Analysis and Synthesis of Pathological Vowels

Prospectus

Brian C. Gabelman

6/13/2003

OVERVIEW OF PRESENTATION

I. Background

II. Analysis of pathological voices

III. Synthesis of pathological voices

IV. Summary

What is a pathological vowel?

May be caused by physical or neural problems. Characterized by substantial and complex NONPERIODIC signal components.

What is this project?

Methods for modeling, analysis, and synthesis of pathological vowels incorporating novel approaches in:

- System identification
- Parameterization of non-periodic components (AM, FM, and noise)
- Synthesizer designs for realtime and offline use

Why do it?

- Create a non-subjective basis to compare pathological voices for:
  1. Improved diagnosis
  2. Tracking changes in a patient's voice
- Generate voice samples with known levels of variations (noise, roughness, etc.) for:
  1. Evaluation of model parameters
  2. Evaluation of listener variability
  3. Evaluation of importance of levels of pathological features.

What has been done before?

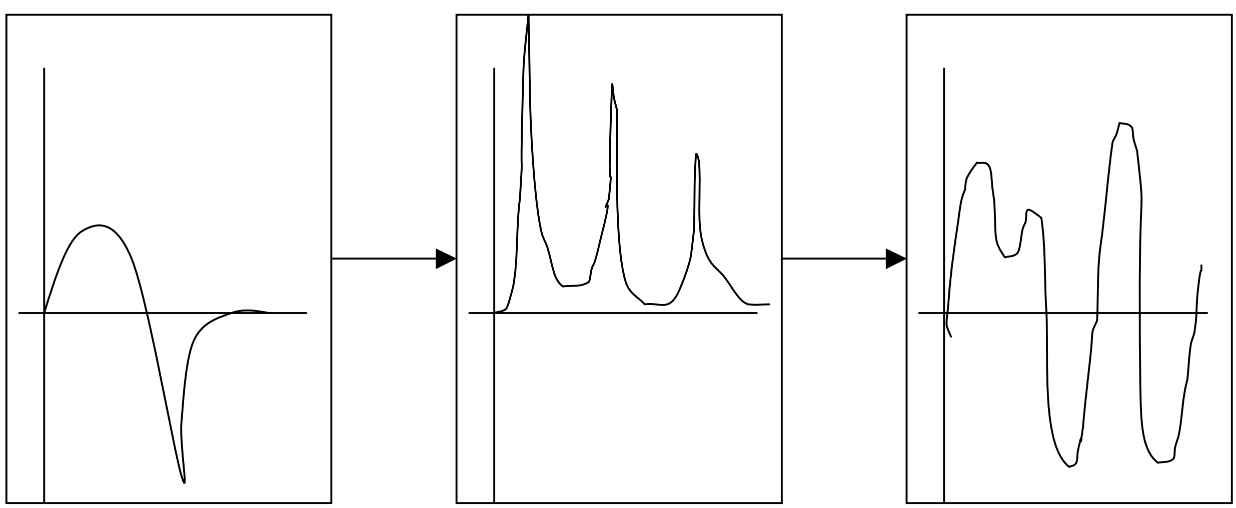
- A well-established theory exists for NORMAL voices
- Recent studies of pathological voices employ perturbation of normal features plus additive noise.
- ES (external source) stimulation of the vocal tract to analyze formants (vowels) since 1942 for NORMAL voices.

For the theoretical/analytical aspect of the project, an expression of the hypothesis of the dissertation in one sentence is:

"By means of FM and AM demodulation techniques, estimation of nonperiodic features of pathological vowels may be improved."

SOURCE - FILTER MODEL OF SPEECH

GLOTTAL SOURCE WAVEFORM	VOCAL TRACT FREQ. RESP.	RESULTING VOICE SIG.
(time domain)	(freq domain)	(time domain)



g(t)	f(t)	v(t)
G(s)	F(s)	V(s)

g(t) conv. f(t) = v(t) (time domain)

G(s) x F(s) = V(s) (freq domain)

Steps for analysis:

PERIODIC ANALYSIS:

1. FORMANT DETERMINATION
   Uses LP (linear prediction) to model vocal tract as cascaded 2nd order digital resonators. External source testing is shown to augment or replace LP for pathological vowels (Inv).

2. SOURCE MODELING
   Uses inverse filtering and least squares optimization to fit source waveform to a standard model (LF).

NONPERIODIC ANALYSIS:

3. ANALYSIS OF PITCH VARIATION
   Uses high resolution pitch tracking to measure detailed nonperiodic frequency variation. Variations are segmented into low and high frequeny FM with Gaussian form.

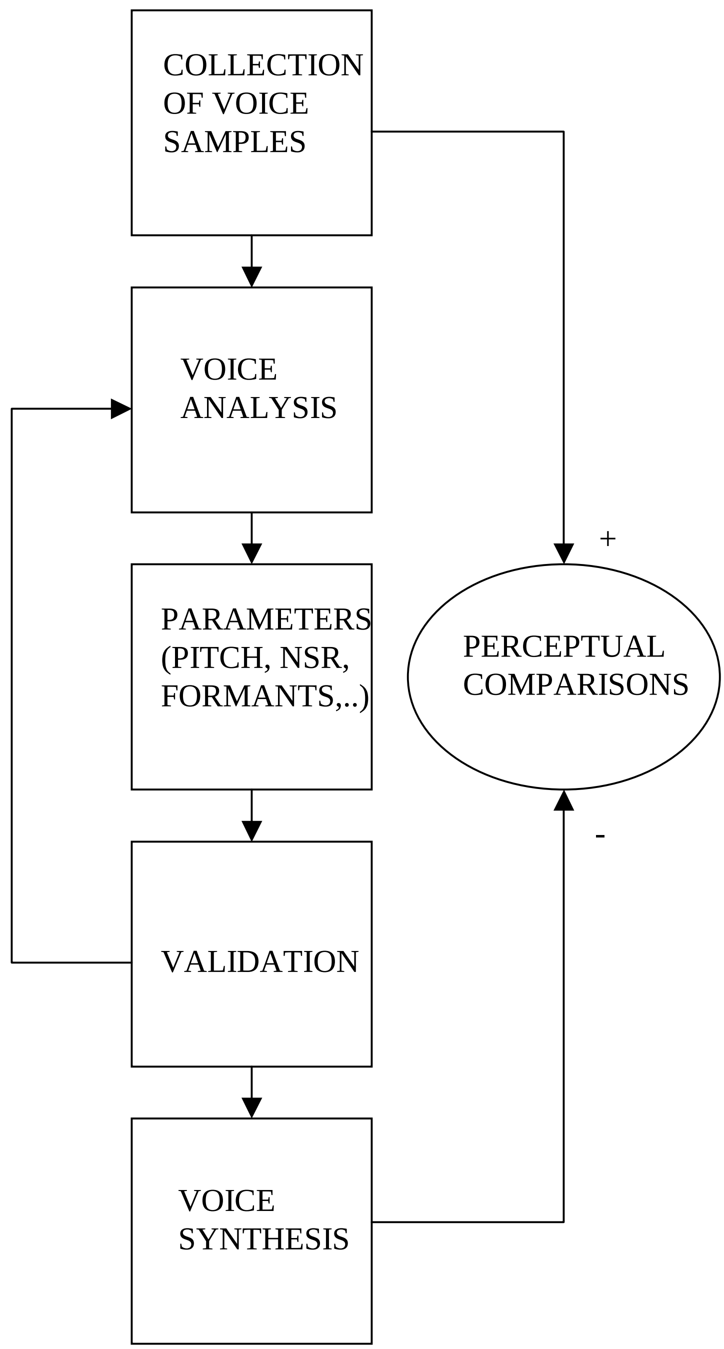
4. FM DEMODULATION
   Pitch variations are removed from original voice to achieve accurate noise estimation (step 7) (Inv.)

