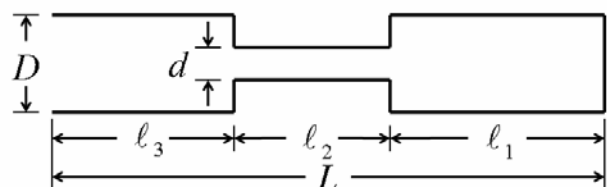
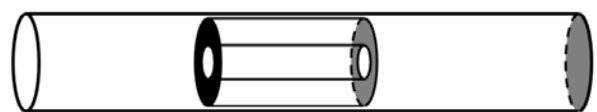
### 2.2 スライド式3音響管モデル

Fant [7]は3つの一様な音響管を縦続に接続した3音響管モデルを電気回路として構築し，声道共鳴のシミュレーションを行っている．Arai (2006) [8]はその3音響管モデルの物理模型を，1本の太い管とその中をスライドする狭窄用の細く短い管で実現している．この3音響管モデル（以下，スライド式3音響管モデル）の外観を Fig. 4 に示す．この図の (b) は外略図と断面図であるが，スライド部が左右にスライドすることによって，3つの音響管のうち両側の2つの音響管の長さが変わる．

図のように管の全長を  $L$ ，両側の太い管の断面積を  $A = \frac{\pi}{4} D^2$  ( $D$  は直径)，狭窄の断面積を  $A_2 = \frac{\pi}{4} d^2$  ( $d$  は直径)，狭窄の長さを  $\ell_2$



(a)



(b)

Fig. 4 Sliding three-tube (S3T) model with an electrolynx attached to the closed end (a) and a schematic view of its mid-sagittal cross-section (from [8])

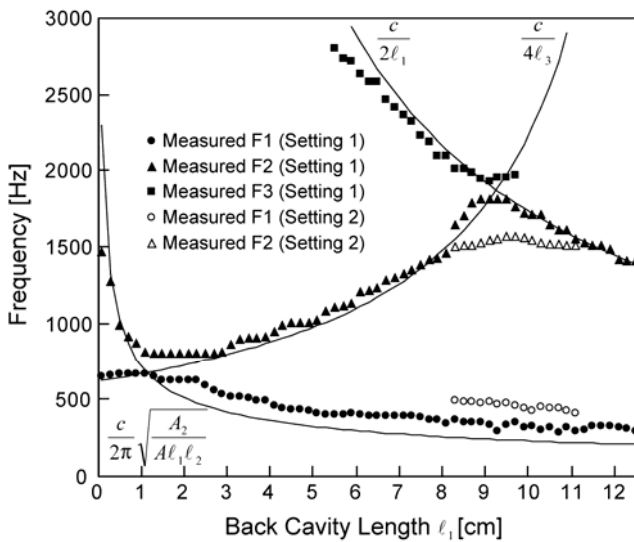


Fig. 5 Measured formants and underlying resonances produced by the S3T model (from [8])

とすると，両側の管の長さ  $l_1$  と  $l_3$  は，

$l_1 + l_2 + l_3 = L$  という条件の下で変化する．

$l_1$  の長さに対する共鳴周波数の変化を Fig. 5

に示す．実線は各曲線の横に示された近似式によって計算された理論値，点は実際に計測された実測値である．この図では  $D = 34$  mm,

$L = 175$  mm,  $l_2 = 50$  mm で，Setting 1 では

$d = 10$  mm, Setting 2 では  $d = 24$  mm である．

Fig. 5 から分かるように，適切な位置に適切な断面積の狭窄を置くと，所望の母音を作り出すことができる[8]．

### 2.3 その他

その他，例えば音響管内の共鳴現象は，「クントの実験」によって可視化することができる[9]．

## 3 音声信号の可視化

音声信号を時間波形として可視化する試みが音響教育でも行われている．マイクロフォンとオシロスコープを使えば，比較的容易に音声信号を目で確認できる．その他，電気を使わずに光を用いた単純な方法もある．例えば，振動膜に鏡を装着し，光源からのレーザー光をその鏡に照射し，反射光を遠くのスクリ

ーンに投影することによって，音声などによるわずかな振動を大きく映し出すというものである．

音声信号の周波数特性を見るには，スペクトル分析をすることになる．音響教育の現場では，リアルタイムでスペクトル表示させることも有効である[2]．また，音声信号に含まれる周波数成分の時間的変化を追うには（サウンド）スペクトログラムが一般的である．スペクトログラムを作るための装置はサウンドスペクトログラフと呼ばれるが，最近ではPC上で手軽に求めることができる．スペクトログラムもリアルタイムで表示することの教育的意義は大きい．

