

Disclaimer

- The material provided in this document is not my original work and is a summary of some one else's work(s).
- A simple Google search of the title of the document will direct you to the original source of the material.
- I do not guarantee the accuracy, completeness, timeliness, validity, non-omission, merchantability or fitness of the contents of this document for any particular purpose.
- Downloaded from najeebkhan.github.io

Text to Speech Synthesis

By
Najeeb
Speech Signal Processing Lab
2013-12-6

Outline



Outline

- Speech
- Text to Speech Synthesis
- Applications
- How TTS works
- Demos of Commercial TTS systems
- Singing Voice Synthesis
- Music Score Editor for SVS



Speech



Speech

Levels of Speech

6. Pragmatic

5. Semantic

4. Syntactic

3. Morphological

2. Phonetic

1. Acoustic

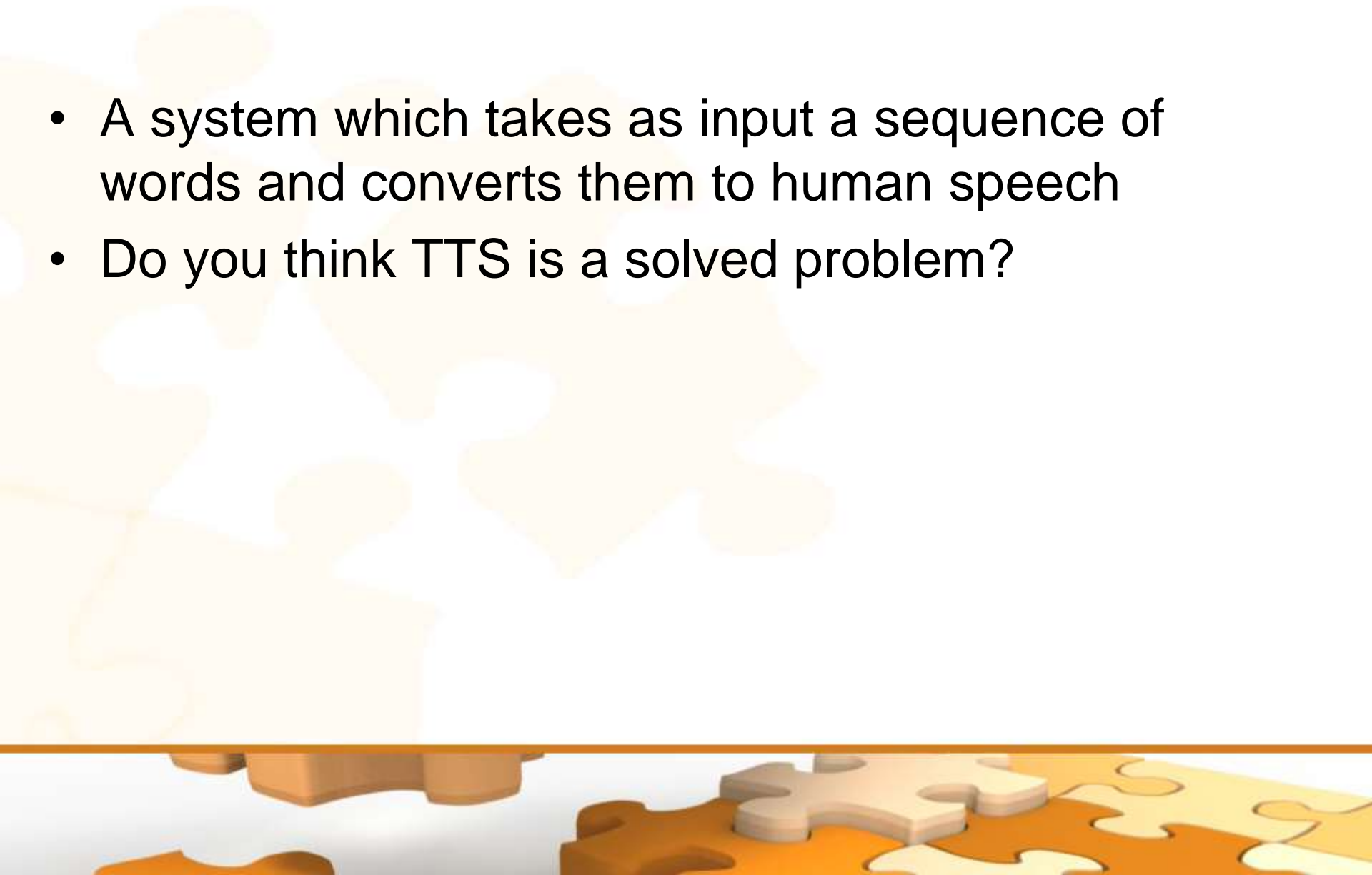


Text To Speech Synthesis



Text To Speech Synthesis

- A system which takes as input a sequence of words and converts them to human speech
- Do you think TTS is a solved problem?



Applications



Applications

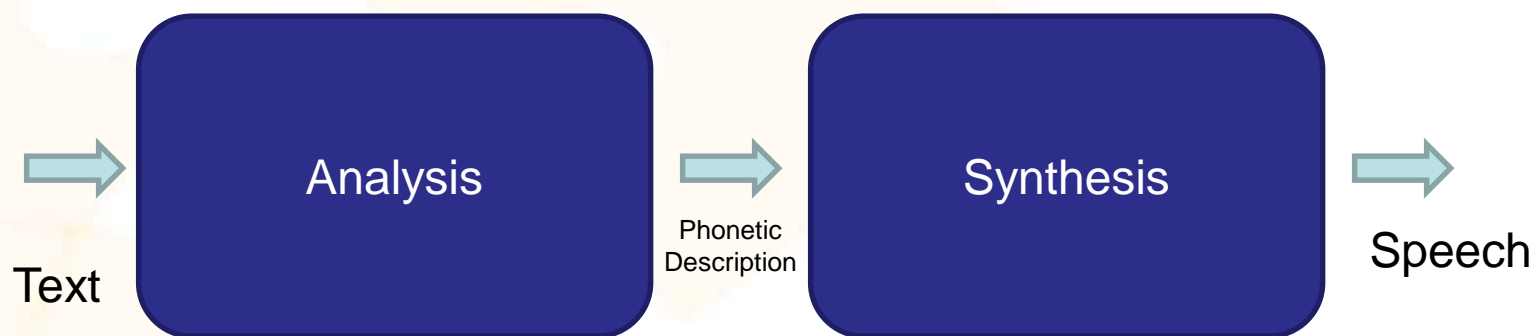
- For applications involving access to very large or rapidly changing databases an TTS system driven from a text database has many advantages
 - Proof-Reading of documents
 - Speaking and Reading aids for the disabled
 - Speech output for intelligent machines



