

音源とフィルタの相互作用を 声道模型を使って示す試み*

○荒井隆行 (上智大・理工)

1 はじめに

音源フィルタ理論は線形モデルを仮定している一方、実際の声帯振動は声道の影響を受けることが指摘されている(例えば, [1]). このような音源とフィルタの相互作用は、声道模型などを用いて物理現象として示すことも可能であり、教育的にも有用である[2].

音声の生成面における音源とフィルタの相互作用を狭義の意味で捉えたとしたら、広義の意味では知覚面における音源とフィルタの相互作用というものも存在する. これはつまり、どのような形であれ音源とフィルタが組み合わさって音声を実現されたとき、その音声を我々が知覚した際、音源の変化が知覚上においてフィルタの変化として知覚されたり、またその逆など、相互に干渉するという見方である.

そこで、本稿では生成面の議論を進展させ、相互作用を多角的に捉えてみた.

2 音声生成における相互作用

2.1 リード式音源と声道模型

Fig. 1 に、リード式音源と母音/u/の声道模型の組み合わせによって作られた母音に対するサウンドスペクトログラム[3]を示す.

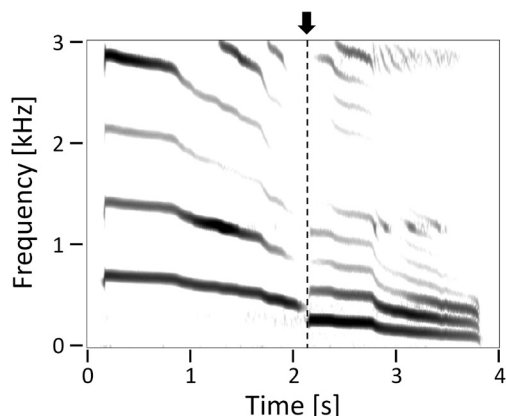


Fig. 1: Sound spectrogram of an output signal when a reed-type sound source is fed into a vocal tract model (from [3]).

この図を見ると、基本周波数が減少するのに従って、倍音構造も減少する様子がわかる. 一方、矢印が付与されているタイミングにおいて不安定な状態を通過しており、これはその時点で基本周波数が声道模型の第1フォルマントをまたいだことに起因するものと考えられる.

2.2 チューブ発声法

音声治療の臨床においてチューブ発声法[4]が行われることも多いが、この手法は音源とフィルタの相互作用を効果的に用いている. この場合、ストローを加えて声帯を振動させて、場合によっては声帯の振動数を上下させる. ストローの直径は10 mm前後のものからそれ以下のものが使用されるが、通常は発声中にストローの太さは変わらない.

そこで、チューブの太さを変えられるような可変式のチューブを製作した(Fig. 2). この図を見るとわかるように、チューブは断面が長方形をしており、上面が上下に動く可変式になっている. 最大位置では、断面が10 mm×10 mmになるように設計されており、太いストローから細いストローまでが模擬される.

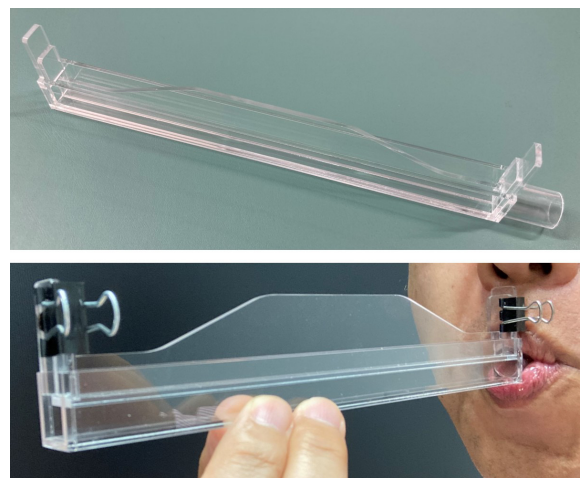


Fig. 2: Tube with variable cross-sectional area for the semi-occluded vocal tract exercises.

* Demonstrating interactions of the source with the vocal tract using vocal-tract models, by ARAI, Takayuki, (Sophia University).



Fig. 3: Head-shape model with a nasal cavity. The degree of nasalization can be changed by rotating the valve [7].

