RAMCESS'07

Realtime & Accurate Musical Control of Expression in Singing Synthesis

Table of Contents

hallenges

Voice quality analysis/modification for expressive/emotional speech synthesis.

- Analysis: High quality source-tract decomposition is difficult : a long time challeng
- Synthesis: Control of low-level source parameters to achieve soft, tense, harsh... voices : research at its early stages

Real-time control for expressive singing synthesi

- Design of appropriate UI and mapping functio

evious Work

RAMCESS 1.0 :

Real-time expressive singing synthesis

Model-based voice quality modification

Vowel-only synthesis

QuickTime™ and a
mpeg4 decompressor
are needed to see this picture.

his Year's Aims

- nclude coarticulation
- mprove naturalness
- mport new voices : Able to synthesize a specific person's voice

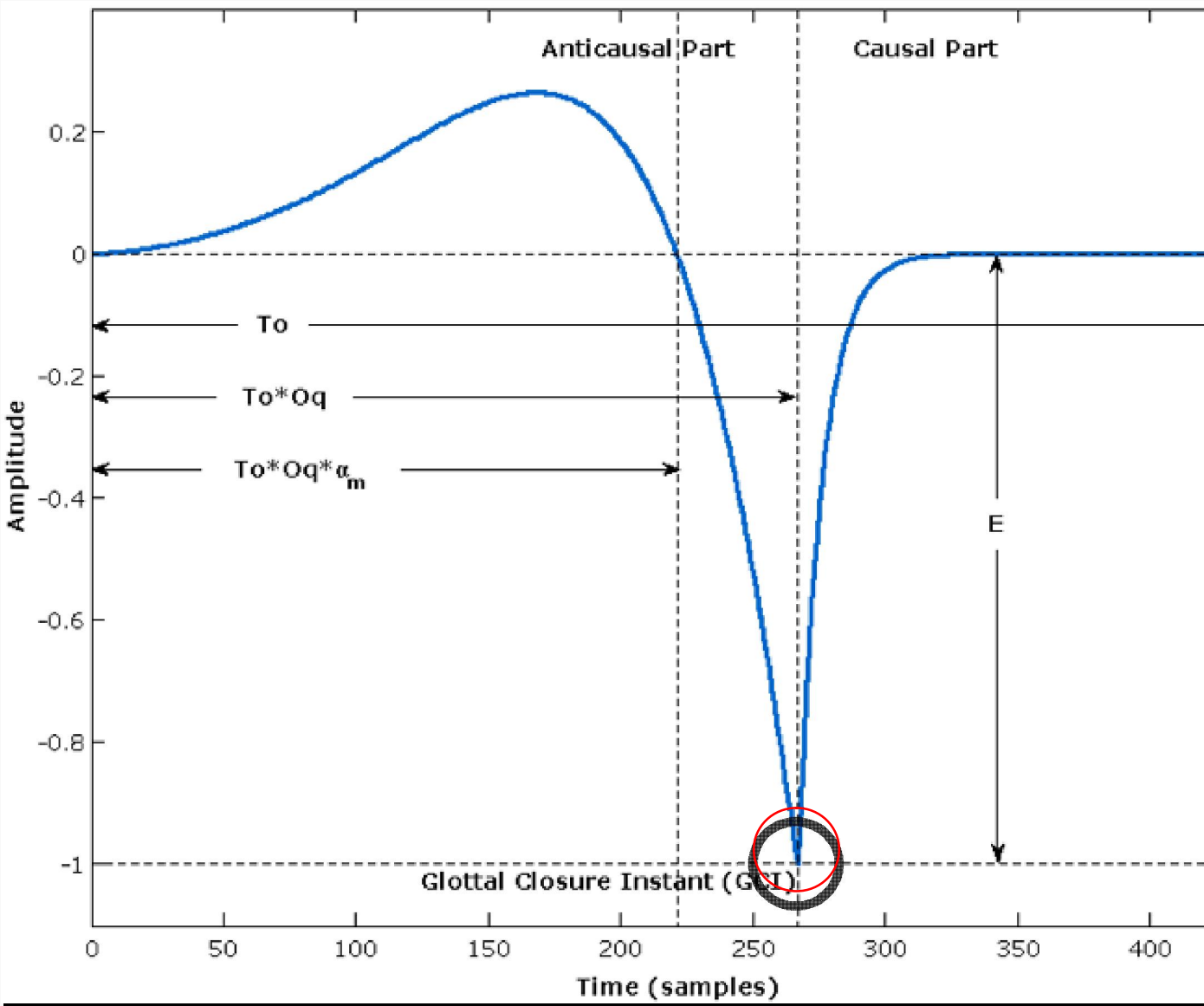
tabase

- Small size : ~ 2 seconds of recording
- Includes unvoiced consonants, vowels nd transitions
- Constant pitch
- Constant voice quality (perceivably)

ckground

LF Model : A model for glottal source

- ***G***lottal ***C***losure ***I***nstant : The main excitation point of and important feature for describing a