Functional outline of a TTS system



Functional outline of a TTS system



Functional outline of a TTS system

Analysis

Symbols to
Standard
Form

Phonetic
Transcription

Parsing

Semantic
Analysis

Functional outline of a TTS system

Symbols to Standard Form

A preprocessor is used to convert symbol strings such as \$3.17B to text

Analysis

Symbols to
Standard
Form

Phonetic
Transcription

Parsing

Semantic
Analysis

Functional outline of a TTS system

Analysis

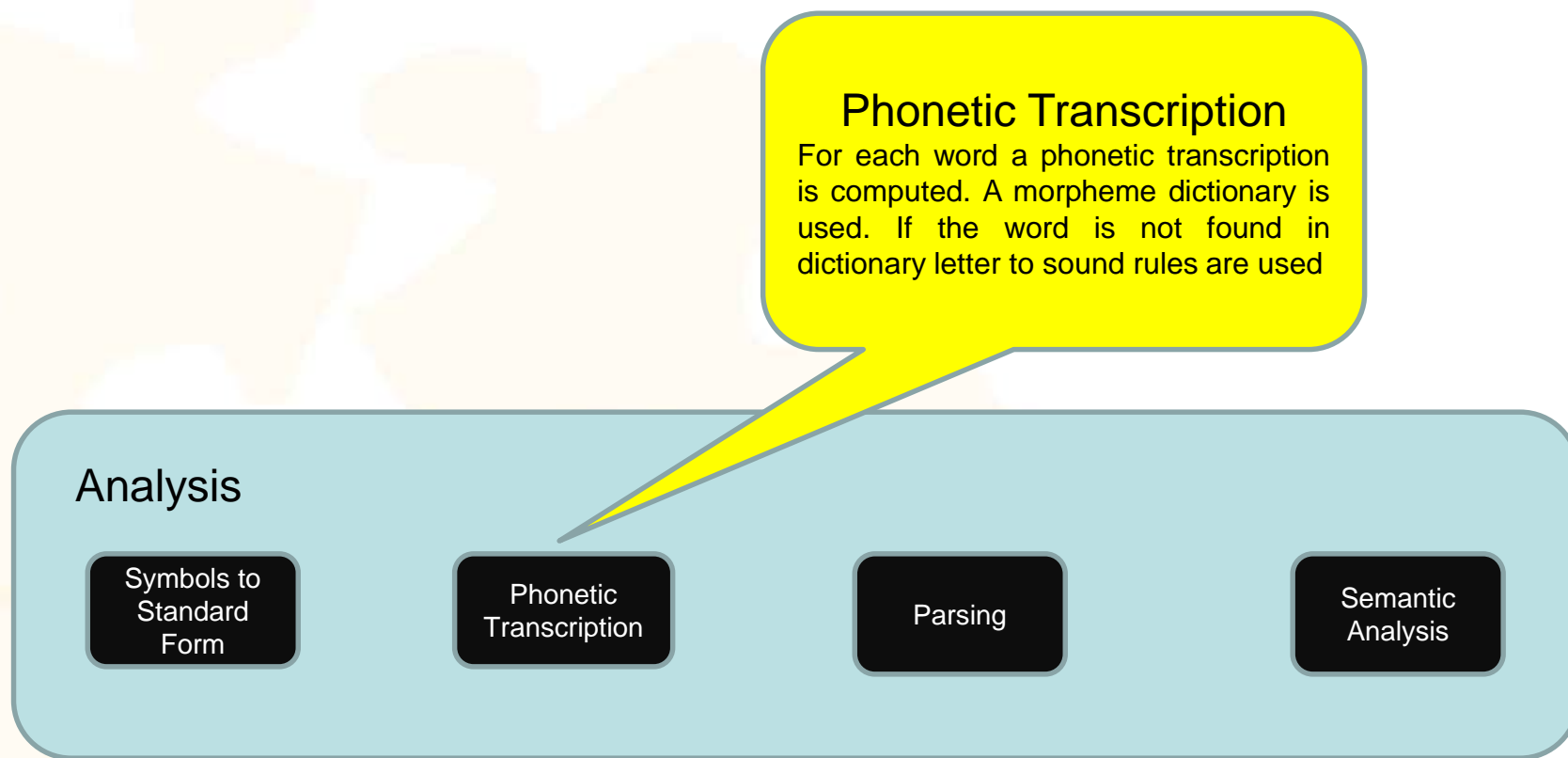
Symbols to
Standard
Form

Phonetic
Transcription

Parsing

Semantic
Analysis

Functional outline of a TTS system



Functional outline of a TTS system

Analysis

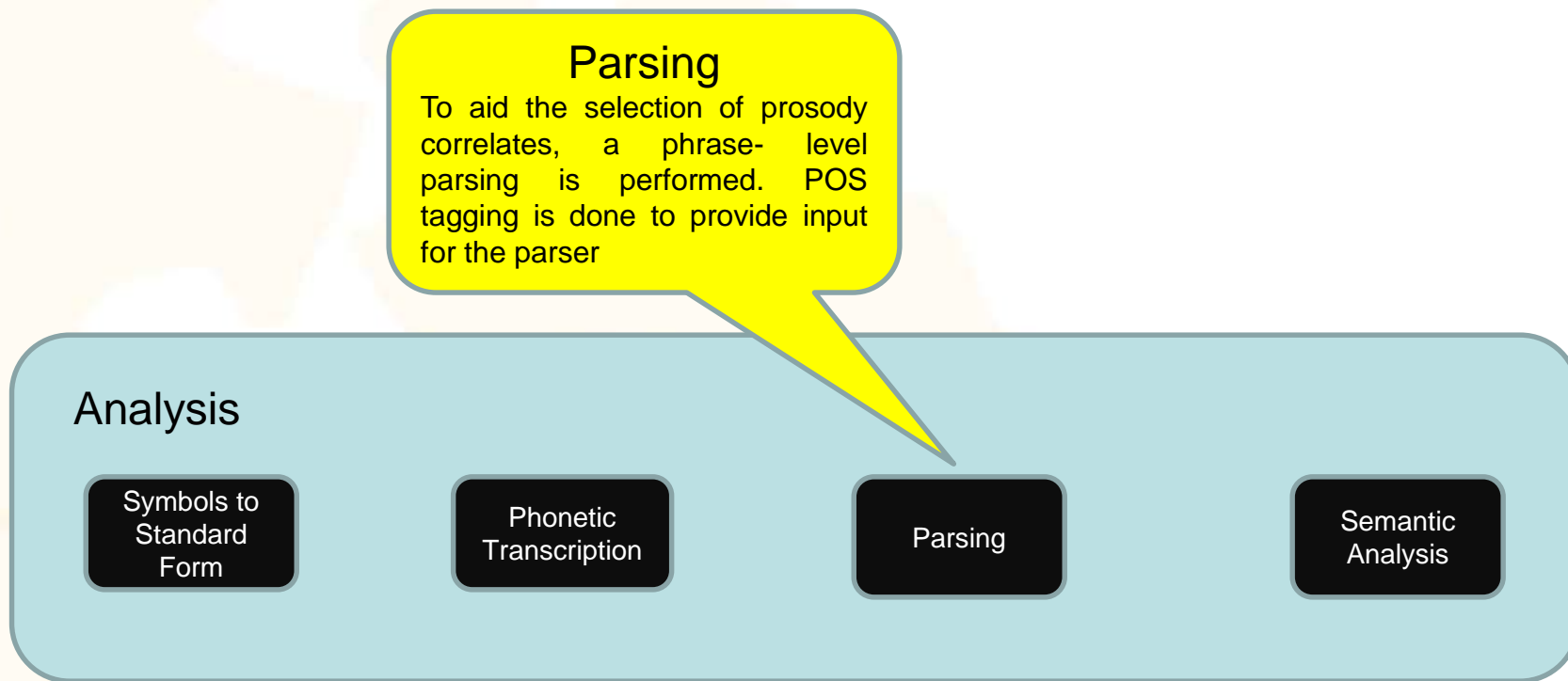
Symbols to
Standard
Form

Phonetic
Transcription

Parsing

Semantic
Analysis

Functional outline of a TTS system



Functional outline of a TTS system

Analysis

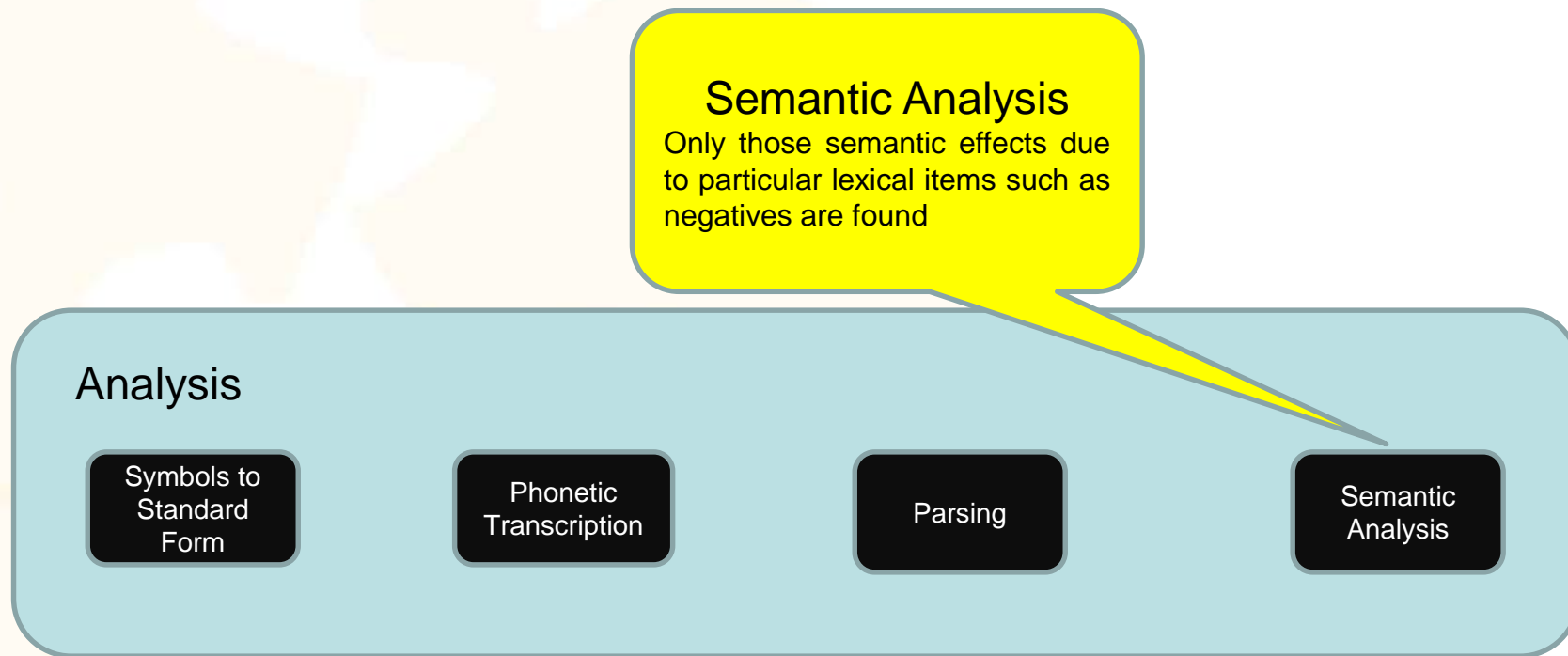
Symbols to
Standard
Form

Phonetic
Transcription

Parsing

Semantic
Analysis

Functional outline of a TTS system



Functional outline of a TTS system

Analysis

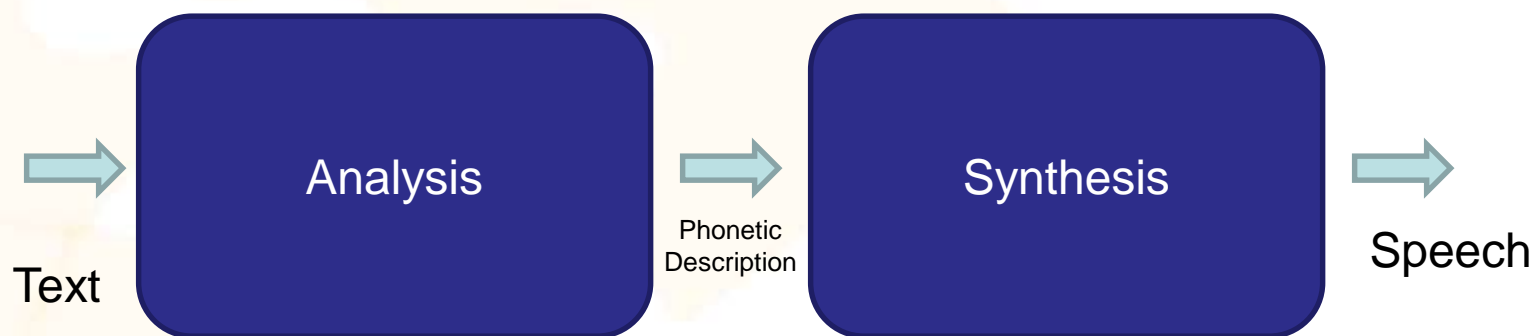
Symbols to
Standard
Form

Phonetic
Transcription

Parsing

Semantic
Analysis

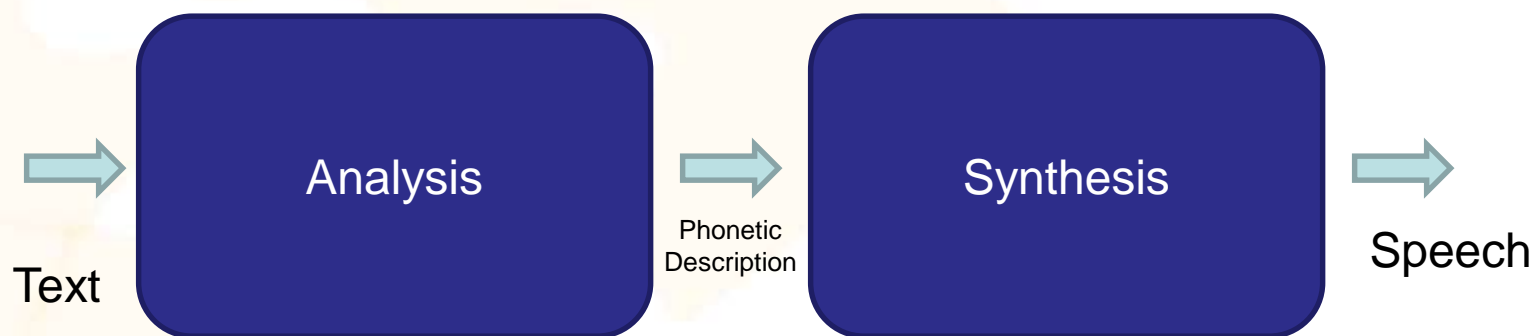
Functional outline of a TTS system



Functional outline of a TTS system



Functional outline of a TTS system



Functional outline of a TTS system

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system

Timing

Prepausal duration and polysyllabic shortening are determined plus the basic duration of each segment

