

# 音響と調音のインタフェース\*

Kenneth N. Stevens (MIT)\*\*・訳：荒井隆行 (上智大学)\*\*\*

43.70.-h, Bk

## 1. はじめに

音声生成における基本的な音響過程は、「音源を作る」と「その音源を声道によってフィルタリングすること」である [1, 2]。通常、呼吸器系の活動により気流が生み出され、その気流が喉頭又は喉頭から口唇までの声道で作られる狭い「狭窄 (constriction)」を通過する際に変調 (modulation) を受けることによって音源が作られる。音源に対するフィルタリングは、声道の口腔・鼻腔間の開閉を含め、喉頭から口唇までの声道形状を変えることによって制御される。これら「音源」と「フィルタリング」の音響特性は、様々な調音パラメータを調節することにより変化し、その調音パラメータには喉頭の位置、口唇や舌端、舌体、その他の組織の位置や形状、特定の組織の硬さなどが関係する。こうした調節の中には、知覚上弁別的な、比較的安定した音響特性をもたらすものもある [3]。この特定の調節が、素性・分節・語といった離散的な単位から成る「音の目録」を言語が選択する際に、役割を果たしている [4, 5]。

本論文では、まず音源の様々な作られ方を説明し、その音源の声道によるフィルタリングを概説する。音源とフィルタに関するこの議論の一環として、音声言語で使用される特定の音の種類を、生成機構の観点と形成される音響パターンや知覚の観点から分類する。更に最終節で、音のこうした様々なクラスが連続音声の中でどのように実現され、配列されるのかを述べる。

## 2. 音声における音響源

音響源は言語内で生じる弁別的な音クラスの目

録を形成する役割を果たしており、四つの種類がある。喉頭又はその周辺で生成される音源が2種類と、喉頭から口唇までの声道における狭め(狭窄)周辺における気流の結果として生成される音源が2種類である [6, 7]。

喉頭周辺の2種類の音源は、(1) 声帯が振動する際に生成される気流のパルスが準周期的に連続する「発声 (phonation) 音源」と呼ばれるものと、(2) 声門又は声門の真上を通る気流の乱れによる「帯気 (aspiration) 雑音」と呼ばれるものである。後者は、準周期的音源すなわち発声音源との共存も可能である。

声門から口唇までの間に存在する狭窄よりも上流側で圧力の蓄積があると、気流における乱れや狭窄の下流側の障害物における乱流によって摩擦 (frication) 雑音と呼ばれる音源が生み出される。子音性の狭窄が数十 ms にわたって定位置に保たれると、連続的な音が作られる。また、気道が完全に閉鎖されてから解放されると乱流雑音の短い破裂が作られる。後者では、短い過渡 (transient) 音源が解放直後の1~2 ms の間に生成される場合がある。過渡音源は、いわゆる“吸着音 (click)”の解放の際にも作られる。この場合、過渡音源は、二つの調音器官(通常は舌端と舌体)の間の小さな室に閉じ込められた空気が陰圧になり、解放されるときに生ずる。

### 2.1 発声と帯気音源

声帯の左右の対が相対的に接近した位置にあり、肺から十分な圧力があると声帯が振動し、発声音源が生まれる。時間と共に変化する「声門を通る空気の体積速度」は、体積速度源としてモデル化される。通常発声 (modal phonation) に対する体積速度の波形は、図-1(a) のような形となる。この例では、声門を通る平均流量は約  $125 \text{ cm}^3/\text{s}$  で、1 サイクル内のピーク流量は  $600 \text{ cm}^3/\text{s}$  である。これらは、肺圧が約  $8 \text{ cm H}_2\text{O}$  のときの成人男性に

\* The acoustic/articulatory interface.

\*\* Kenneth N. Stevens (Massachusetts Institute of Technology, Cambridge, Mass., USA)

\*\*\* Translation: Takayuki Arai (Sophia University, Tokyo, 102-8554)

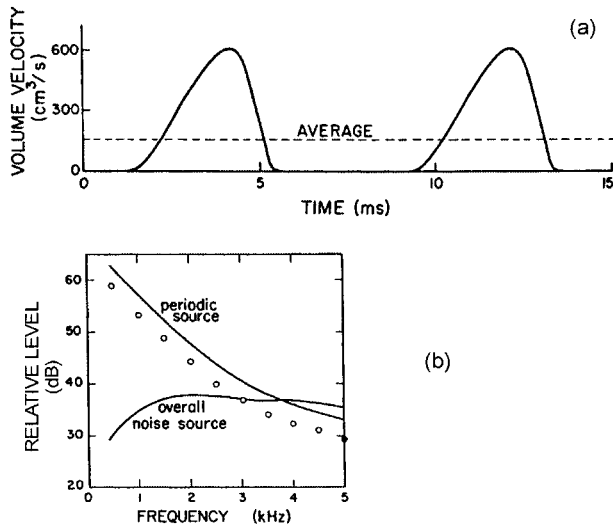


図-1 音声生成に関する声門音源の音響的側面

(a) 声門における発声音源の体積速度の概略(典型的な男性の声、横軸は時間)。(b) 発声音源のスペクトル包絡と声門における帯気雑音源のスペクトルの比較。白丸は基本周波数が 125 Hz の際の 4 倍音ごとの相対レベル(文献 [3] より)。

