# Module 1

## A brief look at speech production and perception

# Roadmap

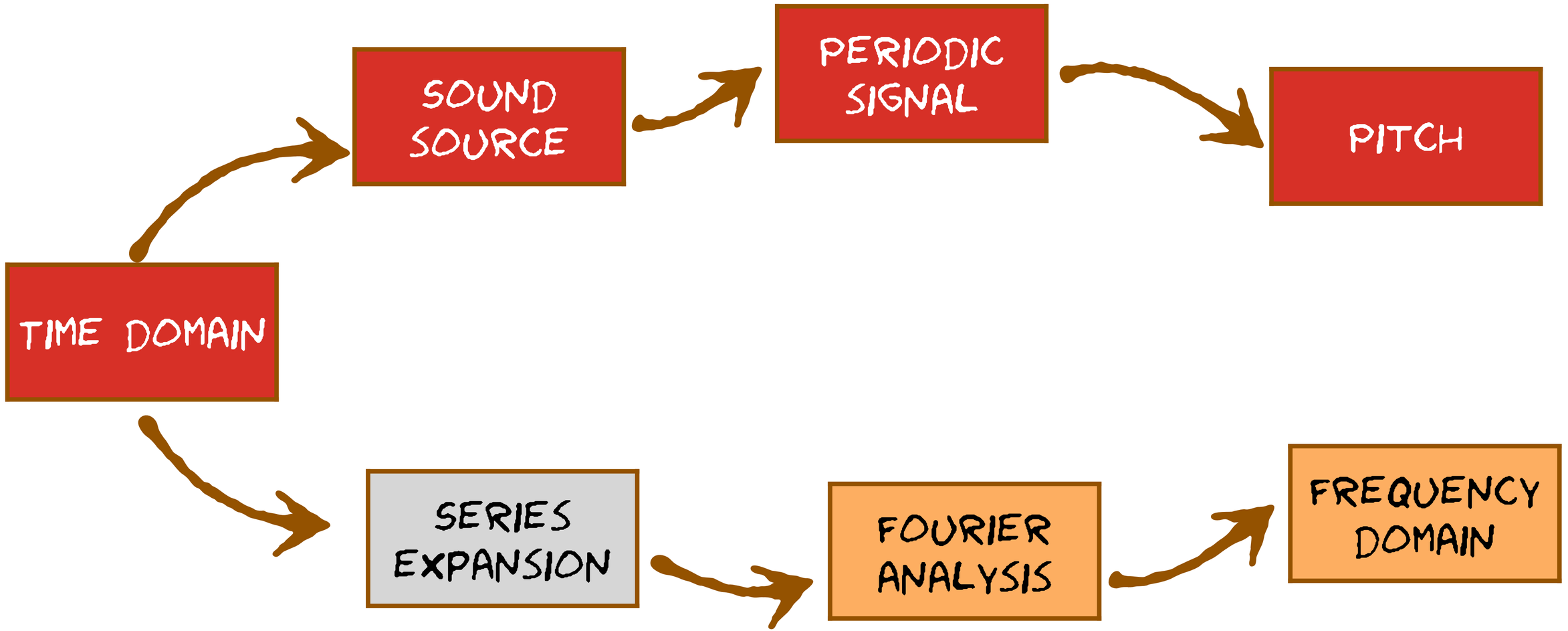
- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

- Block 1 Week 2
  - Module 1: a brief look at speech production and perception
- Block 1 Week 3
  - Foundations: signals
  - Module 2: speech signals and the source-filter model
- Block 1 Week 4
  - Foundations: phonetics

# Today's topics - Module 1: a brief look at speech production and perception

| | THEORY | | | | | APPLICATION | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | SPEECH | | | SIGNAL PROCESSING | PROBABILISTIC MODELLING | SPEECH SYNTHESIS | | AUTOMATIC SPEECH RECOGNITION | | | |
| | SIGNALS | PRODUCTION | PERCEPTION | | | FRONT END | WAVEFORM GENERATION | FEATURE EXTRACTION | PATTERN MATCHING | HIDDEN MARKOV MODELS | CONNECTED SPEECH |
| CONCEPTS | TIME DOMAIN | SOUND SOURCE | PITCH | DIGITAL SIGNAL | DESCRIBING DATA | TOKENISATION & NORMALISATION | WAVEFORM CONCATENATION | SERIES EXPANSION | EXEMPLAR | GENERATIVE MODEL OF SEQUENCES | HIERARCHY |
| | PERIODIC SIGNAL | HARMONICS | COCHLEA | SHORT-TERM ANALYSIS | DISCRETE & CONTINUOUS VARIABLES | PRONUNCIATION | DIPHONE | FEATURES | DISTANCE | | SUB-WORD UNIT |
| | FREQUENCY DOMAIN | VOCAL TRACT RESONANCE & FORMANTS | MEL SCALE | SPECTRAL ENVELOPE | JOINT, CONDITIONAL, BAYES' FORMULA | PROSODY | | FEATURE ENGINEERING | SEQUENCE | HIDDEN STATE SEQUENCE | N-GRAMS |
| MODELS & DATA STRUCTURES | FILTER | RESONANT TUBE | FILTERBANK | IMPULSE TRAIN | GAUSSIAN | FINITE STATE TRANSDUCER | | FEATURE VECTOR | SEQUENCE OF FEATURE VECTORS | HIDDEN MARKOV MODEL | |
| | IMPULSE RESPONSE | SOURCE-FILTER MODEL | PHONEME | PITCH PERIOD | GENERATIVE MODEL | DECISION TREE | | | GRID | LATTICE | GRAPH |
| ALGORITHMS & ANALYSIS | | | | FOURIER ANALYSIS | FITTING A GAUSSIAN TO DATA | HANDWRITTEN RULES | OVERLAP-ADD | MFCCS | DYNAMIC PROGRAMMING (DTW) | DYNAMIC PROGRAMMING (VITERBI) | COMPOSITION ("COMPILING") |
| | | | | CEPSTRAL ANALYSIS | CLASSIFICATION | LEARNING DECISION TREES | TD-PSOLA | | | BAUM WELCH | APPROXIMATION (PRUNING) |

