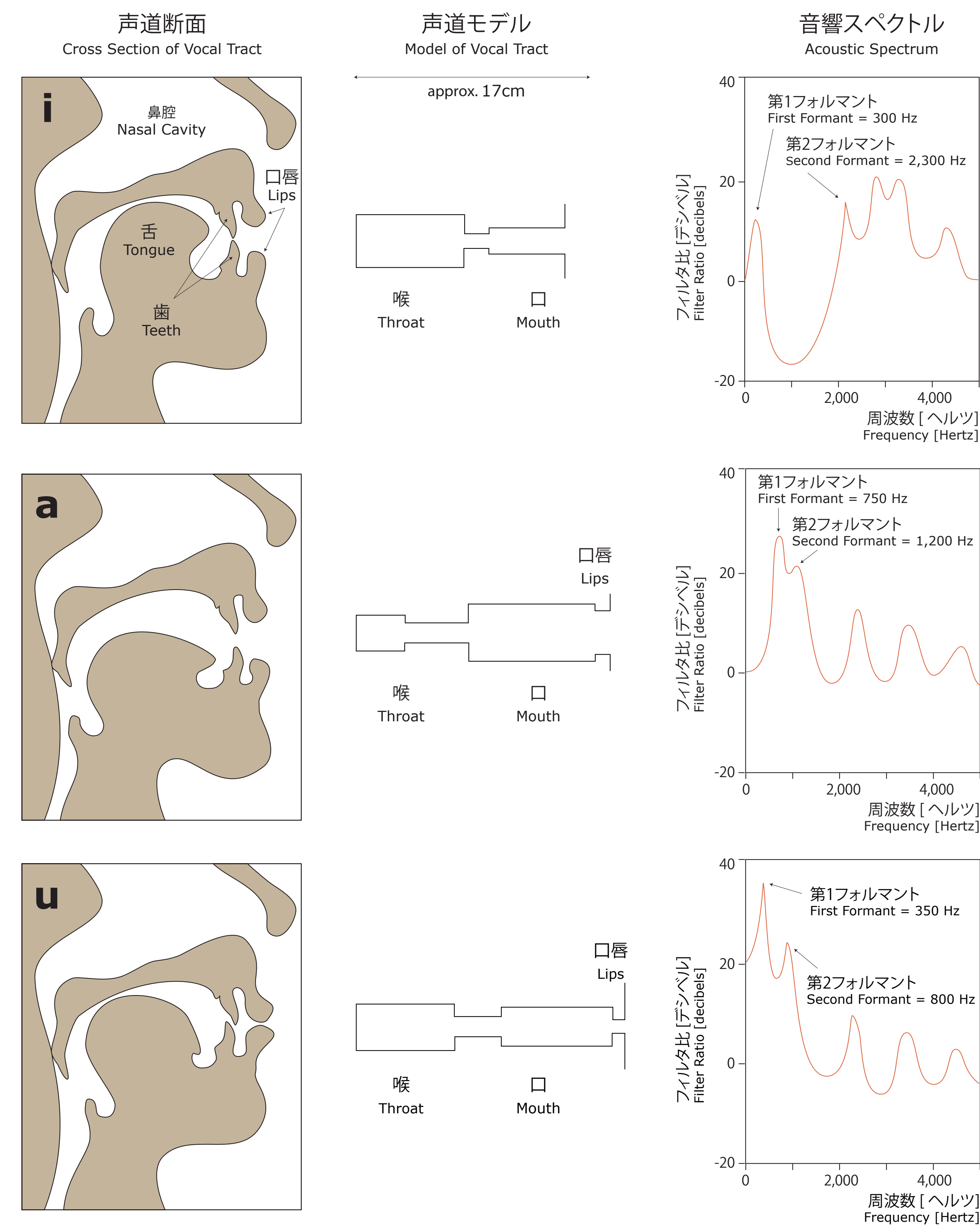


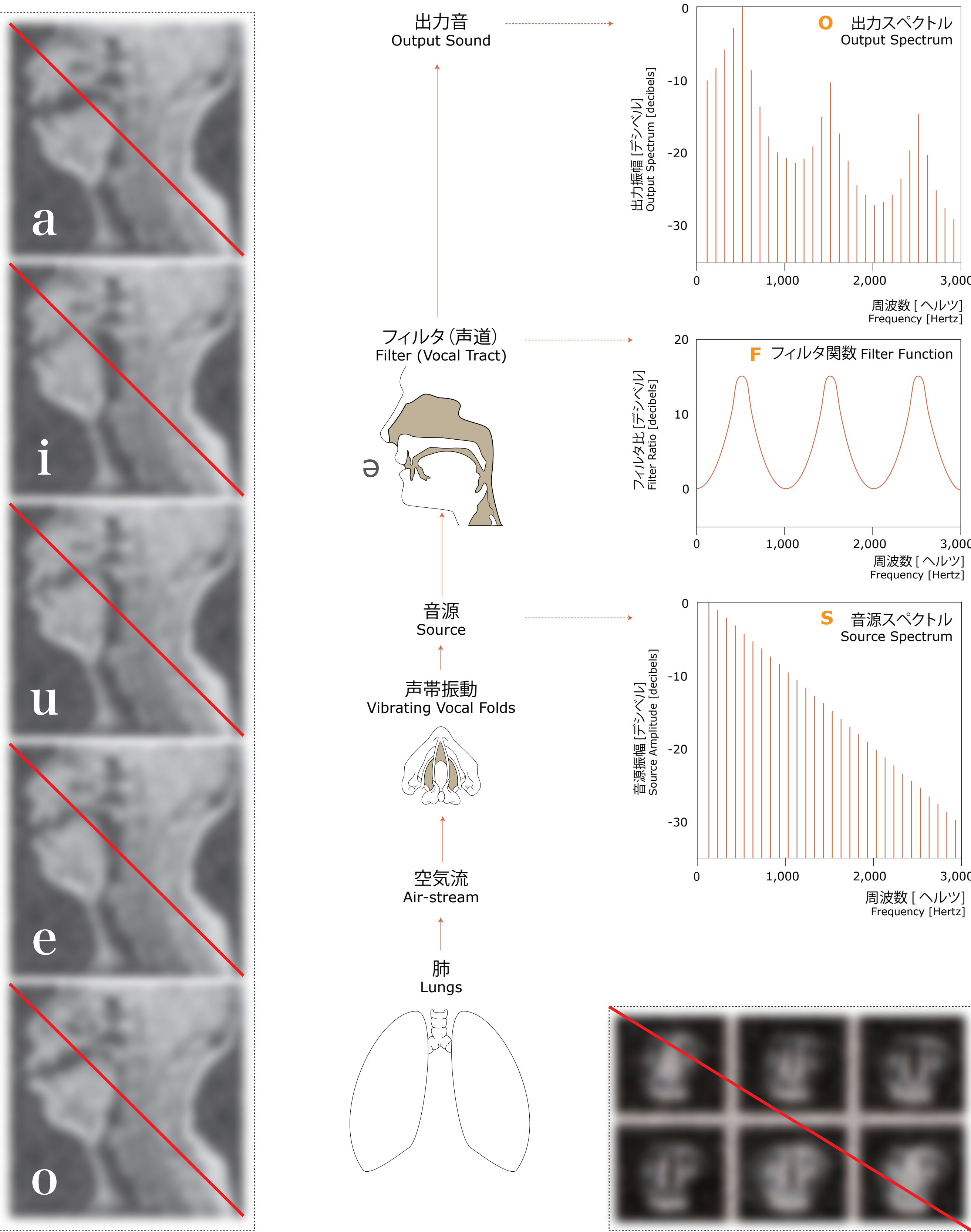
## 声道モデル Vocal Tract Model



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/ の声道の概形が図の左側に描かれています。中央の図は、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、特別に**フォルマント**と呼ばれています。

Drawings made from MRI pictures can be used to estimate cross-sectional areas of the vocal tract, and these estimates can then be used to compute the filter transfer function of the tract. The configurations of the vocal tract for the vowels /i/ as in “feet”, /a/ as in “father”, and /u/ as in “boot” are depicted at the left. A corresponding model of the vocal tract appears in the center column, and the acoustic spectrum for each vowel is given at the right. Psychological studies of speech perception demonstrated that it is this general shape of the spectrum that listeners use to identify which vowel was spoken. The resonant frequencies of the vocal tract are so important for the recognition of vowels, they have been given a special name: they are called **formants** [Miller1981].