対する典型的な通常発声の値である。この波形で特に重要なのは、声帯が閉じる際にかかなり急激に気流が遮断されることである。閉鎖時に波形の傾きに生じる不連続は、声帯振動の 1 サイクル内で声道を励振させる主たる駆動源となっている [8]。図-1(b) の上側の曲線と白丸は、この図-1(a) の波形の 1 次微分に対するスペクトルであるが、高周波域において振幅が減少する様子が示されている。また、低周波域においては、スペクトルが平坦になる(この図で、第 4 倍音より低い周波域)。図-1(a) の基本波形に対して実際の波形では、異なる話者、あるいは異なる音声環境・韻律環境における発声に対して、多様な形が観察される。例えば、閉鎖がさほど急激でない場合(高周波振幅の低下につながる)、振動の 1 サイクル内で声帯が完全に閉鎖されない場合、また、各サイクルの波形が不規則である場合などである。

発声音源の基本周波数は、声帯の緊張つまり硬さを調節することで変えることができる。この声帯の緊張は、喉頭の内在筋と、胸骨及び頭や首といった喉頭以外の場所と喉頭とを結んでいる外来筋の双方により制御される。また、基本周波数は、声門下圧(subglottal pressure)にも依存する。通常の声生成では、基本周波数の範囲はたいてい 1~2 オクターブである。発声音源の振幅は、閉鎖時における波形(図-1(a))の時間微分の振幅によっ

てしばしば表現される。つまり、声門が閉じる際、波形の傾きに生じる不連続性の大きさに対応する。話者は、声門の開き具合を変えたり、声門下圧を高めたり弱めたりすることによってこの振幅を調整できる。

声帯は、声門上下圧(transglottal pressure)の差が特定の閾値を超えるときにのみ振動する [9]。振動を始めるには、通常、4~6 cm H<sub>2</sub>O の声門上下圧差が必要である。声帯がすでに振動している場合に発声を止めるためには、声門上下圧差を 2~4 cm H<sub>2</sub>O かそれ以下に低下させる必要がある。母音生成時の声門下圧は 7~10 cm H<sub>2</sub>O が一般的であり、結果として発声閾値(phonation threshold)より十分高くなっている。詳細は以下(2.2 節)で述べるが、障害音を生成する際は口腔内圧(intraoral pressure)が増加するので、声門上下圧差は減少する。従って、子音区間内では声帯振動が抑えられる。

声帯の緊張度も発声閾値に影響を与えるが、それも話者が制御することができる [10]。声帯の緊張度が比較的lowければ(すなわち緩んでいれば)、声帯振動の開始及び停止に対する声門上下圧差の閾値は、前述した範囲の下端にくると予測される。一方、比較的緊張度が高い場合、発声閾値は高くなり、前述の範囲の上端に位置することになる。

声帯の左右の対が互いに近接する場合や比較的離れて位置する場合にも、発声は抑制される。成人話者の場合、声門断面積を約 12 mm<sup>2</sup> にまで広げれば発声を止めるのに通常は十分である。声門の平均断面積は約 4 mm<sup>2</sup> で、これは広がった声門と閉じた声門の間になるが、このときに振幅が最大となるような通常発声となる。これは、図-1(a)に示した典型的な通常発声の条件にほぼ一致する。

声門が広がっている場合、声門を通過する気流の乱れによる雑音が帯気雑音源を作るが、そこには声帯振動による周期性はない。帯気雑音源は周波数に伴って上昇するような連続スペクトルを有し、その上昇は 2~4 kHz の周波数域で幅広いピークを迎える。図-1(b)では、この帯気雑音源のスペクトルと周期的な発声音源のスペクトル包絡を比較している。高周波域では両者のスペクトル振幅はほぼ同じである。帯気音中の気流は一般的に 1,000 cm<sup>3</sup>/s かそれ以上である。声門が部分的に広がっていると、氣息声の場合のように発声音源