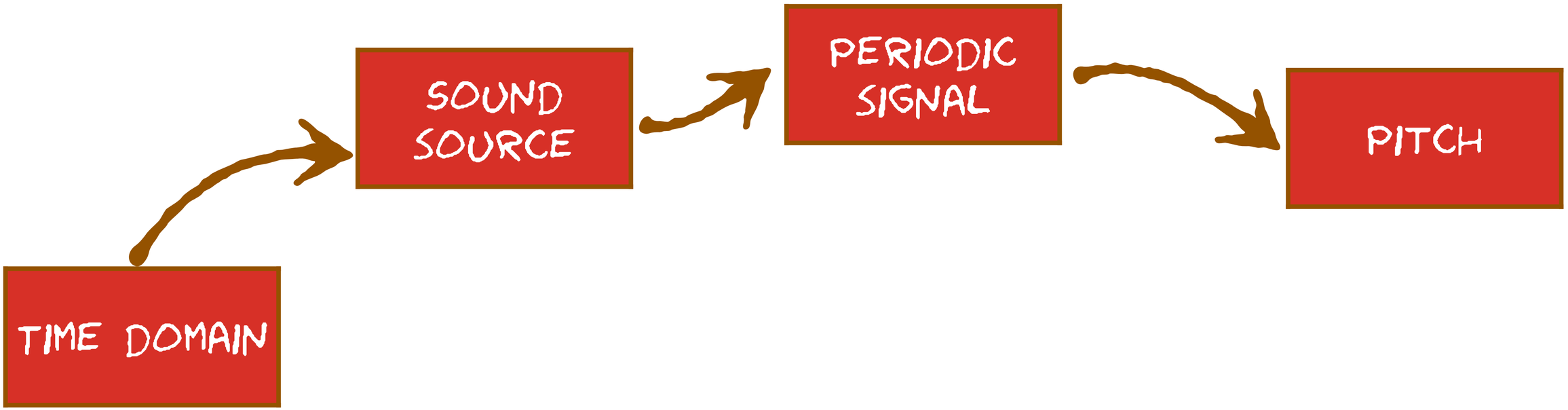
# Today's topics - Module 1: a brief look at speech production and perception



# How speech is produced and perceived

- Speech production
  - sound source, vocal tract
- Speech perception
  - the auditory system