5. ANALYSIS OF POWER VARIATION
   Uses power tracking to measure detailed nonperiodic loudness variations. Variations are segmented into low and high frequency AM with Gaussian form.

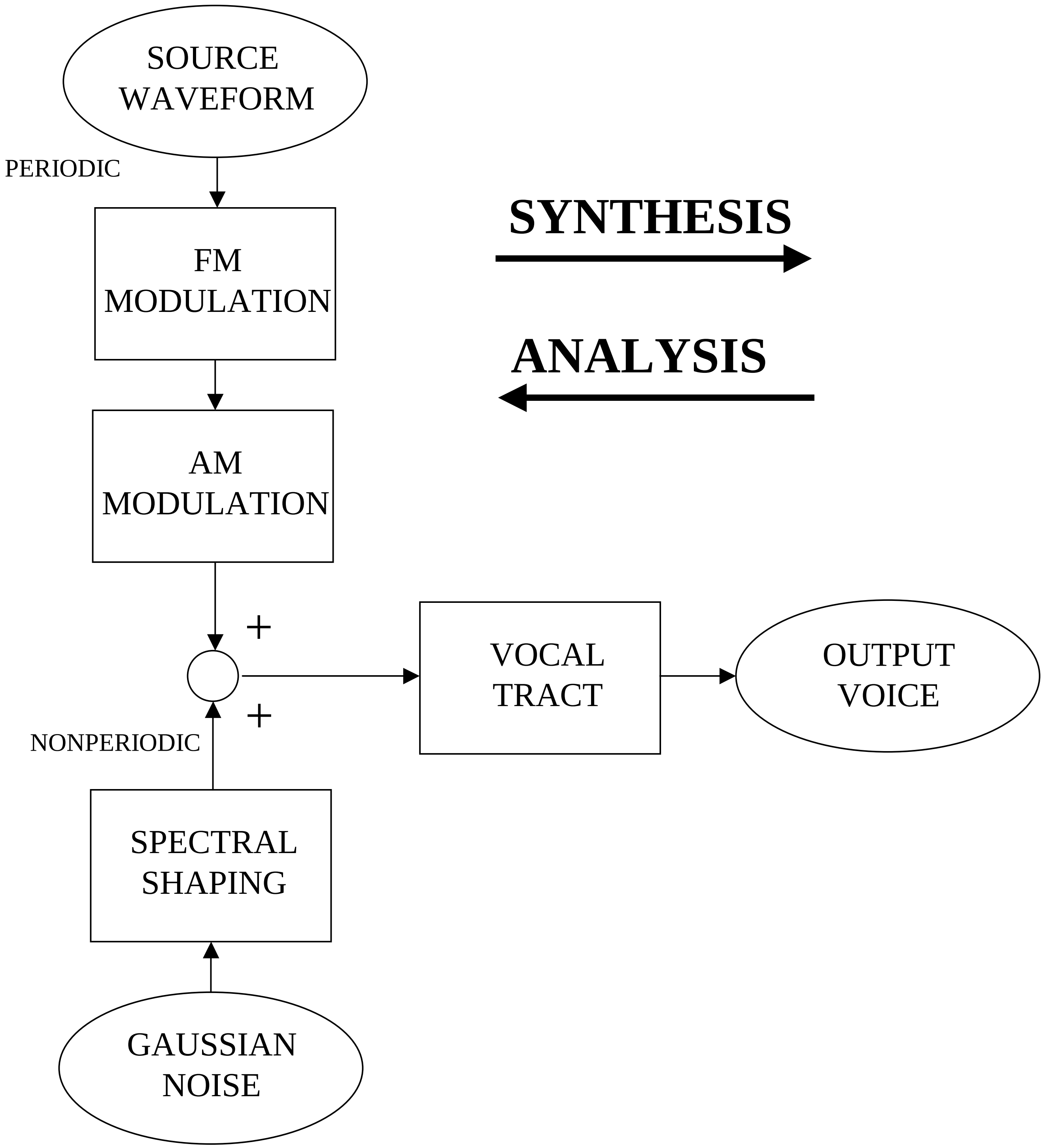
6. AM DEMODULATION
   Power variations are removed from original voice to achieve accurate noise estimation (step 7) (Inv.)

7. ASPIRATION NOISE
   Frequency domain methods are used to separate aspiration noise component. The noise is spectrally modeled.[Gus de Krom]