と帯気音源の両方が共存し得る。

## 2.2 狭窄と過渡音源

子音のうち阻害音 (obstruents) と呼ばれるクラスは、気道に狭窄を作ることによって生成される。この狭窄はかなり狭いので、空気が肺から声道に送り込まれる際にその前後で圧力低下が生じ、狭窄から上流側の声道では大気圧に比べて圧力が上昇する。狭窄が完全な閉鎖を形成すると閉鎖子音となり、定常的な狭い開口部を形成すると摩擦子音となる。いずれにしても、狭窄を通過する急速な空気の流れがあり、この気流の乱れが摩擦雑音を作る。摩擦音源は、声道内の音圧源としてモデル化できる [2]。摩擦子音ではこの雑音が (時間的に) 連続となり、一方、閉鎖子音では閉鎖解放直後に短時間の雑音の破裂が生じる。乱流による雑音のほとんどは、気流が狭窄から下流側の障害物に衝突し、その付近で発生するのが一般的である。雑音源の振幅とスペクトルは狭窄の狭さと狭窄の前後の圧力差に依存している。摩擦雑音源の時間微分のスペクトルは、図-1(b) に示した帯気音源のスペクトルに類似するが、この音源のスペクトルの細かい形状は、気流と狭窄の狭さにある程度依存する。阻害音では、調音位置が異なると狭窄が異なる位置にできることになり、音源の場所も変わってくる。

摩擦子音では普通の発声時よりも気流を多く必要とするため、声門もより広く開く。声門がある面積で開いているとき、(声門から口唇までの間にある) 狭窄付近の摩擦雑音源の振幅は、狭窄の面積が声門の面積の約 1/2 になると最大になる [7]。この狭窄の大きさは調音器官によって調節されるが、その調節は調音器官内の筋肉の制御ばかりでなく、口腔内圧が上昇することで調音器官にかかる力によっても影響を受ける。舌端で作られる狭窄を伴う摩擦子音 (例えば、多くの言語に存在する子音/s/) の場合、下顎を持ち上げて発することが多い。そうすることで、下の前歯 (狭窄から下流側に 1 cm, もしくはもう少し先に位置する) に狭窄からの気流が衝突する。このような摩擦音は、粗擦音 (stridents) として分類される。この条件下では、音源の振幅が増すばかりでなく、狭窄から下流側にある音響腔をより強く励振する位置に音源が存在することになる。

閉鎖子音の解放の際、狭窄の断面積は時間と共

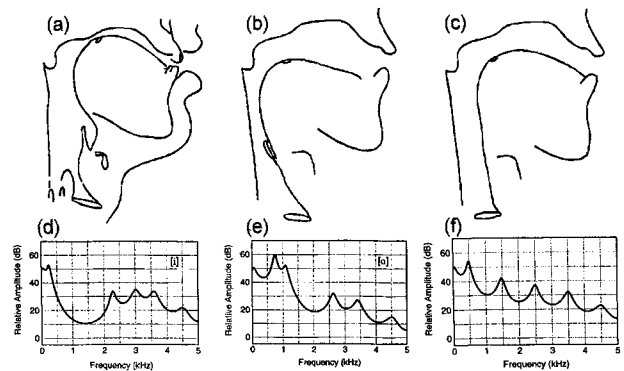


図-2 上段は3母音に対する声道の正中矢状断面を示す。(a) /i/, (b) /a/, (c) 断面積が声道の長さに沿って一様な母音。下段の (d), (e), (f) は各声道形状に対応する放射音のスペクトル包絡を示す。

に増加するが、口腔内圧や狭窄を通過する気流による力からも影響を受ける。その結果、断面積の増加は、短い時間いったん止まるか、あるいはゆるやかになり、その後再び急速に増加する。この初期の区間は、閉鎖子音を作る調音器官によるが、たいてい 5~30 ms である。この間、最初の短い過渡部に続いて、摩擦子音に類似した摩擦雑音の破裂が存在する。過渡音源は、閉鎖子音が解放される際に作られる体積速度のインパルスによってモデル化される。

## 3. 声道によるフィルタリング

### 3.1 声門での音源に対するフィルタリング:母音

はじめに母音の生成について考察する。母音では声門での発声音源が唯一の音源であり、声門から口唇までの声道に狭窄はない。ここでは更に、声道の断面積が最も狭くなる場所でも声道を通る気流を妨げない、あるいは発声音源の働きに対して大きく干渉しないと仮定する。二つの異なる母音/i/と/a/が生成されるときに声道の正中矢状形状を、図-2(a)と2(b)に示す。また、断面積が声道の長さに沿って一様と仮定した場合の形状も図-2(c)に示す。

声道による音源のフィルタリングは伝達関数によって表すことができ、この伝達関数は「音源の体積速度」に対する「口唇での体積速度」のスペクトルの比として定義される。この伝達関数は 0 Hz において 1 (つまり 0 dB) であるが、声道の固有周波数に対応するピーク位置の組み合わせによって特徴づけられる。これらの固有周波数がホルマント (formant) である。ホルマント周波数は声道形状