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system

Synthesis

Timing

Fundamental
Frequency

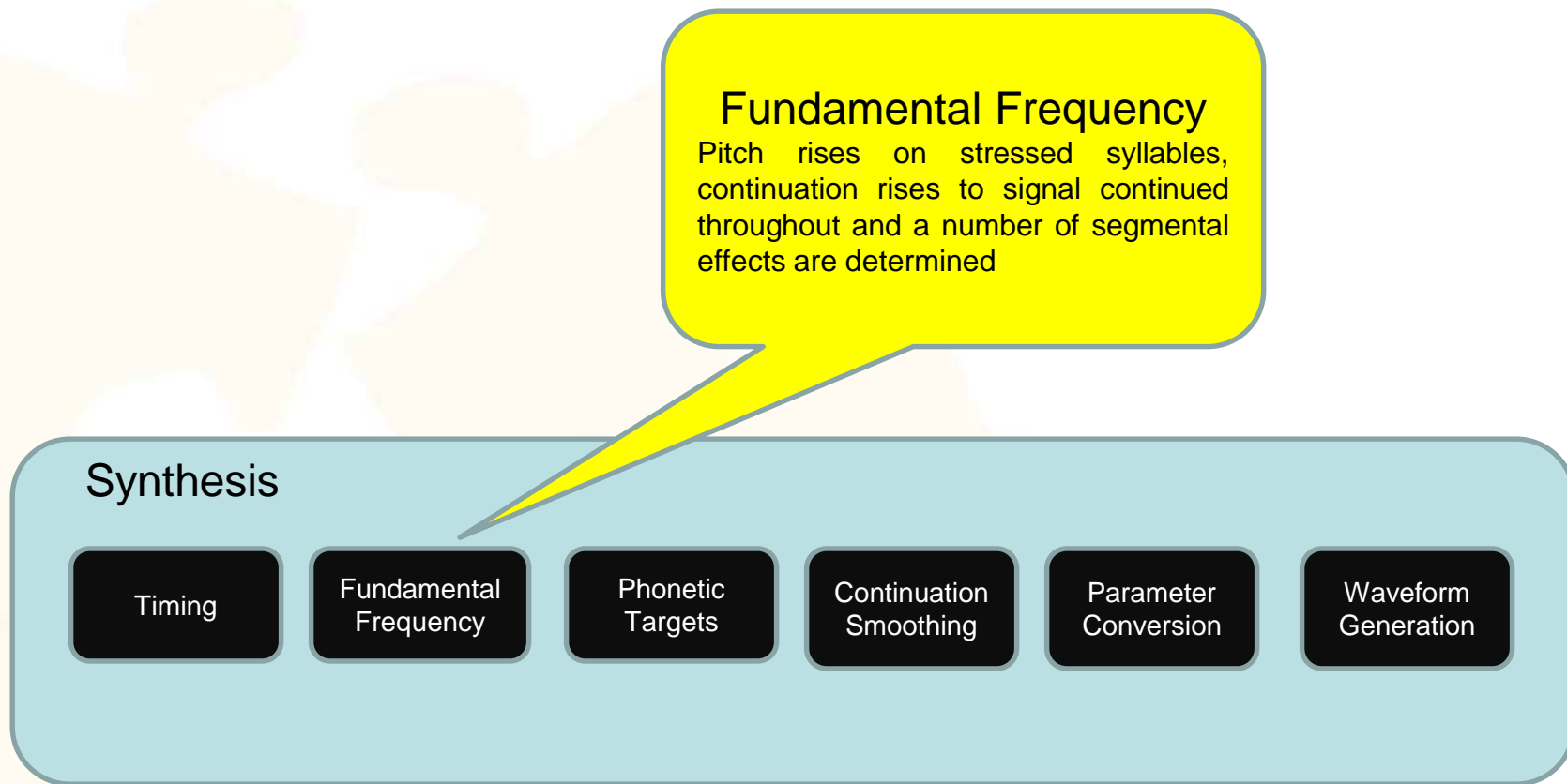
Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system



Functional outline of a TTS system

Synthesis

Timing

Fundamental
Frequency

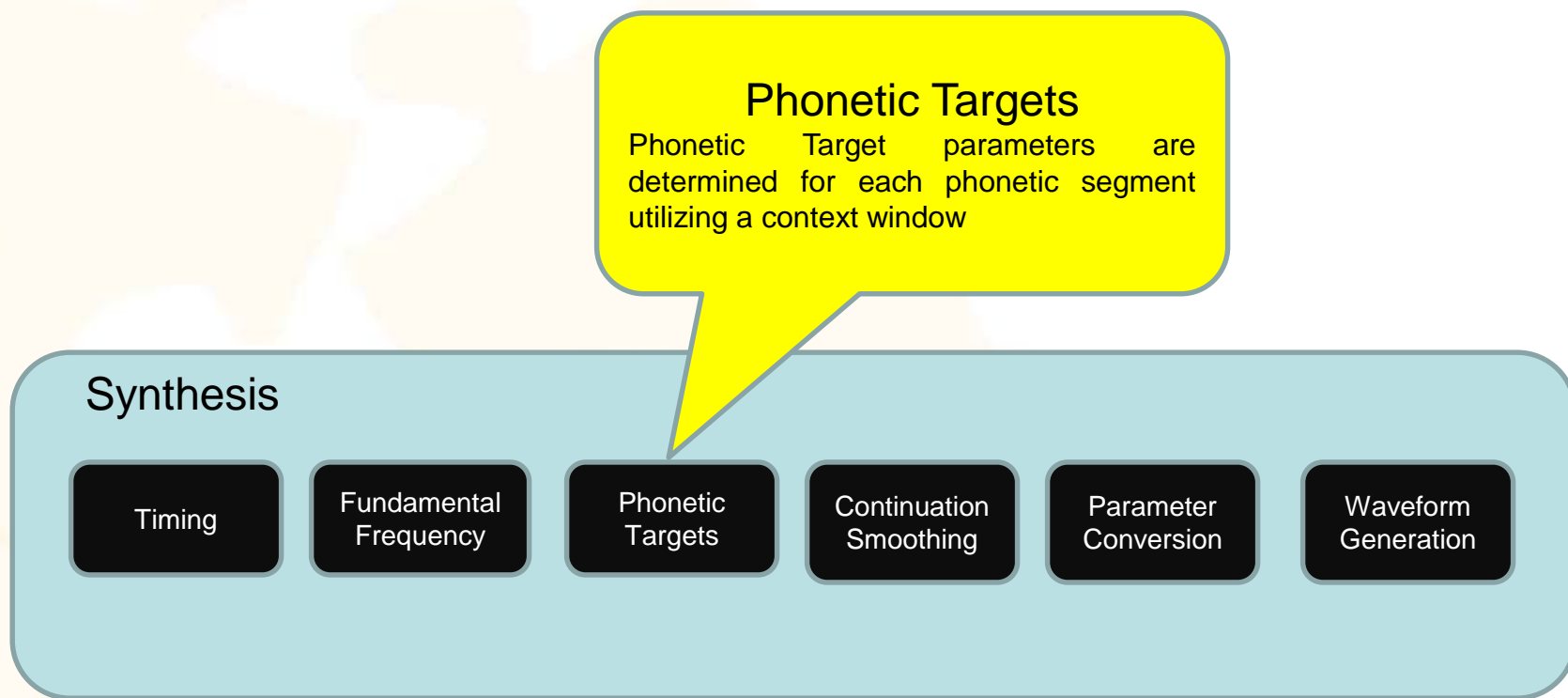
Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system



Functional outline of a TTS system

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system

Continuation Smoothing

The target values are smoothed to get a full set of parameters every frame

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system

Synthesis

Timing

Fundamental
Frequency

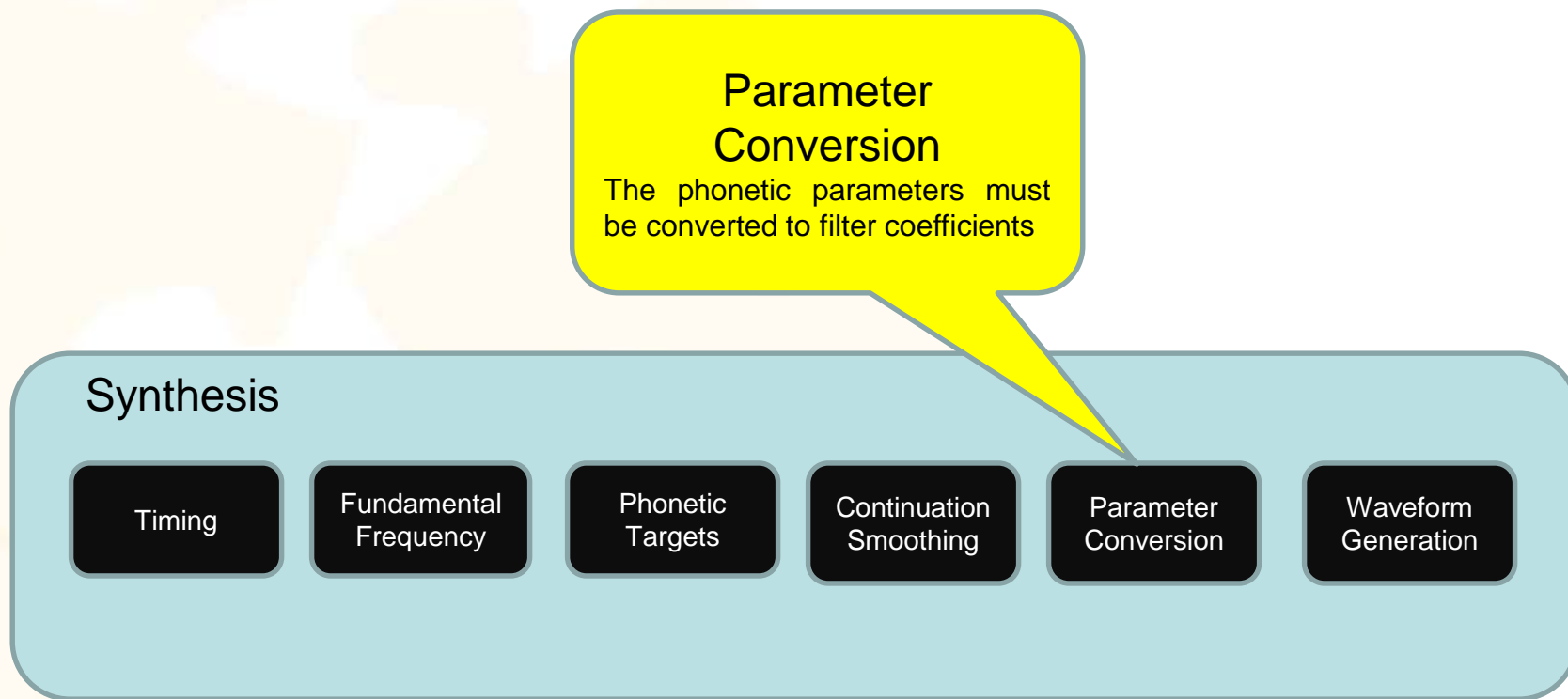
Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system