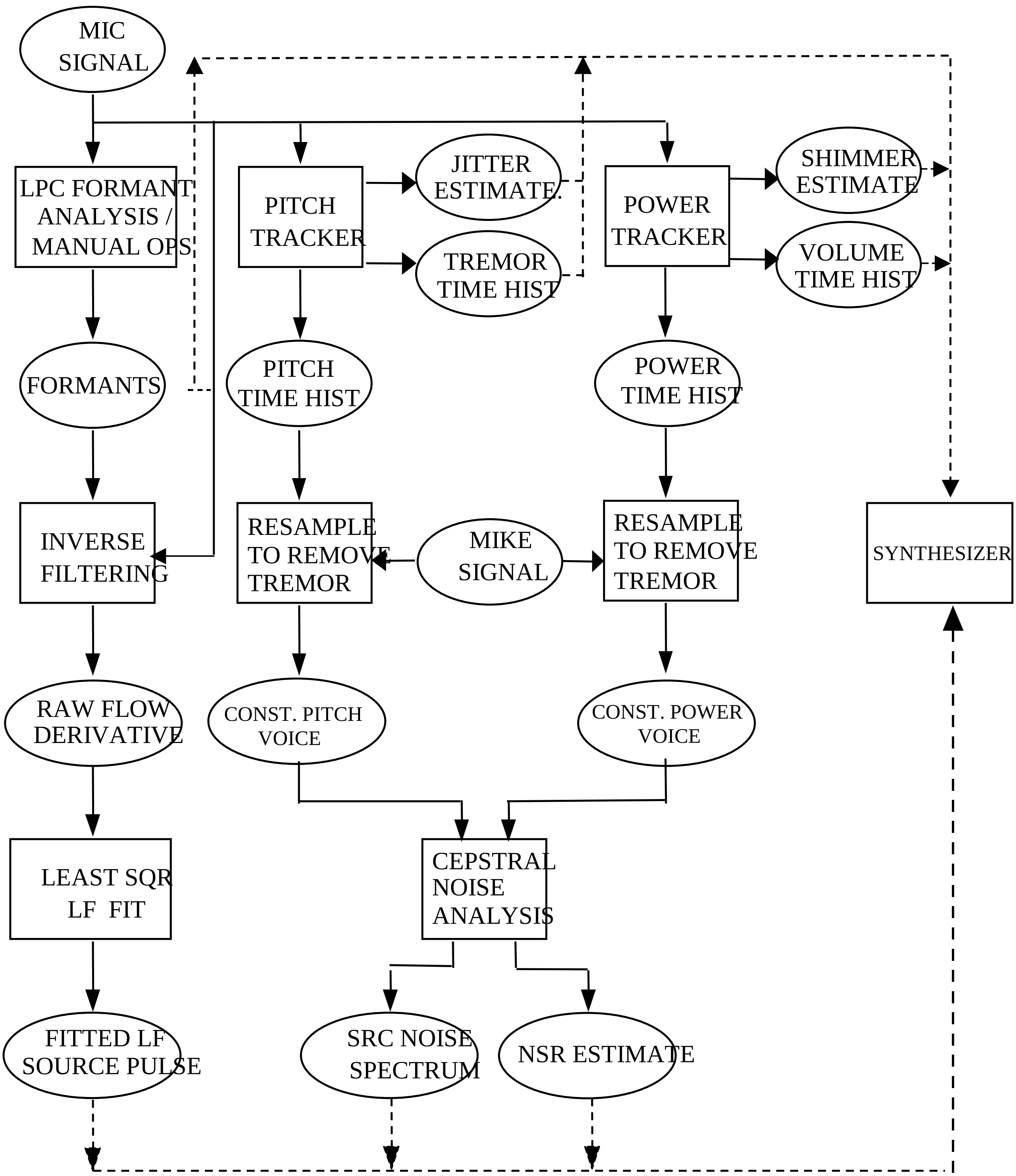
ANALYSIS BY SYNTHESIS



ANALYSIS/SYNTHESIS MODEL OVERVIEW

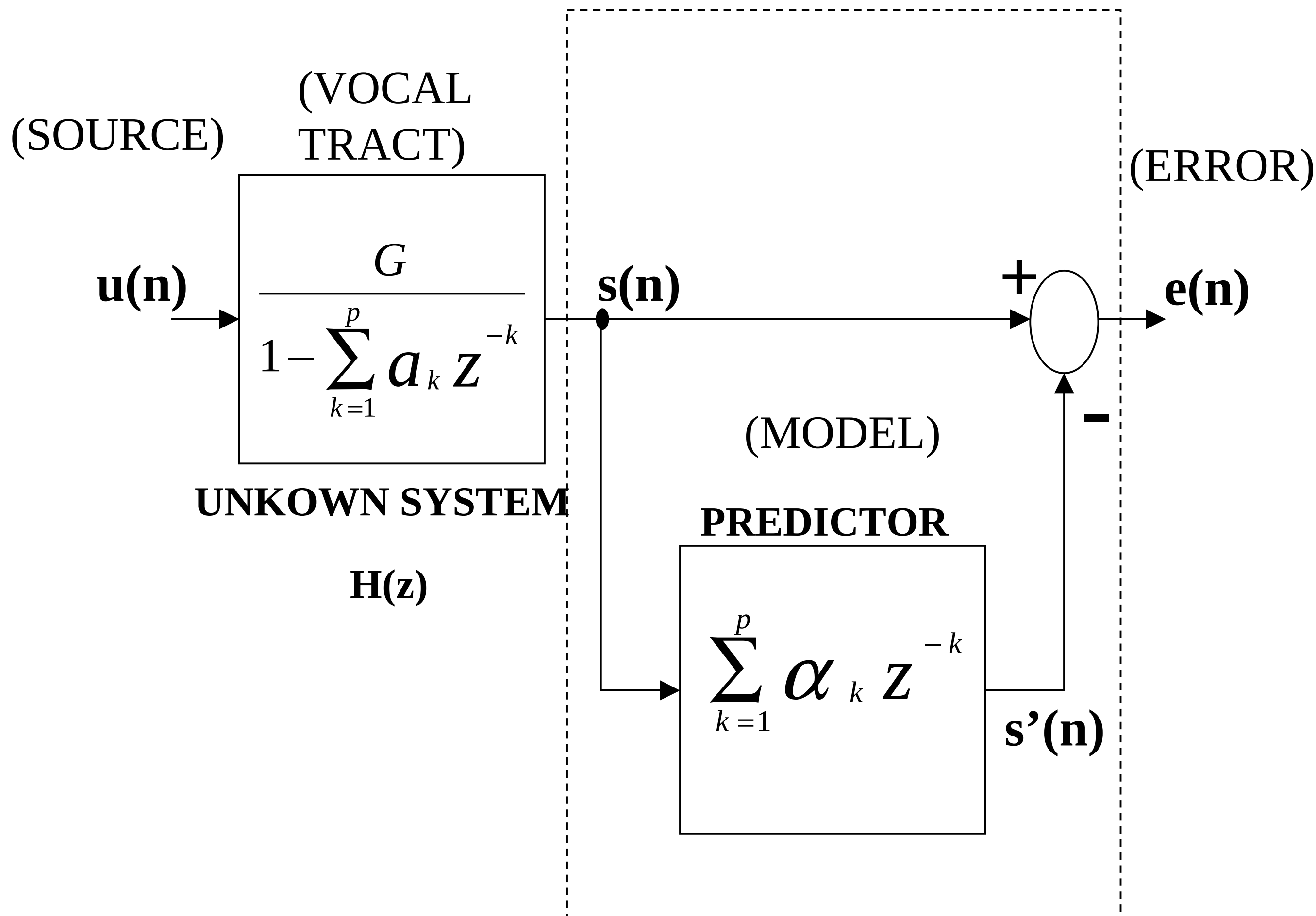


OVERVIEW OF PROJECT OPERATONS



LINEAR PREDICTION

Estimates the vocal tract as an all-pole filter by minimizing the error between actual and model-predicted signals.