3 音声知覚における相互作用

3.1 開鼻声 vs. 嚙声

我々は以前より、開鼻声の聞こえが嚙声の有無によって変化する問題について取り組んできた[5]。例えば、鼻音化による極零対の周波数差を変化させることによって、フォルマント合成された合成音声を用いた実験例がある[6]。嚙声の程度は、OQ (open quotient) の値を変化させたり、高周波域における雑音成分の量を変化させることによって主に氣息性を変えた。音声合成は線形モデルに基づいているが、鼻咽腔結合の程度を変化させて開鼻声の聴覚印象を判定してもらった結果、嚙声を伴う場合に開鼻声の判定が低くなった。

Fig. 3 は頭部形状模型であり、この模型には鼻腔も備わっている[7]。そして、鼻咽腔結合の度合いはダイヤルを回すことによって弁の角度が変化し実現される。嚙声の有無を伴う音源を用いることによって、視覚的にも聴覚的にも現象を確認できるため、教育的にも有用である。

声道フィルタに関する音響的キューは、同時に音源に伴う音響的キューであることも多い。従って、音声を知覚することで、音源とフィルタを完全に分離して知覚しているわけではない。その結果、言語の長い歴史において、sound change が生じることにもつながっている[8]。

3.2 母音の質

以前、我々が報告した別の実験で、Fig.4 に示すような長さが異なるスライド式声道模型4種類に、4つの異なる基本周波数の音源を組み合わせ、16パターンによる出力音声について母音の知覚実験を行った[9]。その結果、同じ声道伝達関数であっても音源によって異なる母音として知覚されることが示された (Fig. 5)。この場合、音源の違いは基本周波数による点が主であったため、その高低を手がかりにした声道長に対する知覚上の正規化が行われ、その際に異なる母音に知覚されるケースも生じたものと解釈される。



Fig. 4: Four acoustic tubes with different length (from [9]).

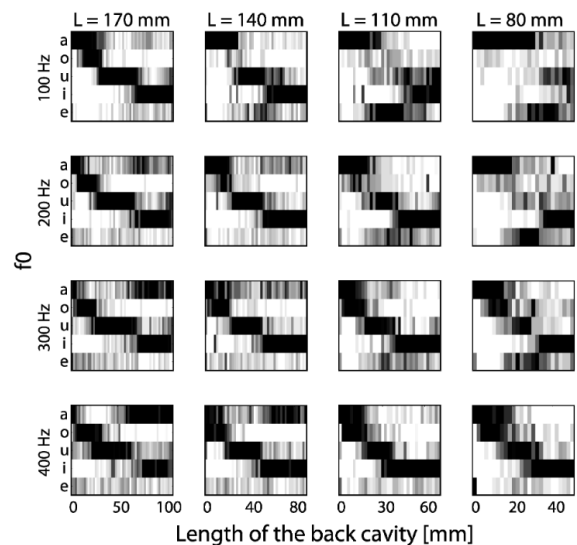


Fig. 5: Results of the vowel identification test with sliding three-tube models with four different lengths (L). Each row corresponds to the fundamental frequency. The darkness represents the identification rate (from [9]).

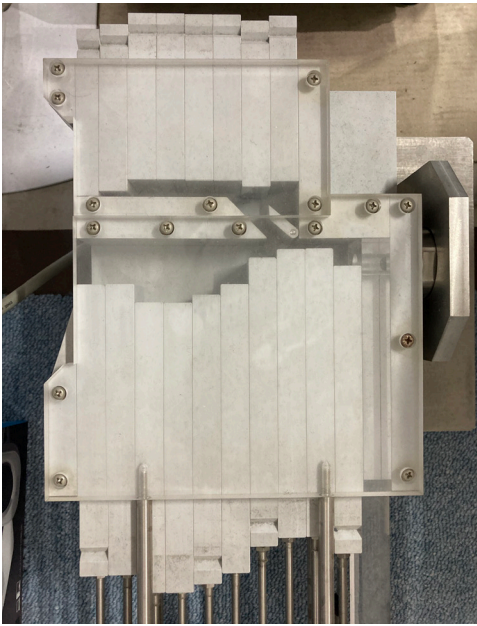


Fig. 5: PC-controlled version of the Umeda and Teranishi's model [10].