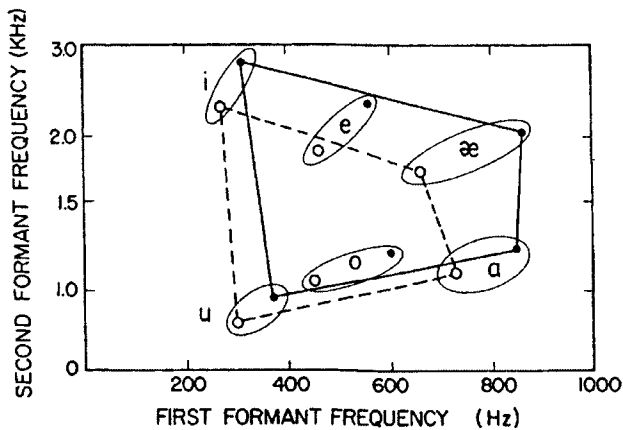


図-3 アメリカ英語の6母音に対する  $F_1$ - $F_2$  図  
白丸とそれを結ぶ破線は成人男性話者のデータ，黒丸（実線）は成人女性話者（文献 [7] より）。

に依存し，最初の二つのホルマントにその形状を特定するために必要な情報のほとんどがある。長さ  $l$  の声道に対して，ピークの平均間隔は  $c/2l$  である ( $c$  は音速)。(厳密に言えば，長さ  $l$  には声道の開口端における放射インピーダンスから生じる補正を含めなければならない。)従って， $l = 16$  cm (成人女性の典型的な声道長) に対して，3 kHz までの周波数域において通常三つのホルマントが存在する (体温での音速を  $c = 34,400$  cm/s とした場合)。図-2(a) と 2(b) に示した母音 /i/ と /a/ に対するスペクトルの例を，図-2(d) と 2(e) に示す。図-2(f) は，理想化された一様な声道 (図-2(c)) に対するスペクトルである。この場合，ホルマントは等間隔に配置され，スペクトル上でピークと谷の振幅差は，第1~3ホルマント ( $F_1 \sim F_3$ ) の周波数域で 10~20 dB の範囲にある。

$F_1 \sim F_3$  が取り得る周波数には，その範囲を制限する幾つかの基本的制約が存在する。このような制約は，共鳴器の音響理論や解剖学的に声道が取り得る形状範囲の制約に基づいている。英語の基本的な6母音に対して，図-3に  $F_1$  及び  $F_2$  の周波数範囲をグラフで示す。データは話者を男女別に示してある。一般的に， $F_1$ - $F_2$  平面上でこれらの母音を結ぶと，任意の言語の母音体系における  $F_1$  と  $F_2$  の限界値が線で示される。つまり，この線によって定義される4角形上かその内側に  $F_1$  あるいは  $F_2$  が位置することになる。

$F_1$ - $F_2$  図を見ると，言語に応じて母音分布を支配するさらなる傾向があることが分かる。その一つは，母音は  $F_1$ - $F_2$  平面において散らばる傾向に

あるというもので，これにより母音は知覚上かけ離れる [11]。もう一つは，個々の母音を生成するのに厳密な調音を要求してはならないというものである。母音の  $F_1$ - $F_2$  平面には，特定の  $F_1 \cdot F_2$  値を達成するのにいくらか不安定な領域があり，調音の微妙な変化が  $F_1$  や  $F_2$  の大きな変化につながることもある。

声門下共鳴の第2共鳴周波数は成人では 1,350~1,500 Hz の範囲にあり [12]，これが  $F_2$  の高い母音と  $F_2$  の低い母音の間に自然な境界線をもたらす。同様に，声門下共鳴の第1共鳴周波数である約 600 Hz の周波数域は， $F_1$  の高い母音とそれ以外の母音の間に自然な境界線をもたらす (この場合，母音クラスは素性 [+low] と [-low] で識別される [5])。多くの話者で，声道共鳴と声門下共鳴との間にある程度の音響結合が必ず起こる。このため， $F_2$  が 1,400 Hz 付近，又は  $F_1$  が 600 Hz 付近で母音を生成しようとする時，これらのホルマントが声門下共鳴に近くなり， $F_1$  や  $F_2$  に対応するスペクトルのピークの周波数がいくぶん不安定になることがある。これらの周波数が，[+low] と [-low] の母音間，そして [-back] と [+back] の母音間における境界線を定義しているように思われる。図-2(a) と 2(b) に正中矢状断面を示した二つの母音の場合，母音 /a/ は [+low, +back] に，/i/ は [-low, -back] となろう。

図-2(a) 及び 2(b) の2母音に対して，そして実際には図-3に示した  $F_1$ - $F_2$  の周囲に位置する6母音すべてに対し，舌体の主な部分の概略的な形状がほぼ同じであることが観察されている。この舌体の形状を前後方向と高低方向に操作することにより，異なる母音が生成される。これらのうち2母音 (/o/ と /u/) の場合，全体的な声道形状に加えて円唇化という補足的な調節が行われる。その結果， $F_1$  と  $F_2$  が下方へシフトし，母音の位置が母音図の周囲に向かって移動する。

