第2章　音声

人間はその言語能力によって文化を築き，社会，芸術，科学技術を発展させてきました．このことにより，人間は他の生物とは一線を画する存在となっています．犬，鳥，鯨なども声を使ってコミュニケーションしますが，意味の区別がある声の種類は 5〜30 程度だといわれています．また，人間のように声を組み合わせて新しい意味を作ることはできないといわれています．

人間の音声言語は，音素が組み合わされて形態素（語）を構成し，形態素（語）が組み合わされて文が構成されるという二重構造になっています（図 2.1）．このことによって，比較的少数の音素[1]を任意の個数並べることにより数千〜十数万の語を構成することができ，それらの語を任意の個数並べることによって，事実上無限の文の生成が可能です．これを**二重文節**といい，他の生物には無い人間の言語だけの特徴だといわれています．

図 2.1: 言語の二重分節．文は形態素（語）に分解され，形態素（語）は音素に分解される．

19 世紀以降の科学技術の発展により，言語の実体としての音声の研究 [1, 2, 3] は急速に進展してきました．それは，電気通信技術の発展と，その影響による音響機器，電子回路，放射線機器，およびコンピュータの発達の恩恵を受け，それらのツールを用いて音声信号を分析したり合成することが可能となったことによります．

音声研究の歴史において，研究対象は上で述べた二重文節の小さな単位から大きな単位へと発展してきました．音素の性質は発声器官の形とその運動によって決まるので，X 線撮影で撮影した発声器官の形を測定する基礎研究が古くから行われていました．発声器官の形やその変化の仕方が分かれば，音声の周波数特性やその変化の仕方を計算することができます．この章では，発声器官について説

[1]最も少ない言語で 11，最も多い言語で 141，多くの言語は 20 から 40 種類 [10].

明し，音声生成の音響工学的な解釈，日本語の主な母音，子音の性質について解説します．

2.1　音声生成のしくみ

我々が話す時には，肺からの空気を制御して少しずつ吐き出し，喉頭と声道（喉，口，歯，口唇，一部の音では鼻を併用）を使って発声します (図 2.2)．声道にさまざまな狭めを作ることによって，さまざまな音色の音声が生成されます．最も重要な音源は喉頭（喉仏）であり，ここには声帯があります．適切な張力が

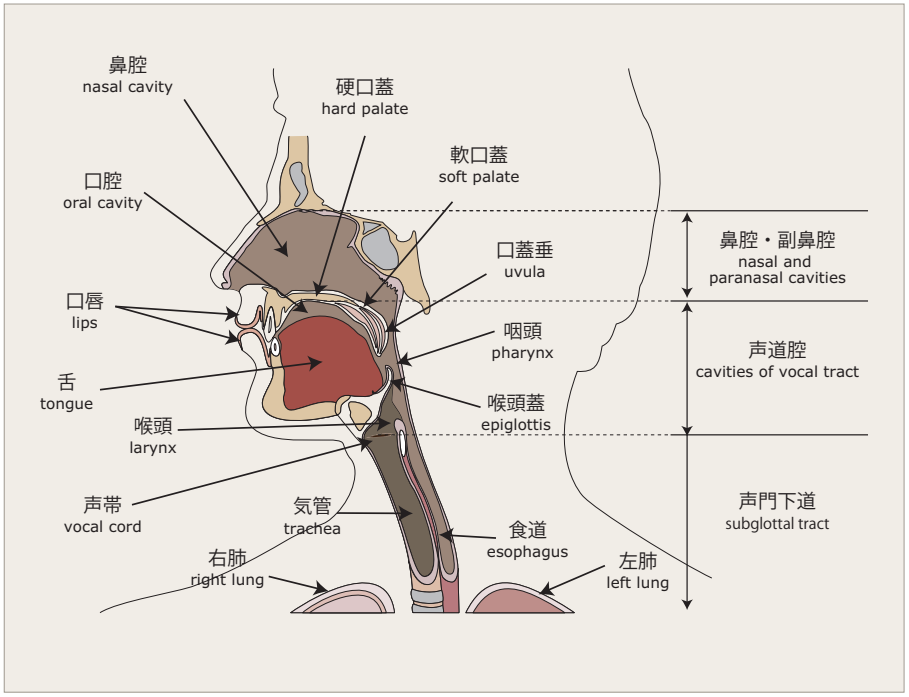

図 2.2: 発声器官概略図（頭部正中矢状断面）．音声器官は肺，喉頭，咽頭，口，鼻からなります．声道は咽頭，口からなり，その形は母音の種類によって異なります．声道の形は口唇，顎，舌，喉頭の位置や形によって変化します．

掛かった声帯の間を呼気が通ると，声帯は細かく振動し，断続的な空気流を声道に送り込みます (図 2.3)．この声帯音源の響き方は声道の形によって変わります．声道は，ある周波数の振動を強め，一方で別の周波数の振動を弱めるようなフィルタとして作用します．全ての音声の音が声帯音源を含んでいる訳ではありません．声帯音源を含んでいる音声を有声音，そうでない音声を無声音といいます．

2.1.1　音源—フィルタ理論

音声の生成において，肺は動力源として，振動する声帯は発振器として，声道は共鳴器として働きます．音源—フィルタ理論によれば，音源は肺からの空気流によって振動する声帯によって発生する声帯音源です．音源の音響スペクトルは周波数が高くなるにしたがって振幅が減少します（S）．フィルタに相当するのは声道であり，音源の周波数特性に変化を与えます．その結果フィルタ関数が得ら