Functional outline of a TTS system

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system

Waveform Generation

The synthesizer utilizes the coefficients to generate the speech waveform

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system

Synthesis

Timing

Fundamental
Frequency

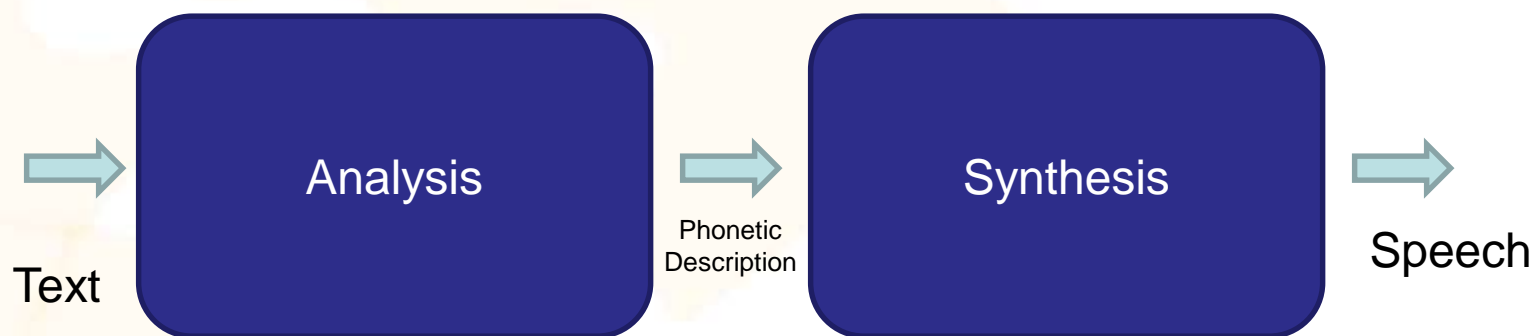
Phonetic
Targets

Continuation
Smoothing

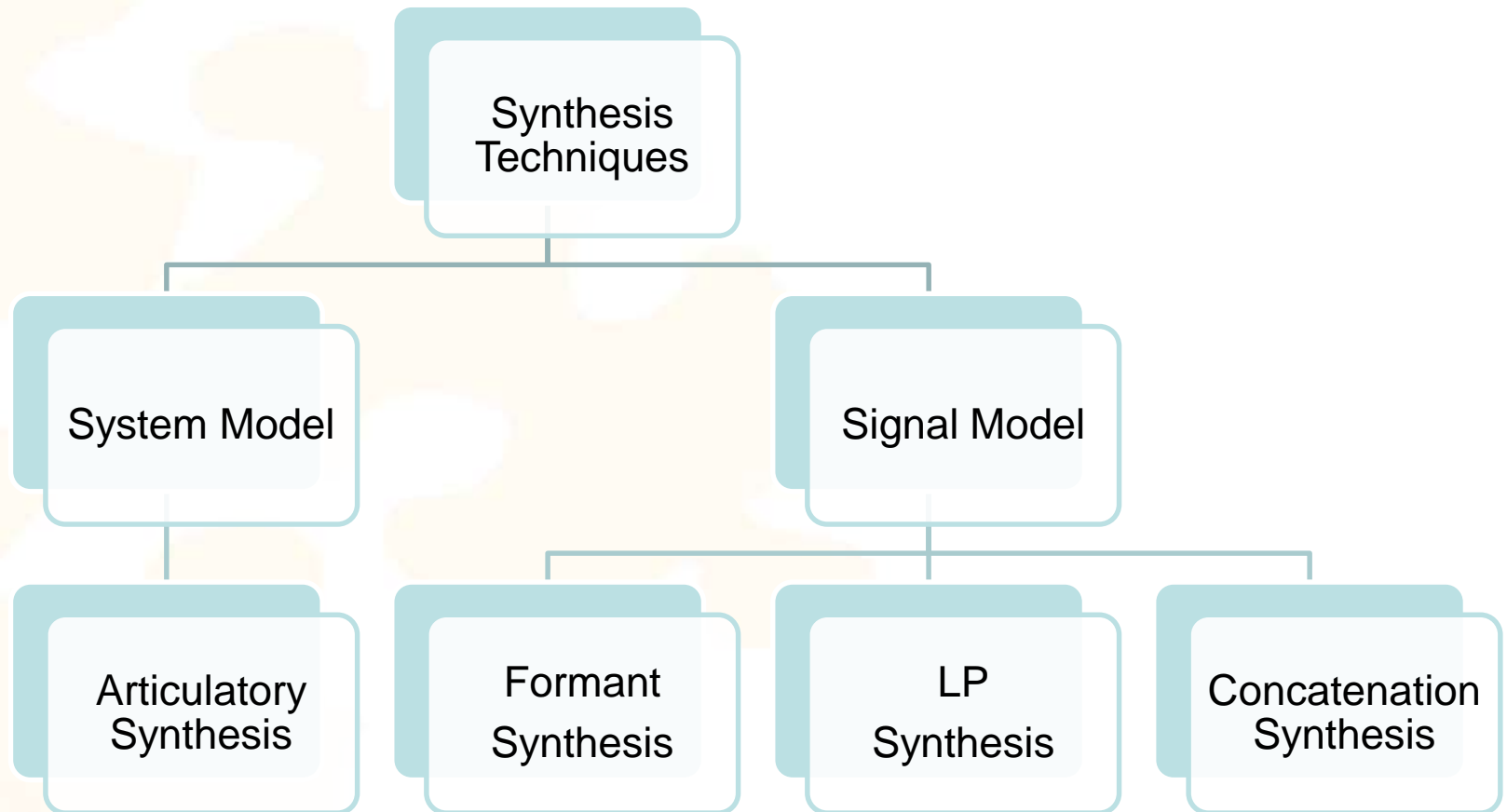
Parameter
Conversion

Waveform
Generation

Functional outline of a TTS system

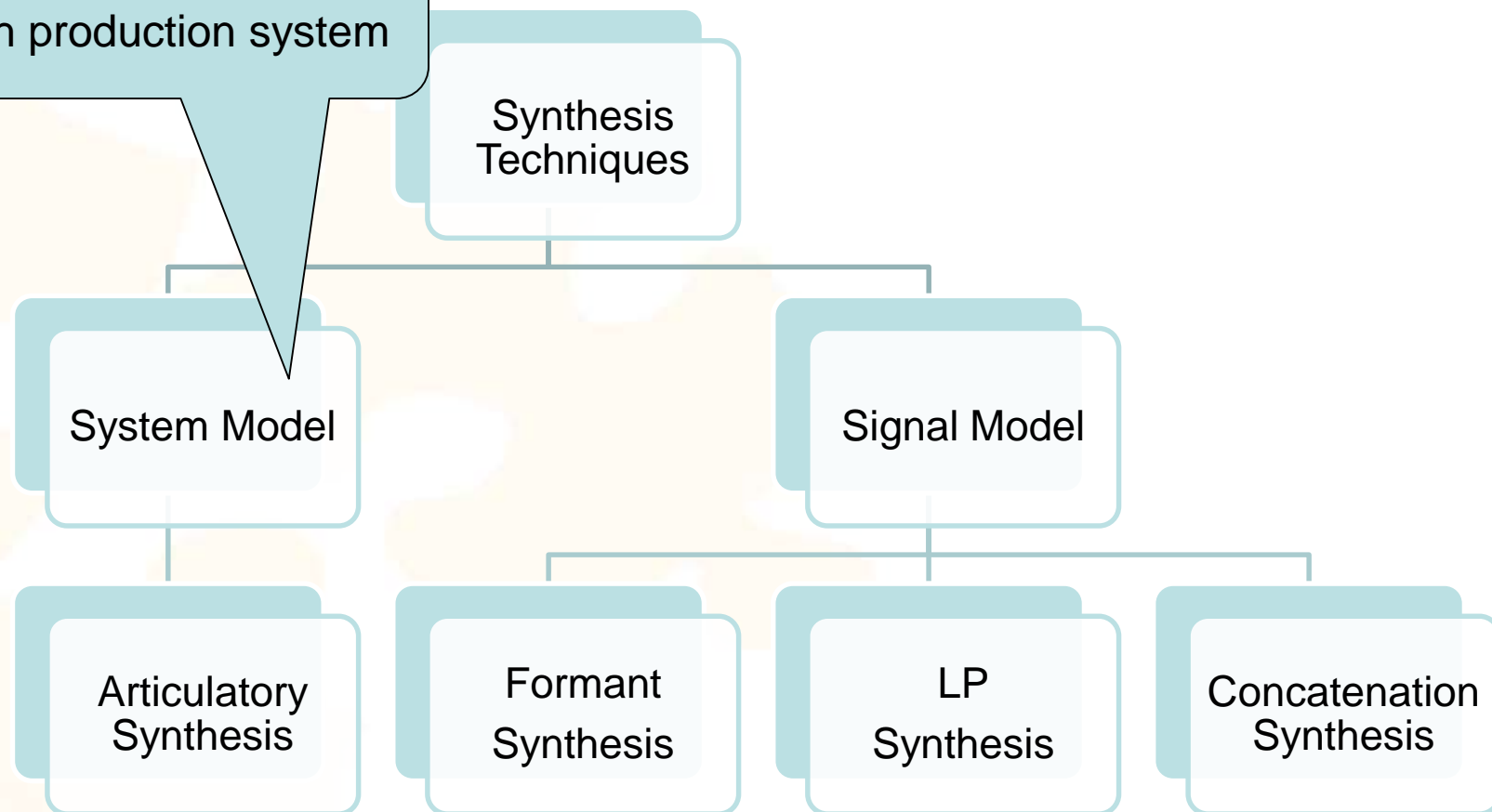


Synthesis Techniques

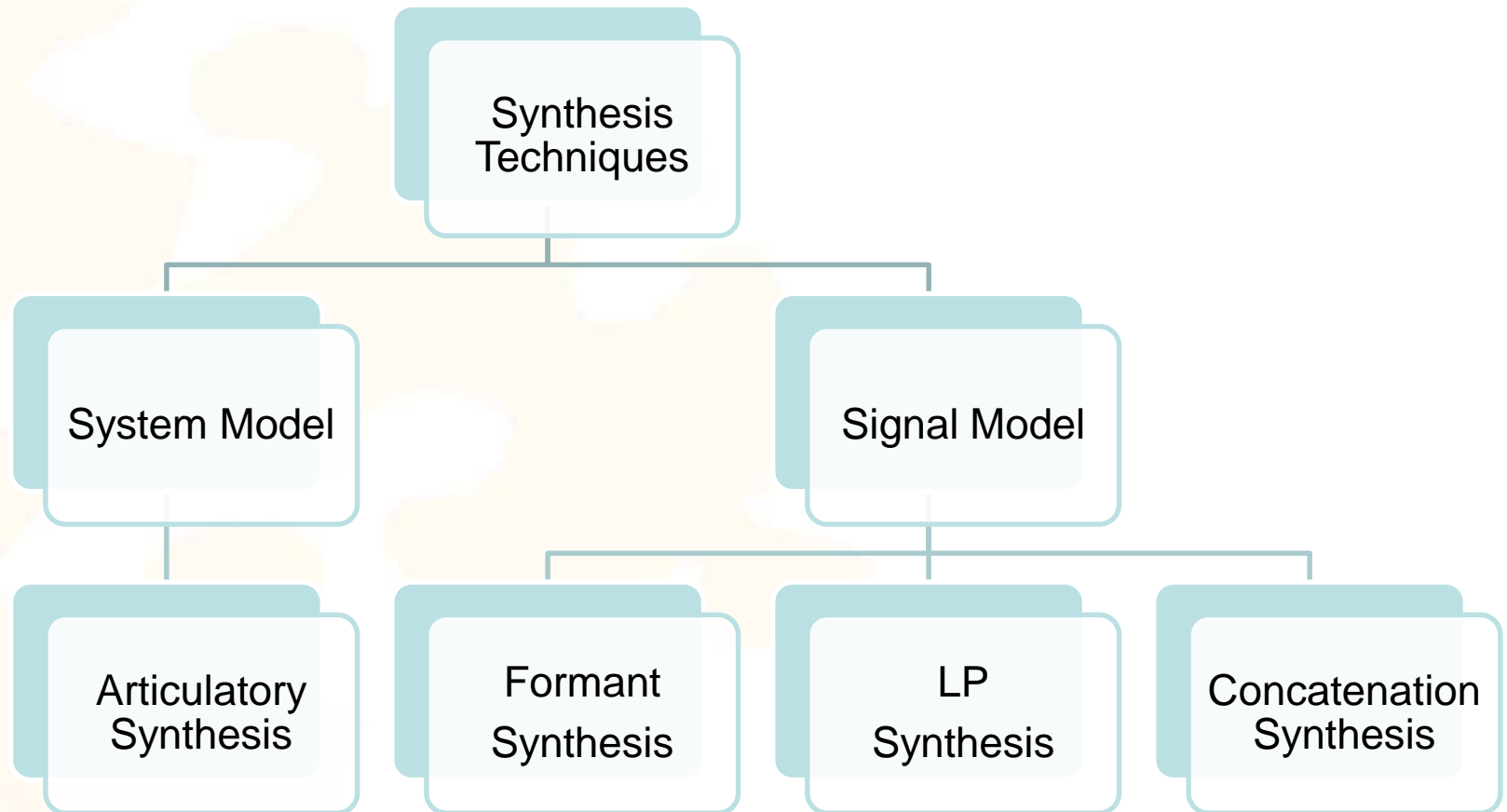


Synthesis Techniques

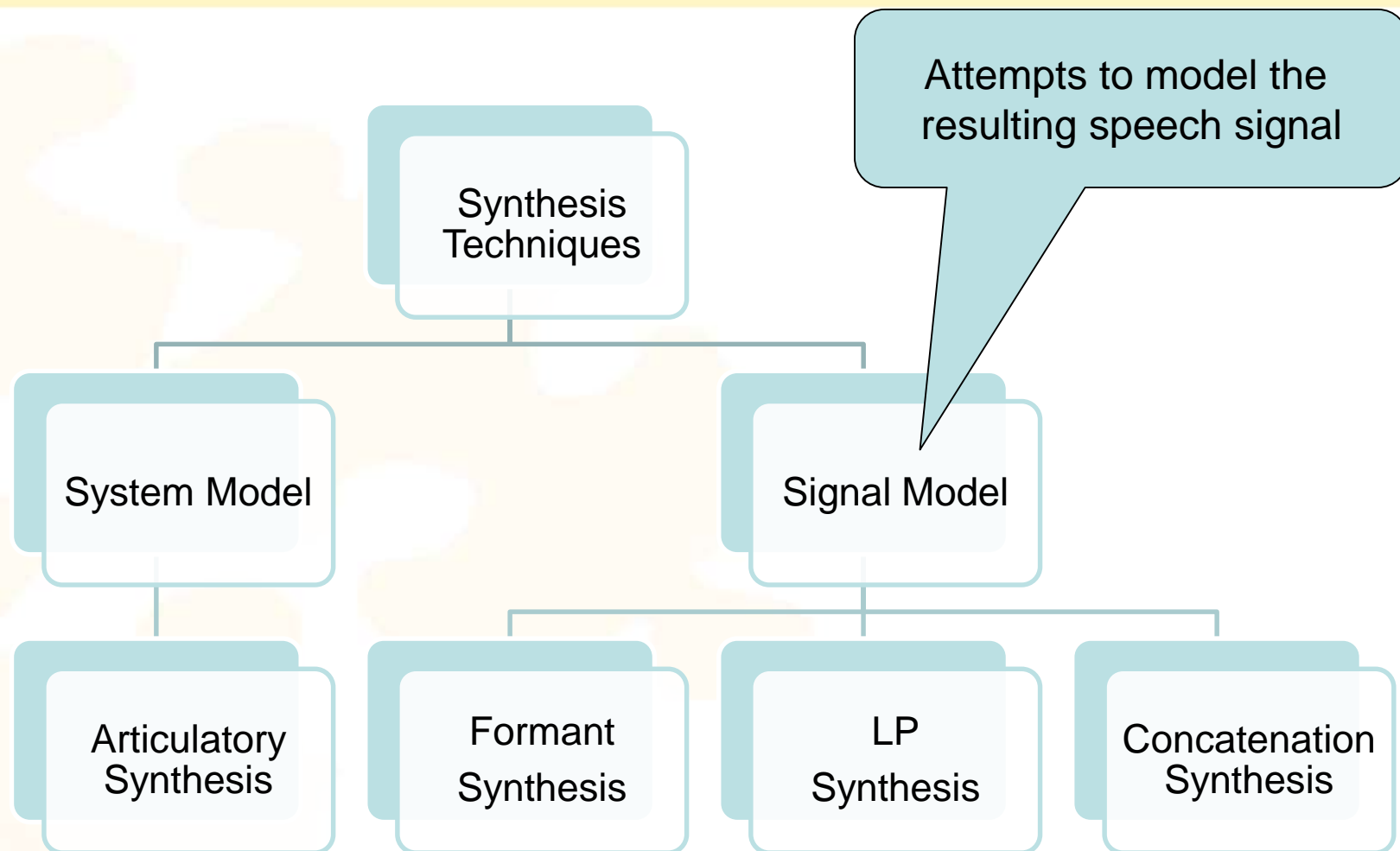
Attempts to model the human speech production system