さらには、音源信号のスペクトル包絡に顕著な山や谷がある場合とない場合とでも、母音知覚への影響が変化する。そのように考えると、話者が意図する母音を聴取者が知覚できないケースも起こり得る。また、音源の基本周波数が高い場合、スペクトル上で倍音の間隔が広がり、フォルマントが捉えづらくなる。しかし、基本周波数が一定の場合に比べ変化する場合には、倍音周波数の変化によってフォルマントを捉えやすくなり母音の知覚も改善する。

4 音声のダイナミクス

我々の別の研究では、Fig. 5 に示すように梅田・寺西による声道模型に対し、複数のアクチュエータを接続して声道形状を時間と共に変化させるような PC 制御式のシステムを開発した[10]。そのシステムにおいて、音源信号と声道形状変化の制御とにおいて、いかに同期させるかが問題であった。人間にとっては、両者の時間的な同期制御は精巧に実現されているものの、機械式の場合にはそのずれが問題となることがしばしば起こり得る。

超分節的な韻律情報は音声知覚において句境界の判断や分節化に大きく貢献している。音源とフィルタのダイナミクスにおいても、相互作用は起きる。例えば、「ありがとう」というフレーズに対して、正しい

タイミングで LHLLL のピッチアクセントを伴って発話された場合の認識率に比べて、基本周波数が一定の場合、その認識率は下がる場合がある。特にそれは、雑音・残響環境であったり、発話が不明瞭な場合などがその例である。

5 まとめ

本稿では、音源とフィルタの相互作用を声道模型を中心に示す試みを紹介した。音声生成の面からは、リード式音源と声道模型の間で見られる相互作用に加え、チューブ発声法についても触れた。また、生成のみならず相互作用の定義を拡張し、広義の意味で音声知覚の側面からも考察した。声道模型を用いて示すことは、直感的でわかりやすく教育的な点からも有効であることから、著者は以前から授業などでも取り入れている。特に、音源フィルタ理論を教える際に、線形モデルが程よい近似となる場合に加えて、音源とフィルタの相互作用も同時に教えるよう心掛けている。その他の議論も含め、その導入例もまだ確立されていない面も多いため、今後はより深い議論につながるよう進めていきたい。

謝辞

チューブ発声法については、城本修先生のご指導をいただきました。内容の一部は、JSPS 科研費 21K02889 の助成を得ました。

参考文献

- [1] I. R. Titze and B. H. Story, "Acoustic interactions of the voice source with the lower vocal tract," *J. Acoust. Soc. Am.*, 101(4), 2234–2243 (1997).
- [2] T. Arai, "Education in basic acoustics for phonetics and speech science," *J. Acoust. Soc. Am.*, 152(5), 2746–2757 (2022).
- [3] T. Arai, "Vocal-tract models in phonetic teaching and research," in *The Routledge Handbook of Phonetics*, W. F. Katz and P. F. Assmann, Eds., Routledge, New York (2019).
- [4] 城本修, "症状対処の音声治療: チューブ発声法," ST のための音声障害診療マニュアル, 廣瀬肇監修, インテルナ出版 (2008).
- [5] 今富摂子, 荒井隆行, 加藤正子, "開鼻声の

- 聴覚判定における嘎声の影響: 音源フィルタ理論による検討,” 音声言語医学, 44, 304–314 (2003).
- [6] T. Arai, “Cue parsing between nasality and breathiness in speech perception,” *Acoust. Sci. Technol.*, 27(5), 298–301 (2006).
- [7] T. Arai, “Education system in acoustics of speech production using physical models of the human vocal tract”, *Acoust. Sci. Technol.*, 28(3), 190–201 (2007).
- [8] J. J. Ohala, “The phonological end justifies any means,” *Proc. of Int. Congr. Linguistics*, 232–243 (1982).
- [9] T. Arai, “Vowels produced by sliding three-tube model with different lengths,” *Proc. of INTERSPEECH*, 2190–2193 (2012).
- [10] T. Arai, “Mechanical vocal-tract models for speech dynamics,” *Proc. of INTERSPEECH*, 1025–1028 (2010).
-