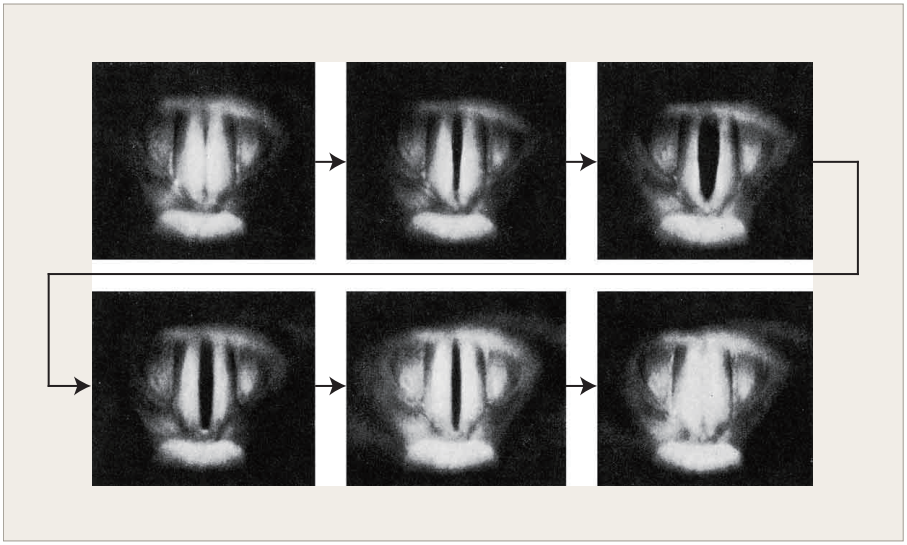

図 2.3: 声帯の振動の様子（1 周期，約 8 ms)[3]

れます（F）．音源にフィルタが適用されると，母音の出力スペクトル（O）に共振ピークが生じます（図 2.4）．

2.1.2　声道モデル・母音の生成

音声を生成しているときの発声器官の形を正確に知ることは，音声の工学的研究が始まった当初から重要な問題でした．発声器官の形状が分かれば，音響理論によって共鳴の特性すなわち声道フィルタの特性を計算することができるからです．X 線を用いて撮影したり，舌や口蓋に電極を貼付けたりして声道の形を測定することが行われてきましたが，最近は磁気共鳴画像診断装置（Magnetic Resonance Imaging: MRI）が用いられています．図 2.5 は話者が MRI に仰向けの姿勢で入り，日本語の 5 母音を発声した時の正中矢状面の画像です[2]．母音によって，口の開き具合と舌の形（位置）に特徴があるのがわかります．例えば，/a/は/i/に比べて口を大きく開き，舌を奥のやや低い位置に構えています．舌が最も盛り上がっている（したがって声道が最も狭くなっている）位置が口腔の前側か後側かに注目すると，前側であるのが/i/と/e/，後側であるのが/a/，/u/，/o/です．前者は**前舌母音（front vowel）**，後者は**後舌母音（back vowel）**と呼ばれています．

X 線写真や MRI 画像から得られた発声器官の形状から声道断面積を推定し，それを用いて声道のフィルタ特性を計算することができます．英語の母音，”feet” の /i/，”father” の/a/， “boot” の/u/の声道の概形が図 2.6 の左側に描かれています．図の中央は，声道を少数の円筒を接続したモデルとして表したモデル，右側は対応する音響スペクトルです．聴覚実験により，人間は音響スペクトルの概形（特に共振周波数）によって，母音の種類を識別していることが分かってい

[2]**転載厳禁**　この MRI データは，ATR 人間情報科学研究所が独立行政法人情報通信研究機構からの研究委託「人間情報コミュニケーションの研究開発」に基づいて収録し，公表した『ATR 母音発話 MRI データ』の一部です．本データの使用および成果の発表は，株式会社 ATR-Promotions との使用許諾契約に基づいております．