言語に存在し得る母音の目録を拡張するために，言語は様々な方法を用いて比較的固定した舌体形状に加え，ある種の「修正」を導入する (例えば上記の2母音の場合は円唇化)。以下に，こうした四つの「修正」を述べる。

- (1) 「円唇化」の素性を追加する (前舌母音や /a/ に対して「円唇化」を追加，あるいは /o/ と /u/ に対して「円唇化」を除去)。このように円

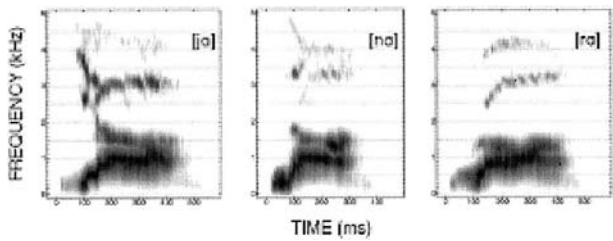


図-4 共鳴音の音響特性を示す三つの CV 音節 /ja/, /na/, /ra/ のスペクトログラム (女性話者)。

唇化の度合いを変えると、図-3 の母音空間の内側に新たな母音の集合が作られる。

- (2) 舌根付近の声道断面積が小さくなるように、舌根を操作する。この調節は、一般的に非低母音に対してだけ行われ、結果として  $F_1$  が上昇する。
- (3) 舌尖よりも後ろの声道口腔部で舌の形状を操作し、この部分での声道断面積を広くする。この調節の結果、音響的には主に  $F_3$  が修正される。
- (4) 鼻咽腔結合を実現する。この調節の主な音響関連量は、 $F_1$  の周波数域における母音スペクトルの平坦化である。

母音目録の拡張を目的としたこれらの基本操作はすべて、その他の調音器官による調節を取り入れることで微調整され得る。これは、修正された母音をその他の隣接する母音に対して知覚的に目立たせるためである。

### 3.2 声門での音源に対するフィルタリング：非母音性分節音

音源が声門にあっても、声門から口唇までの声道形状が標準的な母音の形状（大きな狭窄を持たず気道が一つ）とは逸脱している音、又は音のクラスも幾つか存在する。この種の音は通常、母音に隣接し、低・中周波域における振幅が母音よりも小さい。3種類の共鳴音 (sonorants) の例として、音節 /ja/, /na/, /ra/ のスペクトログラムを図-4 に示す。

共鳴音の一つのクラスであるわたり音 (/w/ や /j/) は、多くの言語に存在する。/w/ や /j/ の音は高母音の /u/ や /i/ と似た声道形状で生成される。ただし、口腔内の狭窄が母音の場合より狭くなっている。音源の振幅は小さくなるが、これはこの狭窄が声門音源に影響するからのものである。

また、共鳴音の別のクラスに、流音 (英語の /l/

と /r/) がある。この音に対しては、声門から口唇までの気道に分岐管が存在する (分岐管は舌の端付近か裏側)。母音に似たスペクトルを持つが、分岐管があるため (母音には分岐管がない) いくぶん複雑なスペクトルになり、小さなピークが加わったり、 $F_3$  の周波数や振幅が母音に比べ変化を受けたりする。

共鳴音のクラスで更に一般的なものは鼻子音である。鼻子音は、鼻咽腔結合と同時に、口腔が完全に閉鎖されることによって生成される。口腔閉鎖は、口唇、舌端又は舌体で作られる。主たる音の経路は、声門から咽頭部を通り鼻腔に抜ける。この主経路に対し、鼻咽腔結合部から口腔閉鎖までの経路が分岐管を形成する。分岐管の長さは、鼻子音の調音位置により異なる。閉鎖区間において鼻子音の特徴を示す音響的特性は、低周波域 (250 Hz 付近) におけるスペクトルの山と、鼻咽腔結合時に鼻腔共鳴を表す 900~1,000 Hz 付近の山である [13, 14]。鼻子音と母音の境界では、 $F_2$  付近 (子音の調音位置に依存) でスペクトル振幅が突然変化する。この不連続性は、口腔が突如分岐管になり、口唇開口部から音が放射されるようになることで生じる。鼻子音による鼻咽腔結合は通常、隣接する前後の母音にも及ぶため、母音が鼻音化される。

共鳴音として分類される他の音に /h/ もある。この音は、比較的開いた声道と、帯気雑音源を生成するほどに拡大された声門によって作られ、発声音源との組み合わせも可能である。他の共鳴音と同様、声門から口唇までの間で圧力の上昇はほとんど (もしくは全く) ない。