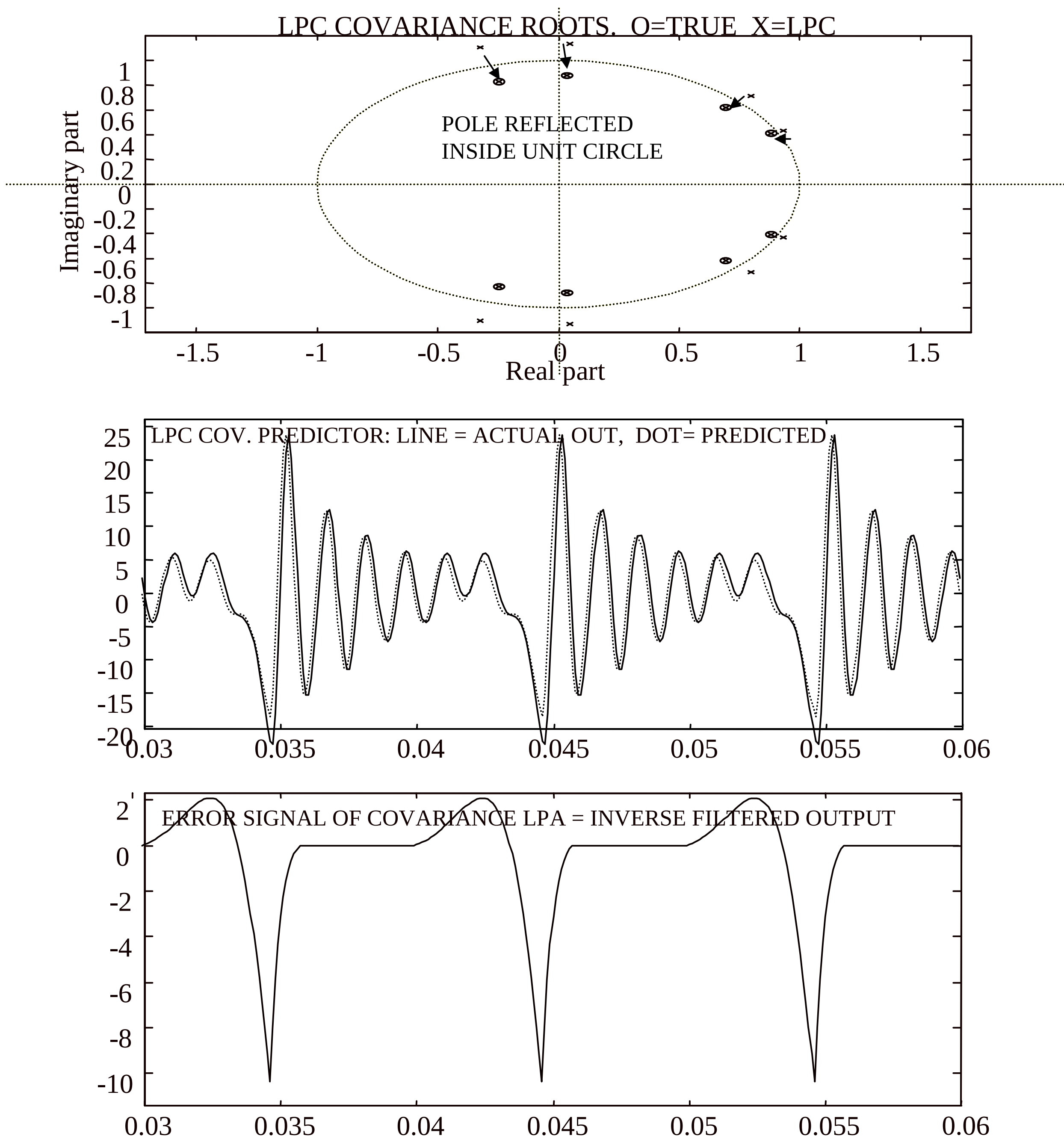


ESTIMATED INVERSE SYSTEM

$$\textbf{A(z)} = 1-\sum_{k=1}^{p}\alpha_k z^{-k} \approx \textbf{G/H(z)}$$

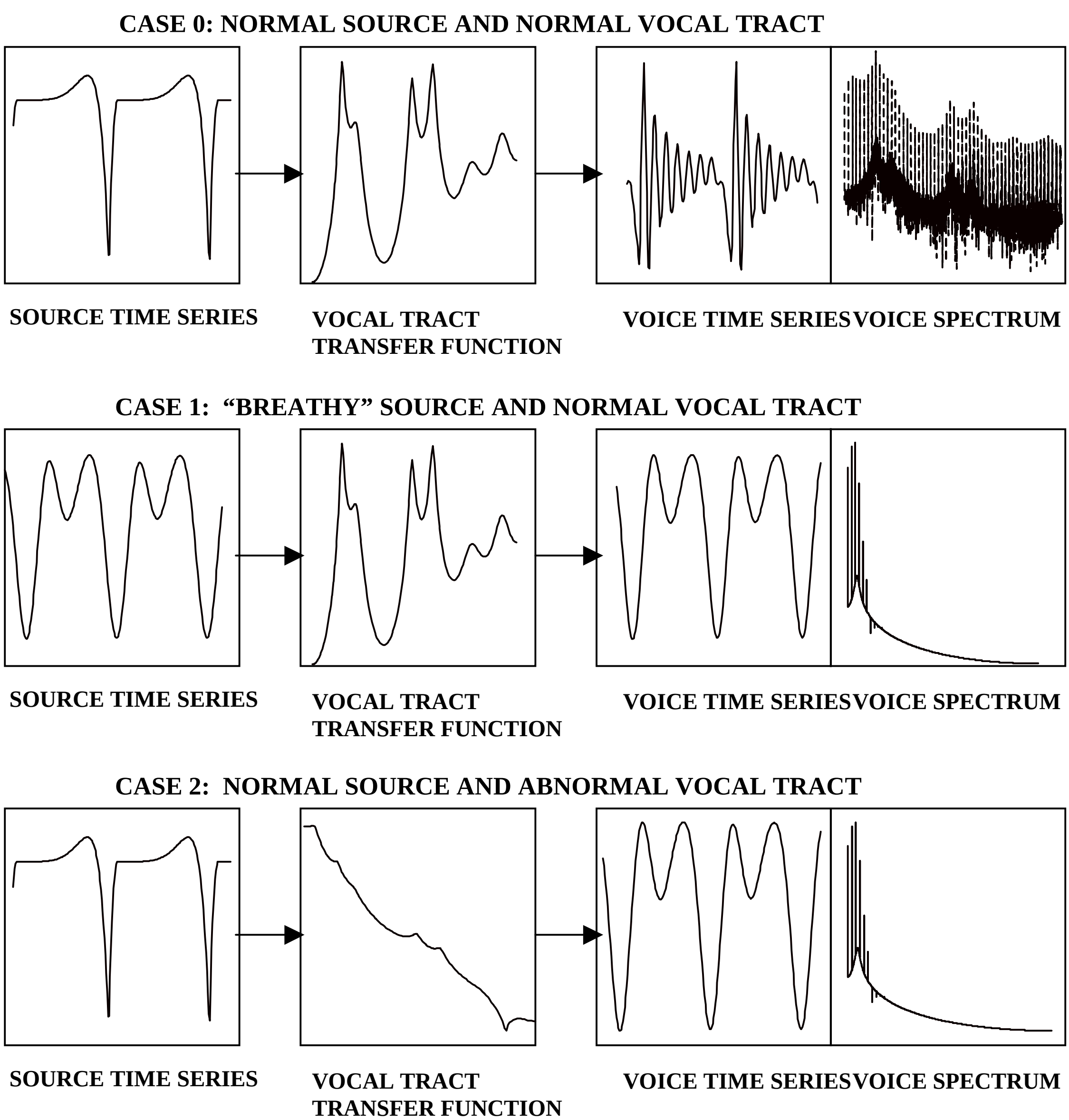
IDEALIZED LP RESULT

Requires a priori knowledge of system.
More difficult for pathological vowels.