ckground

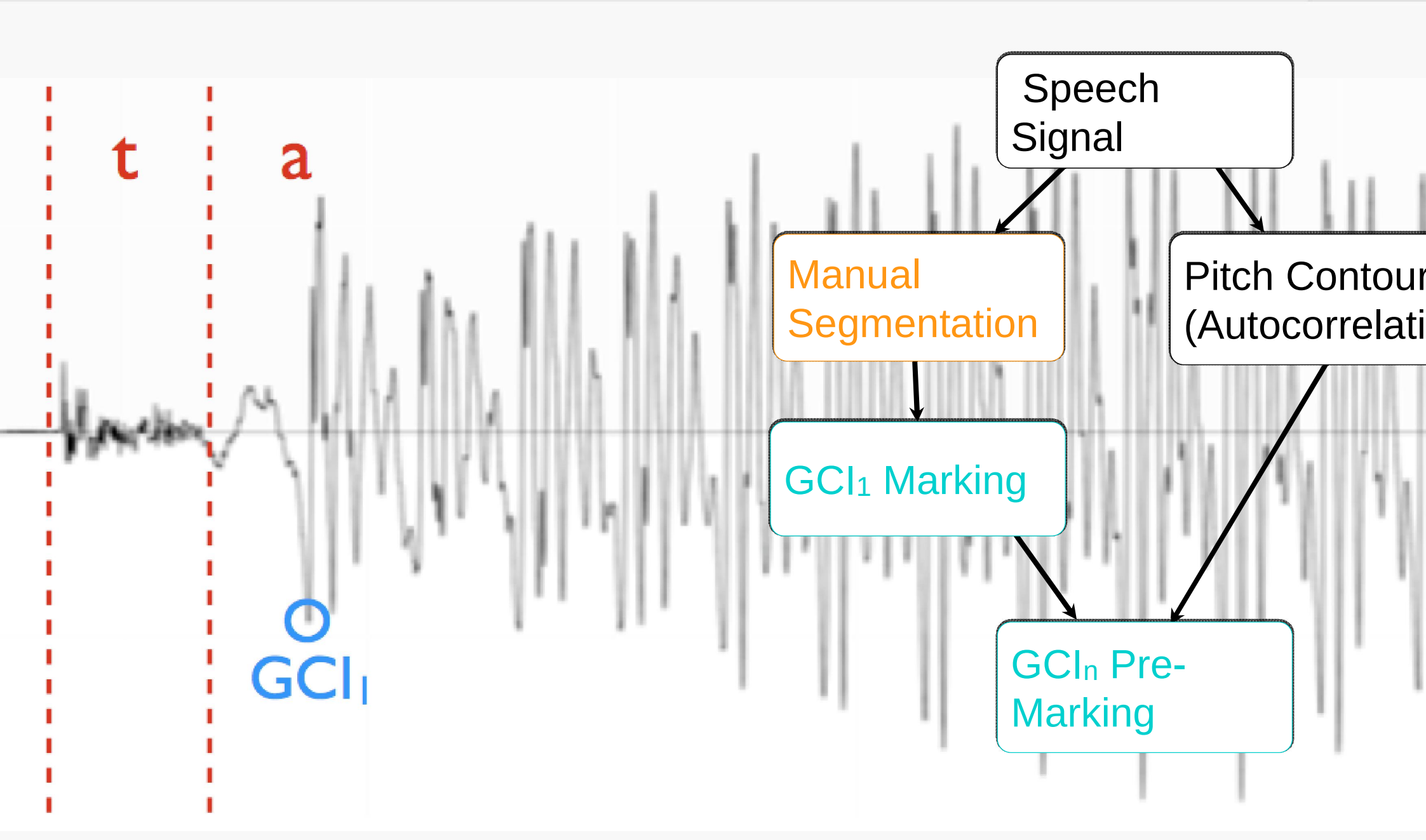
ZZT : Zeroes of Z-Transform

- Source - tract decomposition
- Sensitive to window position
- High quality source estimation
- Bad at spectral tilt