### 3.3 摩擦音源と過渡音源に対するフィルタリング

摩擦音源は、声道の狭窄を通過する急速な気流の乱れにより生成される。この気流の乱れは、狭窄から上流側の声道で圧力が蓄積されて上昇すると起こる。こうした口腔内圧の上昇によって生成される音は、阻害音 (obstruents)、更に厳密には素性 [-sonorant] を持つ子音と呼ばれる。

阻害音は、狭窄の程度に応じて次の3種類に分類される。狭い狭窄と連続的な摩擦雑音を伴う摩擦子音、完全な閉鎖の後に急激な解放を伴う閉鎖子音、ある調音器官による完全な閉鎖の後に急激な解放を伴うが、この調音器官のすぐ後方で摩擦音に似た狭い狭窄が維持される破擦音である。素性に関

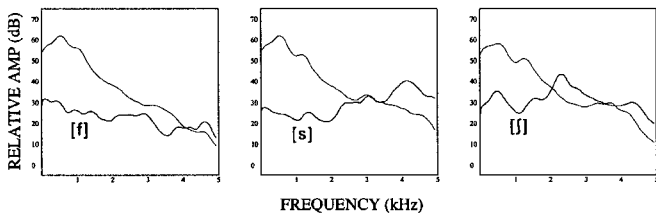


図-5 音節 /fa/, /sa/, /ʃa/ における摩擦子音のスペクトルの例  
隣接母音のスペクトルが重ね描きされている (26 ms の窓を掛けて計算し、スムージングして求めたスペクトル。男性話者)。

しては、これらの子音はそれぞれ [+continuant], [-continuant], [ $\pm$ continuant] として分類される。

阻害音の形態として、雑音源が狭窄内に位置するものと、狭窄からすぐ下流側の出口に位置するものがある。下の前歯のように狭窄から下流側に障害物がある場合、狭窄から発生する気流がこの障害物に衝突するときに乱流雑音が生まれる。いずれにせよ、この雑音源は音圧源として表され、この音源は狭窄付近か狭窄からいくらか下流側に分布する [7]。

一般に摩擦音源は、子音性の狭窄より下流側の声道体積の共鳴 (一つ又は複数) を励振する。狭窄が口唇にあると、本質的に狭窄の前方には腔がなく、音響源はフィルタリングされずに口唇開口部から直接放射される。狭窄が口唇の後方に位置するとき、狭窄と口唇の間に気道ができる。その長さは成人話者で、舌端によって狭窄が作られるときは約 1~3 cm、硬口蓋や軟口蓋に対して舌体を挙上することによって作られる場合は 3~4 cm から約 6 cm までと変化する。前腔の長さ、口唇の丸めの程度によっていくらか影響を受ける。放射特性による長さの補正を考慮し、一様音響管の第 1 固有周波数を求める公式  $c/4l$  を用いると ( $l$  は前腔の長さ)、前腔共鳴は、舌端による子音に対しては  $F_4 \sim F_6$  が、軟口蓋子音に対しては  $F_2$  と  $F_3$  が関連していることが分かる [15]。これらのクラスの阻害音によって生成される雑音のスペクトルでは、こうした周波数範囲にピークがくると予測される。しかし口唇音に対しては、そのような高周波ピークは存在しない。

図-5 は英語の音節 /fa/, /sa/, /ʃa/ の摩擦子音のスペクトルであり、各々の摩擦音のスペクトルの上に隣接母音のスペクトルが重ね描きされている。これらは上記の特性を幾つか示している。口

唇音のスペクトルには、高周波域に大きな山は見られない。/s/ の場合には 4,100 Hz、/ʃ/ の場合には 2,300 Hz に大きな山が見られ、おのおの  $F_5$  と  $F_3$  に対応している。これらの子音のピーク振幅は、同じ周波数付近で対応している隣接母音のピークよりも伸びている。これら二つの摩擦子音のスペクトルでは、 $F_2$  域に小さな山が見られるが、これは摩擦雑音源が狭窄より後方の腔をわずかながら励振していることを示している。

図-5 の /s/ や /ʃ/ のように舌端で作られる幾つかの摩擦子音の場合、舌によって狭窄が形成されるが、顎を挙上することで下の前歯によって狭窄から 1~2 cm 下流側に障害物が作られる。乱流雑音はこの気流中の障害物において作られる。音響理論によると、音圧源が前腔の狭窄側付近にあるよりも開口端付近にあったほうが前腔共鳴はより強く励振される。このような大きい振幅を伴う摩擦子音は粗擦子音と呼ばれ、図-5 に示した /fa/ の例に見られるように、下流側に障害物がない状態で生成された雑音の弱い非粗擦子音と対立する。