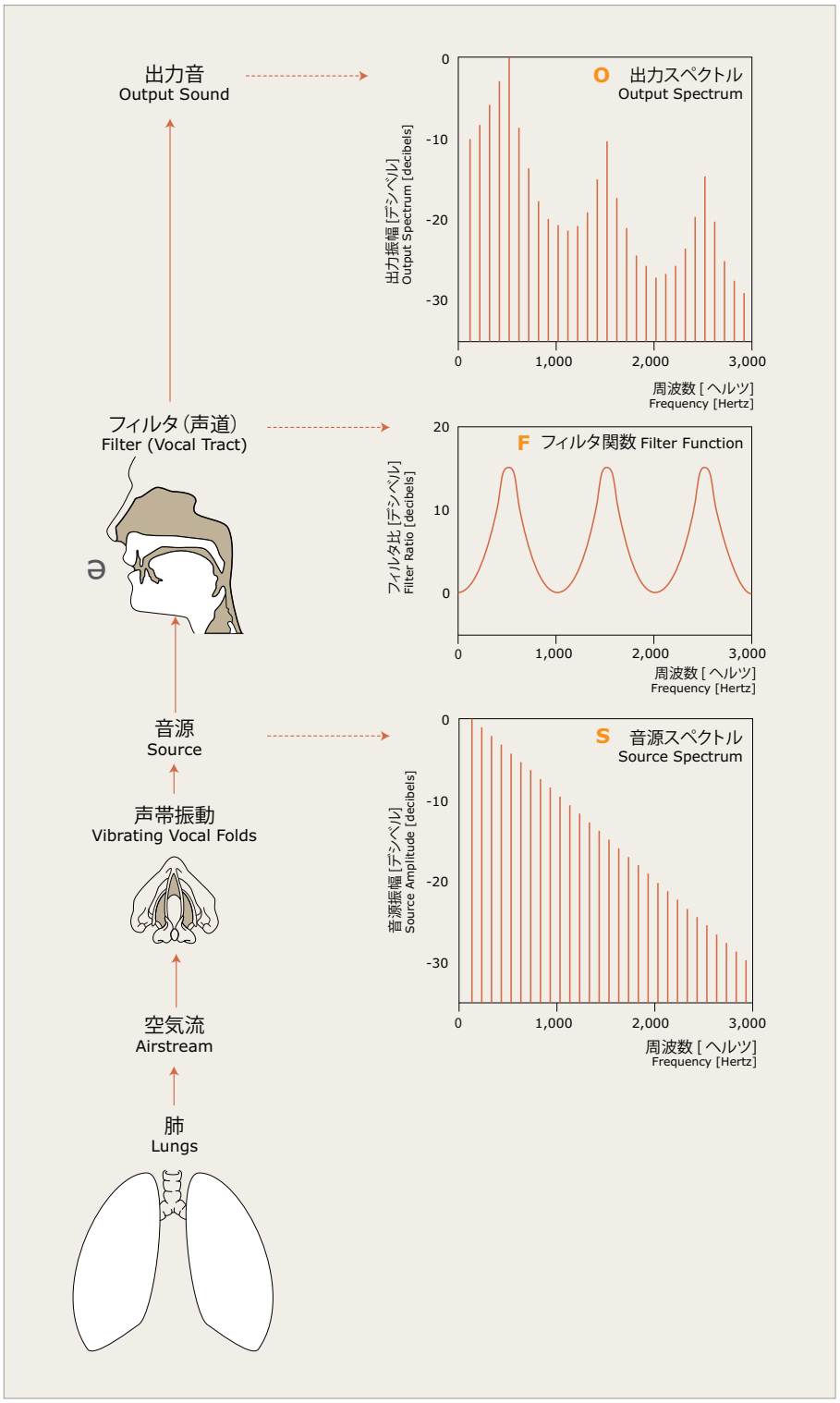

図 2.4: 音源—フィルタモデル．イラストは [4] 掲載のものに基づく．

ます．声道の共振周波数は母音の認識にとても重要なので，特に**フォルマント（formant）**と呼ばれています．

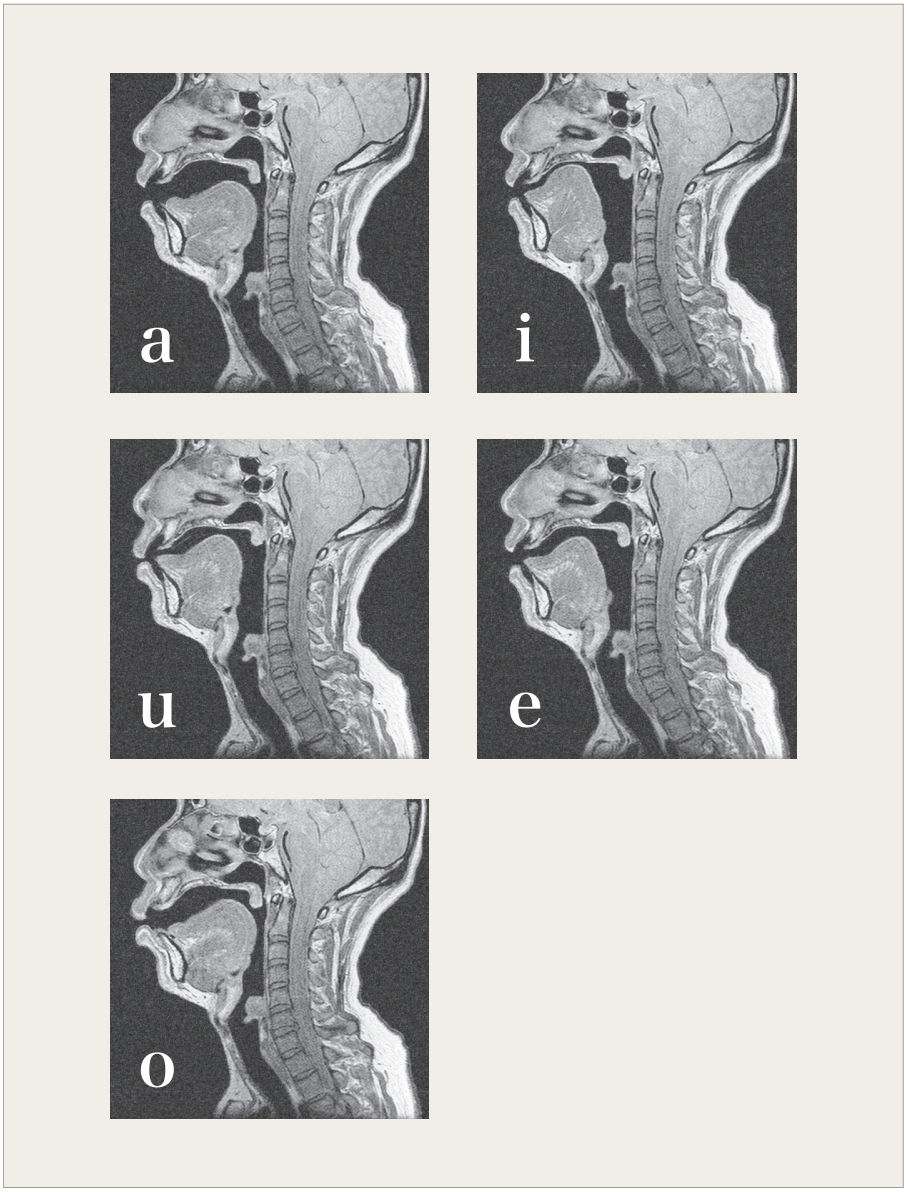

図 2.5: 日本語 5 母音発声時の頭部 MRI 画像（『ATR 母音発話 MRI データ』）転載厳禁

2.1.3 子音の生成

子音は一般的に母音と比べて声道の狭めが強いのが特徴です．声道に強い狭めの位置では呼気による圧力が高まって乱流が生じ，この位置が音源となります．声道に一時的な閉鎖が生じた後の急激な解放によってその部分が音源となる場合もあります．音源の生成のされ方（狭め，閉鎖）や位置（口唇側，声帯側）によって音源の特性はさまざまです．声帯の振動が伴う場合もあります．器官の瞬間的な動きによって生成されるものがあります．

狭めによる乱流が音源となる子音は**摩擦子音（fricative）**と呼ばれ，/ f /, / ʃ /, / v / などがあります．一時的な声道の閉鎖の後の解放によって生じる子音を**破裂子音（plosive）あるいは閉鎖子音（stop）**といい，代表的なものとして/p/, /t/, /k/があります．図 2.7 には，閉鎖音/p/, /t/, /k/の調音位置（唇，歯，軟口蓋）を示しています．