# Topics - How speech is produced and perceived



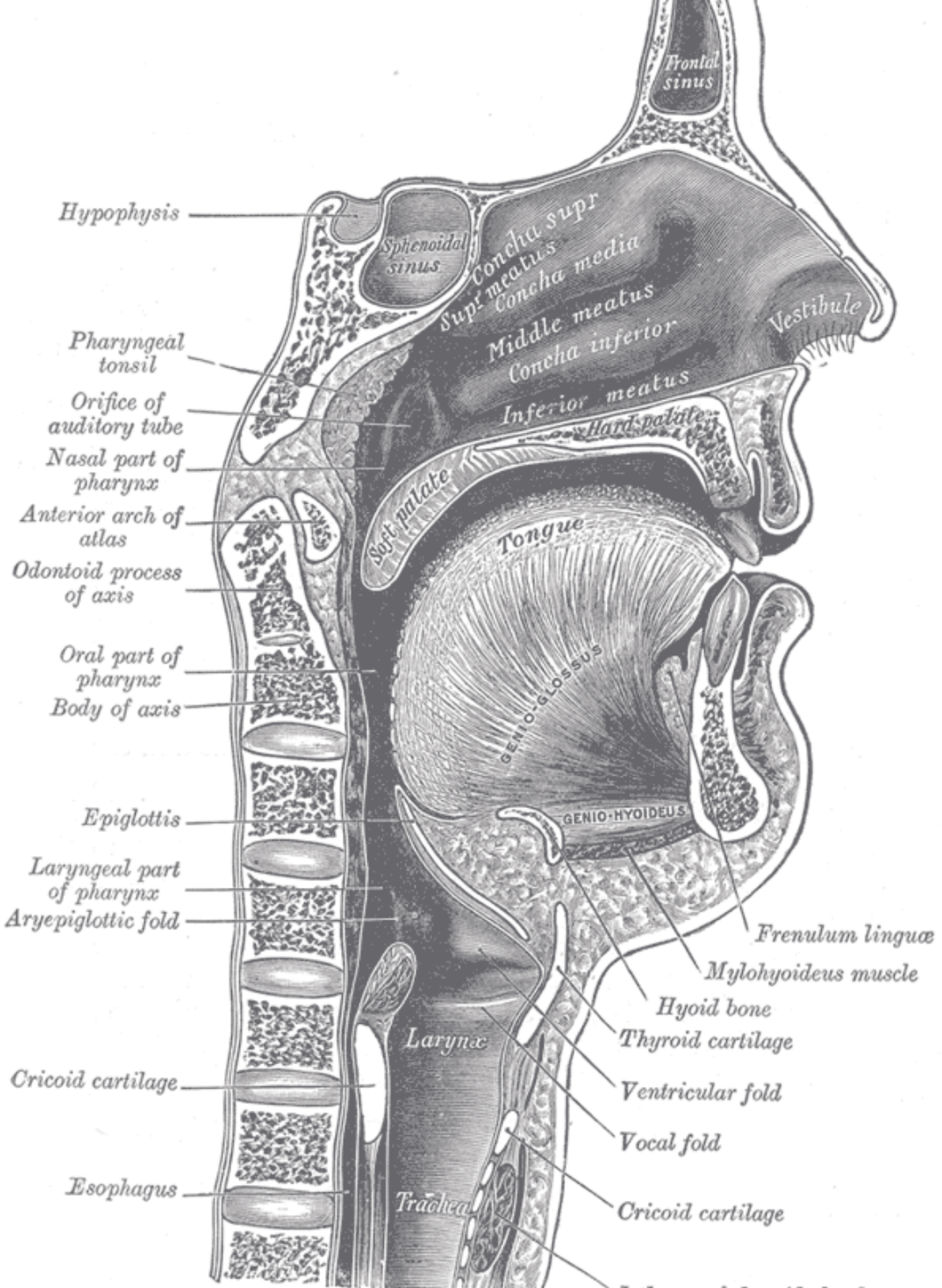
# How speech is produced and perceived

- Speech production
  - sound source, vocal tract
- Speech perception
  - the auditory system

# Vocal tract anatomy

- Vocal tract is a **tube**
- Shape can be **changed** by moving the tongue, jaw and lips
- The **nasal** branch can be connected by lowering the velum

- The tongue is larger than you might have thought - a complex set of muscles
- The nasal cavity is surprisingly large



# How speech is produced and perceived

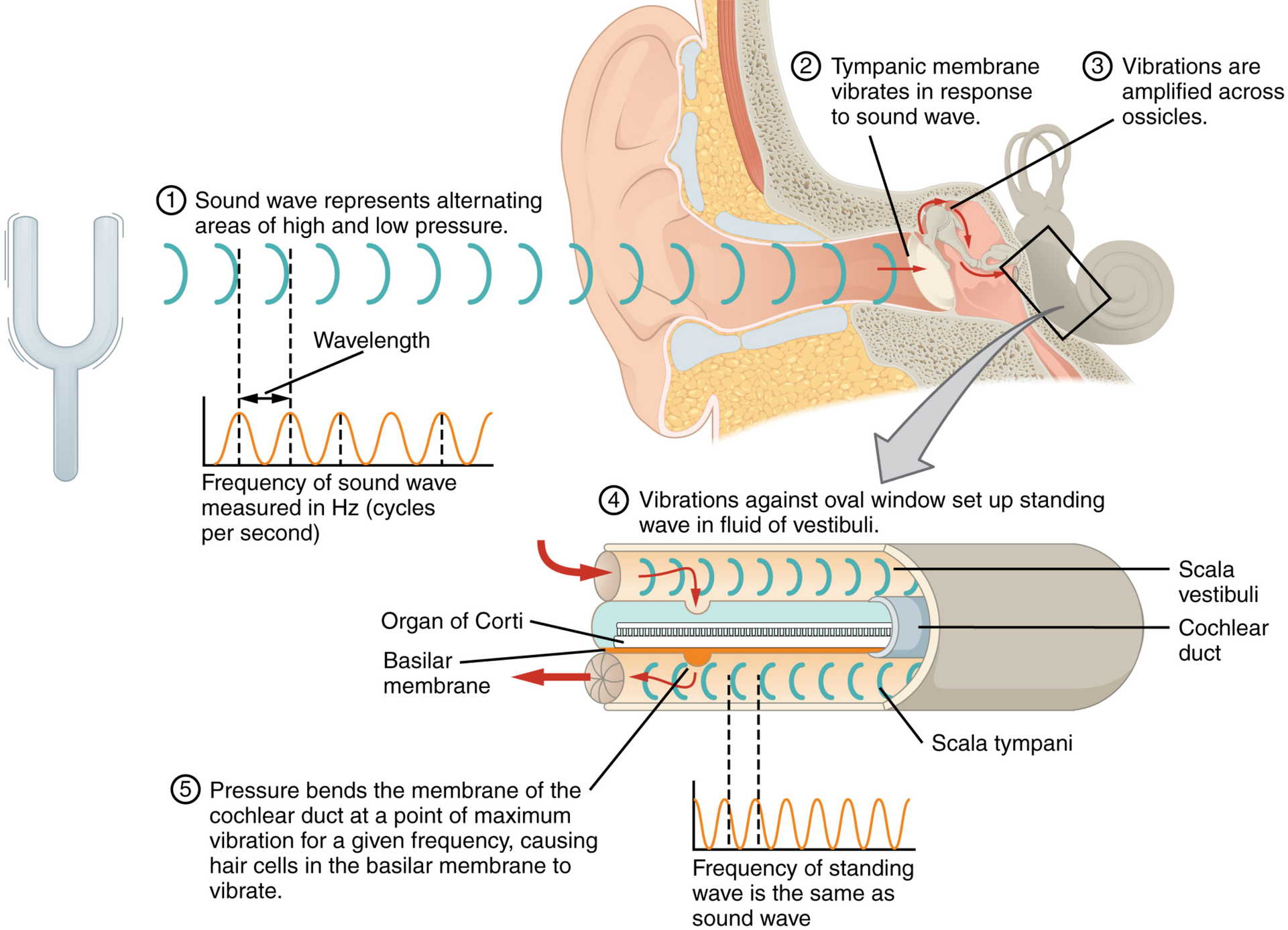
- Speech production
  - sound source, vocal tract
- Speech perception
  - the auditory system