閉鎖子音の生成にも、口唇・舌端・舌体によって形成される狭窄を通る急速な気流、つまり乱流による雑音の生成が関係している。閉鎖が解放される間に起こる摩擦雑音の持続時間は、子音を形成する調音器官の表面の特性 (硬さなど) にも依存し、一般的には約 5 ms (両唇閉鎖音) から約 20 ms (軟口蓋閉鎖音) の範囲である (舌端による閉鎖音ではその中間)。雑音の破裂開始時には、短い過渡部もよく観察される。雑音の破裂のスペクトル特性は、調音位置が同じ摩擦音のスペクトル特性と類似している。破擦子音の場合は、解放時の破裂の直後に連続的な摩擦雑音が続く。

この節で論じた種々の阻害音では、口腔内圧が上昇している間に声帯振動を伴うこともあれば伴わないこともある。多くの言語において、無声子音に対する声帯振動の抑制は、声帯を緊張させることによって実現されることが多い。場合によっては、声門を広げたり口腔内圧を操作することによっても実現される。このような調音動作 (gesture) は、発声閾値を下回るような声門の状態を作り出すために行われる。声帯振動を維持する際には、声帯を緩めたりその他の調整をすることでその維持を達成する。声帯の緊張と弛緩が、有声素性を定義する主たる調音特性であることが指摘

されている [10]。

閉鎖子音の摩擦雑音に含まれる音響特性は、聞き手に対し閉鎖子音の調音位置の手がかりを与えるが、それに加えて閉鎖子音に隣接する母音のホルマント遷移も、同様にさらなる手がかりとなる。特に  $F_2$  が最も強い手がかりである。このホルマント遷移は鼻子音に対しても調音位置の手がかりを与え、同様のことが摩擦子音や破擦子音についてもある程度あてはまる。

狭窄を形作る主だった調音器官 (例えば、口唇、舌端、舌体) に基づいて子音を分類するために、様々な弁別素性 (distinctive feature) が提案されている。そうした弁別素性はどのように調音器官を操作して形作られるのか (例えば、短い/長い狭窄) を示しており、素性ごとにそれぞれ、「基本となる調音器官や調音形状」と、対応する「基本となる音響特性」がある。

#### 4. 連続音声における音声学：多様性の源

前節までにおいて、発話を作る語は分節ユニットの連続から構成され、各ユニットは二項弁別素性の束 (bundle) から成っていることを見てきた。これらの素性の幾つかは特定の調音器官の操作を規定しない音響源の特性を示しており (例えば [+continuant])、これらは調音器官独立素性 (articulator-free feature) と呼ばれる。その他は、どの調音器官をどのように操作して分節を作るかを特定する素性で、これらは調音器官依存素性 (articulator-bound feature) と呼ばれる。1章と2章 (及び文献 [3]) では、「調音パラメータ」と「知覚に関わる音響特性」の関係が量子的な (quantal) 性質を示す例を見た。つまり、調音パラメータがある範囲内にあるとき、調音の変化に対して音響特性が比較的鈍感になる一方、その範囲外では調音の小さな変化に対して音響特性が敏感になる。こうした調音と音響の量子的な関係は、音響源の開始に関する閾値や、声道の境界による調音位置の端点、又は共鳴腔同士が音響的に弱く結合することの影響などから生じる。

「言語において音声対立を実現するのに使われる弁別素性について、(すべてではないものの) その多くはこういった調音/音響の関係によって定義される」という仮説が立てられている。こうした弁別素性は 20~30 種類あり、ある言語はこれら

の素性の一部を使用してその言語における対立を規定する。つまり、心的辞書にある語が「素性の束 (bundles of features) の並び」として記述されると仮定すると、一つの素性が取り得る 2 値の変化がその「並び」を変化させ新しい語を生む可能性を持つ。例えば、語 fan と pan の最初の分節は素性 [continuant] において異なるが、その他の素性は全分節にわたり同じである。

話者が発話を産出するとき、最初の段階として発話に対する計画が脳内でまず組み立てられると仮定される。この計画段階の一つの要素に「生成される語の並び」があり、各語は「素性の束の並び」として脳内に格納される。各素性は特定の「基本動作 (defining gesture)」を規定し、その基本動作によって声道の状態や動きに対して幾つかの側面が順に制御される。複数の声道の動きは素性の束の並びに沿って組み合わせられ、声道から放射される音響信号の様々な側面を生む。

しかし多くの場合、計画段階で素性によって定義される基本動作の組に対して、幾つかの補足的な調音指令が追加される。これらの補足的な調音指令には、句境界や発話の時間、リズム上の側面など、計画段階で規定される韻律情報が含まれる。また、特定の素性の知覚的特徴を強調するために導入される補足的な動作も含まれる。こうした強調動作は言語に依存し、また韻律環境や隣接素性の環境にも依存する [16, 17]。