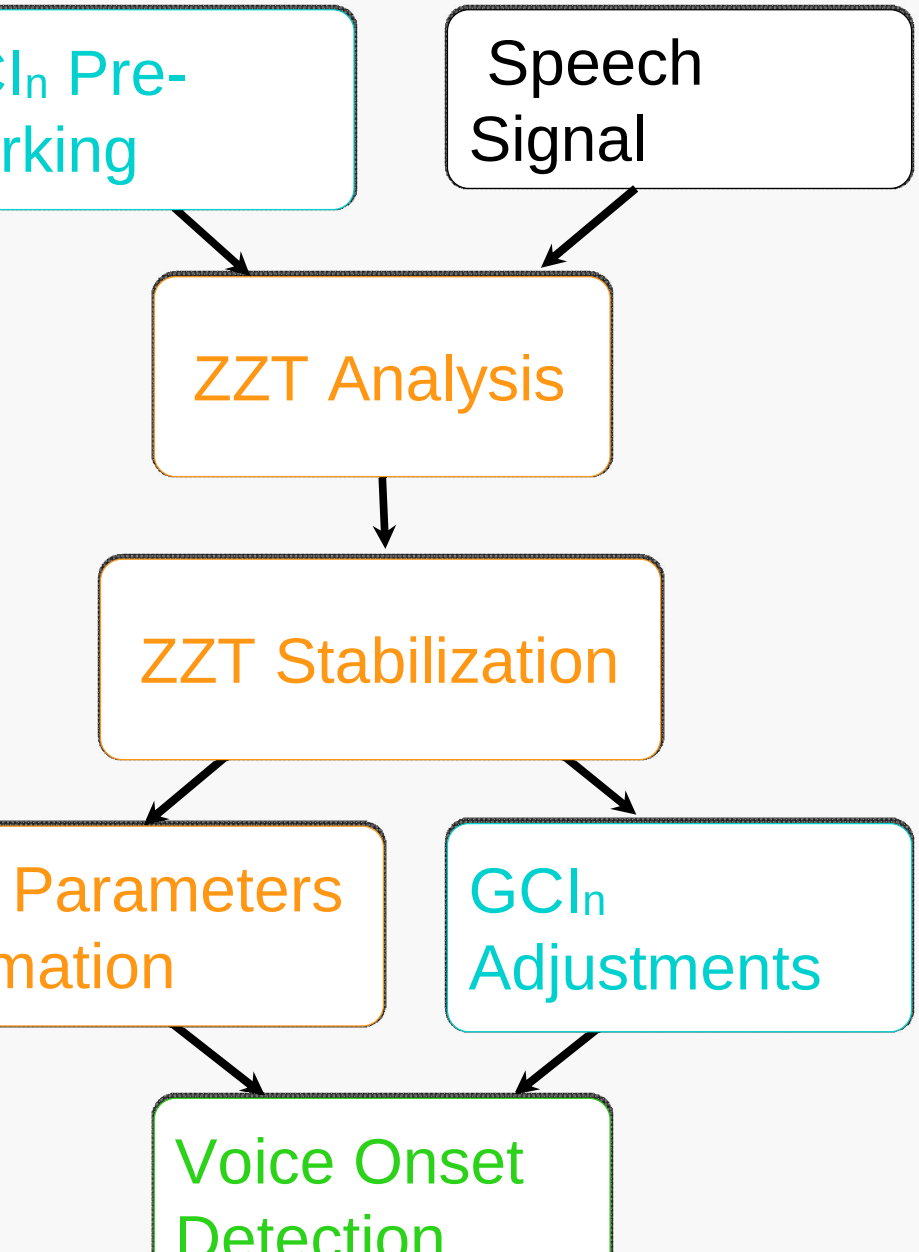
ARX : Auto-Regressive eXogenous

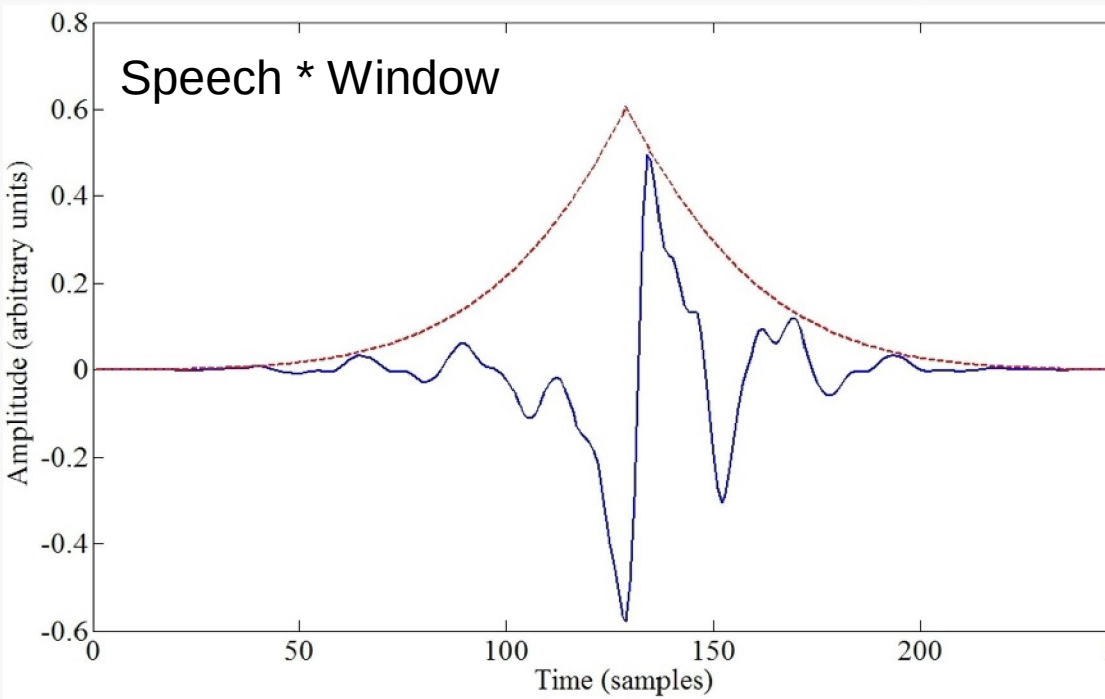
- We use : An ARX model excited by the LF mode
- Good at spectral tilt