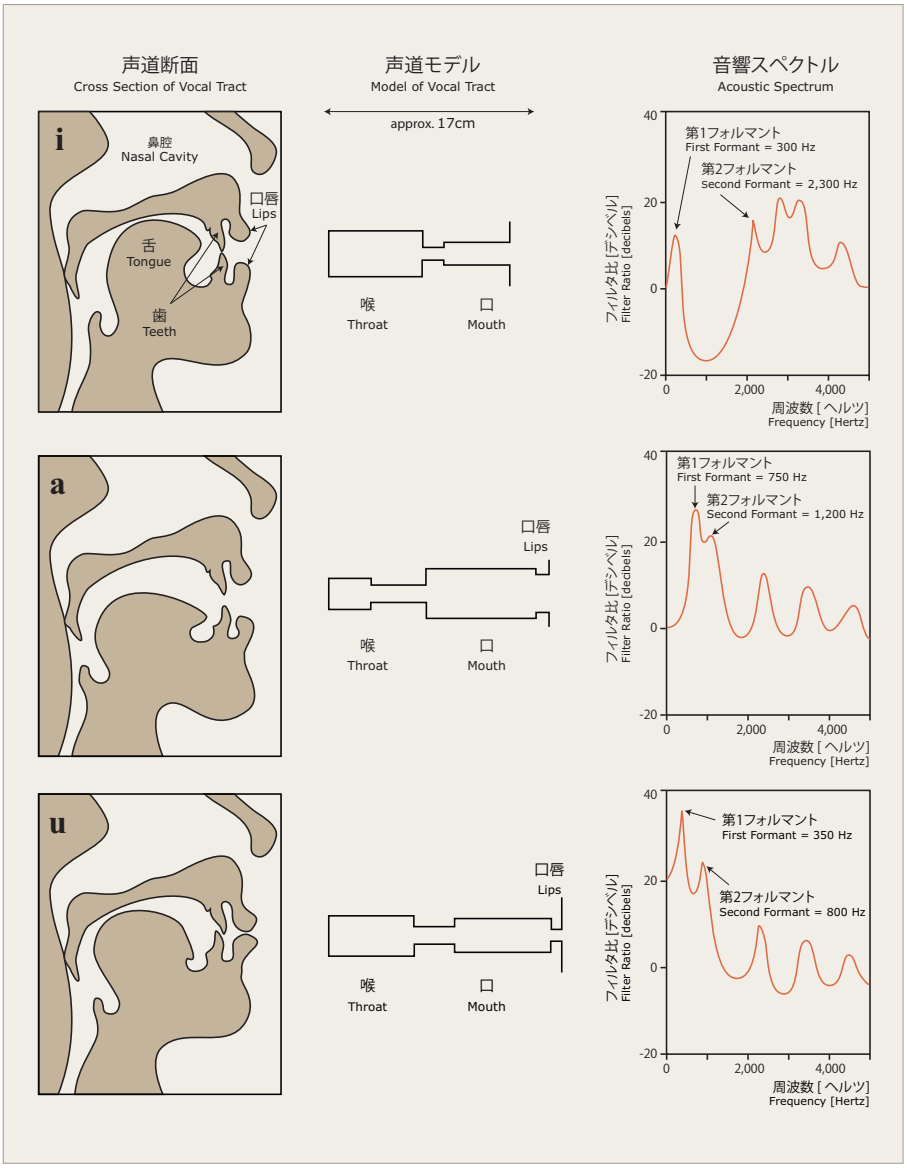

図 2.6: 声道の円筒モデル．イラストは [4] 掲載のものに基づく．

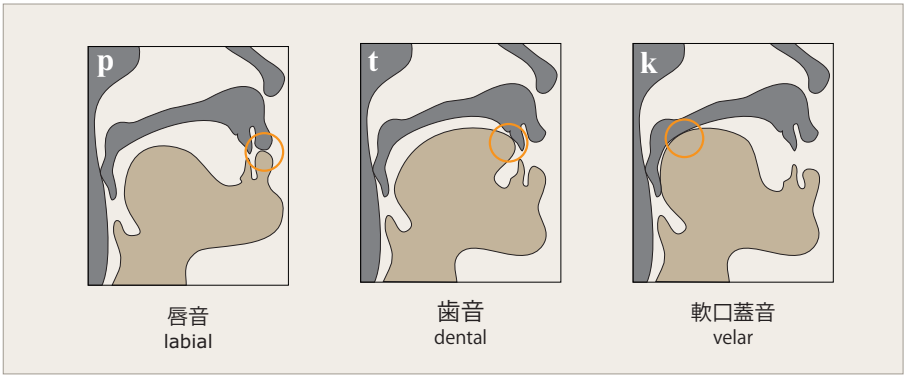

図 2.7: 破裂（閉鎖）子音の調音位置．その調音位置により，/p/は口唇音（labial），/t/は歯茎（dental），/k/は軟口蓋音（velar）と呼ばれます．イラストは [4] 掲載のものに基づく．

2.2 日本語の音素

日本語をローマ字で書き表すと, 例えば, 赤 /aka/, 秋 /aki/のようになります. この 2 つの単語は，ローマ字の第 3 文字の /a/ と /i/ の違いによって区別されます．このように，ある言語において単語の意味の違いに関わる音声の最小単位を**音素（phoneme）**といいます．音素は大きく**母音（vowel）**, **子音（consonant）**に分けることができます．母音は日本語では /a, i, u, e, o/ の 5 種類，子音には /k, s, t, n, h, m, z, d, b, p, N/ などの音素や，**半母音（semivowel）**と呼ばれる /j, w/などがあります．表 2.8 は，日本語の子音を調音位置と調音方法で分類したものです.

音素は**有声音（voiced sound）**と**無声音（unvoiced sound）**に分類されます．有声音は発声の際に声帯の振動を伴うもので，母音などがこれに該当します．一方，無声音には口の中の狭めを空気が通るときの音素（摩擦音/s,sh/など）や，口がいったん閉じて急に開くときの音素（破裂音または閉鎖音/p, t, k/など）があります．これらの音素に声帯の振動が伴うと，それぞれ/z, dz/，/b, d ,g/となります．有声音の中には，音が鼻から出る**鼻音（nasal）**（/m, n/など）というものもあります．半母音は子音とそうではない母音の間の中間的な狭めを形成して発声されるものです.

調音位置	両唇音		歯音		歯茎音		口蓋音		声門音
調音方法	無声	有声	無声	有声	無声	有声	無声	有声	無声
破裂音	p	b			t	d	k	g	
摩擦音	f		s	z	ʃ	ʒ			h
破擦音			ts	dz	tʃ	dʒ			
鼻音		m				n		ŋ	
流音		w						j	
弾音						r			