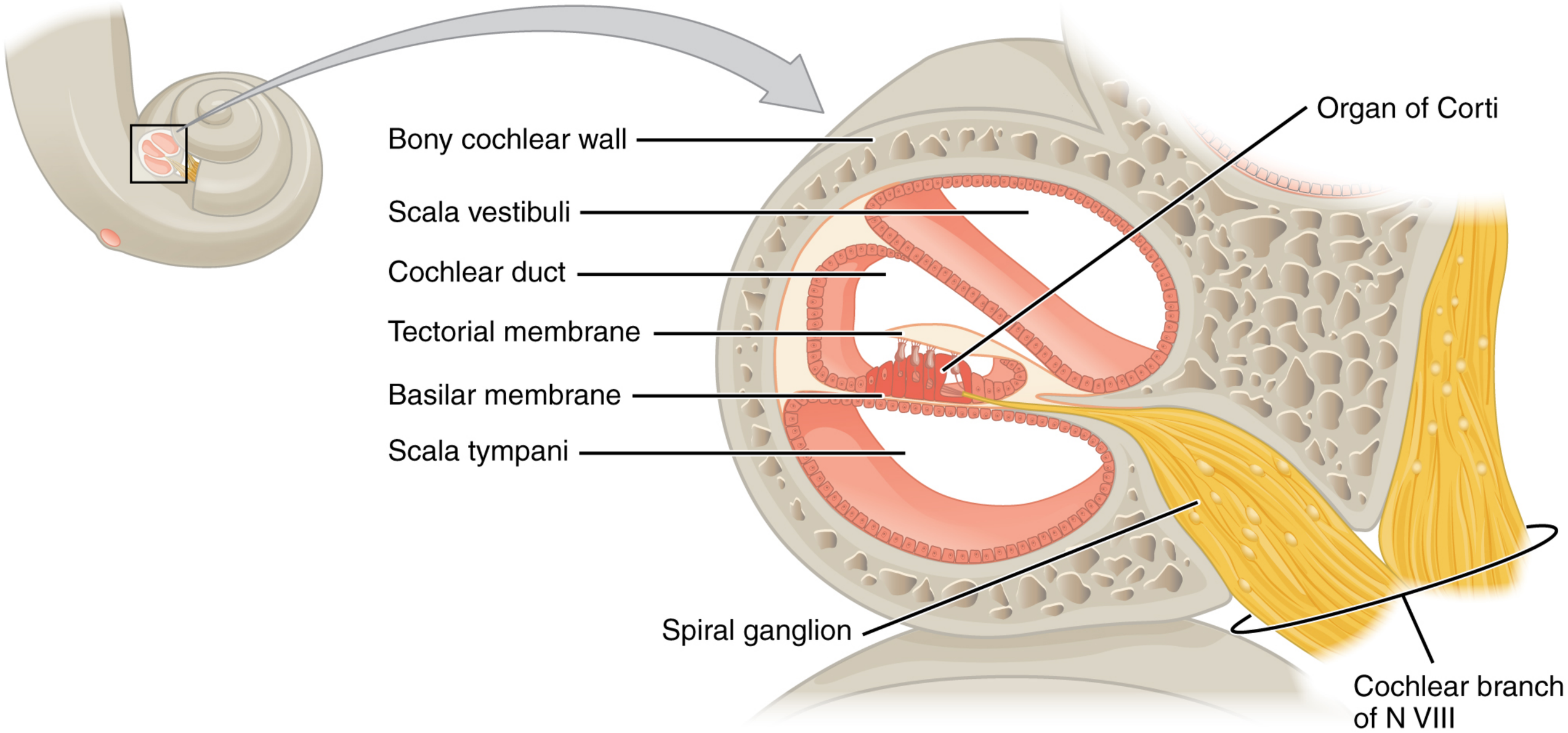
Image credit: Department of Mathematics and Systems Analysis, Aalto University, School of Science