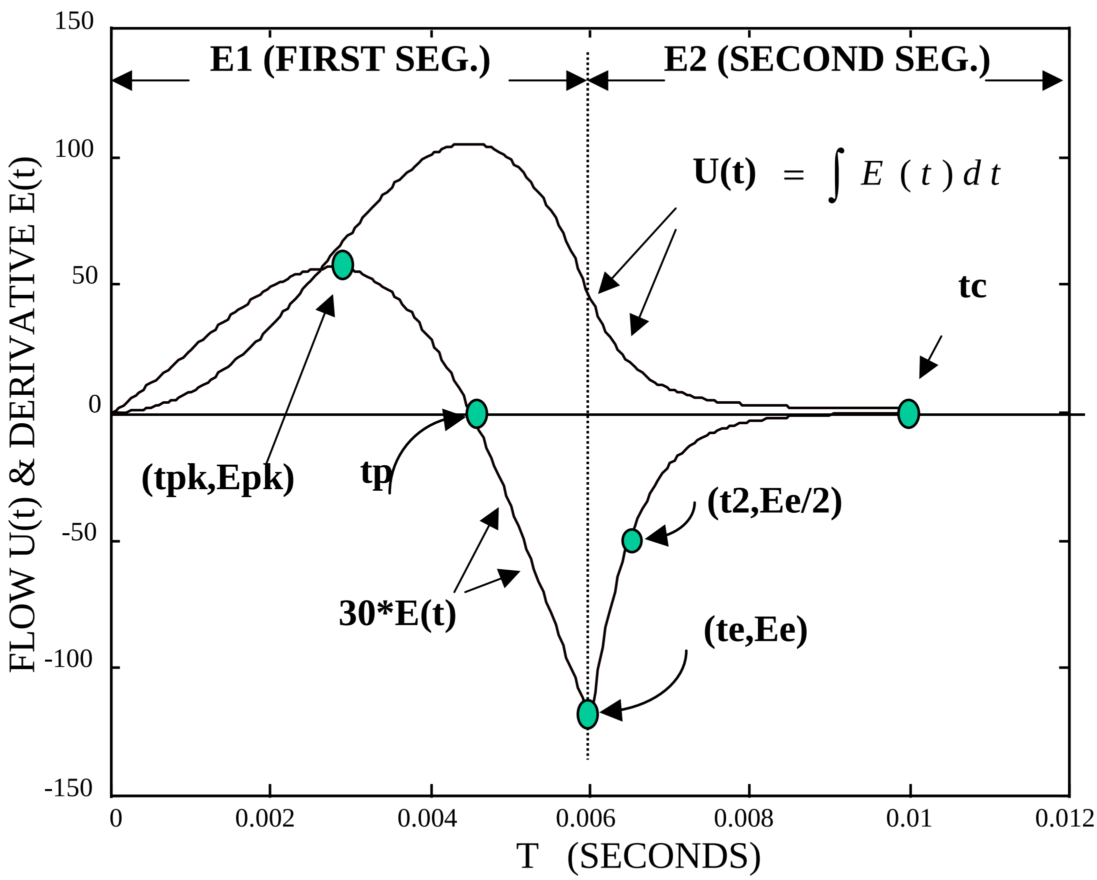
SOURCE-FILTER AMBIGUITY

Source & filter are mixed in final voice.
Unique LP solution may be difficult.



FITTING RAW SOURCE TO LF

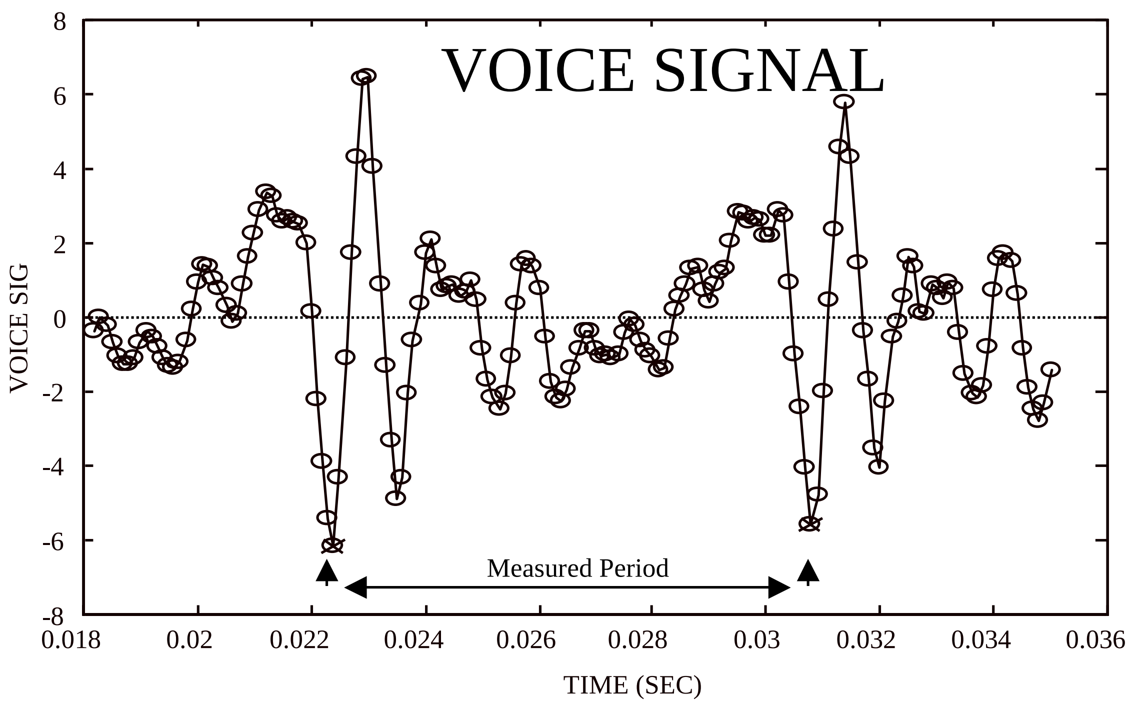
Having established the inverse filtered source, it is fit to the LF model [Qi]

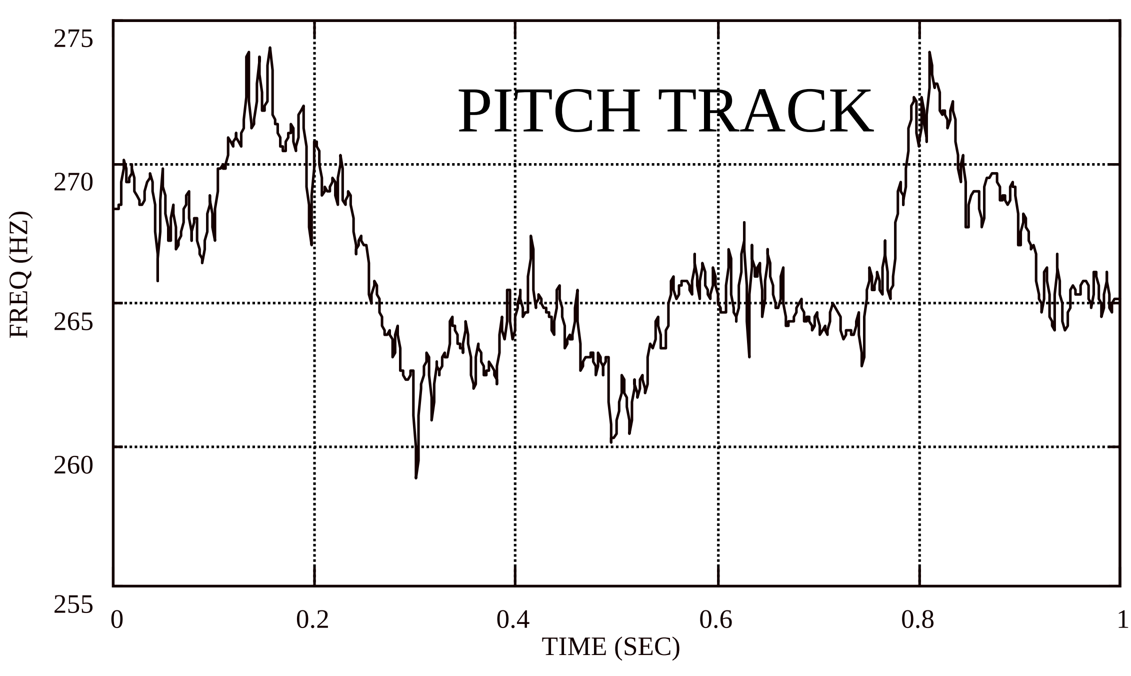


$$E(t) = \begin{cases} E_1(t) = E_0\, e^{\alpha t} \sin \omega_g t, & t \le t_e \\ E_2(t) = -E_e\, e^{-\varepsilon(t-t_e)}, & t_e \lhd t \le t_c \\ E_{2B}(t) = -E_e\, e^{-\varepsilon(t-t_e)} + m(t-te), & t_e \lhd t \le t_c \end{cases}$$