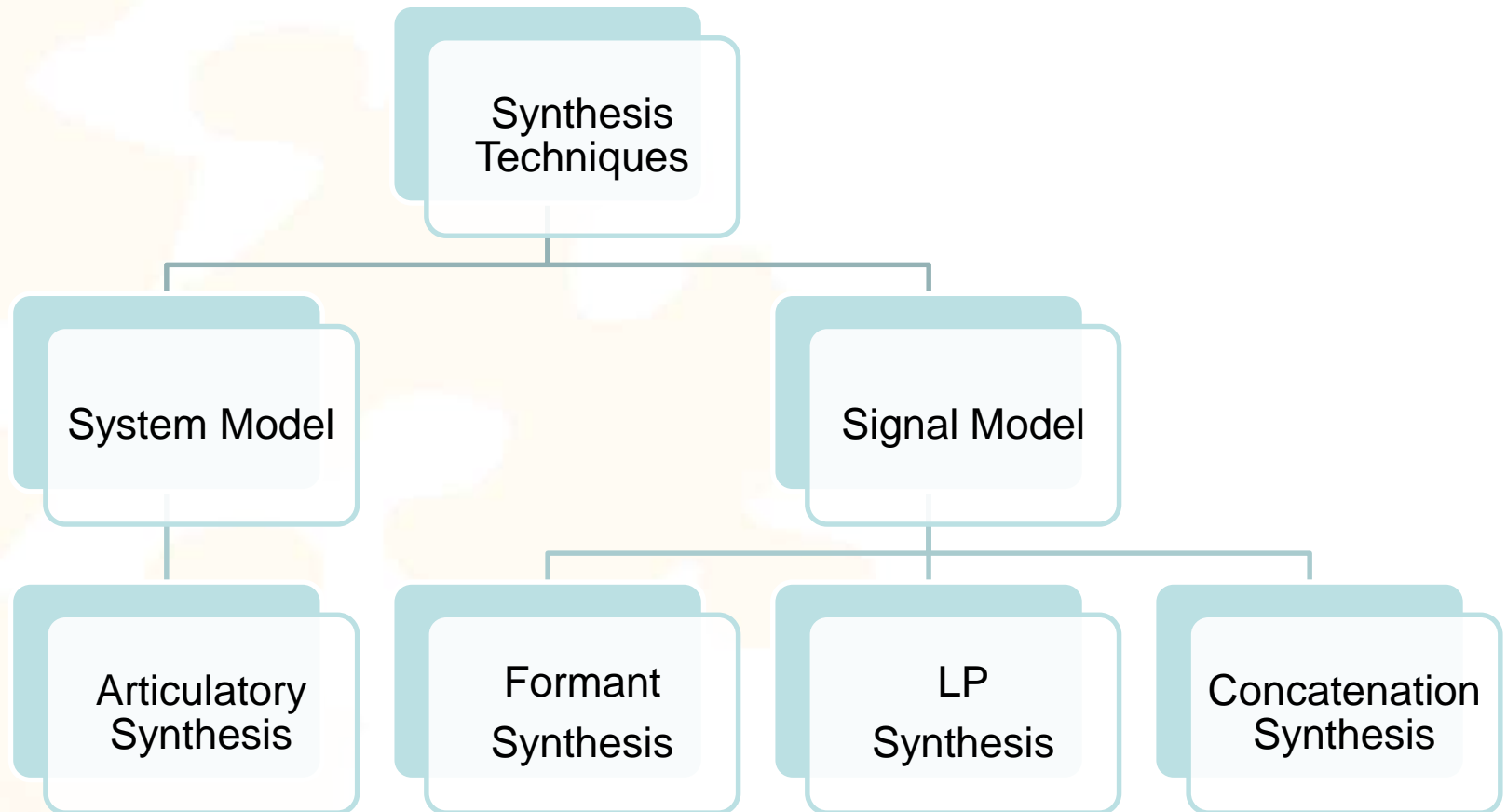
Synthesis Techniques



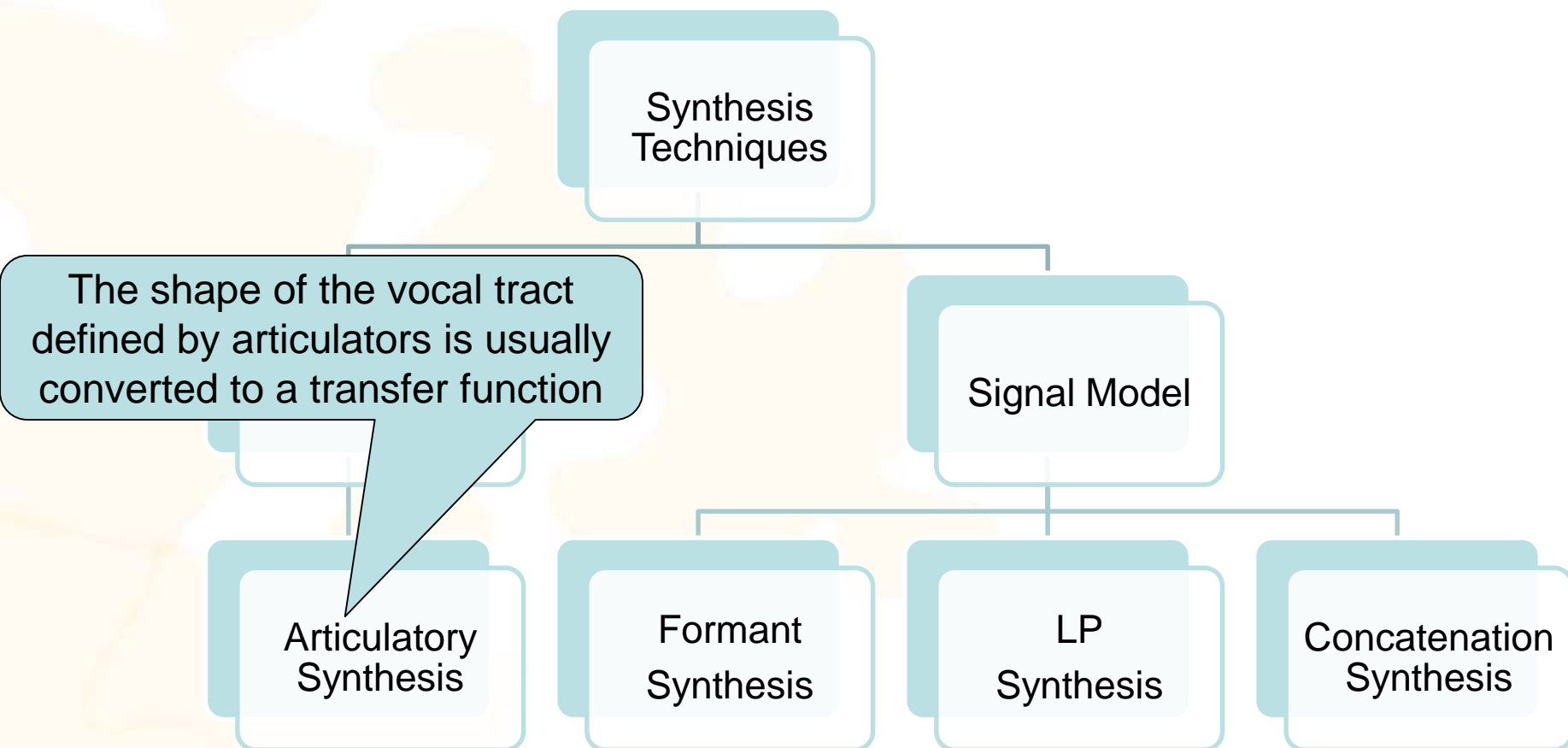
Synthesis Techniques



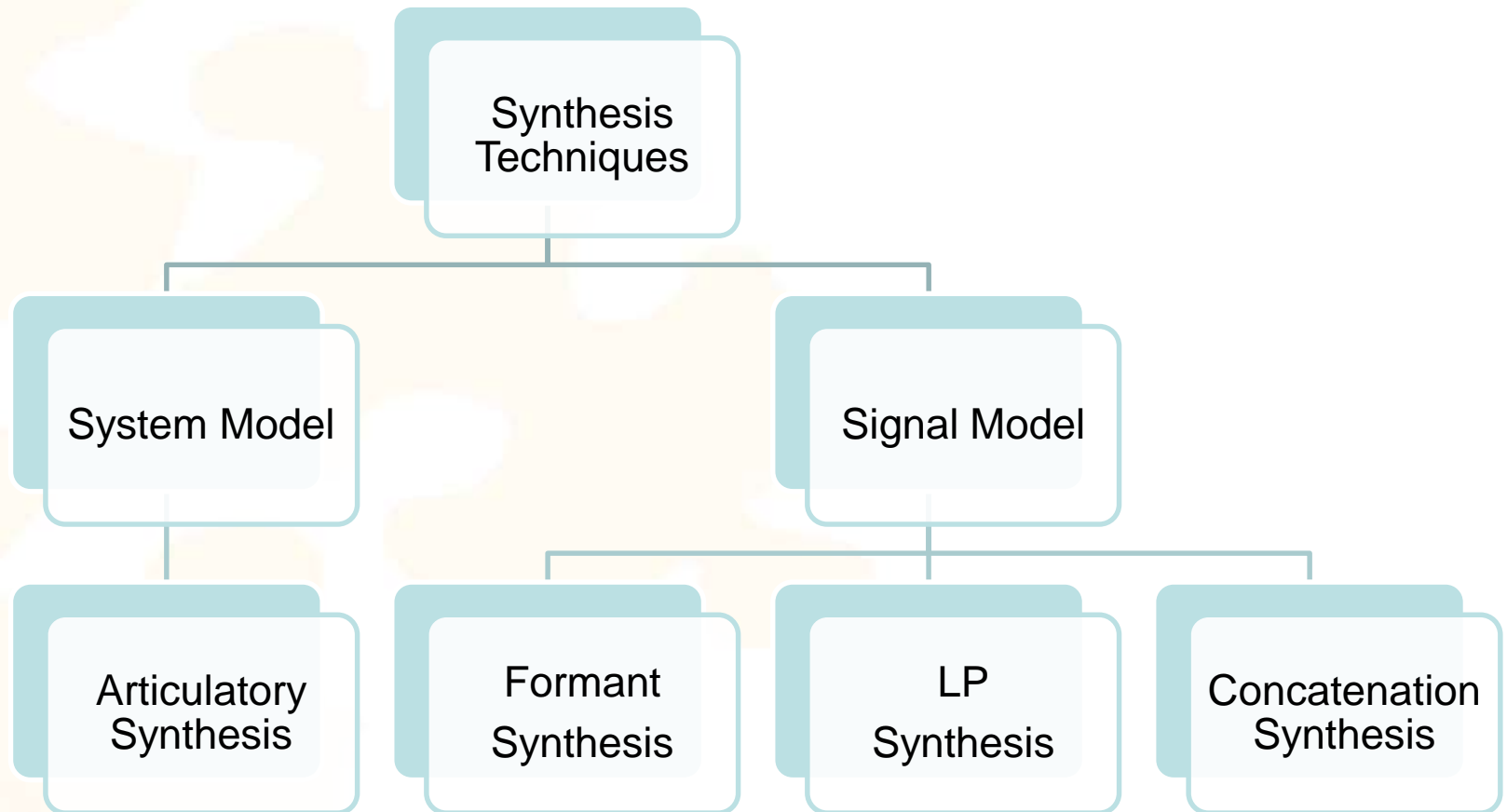
Synthesis Techniques



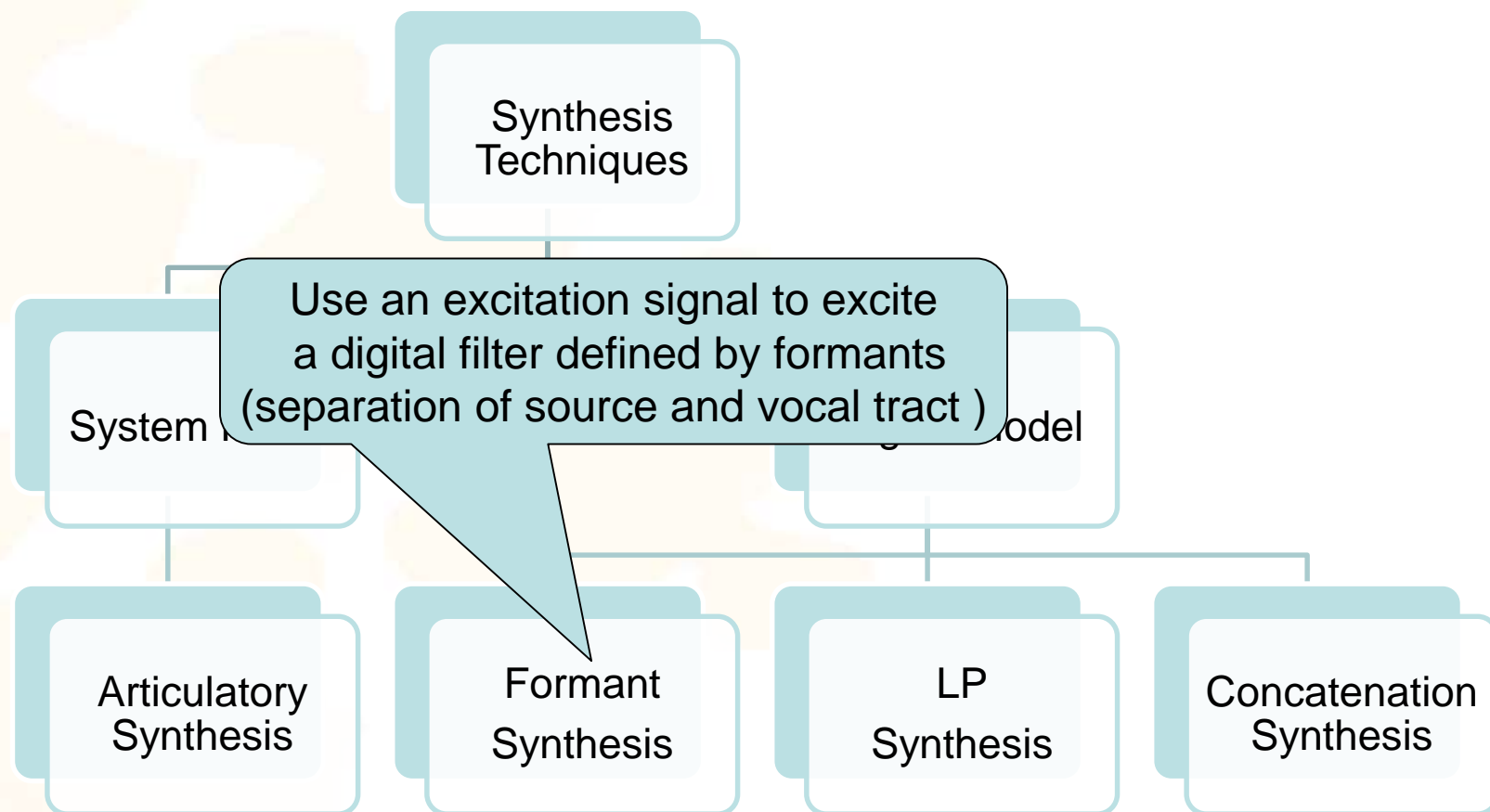
Synthesis Techniques