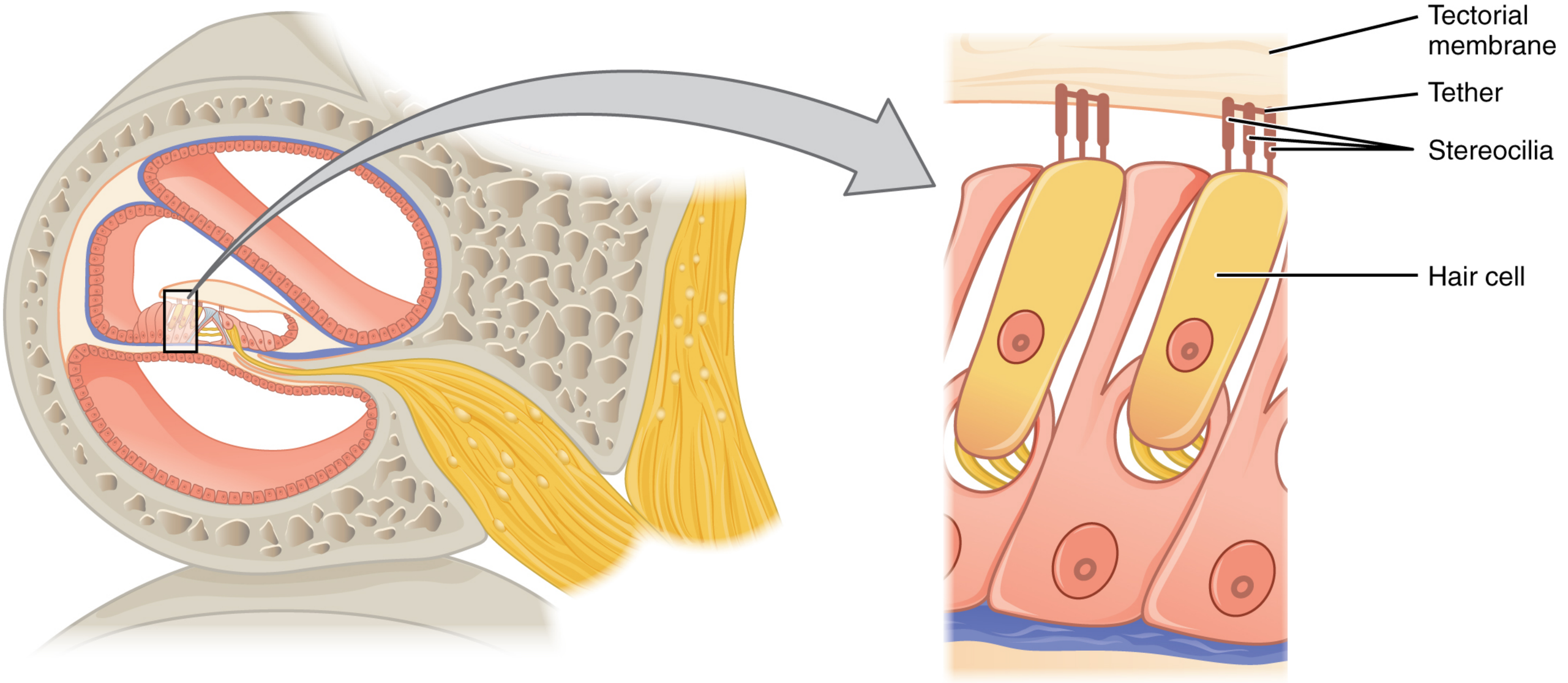
The vocal tract is a resonator. A resonator can act as a **filter**.
So, what is a filter?