図 2.8: 日本語子音の調音位置と調音方法による分類．声道の最も狭い部分を調音位置といいます.”無声”は声帯の振動を伴わないもの,”有声”は声帯の振動を伴うものをいいます.

2.2.1 母音

音波が円筒のような音響管を通過すると，ある周波数成分が強められ，ある周波数成分は弱められるという共鳴現象が生じ，その共鳴周波数は円筒の形に依存します．声帯から唇までの声道は一種の音響管として機能するので，円筒管の場合と同様に共鳴が生じます．声道の形は舌や唇を使ってある程度自由に変えることができます．そうすることによって共鳴のしかたが変わり，異なる音色の母音ができます.

日本語の 5 母音の音声波形（男性）の例を図 2.9 に示しました．波形は母音によって形が大きく異なっていることが分かるでしょう．母音の波形は，同じような波形が繰り返しています．図 2.9 の /a/ の場合は約 9 ms です．この 1 周期を音声の**基本周期（fundamental period）**，基本周期の逆数を**基本周波数（fundamental frequency）**といいます．基本周波数は声の高さと関係があります．基本周波数が低ければ低い声になり，高ければ高い声になります．基本周波数は，性別によって異なりますし，個人差もあります．基本周波数の値は，男性の場合およそ 80〜200 Hz，女性の場合およそ 150 Hz〜400 Hz です．基本周波数はアクセントやイントネーションなどによって影響を受けるので，同じ人の音声でも一定ではありません．

音声スペクトルには音素によって特徴的な山や谷が見られ，山の部分を**フォルマント（formant）**といい，その代表周波数（山のピークの周波数）をフォルマント周波数といいます．周波数の低い方から第 1 フォルマント，第 2 フォルマント，$\cdots$ と呼び，F1，F2，$\cdots$ と表記します．日本語 5 母音（男声）をスペクトログラム表示した図 2.10 に見える濃い横筋がフォルマントに対応しています．スペクトログラム（Spectrogram）はスペクトルを 3 次元のグラフ（時間，周波数，信号成分の強さ）で表わしたものです．一般的には，横軸が時間，縦軸が周波数で，点の明るさや色がその点の時間–周波数位置での信号成分の強さを表しています．F1 と F2 の位置を矢印で示してあります．2 つのホルマントの位置が，母音によってはっきり異なっていることがわかります．