## 4 デジタル・パターン・プレイバック

音声信号を可視化したスペクトログラムは，パターン・プレイバックを用いることで再び音声に戻すことが可能となる．パターン・プレイバック（Pattern Playback）はスペクトログラムから音声信号を復元する機械として，Haskins 研究所のCooperらによって1940年代終わり頃に作られ[10]，その後の音声研究の飛躍的な発展に大きく寄与した[11-13]．特に，この機械で可聴化することによって，スペクトログラムに反映される情報の中でもどのような音響的キューが音声知覚に重要であるかを確認することができる．場合によってはその音響的キューを単純化して表現し，また系統的に変化させながら合成することも可能となる．これによって，閉鎖子音の知覚において後続する母音の第2フォルマントの軌跡が重要であることを説明するローカス理論など多くの実験が行われた[14]．

このパターン・プレイバックは，近年のデジタル処理により容易にデジタル化され得る[15]．Arai *et al.* [16,17]は，デジタル・パターン・プレイバックのアルゴリズムの可能性を再検討し，デジタル信号処理技術を用いて現代版パターン・プレイバックを実現した．第1のアルゴリズムは，デジタル的に発生させた正弦波による倍音構造を駆動源とし，それをスペクトログラムによってAM変調する手法（AM法）で，これはオリジナルの原理をそのままデジタル的に実現したもので，音源フィルタ理論的な考えに基づいている．第2のアルゴリズムは，FFT（高速

フーリエ変換)を用いる方法で, スペクトログラムの濃淡からフレームごとにスペクトル包絡を求め, それを IFFT することによってその時点での声道のインパルス応答が近似される. フレーム周期(シフト幅)をピッチ周期に合わせることによって, 容易に倍音構造を達成することができるようになる.

このデジタル・パターン・プレイバックはその教育的応用の価値は高く, 音声に関わる音響教育における効果が確認されている[17]. 例えば, あるスペクトログラムから個人性を含む余計な成分を削除し低次フォルマントや音声知覚に重要な子音のキューだけを残したような単純化したスペクトログラムからでも十分, 明瞭な音声を再合成することができる. このように, 同じカテゴリーの音に共通な特徴を絞り込むことによって, 音声信号に含まれる知覚に関わる特徴の確認に有効である.

なお, Arai *et al.* [16,17]によるデジタル・パターン・プレイバックのデモは, 英語版 Wikipedia からたどることができる[18].

## 5 おわりに

音声に関わる音響教育における可視化・可聴化について述べた. 今後はこれらを体系的に見ると同時に, 補うべき側面を強化する教材やプログラムの開発を検討したい.

## 謝辞

本研究の一部は日本学術振興会の科研費(A-2, 16203041 ならびに C-2, 17500603, 19500758)および文部科学省私立大学学術研究化推進事業上智大学オープン・リサーチ・センター「人間情報科学研究プロジェクト」から助成を得た.

## 参考文献

- [1] Arai, *Proc. of the Forum Acusticum*, Sevilla, 2002.
- [2] Arai, *Proc. of the International Congress of Phonetic Sciences*, 1, 305-308, 2003.
- [3] 荒井, 日本音声学会全国大会予稿集, 3, 2005.
- [4] Arai, *Journal of the Phonetic Society of Japan*, 5 (2), 31-38, 2001.
- [5] Chiba and Kajimaya, "The Vowel: Its Nature and Structure," Tokyo-Kaiseikan, Tokyo, 1942.
- [6] Arai, *Acoust. Sci. Tech.*, 28 (3), 190-201, 2007.
- [7] Fant, "Acoustic Theory of Speech Production," Mouton, The Hague, Netherlands, 1960.
- [8] Arai, *Acoust. Sci. Tech.*, 27 (6), 384-388, 2006.
- [9] 吉田他, 音講論(春), 399-400, 2002.
- [10] <http://www.haskins.yale.edu/haskins/MISC/PP/pp.html>
- [11] Cooper *et al.*, *PNAS*, 37, 318-325, 1951.
- [12] Cooper *et al.*, *J. Acoust. Soc. Am.*, 24, 597-606, 1952.
- [13] Borst, *J. Audio Eng. Soc.*, 4, 14-23, 1956.
- [14] 荒井他, 音声の音響分析, 海文堂, 1996.
- [15] Nye *et al.*, *Haskins Lab. Status Report on Speech Research*, SR-44, 95-107, 1975.
- [16] 荒井他, 音講論(秋), 429-430, 2005.
- [17] Arai *et al.*, *Acoust. Sci. Tech.*, 27 (6), 393-395, 2006.
- [18] 英語版 Wikipedia, "Pattern Playback," [http://en.wikipedia.org/wiki/Pattern\\_playback](http://en.wikipedia.org/wiki/Pattern_playback)