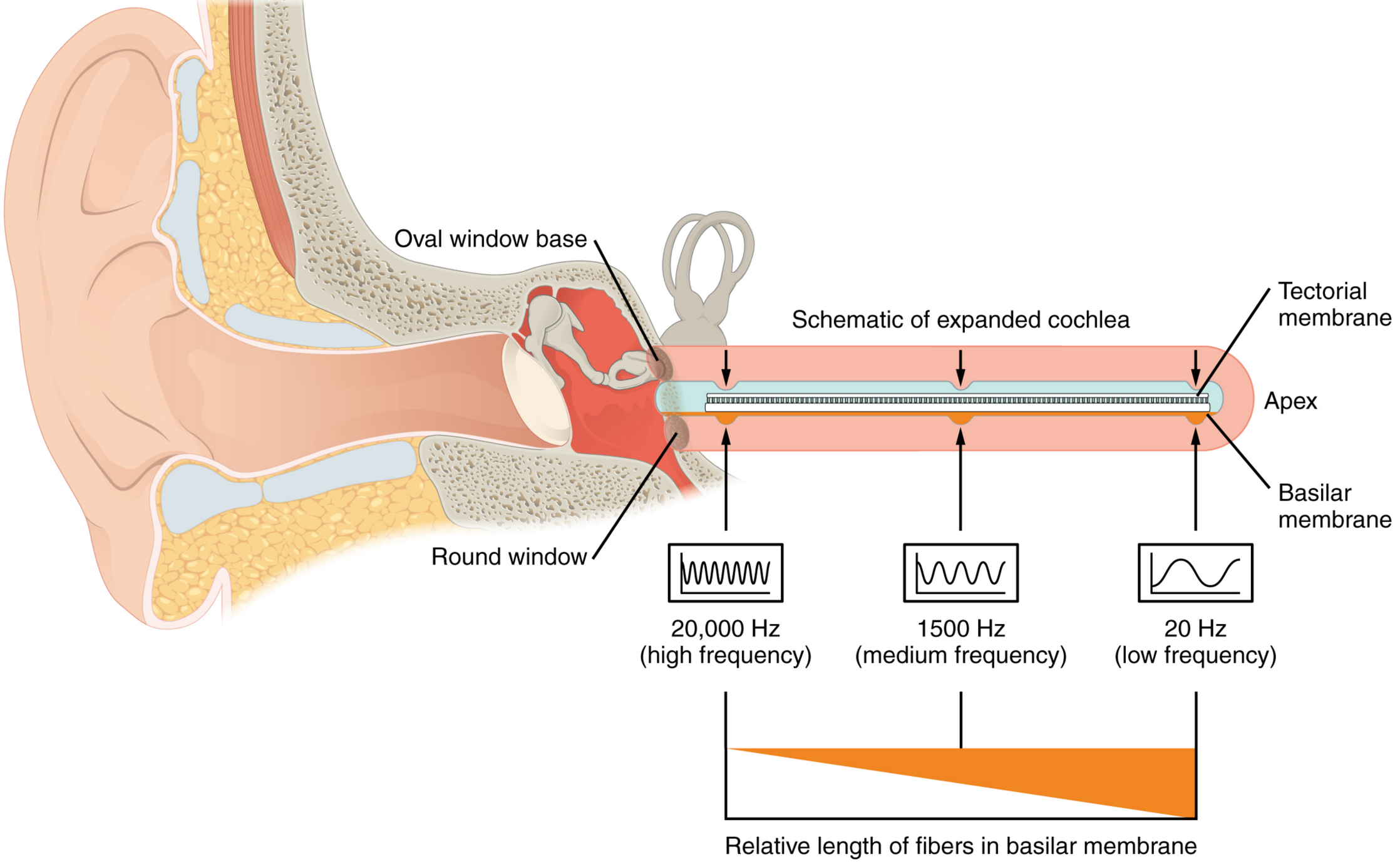
# How speech is produced and perceived

- Speech production
  - sound source, vocal tract
- Speech perception
  - the auditory system