母音による F1，F2 の周波数の違いは，話者によらず一定の傾向があります．図 2.11 は男性と女性各約 30 名の音声を分析して得られた図です [18]．横軸に F1 の周波数，縦軸に F2 の周波数をとり，各母音の測定結果を分布で表わしています．黒丸 (●) は男性，白丸 (○) は女性の平均値を表わしており，それを中心に標準偏差が描かれています．男女間の平均的違いは，母音間の位置関係を保ったまま，おおよそ周波数を縮尺した関係にあることがわかります．

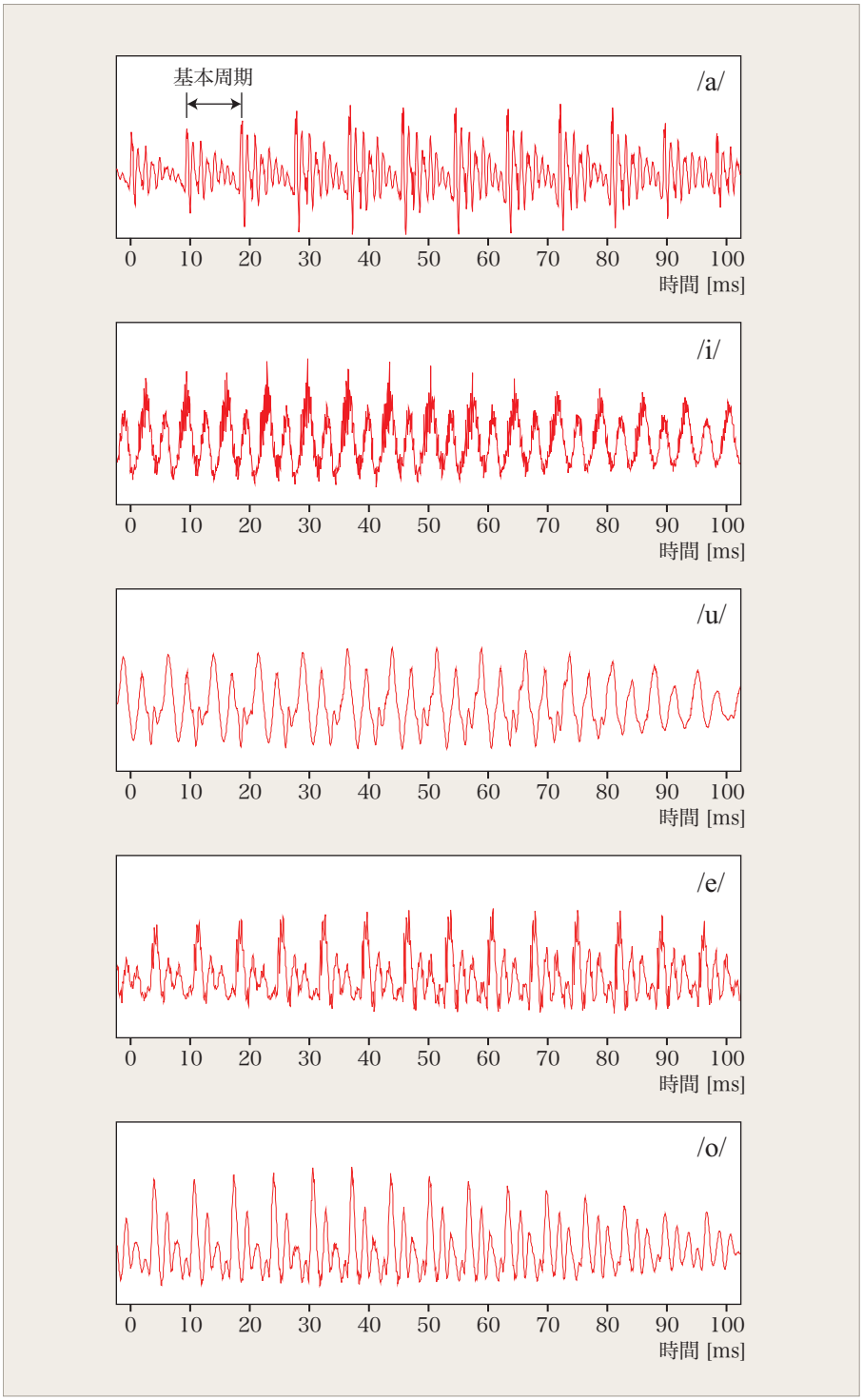

図 2.9: 日本語 5 母音の音声波形（男声）

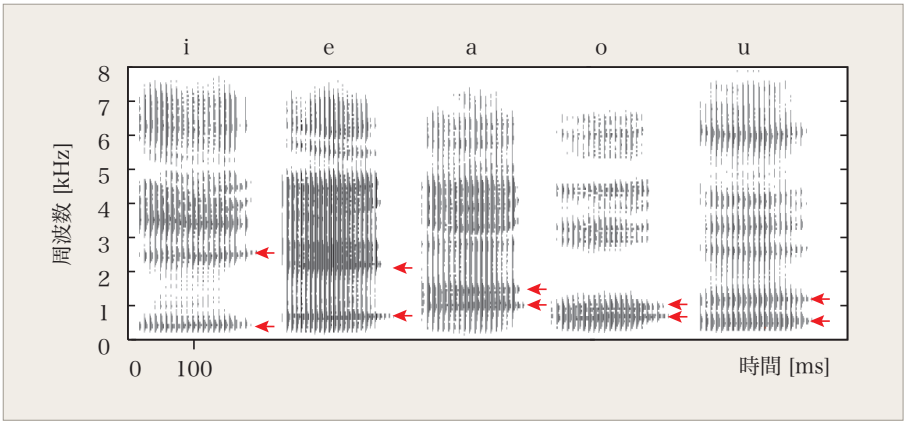

図 2.10: 5 種類の母音（男声）のスペクトログラム表示

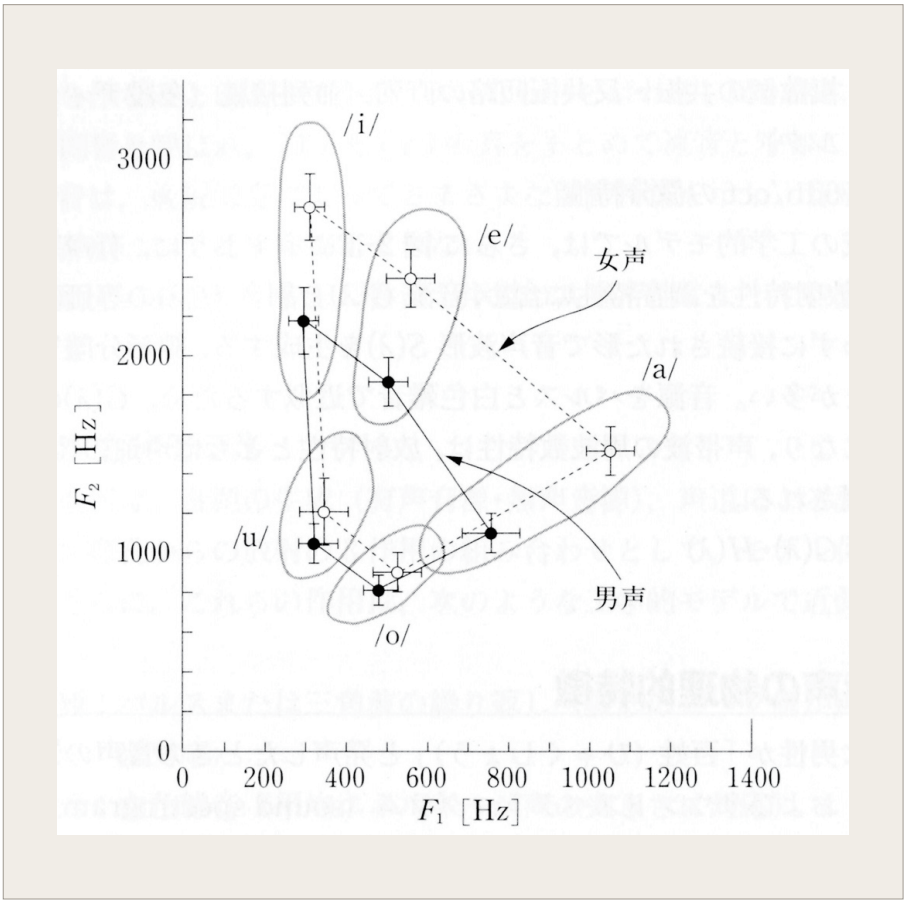

図 2.11: 日本語母音の第 1 第 2 フォルマント周波数の分布 (男性女性各約 30 名)[18]

2.2.2 子音

子音は声道に強い狭めあるいは閉鎖を作ることにより生成されます．子音の音源はこの狭めあるいは閉鎖付近に生じます．声道の閉鎖による子音の場合は，閉鎖が解放された時の音が特徴的です．