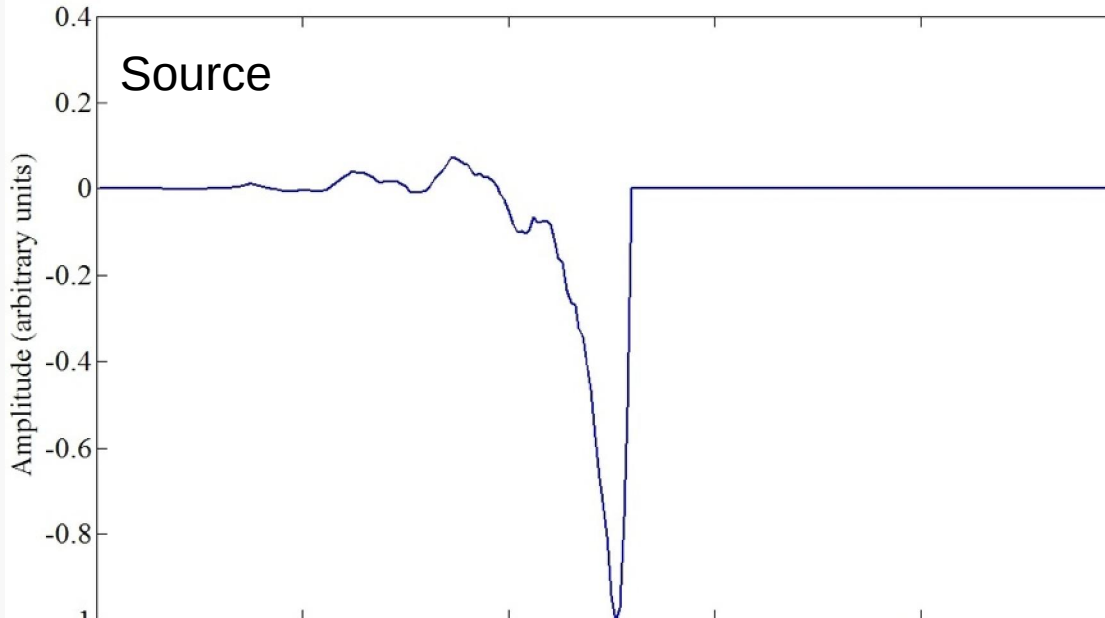
GCI & ALF Estimation



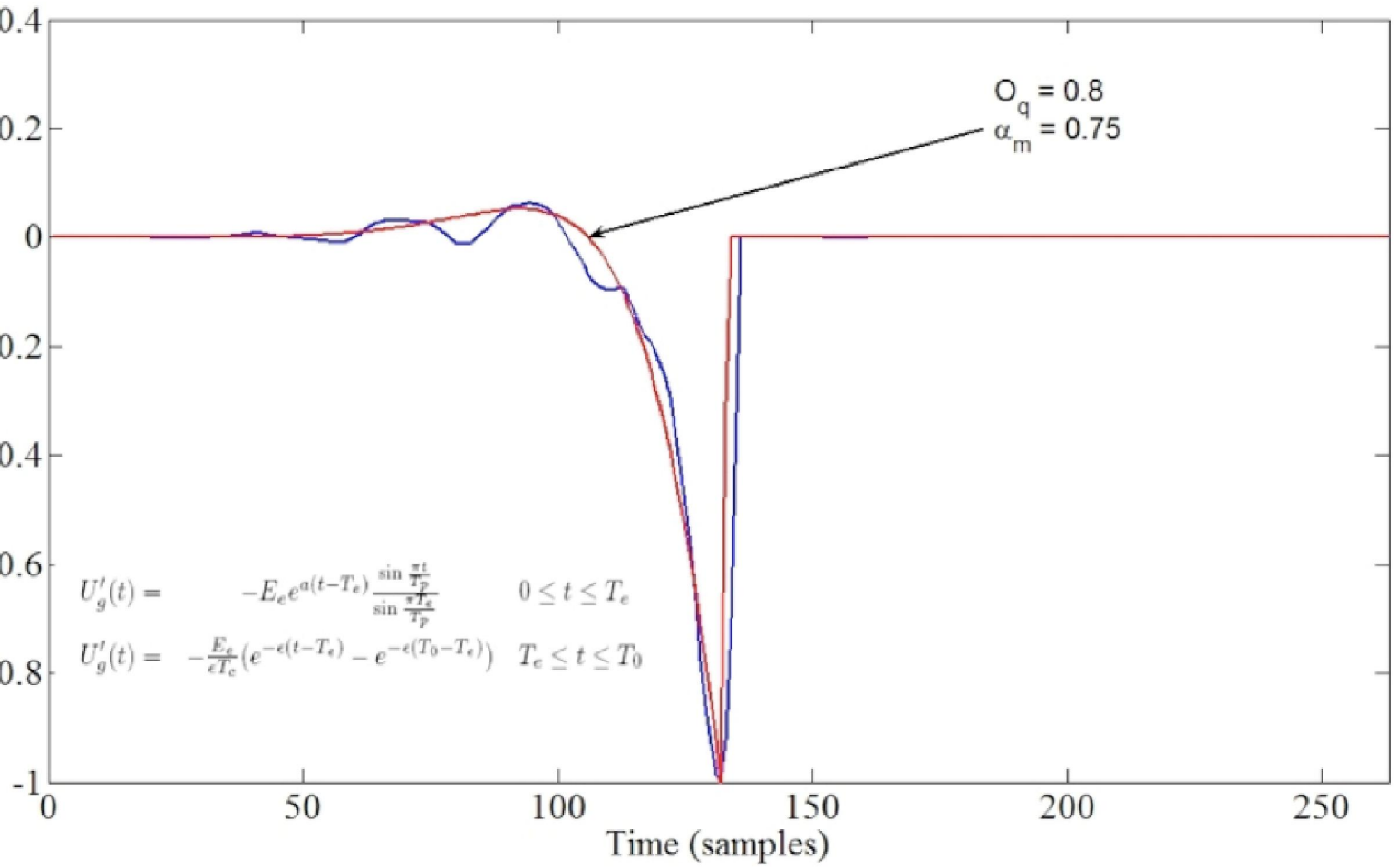
CI & ALF Estimation







