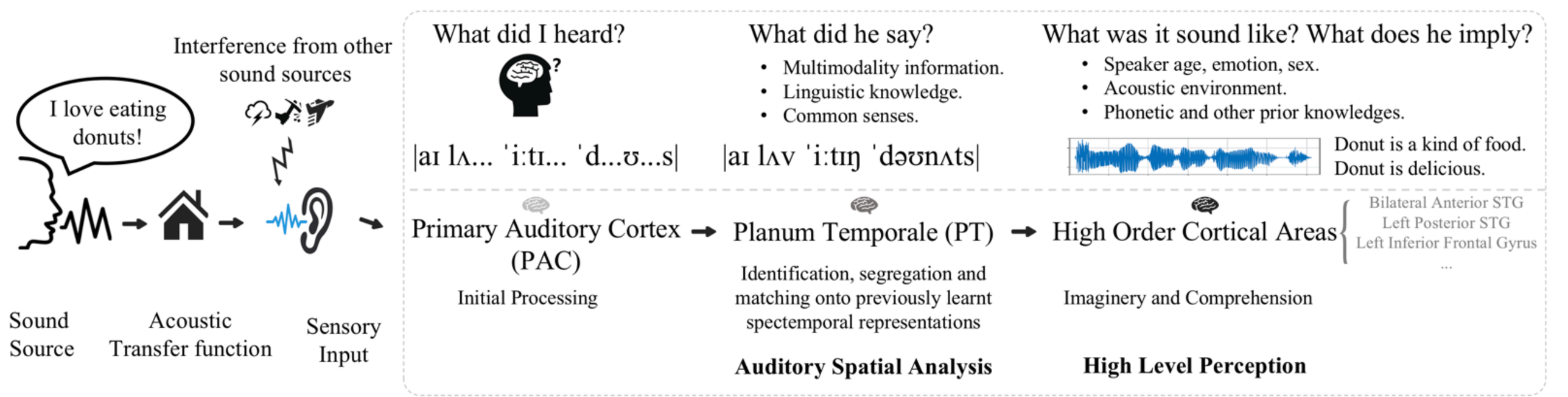
Motivation

How do we percept speech in a noisy environment?