また鼻音のように，声道に鼻腔への分岐が生じる場合があります．このように，子音は，音源の形態や位置，閉鎖の強さ，調音器官の時間的変化の様態、声道の分岐の状態などによって分類されます．

子音は音響理論的扱いが難しく，スペクトルの特徴が複雑です．子音（図 2.8）は母音に比べて種類が多く，非定常な現象によって特徴付けられるものがあります．日本語に限っても子音の性質を網羅的に説明するのは、この実習書の目的の範囲を超えます．

子音の性質についての詳しい説明は関連図書（例えば [3, 11, 19]）に譲ります．この実習で音声認識タスクの実例として説明に用いている数字音声の音響的特徴については付録 A で説明しているので，子音のスペクトルの実例として，課題に取り組む際の参考にしてください．ここでは、他の子音に比べて音響的性質が説明しやすい特徴があり，初心者でも理解しやすい破裂音と摩擦音を取り上げて解説します．

破裂音

日本語の破裂音には無声破裂音/p/，/t/，/k/と有声破裂音/b/，/d/，/g/があります．無声口蓋破裂音の/k/は，先行母音からの過渡部，閉鎖部，破裂部，気音部，後続母音への過渡部からなります（図 2.12）．波形から分かるように，物理的には閉鎖部は無音ですが，この部分も人間が破裂音を知覚する重要な特徴になっています．