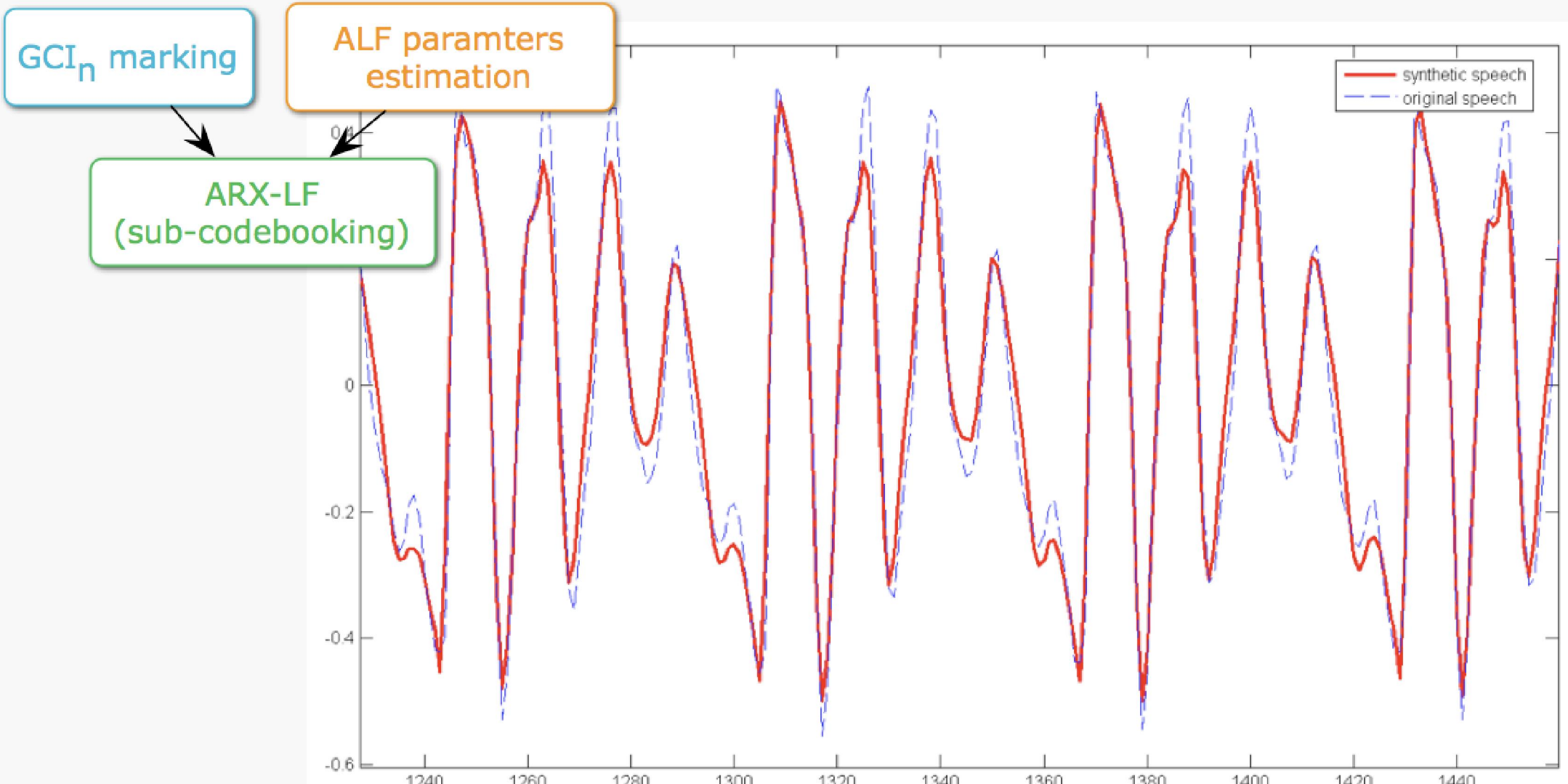
t LF Model to AC part



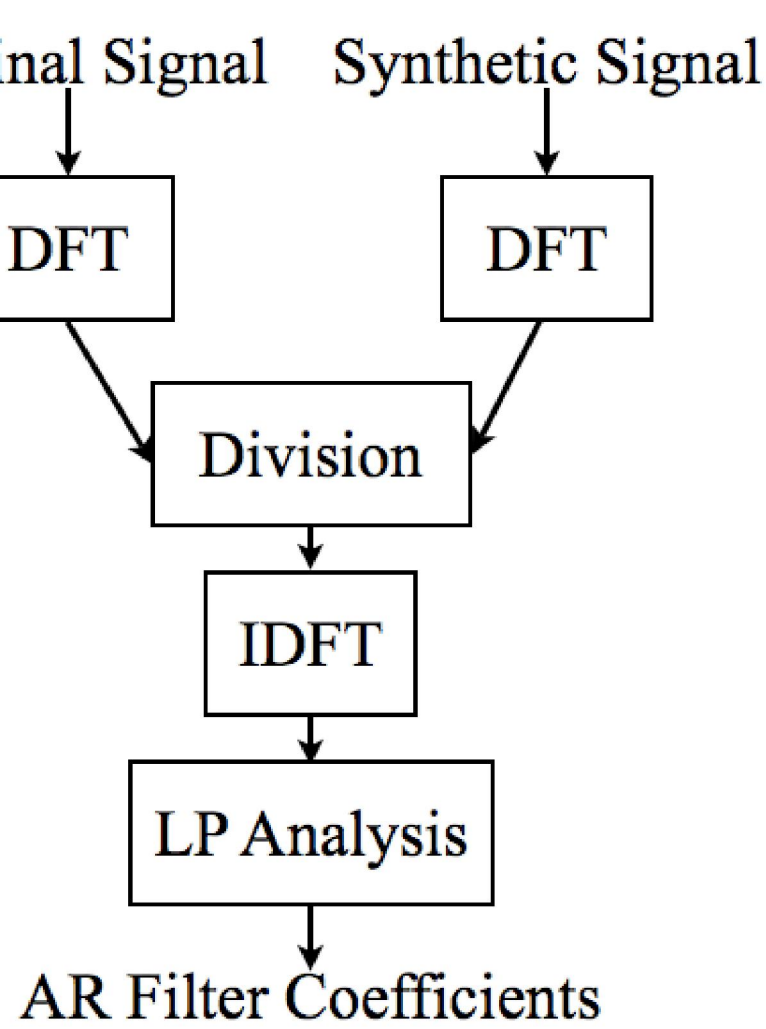
- Achieve by codebo search
- Perform on window signals
- Outputs