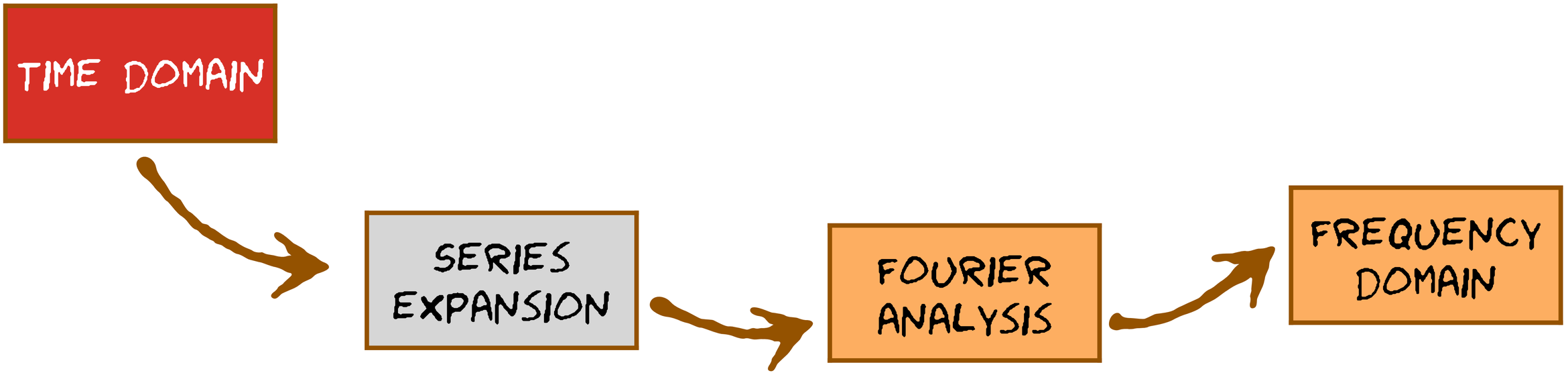




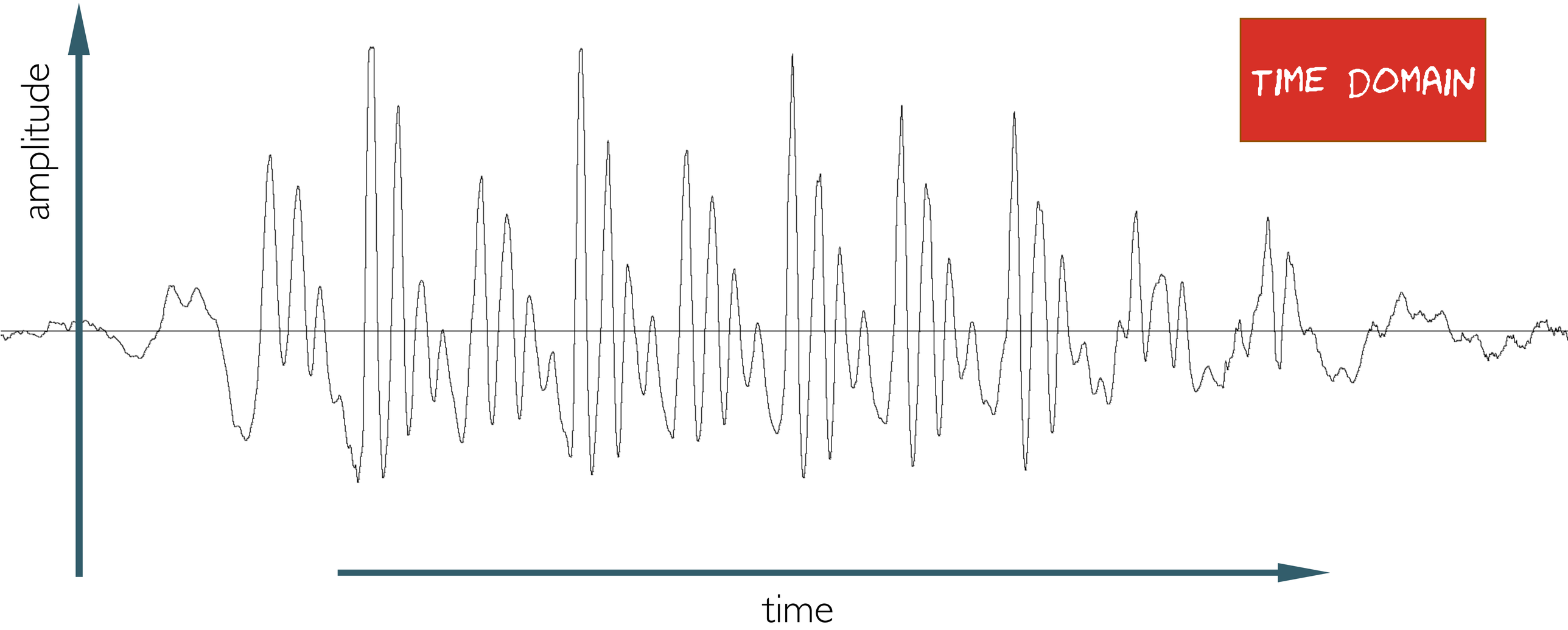




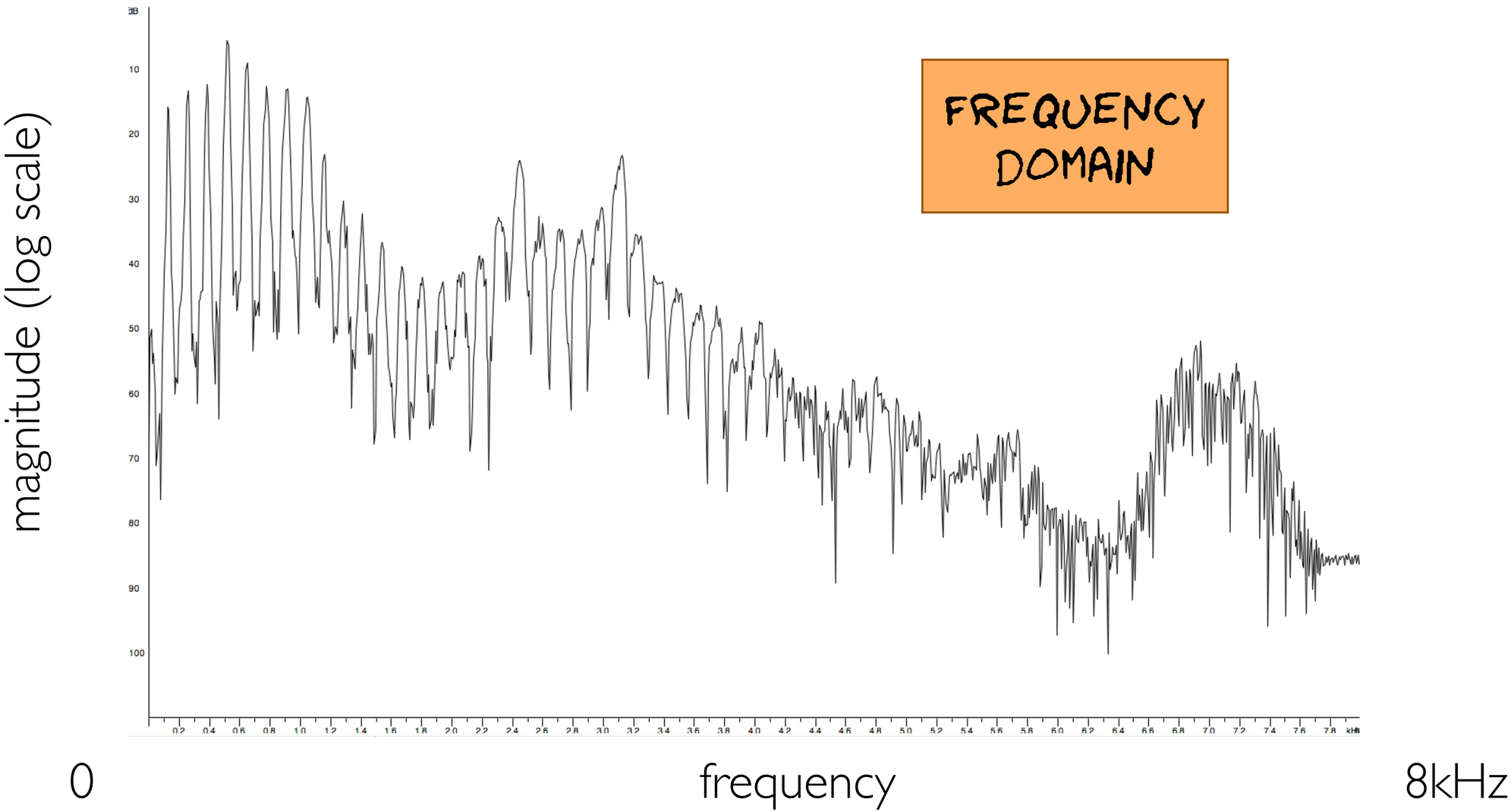
# Topics - from the time domain to the frequency domain