HIGH RESOLUTION PITCH TRACKING

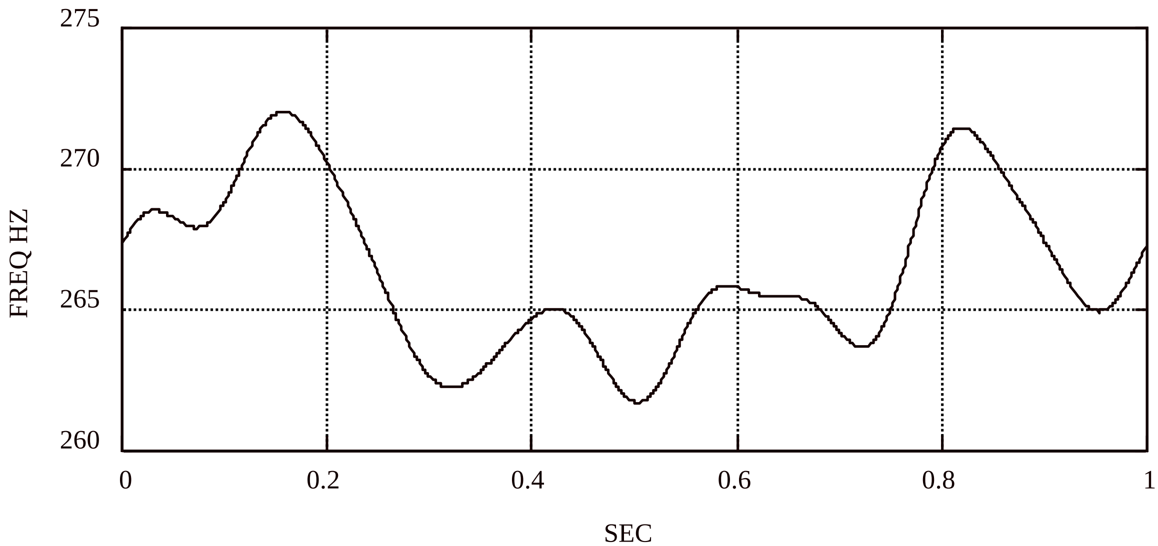
Nonperiodic analysis begins with interpolating pitch tracking.

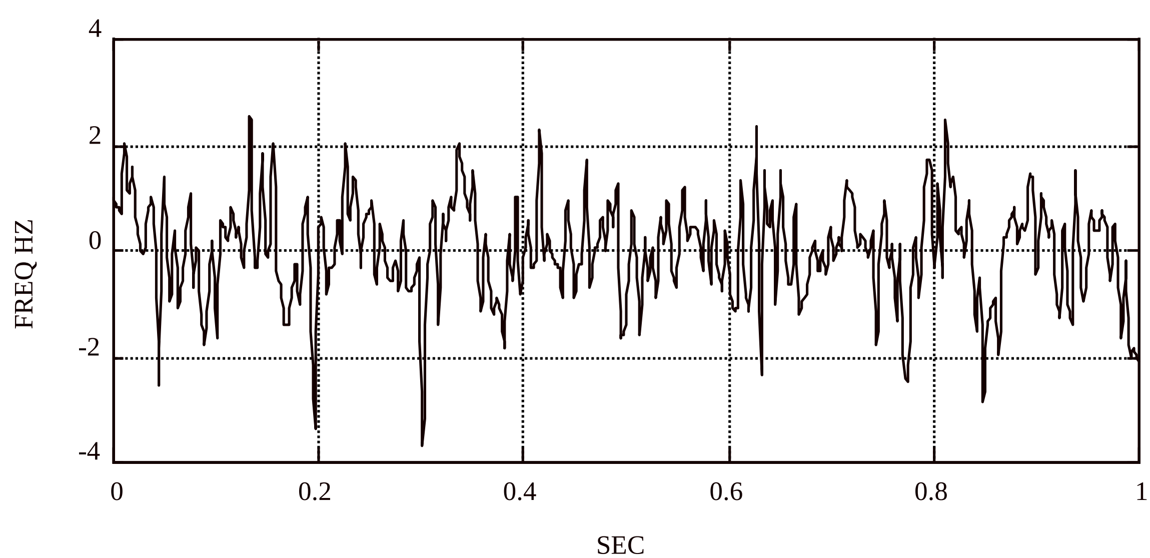




FM DEVIATION SEGREATED TO LOW AND HIGH FREQUENCY

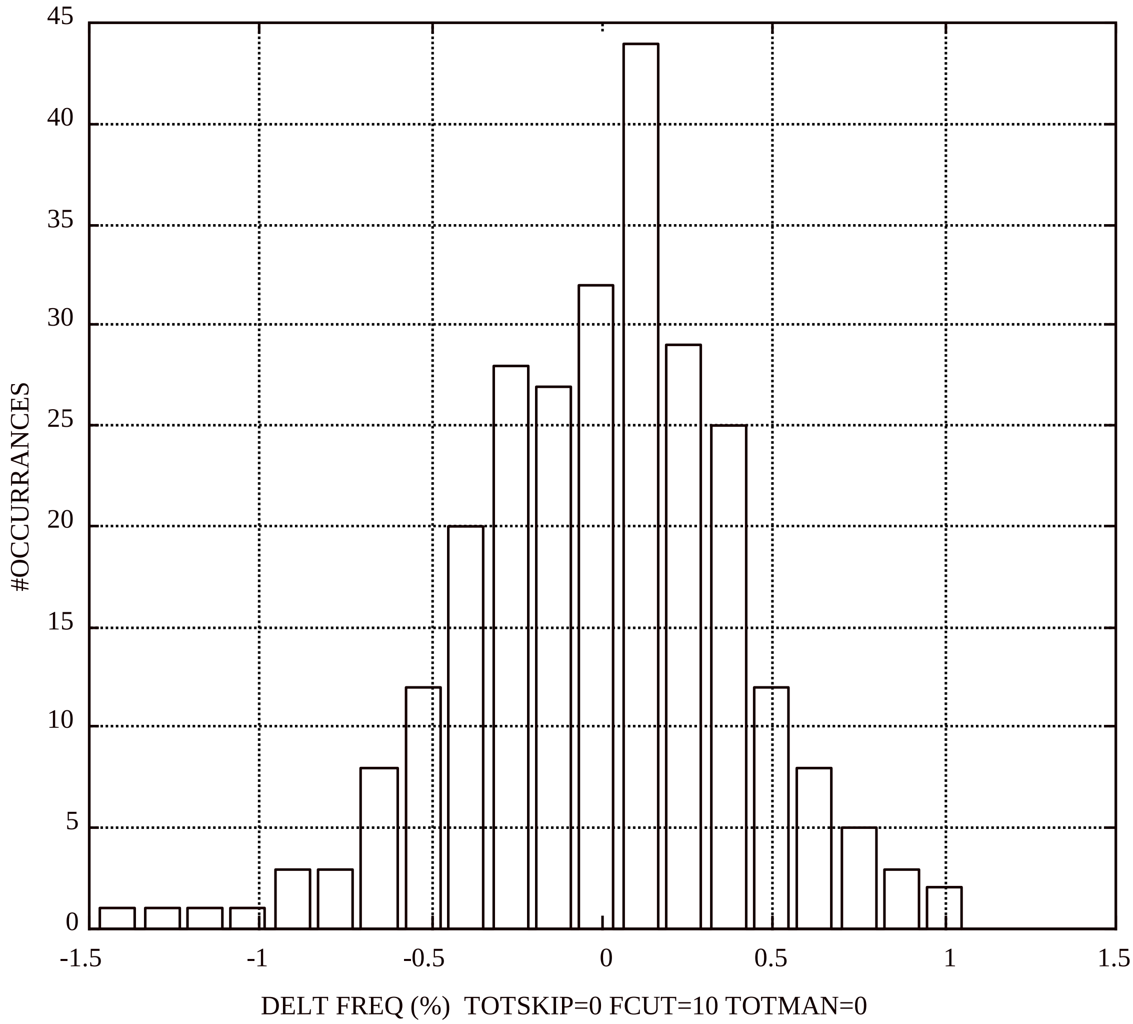
The pitch track is low/hi pass filtered to yield tremor and HFPV (High Frequency Pitch Variation).