英語における強調動作の例を幾つかあげる。(1) 素性 [-voice] の知覚的特徴を強調するため、強勢直前の無声閉鎖子音生成時に帯気を導入する (例えば、/ta/対/da/)、(2) 歯茎閉鎖子音 (舌端で狭めを作る) に対して調音位置の知覚的特徴を強調するため、歯茎閉鎖子音生成時に舌体を前方化する (例えば、/da/対/ba/)、(3) 歯茎摩擦音に対して知覚的対立を強調するため、素性 [+anterior] によって定義される口蓋摩擦子音の円唇化 (例えば、/ja/対/sa/)、などである。最初の二つのケースでは、強調動作によって、素性の基本的な音響特性のほかに、音響の手がかりに寄与する音響特性が加えられる。三つ目の円唇化の例は、素性 [-anterior] の基本的音響特性の知覚的特徴を強化する効果がある。その他、多くの例が様々な言語から引用できる。韻律的影響と強調動作が導入される結果、特定の素性の音響表現において多様性が生まれる



が、それと同時に、特に雑音や不明瞭な発話といった条件下では、素性の音響表現に冗長性が導入される。

連続音声中の素性の音響表現における多様性の原因としてもう一つ重要なのは、隣接する分節中での素性の調音動作のオーバーラップである。このオーバーラップは、話速と発話の明瞭性に依存する。簡単な例として、(1) “top tag” という並びにおける口唇開放や歯茎閉鎖の音響表現の消失、(2) “can’t go” という並びにおいて歯茎閉鎖の音響的証拠がないこと、(3) 発話の最後の無声閉鎖子音における素性 [-voice] に対する直接的な証拠の欠如がある。

多様性のこうした原因（弁別素性に対して補足的な手がかりを生む強調動作の導入や、動作のオーバーラップにより起こり得る音響的手がかりの消失など）によって、連続音声中の語をより確実に識別できるようなモデルの開発は複雑化している。しかし、そうした原因が、多様性に関して理にかなった説明を与えているのである。

#### 謝 辞

翻訳にあたり、訳を一部確認して下さった米国 University of Massachusetts の新谷敬人さんに感謝申し上げます。

#### 文 献

- [1] T. Chiba and M. Kajiyama, *The Vowel: Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo, 1941).
- [2] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).
- [3] K.N. Stevens, “On the quantal nature of speech,” *J. Phonet.*, 17, 3–46 (1989).
- [4] R. Jakobson, C.G.M. Fant and M. Halle, “Preliminaries to speech analysis: The distinctive features and their correlates,” *Acoustics Lab. Tech. Rep. 13*, Massachusetts Institute of Technology, Cambridge, Mass. (1952) (reprinted by MIT Press, Cambridge, Mass., 1967).
- [5] N. Chomsky and M. Halle, *The Sound Pattern of English* (Harper & Row, New York, 1968).
- [6] P. Ladefoged and I. Maddieson, *The Sounds of the World’s Languages* (Blackwell Publishers, Cambridge, Mass., 1996).
- [7] K.N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, Mass., 1998).
- [8] G. Fant, J. Liljencrants and Q.G. Lin, “A four-parameter model of glottal flow,” *STL Q. Progr. Stat. Rep.*, 4, 1–13, Royal Institute of Technology, Stockholm (1985).
- [9] H. Hirose, S. Niimi, K. Honda and M. Sawashima, “The relationship between glottal and the transglottal pressure difference during consonant production,” *Res. Inst. Logop. Phoniater. Annu. Bull.*, 19, University of Tokyo, 55–64 (1985).
- [10] M. Halle and K.N. Stevens, “A note on laryngeal features,” *MIT RLE Q. Progr. Rep.*, 101, 198–213 (1971).
- [11] J. Liljencrants and B. Lindblom, “Numerical simulation of vowel quality systems: The role of perceptual contrast,” *Language*, 48, 839–862 (1972).
- [12] B. Cranen and L. Boves, “On subglottal formant analysis,” *J. Acoust. Soc. Am.*, 81, 734–746 (1987).
- [13] S. Hattori, K. Yamamoto and O. Fujimura, “Nasalization of vowels in relation to nasals,” *J. Acoust. Soc. Am.*, 30, 267–274 (1958).
- [14] M. Chen, “Acoustic correlates of English and French nasalized vowels,” *J. Acoust. Soc. Am.*, 102, 2360–2370 (1997).
- [15] K.N. Stevens, “Acoustic and perceptual evidence for universal phonological features,” in *Proc. 15th Int. Congr. Phonetic Sciences*, Barcelona, Spain, pp. 33–38 (2003).
- [16] S.J. Keyser and K.N. Stevens, “Enhancement revisited,” in *Ken Hale: A Life in Language*, M. Kenstowicz, Ed. (MIT Press, Cambridge, Mass., 2001), pp. 271–291.
- [17] K.N. Stevens, “Features in speech perception and lexical access,” in *The Handbook of Speech Perception*, D. Pisoni and R. Remez, Eds. (Blackwell Publishers, Cambridge, Mass., 2005), pp. 125–155.