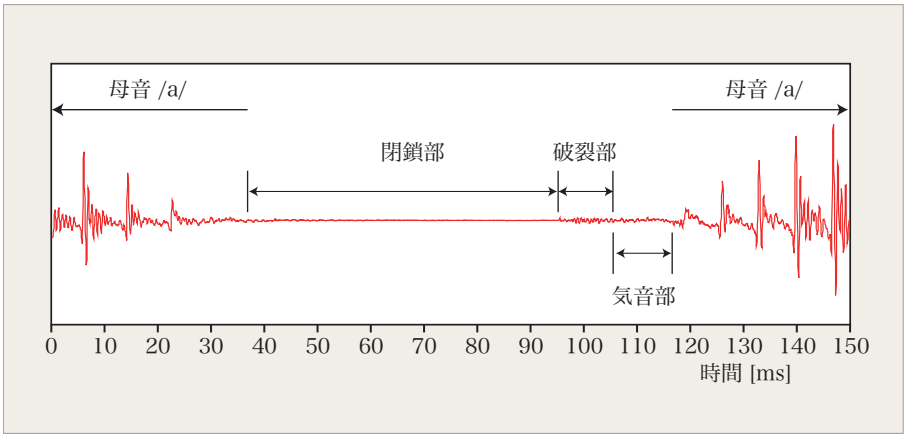

図 2.12: /aka/の無声破裂音/k/の部分の音声波形の例

無声破裂音と有声破裂音の違いや破裂の位置（図 2.7）によって，破裂の瞬間から母音に遷移するまでの時間や母音のフォルマントの過渡部の形が異なります．子音の調音位置と母音の種類によって様相が異なるので，母音部も含めた音声波形およびスペクトルの時間的変化の記述が必要です．

摩擦音

声道の強い狭めの位置に肺からの気流が通過するときに乱流が生じ，それが音源となって摩擦音が生じます．破裂音と異なり，定常的で比較的安定していて，息が続く限り継続することができます．無声歯茎摩擦音/s/は周期性のない雑音性の波形を示します（図 2.13（上））．無声子音では声帯が振動していないため，母音のような基本周期に対応した繰り返し波形は現れません．スペクトル特性には主として調音点（狭めの位置）から唇側の共振特性が現れます．3 kHz 以上の周波数帯域にパワーが現れます（図 2.14）．有声歯茎摩擦音である /z/ の場合は，声帯が振動しているため，母音の波形とは違いますが，振幅の小さい周期性の波形が現れています（図 2.13（下））．/s/と/z/の違いはスペクトルにも現れています（図 2.14）．/z/の場合は/s/に比べて高域成分が少なく，500 Hz 以下の低い周波数成分にパワーが見られます．このような低周波数成分は有声子音の特徴です．

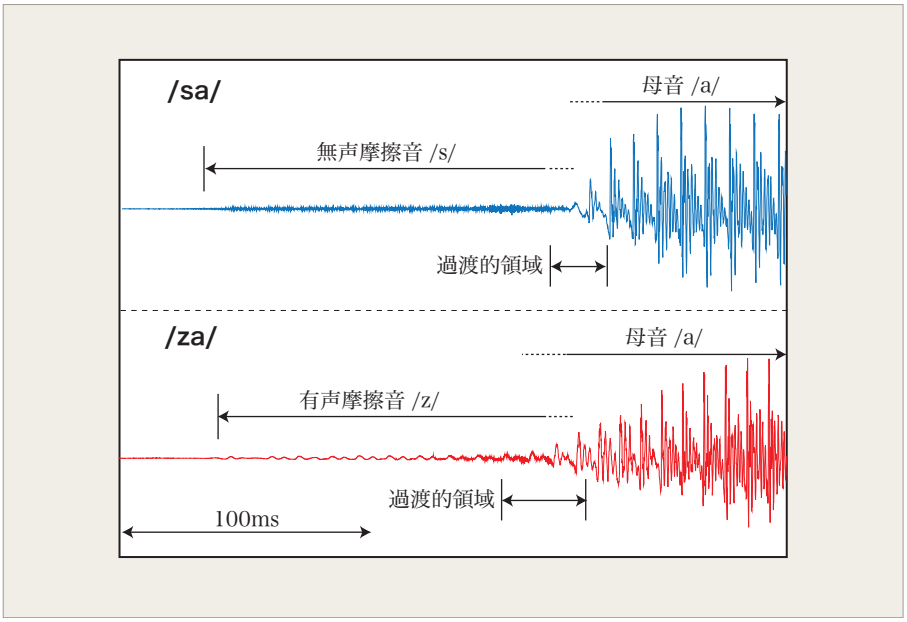

図 2.13: 無声摩擦音/s/（上）と有声摩擦音/z/（下）の音声波形の例

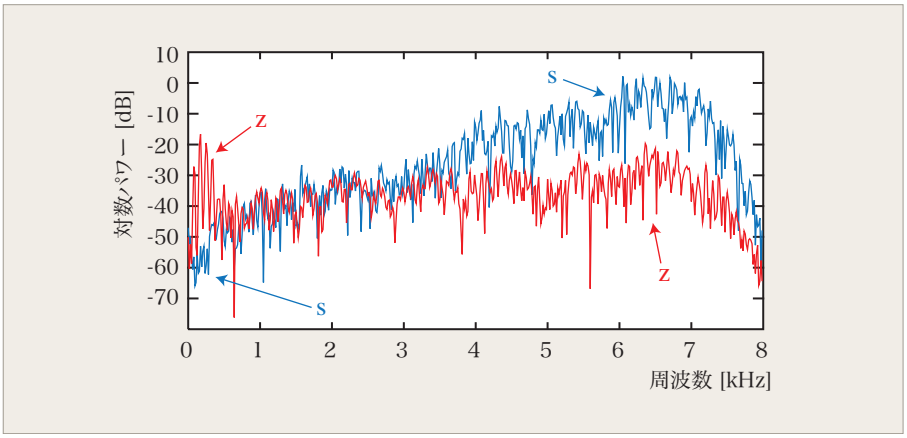

図 2.14: /sa/と/za/の対数パワースペクトルの例