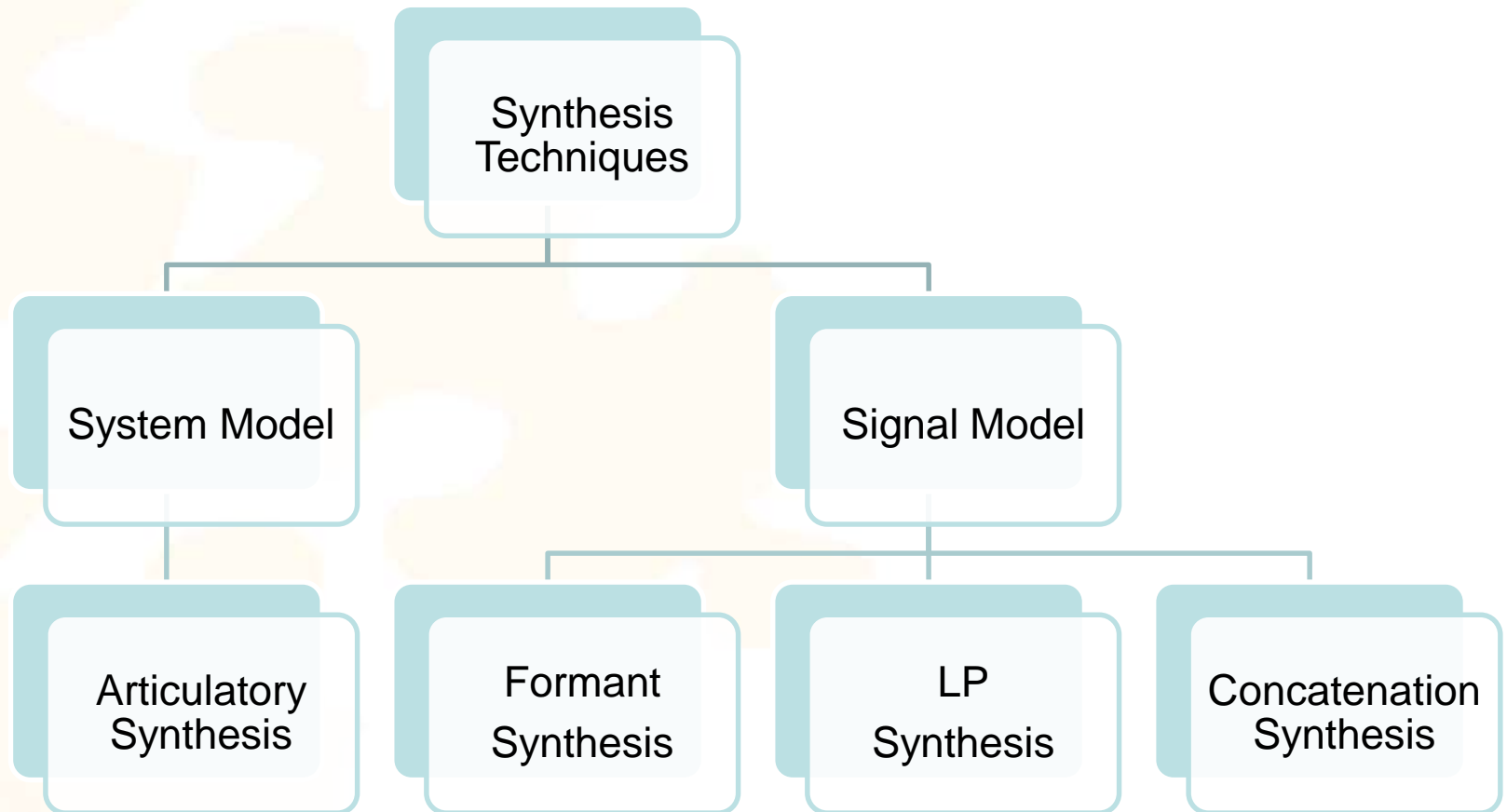
Synthesis Techniques



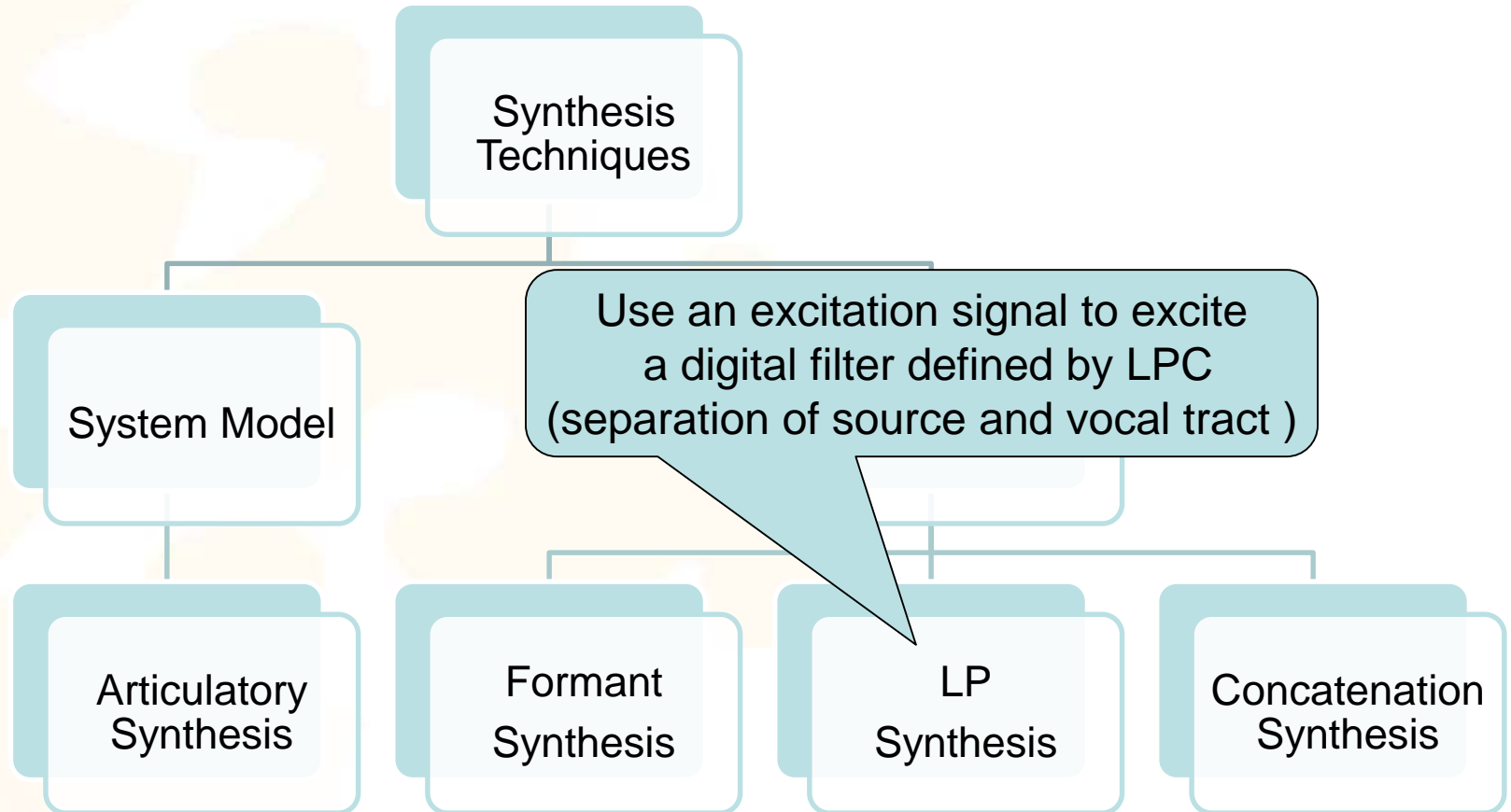
Synthesis Techniques



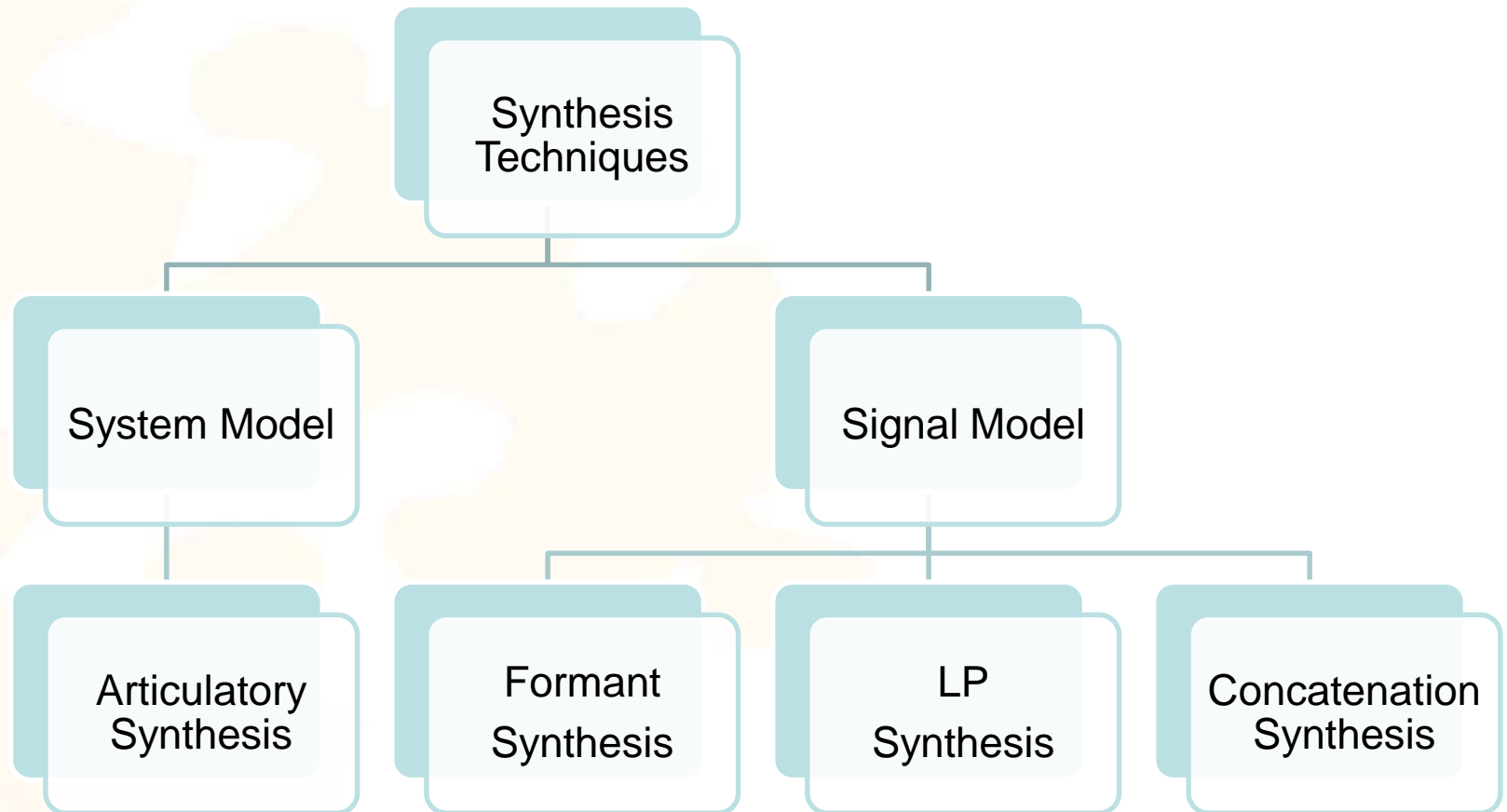
Synthesis Techniques



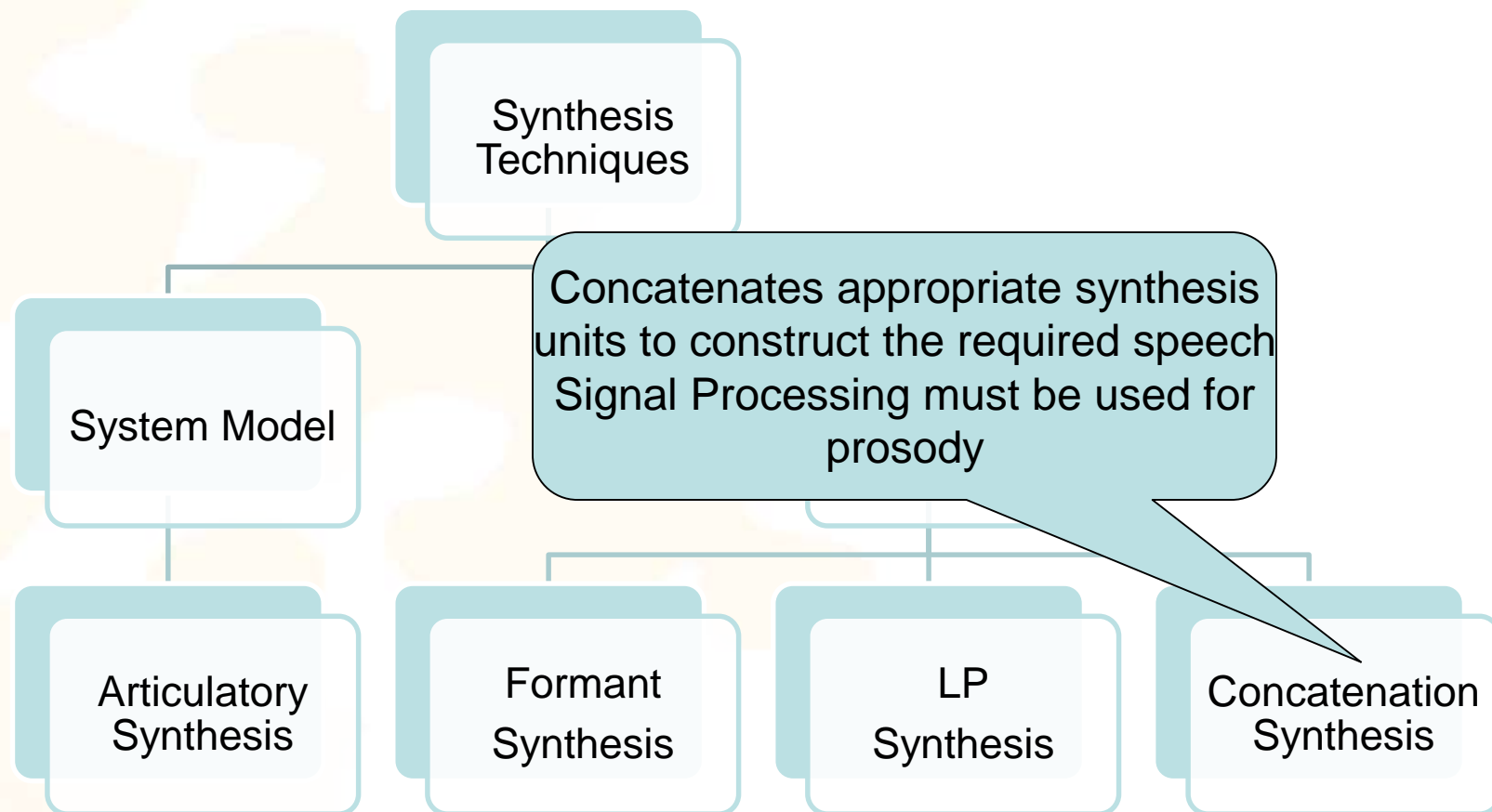
Synthesis Techniques