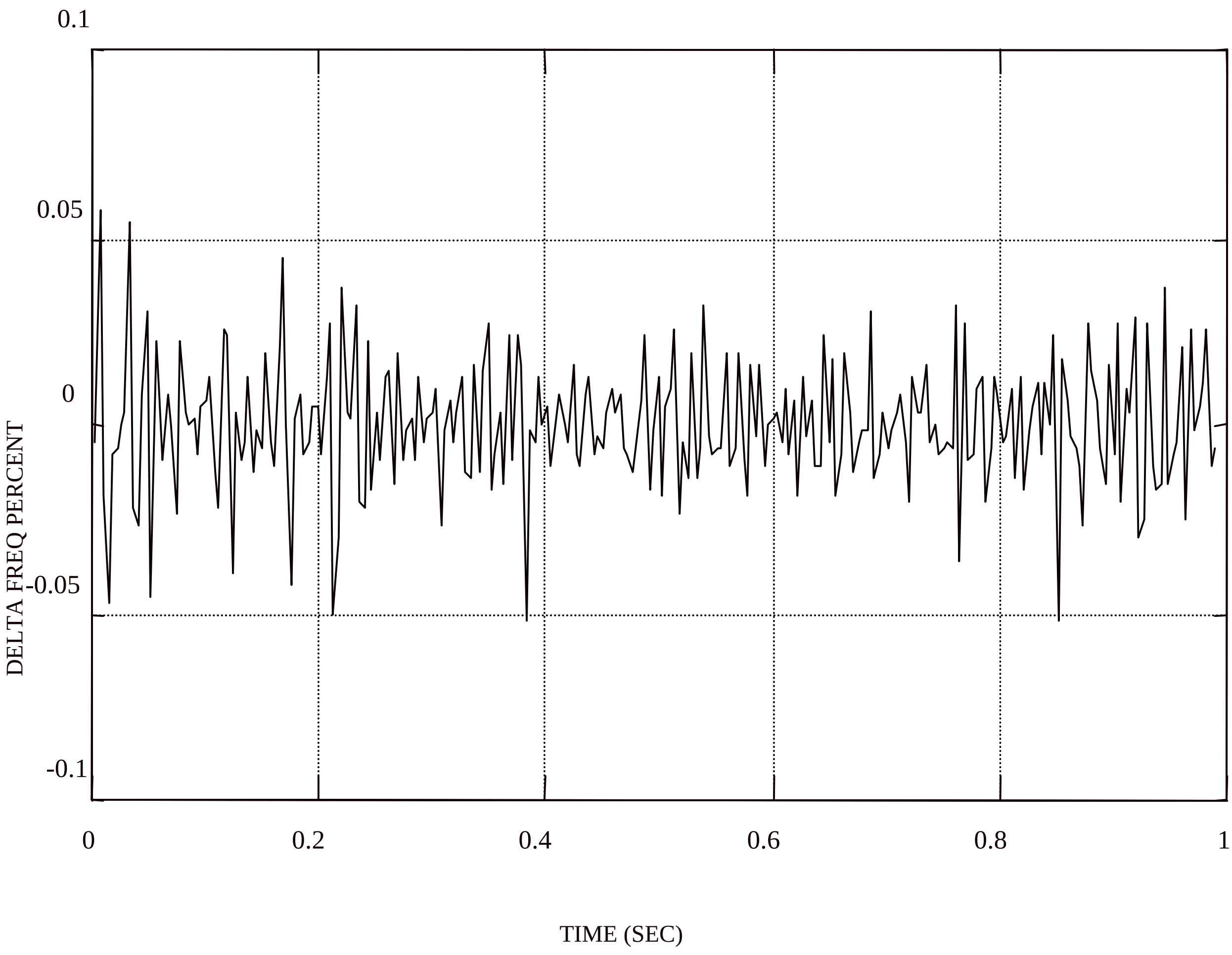
GAUSSIAN HFPV

Successful pitch tracking yields a Gaussian distribution in HPFV. The standard deviation is a convenient measure of HFPV.