## 音源－フィルタ理論 The Source-Filter Theory



日本語5母音発声時の頭部MRI 画像  
MRI pictures of a head while producing five Japanese vowels

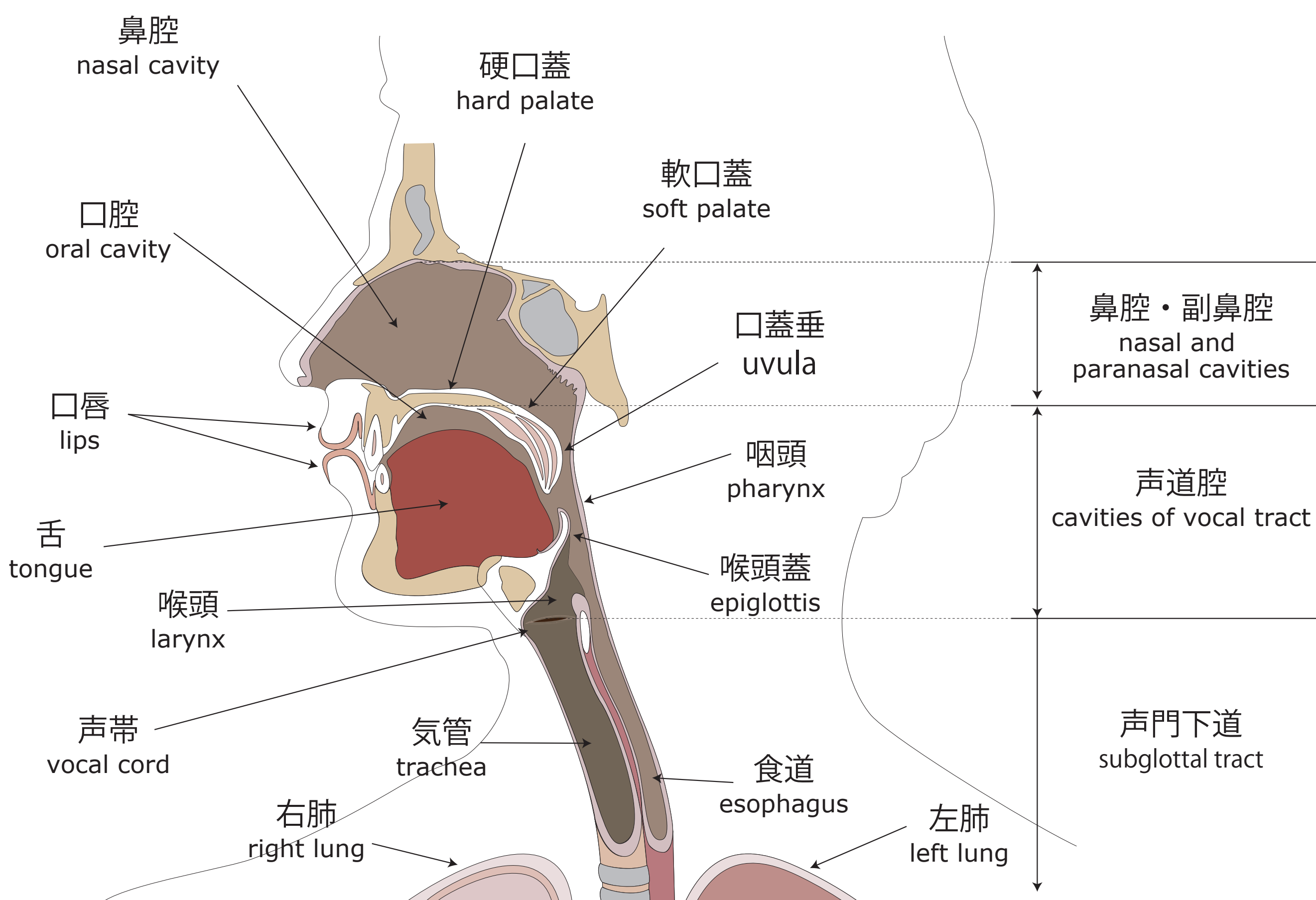
本研究で使したMRIデータは、ATR人間情報科学研究所が独立行政法人情報通信研究機構からの研究委託「人間情報コミュニケーションの研究開発」に基づいて収録し、公表したATR母音発話MRIデータの一部です。本データの使用および成果の発表は、株式会社ATR-Promotionsとの使用許諾契約に基づいております。

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。

In speech production the **lungs** function as a power supply, the **vibrating vocal cords** as an oscillator and the **vocal tract** as a resonator. According to the source-filter theory, the source is a laryngeal tone produced when the vibrating vocal cords interrupt the air-stream from the lungs. The acoustic spectrum for a source shows the amplitude decreasing uniformly with frequency (**S**). The filter is the vocal tract, which modifies the source. The result is called the filter function (**F**). When the source is filtered, the resonant peaks are imposed on the output spectrum (**O**) of the vowel [Miller1981].