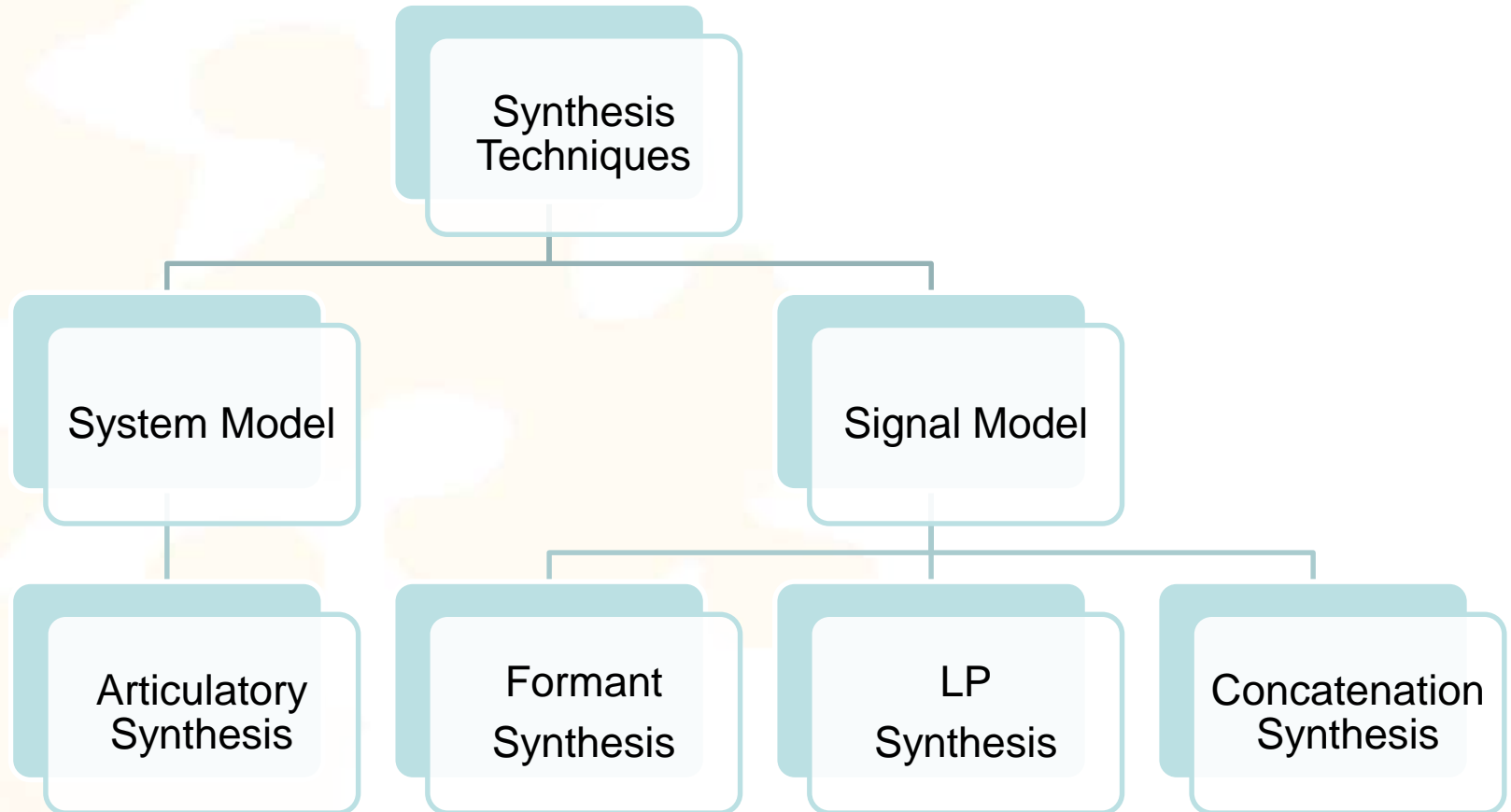
Synthesis Techniques



Synthesis Techniques



Synthesis Techniques



TTS Demo

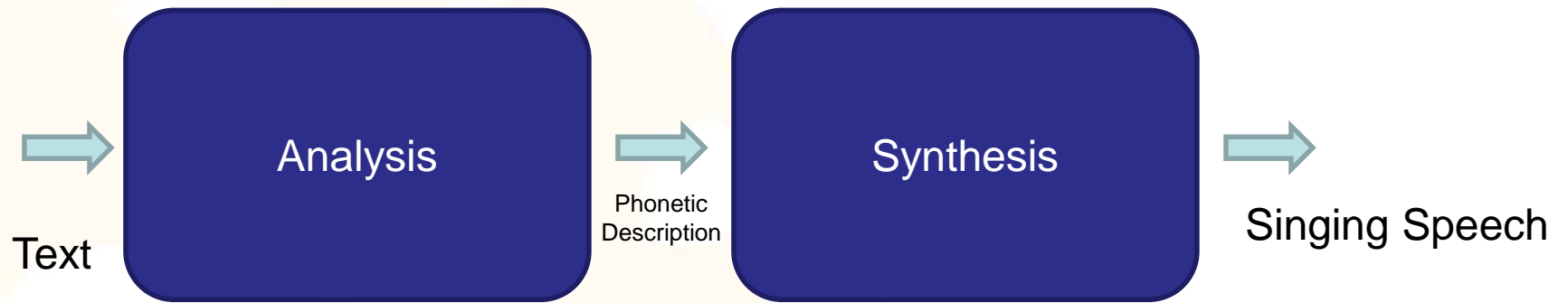
- <http://www.ivona.com/us/>



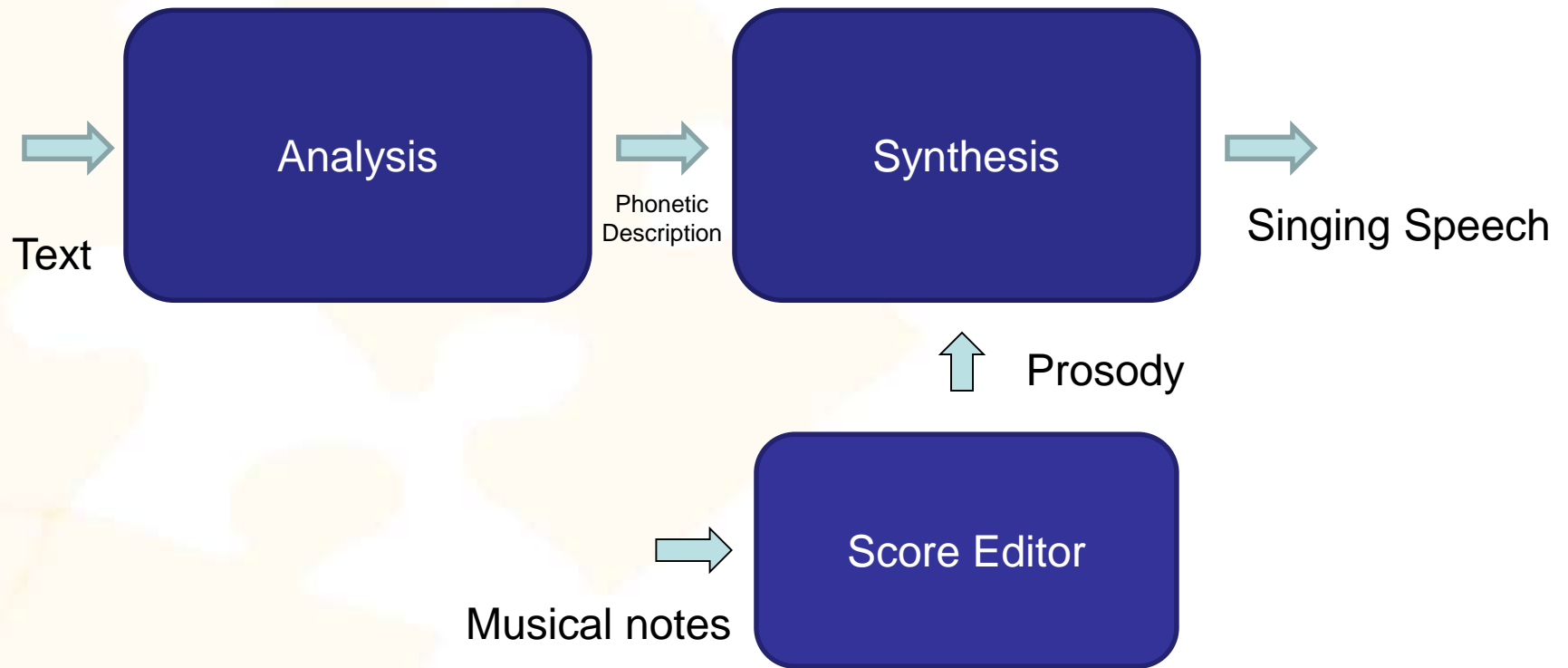
Singing Voice Synthesis



Singing Voice Synthesis



Singing Voice Synthesis



Japanese Commercial SVS system

- Vocaloid
- 'Supercell' was created with Vocaloid and it became a hit album/band in 2009 with more than 100,000 cumulative sales



Music Score Editor