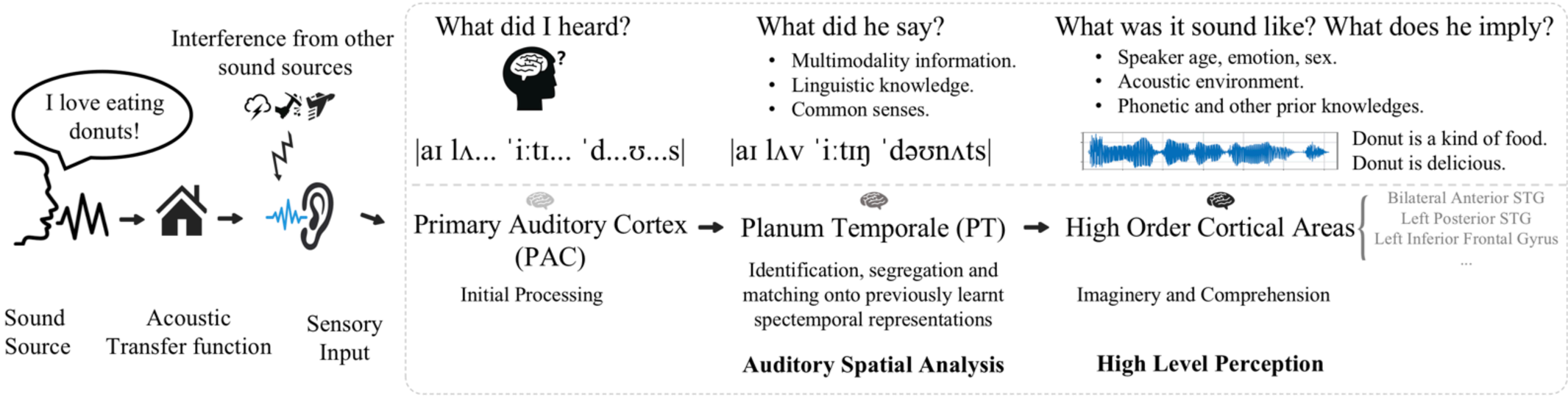

Interference from other sound sources
I love eating donuts!
Sound Source
Acoustic Transfer function
Sensory Input
What did I heard?
What did he say?
Multimodality information.
Linguistic knowledge.
Common senses.
What was it sound like? What does he imply?
Speaker age, emotion, sex.
Acoustic environment.
Phonetic and other prior knowledges.
|aɪ lʌ... ˈiːtɪ... ˈd...ʊ...s|
|aɪ lʌv ˈiːtɪŋ ˈdəʊnʌts|
Donut is a kind of food.
Donut is delicious.
Primary Auditory Cortex (PAC)
Initial Processing
Planum Temporale (PT)
Identification, segregation and matching onto previously learnt spectemporal representations
Auditory Spatial Analysis
High Order Cortical Areas
Imaginery and Comprehension
High Level Perception
Bilateral Anterior STG
Left Posterior STG
Left Inferior Frontal Gyrus
...

3. Can we use the similar two stage processing?

Answer: Yes, we can divide it into two sub-tasks.

Motivation

How do we percept speech in a noisy environment?

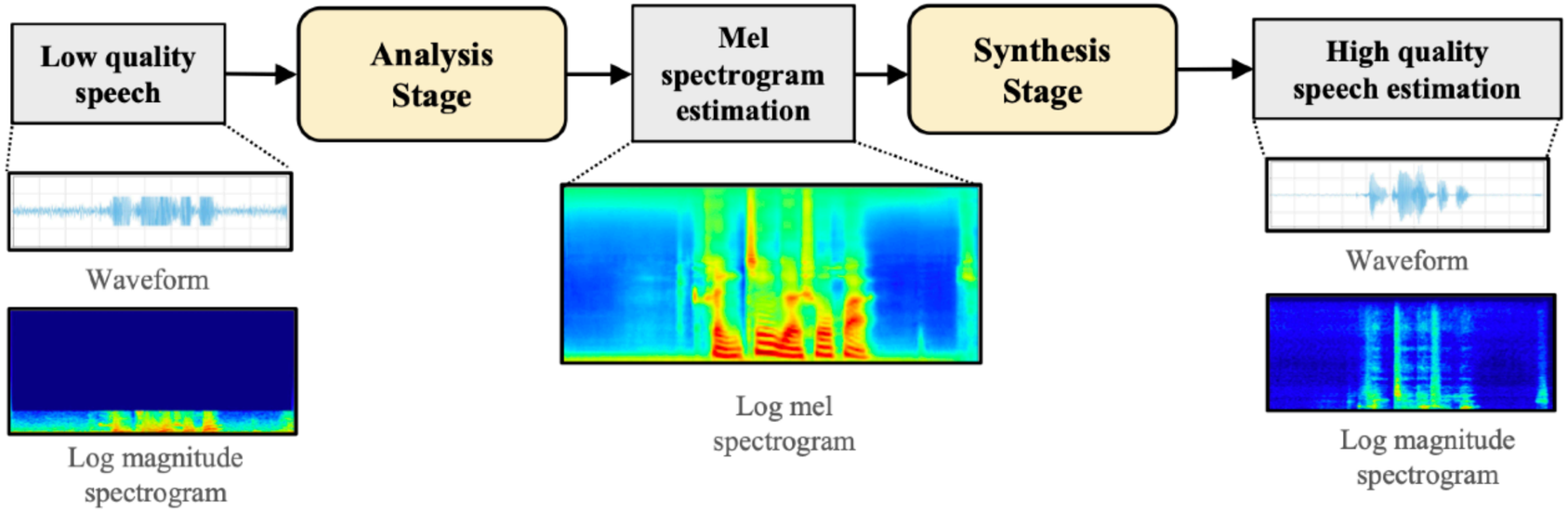


3. Can we use the similar two stage processing?

Answer: Yes, we can divide it into two sub-tasks.

VoiceFixer

Overview



- Two Stage Processing $f : x \rightarrow z, z \rightarrow \hat{s}$
  - Analysis Module: Input distorted mel spectrogram and output restored mel spectrogram.
  - Synthesis Module: Input restored mel spectrogram and output waveform