## 音声生成器官概略図 (頭部矢状断面)

Sketch of speech production apparatus (cross sagittal section)



音声器官は肺、喉頭、咽頭、口、鼻からなる。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る。声道の形は口唇、顎、舌、喉頭の位置や形によって変化する。

The voice-producing mechanism is made up of the lungs, the larynx, pharynx, mouth and nose. The vocal tract, consisting of the pharynx and mouth, is a resonator whose shape determines vowel sounds. The shape can be modified by changes in position and shape of the lips, jaw, tongue and larynx.

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、一部の音では鼻も）を使って発声します。音声は声道にさまざまな狭めを作ることで生成されます。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。適切な張力が掛かった声帯の間を呼吸が通ると、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。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有声音、そうでない音声が無声音といえます。

When we talk, air from our lungs is expelled in a controlled way through the larynx and the vocal tract — the throat, mouth, teeth, lips and, for some sounds, nose. Sound is produced by making constrictions in this system. The most important source of sound is in the larynx (so called Adam’s apple), where the vocal cords can be brought together with just the right tension to vibrate — to release a rapid series of puffs of air into the vocal tract. How this vocal cord vibration sounds depend on the shape of the vocal tract, which acts as a filter (a resonator that reinforces some frequencies of vibration and attenuates others). Not all speech sounds involve this vocal cord vibration. Those that do are called voiced sounds. These that do not are called unvoiced sounds [Miller1981].

参考文献 References  
[Miller1981] George A. Miller, Language and Speech, W. H. Freeman and Company, 1981.  
[Flanagan1972] James L. Flanagan, Speech Analysis Synthesis and Perception, Springer-Verlag, 1972.

## 機械による音声言語の認識 Recognition of Spoken Language by Computer

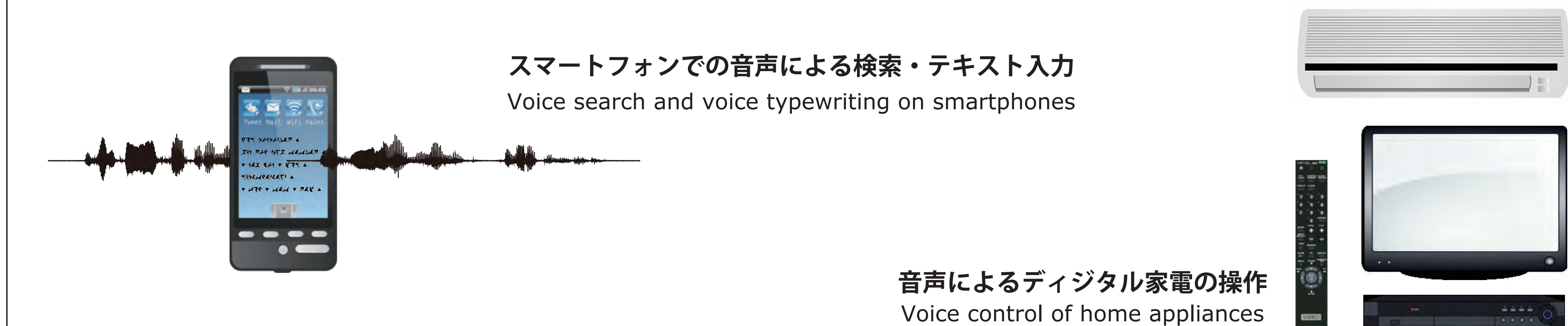


音声言語は人間同士のコミュニケーションにおいて**最も重要なメディア**です。言葉を話したり聞き取ることは以前は人にしかできないことでしたが、最近では、ある程度機械にもできるようになって来ています。音声言語は人間と機械との相互作用やコミュニケーションシステムにおいて有用であると期待されています。音声言語の処理技術は**今後も大きな発展が期待される分野**であり、コンピュータネットワークを背景とした未来のコミュニケーションを支える重要な技術の一つと考えられています。

Spoken language is **the most important medium of communication** between humans. To speak and understand the spoken language used be possible only to human, but it has become to some extent feasible for machines these days. Spoken language is also expected to be useful in human-machine interaction and communication systems. It is expected that automatic processing of spoken language will **continue to be one of the technology areas in the future, in which a considerable progress will be made**. Recognition and processing of spoken language by computer will play an important role in supporting future communication with the computer network in the background .

### 自動音声認識によって実用化されている応用技術

Applications of automatic spoken language processing technologies that are being commercialized



音声によるデジタル家電の操作  
Voice control of home appliances

自動音声認識について  
about automatic speech recognition

話した言語音を機械で認識すること  
computers that understand spoken language