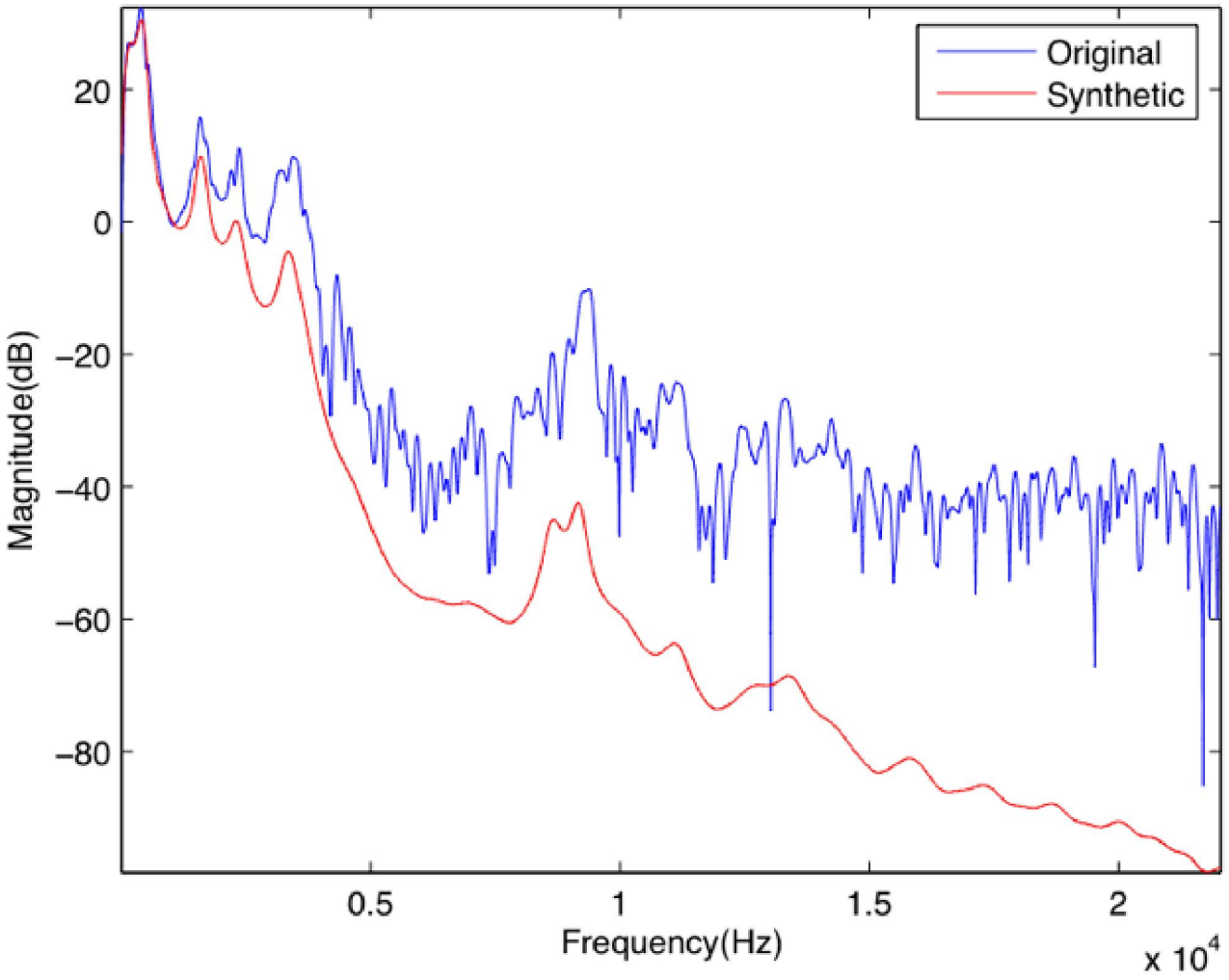
RX for filter estimation

Joint estimation method for source and tract

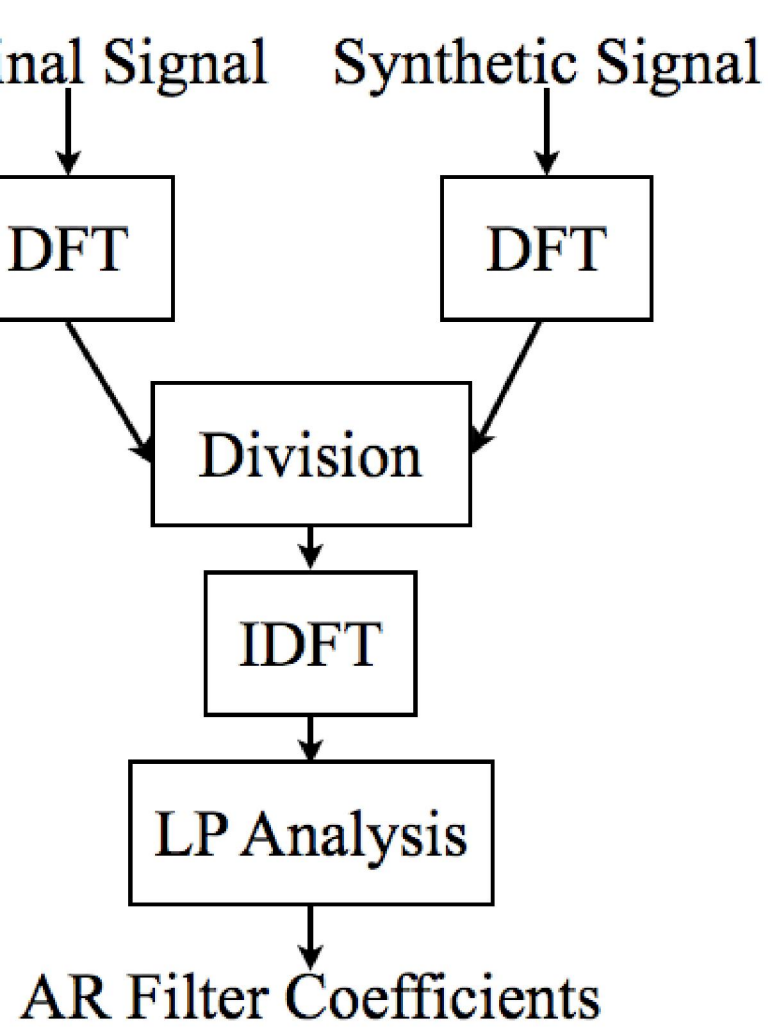


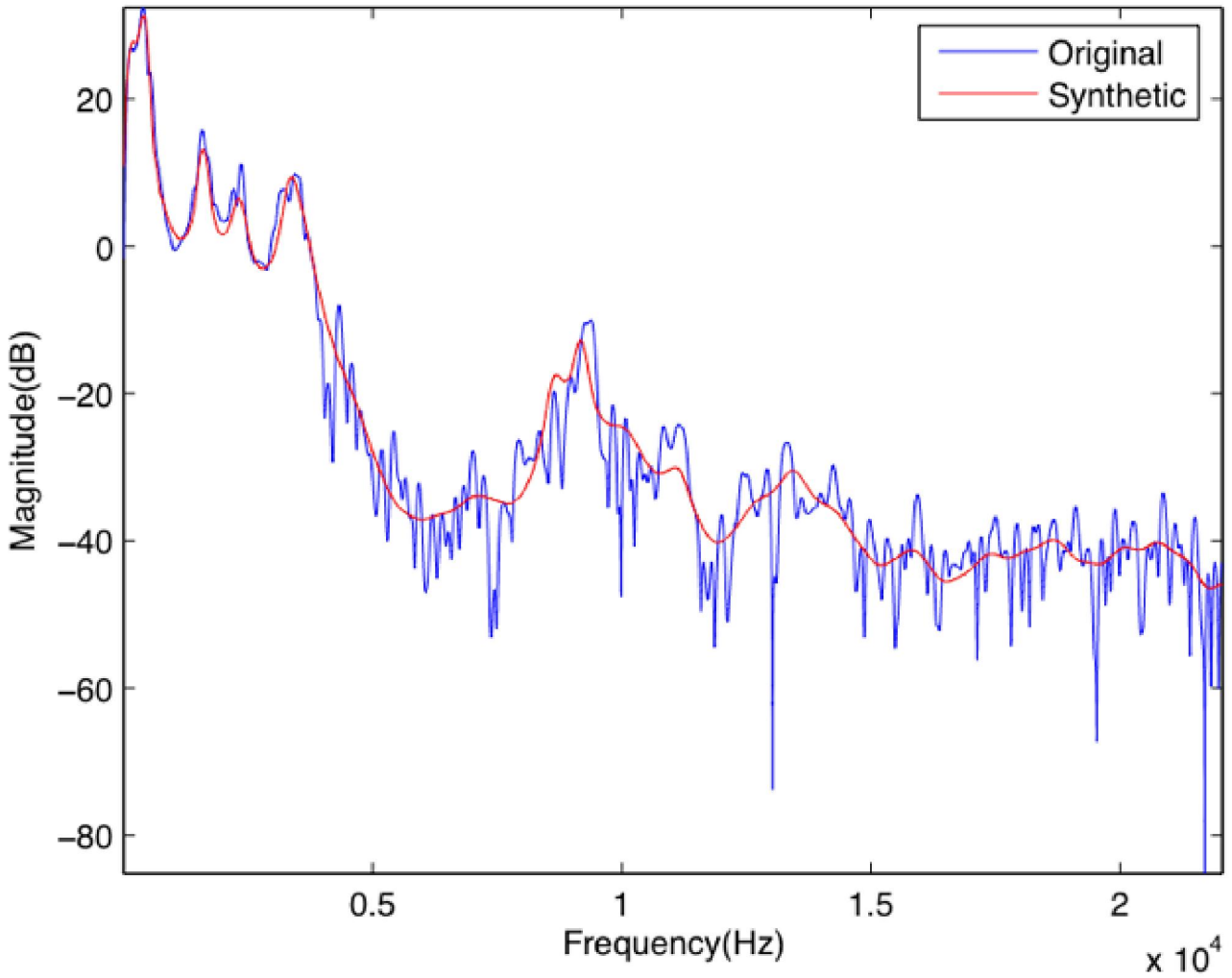
ompensation



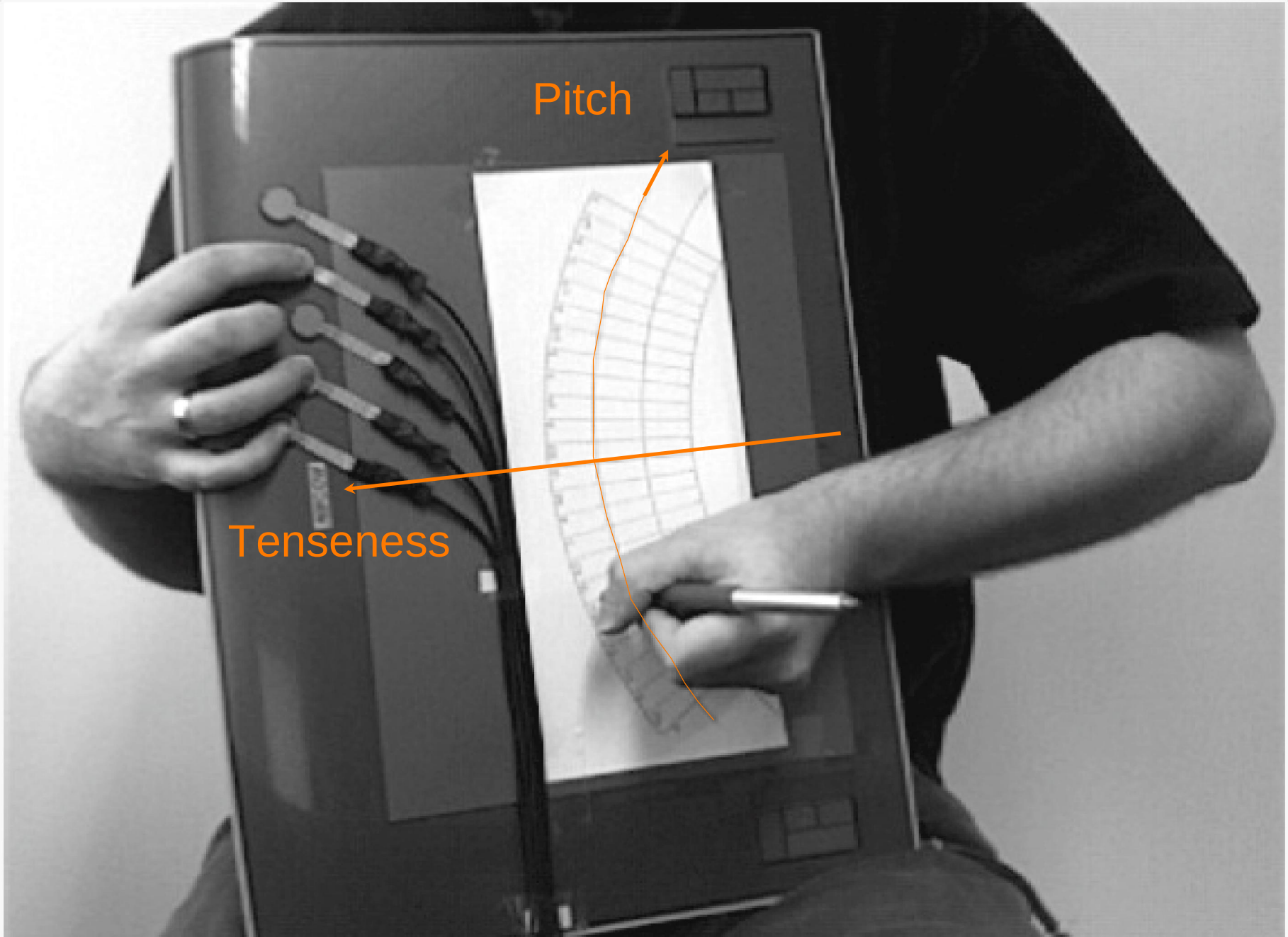


ompensation

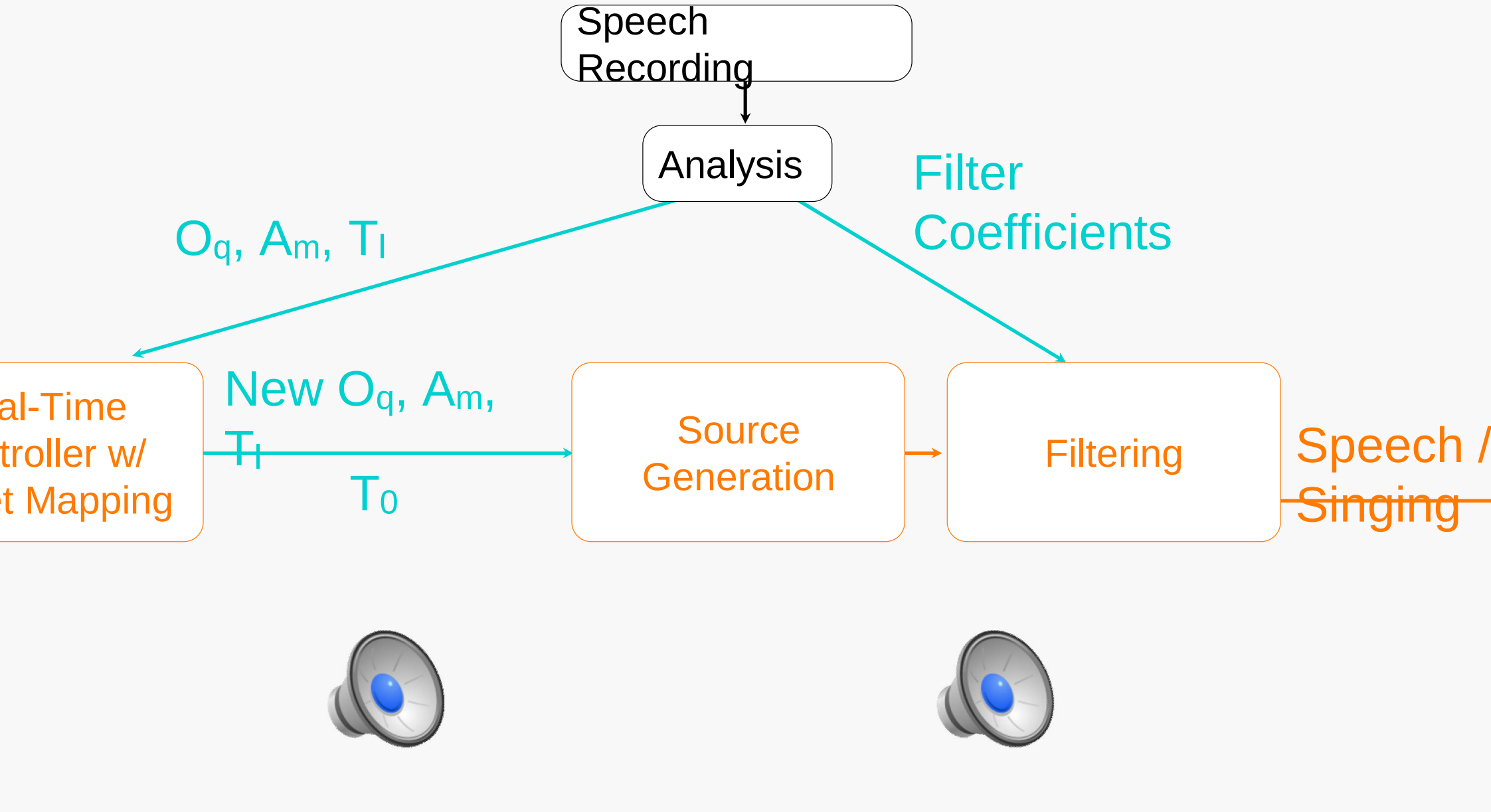




nthesis



nthesis


Speech Recording
Analysis
Filter Coefficients
O_q, A_m, T_l
al-Time troller w/ t Mapping
New O_q, A_m, T_l
T_0
Source Generation
Filtering
Speech / Singing

nthesis

QuickTime™ and a
Photo - JPEG decompressor
are needed to see this picture.

ifficulties in Synthesis

- Frame-by-frame analysis-synthesis results in
  - ough speech
    - Smoothing applied
- Mapping of tenseness level to source
  - parameters
    - Obtained by trial and error
- Real-time residual generation and mixing
  - A method is devised and tested offline with
    - positive results. Not integrated into the real-time

ture Work

- mprove transitions
- Automating analysis
- Devising a control interface for syllable switching
- Testing with different voices
- Testing in musical performance

Thank you for your attention.

Questions?