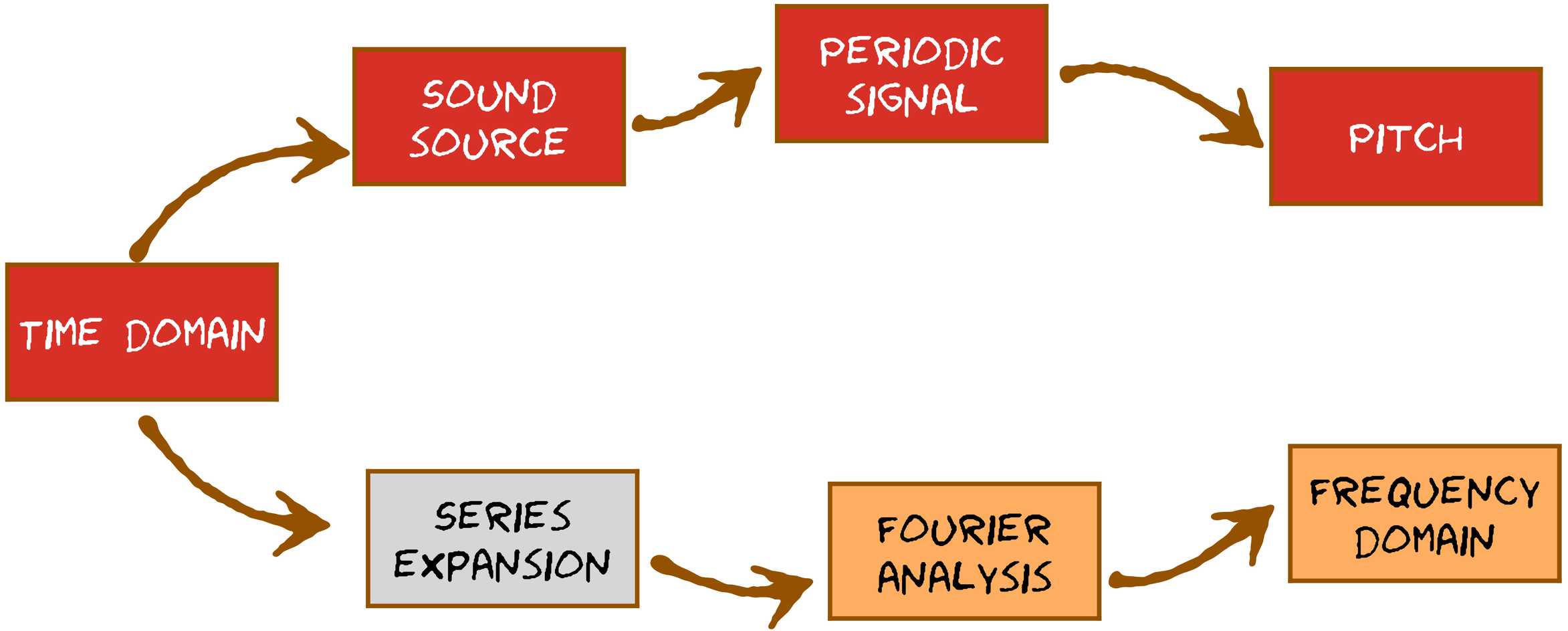
# Waveform



# Spectrum


magnitude (log scale)
dB
10
20
30
40
50
60
70
80
90
100
FREQUENCY DOMAIN
0
frequency
8kHz

# Today's topics - what we covered



# What next?

- Signals
  - Fourier analysis in more depth
  - harmonics
  - spectral envelope

In next week's foundation class

- The source-filter model
  - vocal tract resonance & formants
  - filter, impulse response

In Module 2