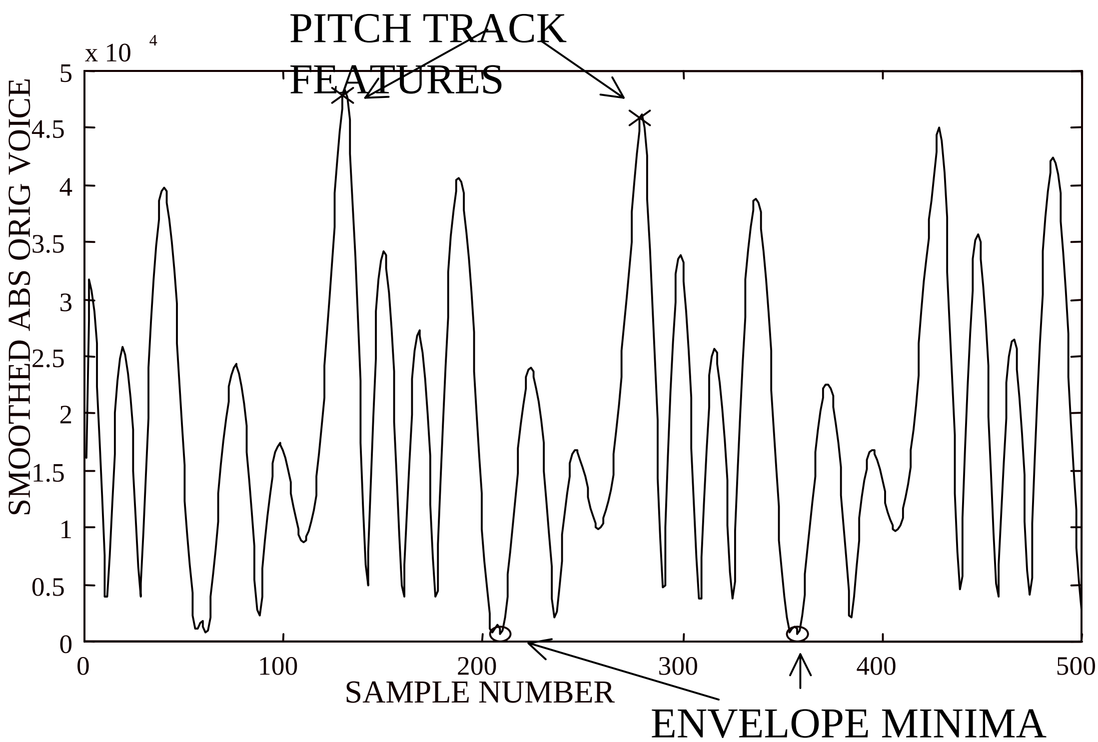
FM DEMODULATION

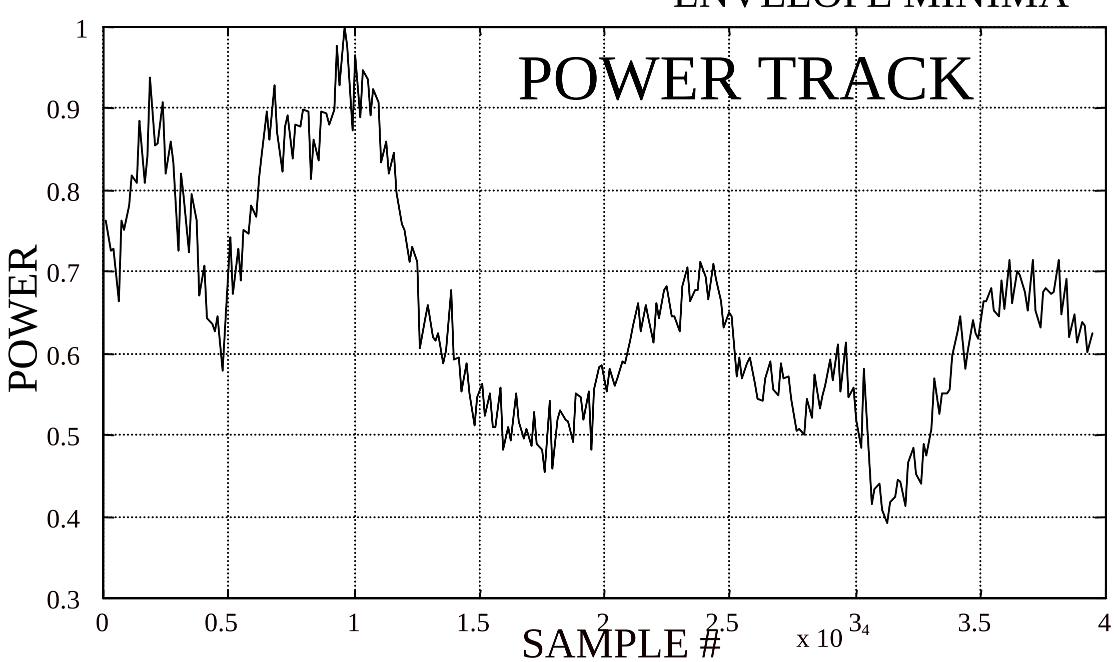
The pitch track may be used to demodulate the original voice to obtain a version with almost no pitch variation; re-tracking verifies constant pitch (<0.1%).



POWER TRACKING

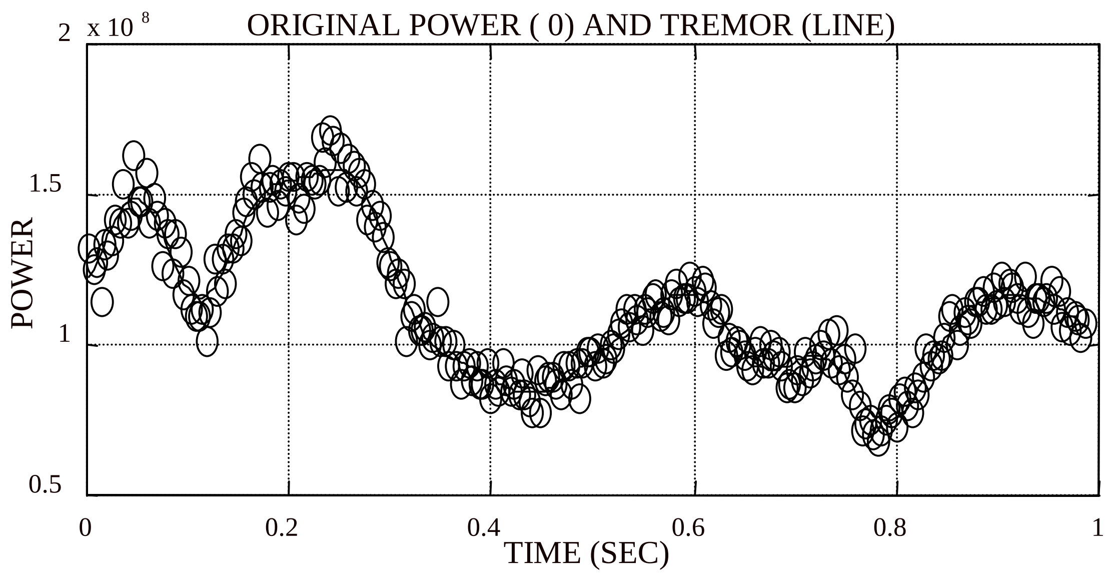
Analogously to pitch tracking, voice power is tracked.

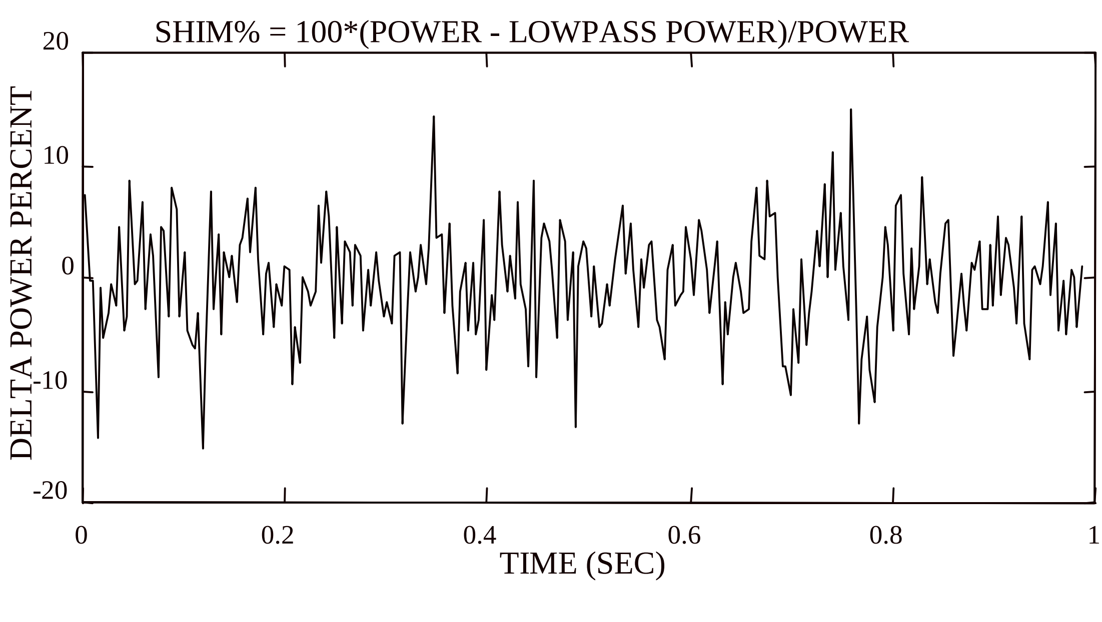




POWER SEGREGATED TO LOW AND HIGH FREQUENCY

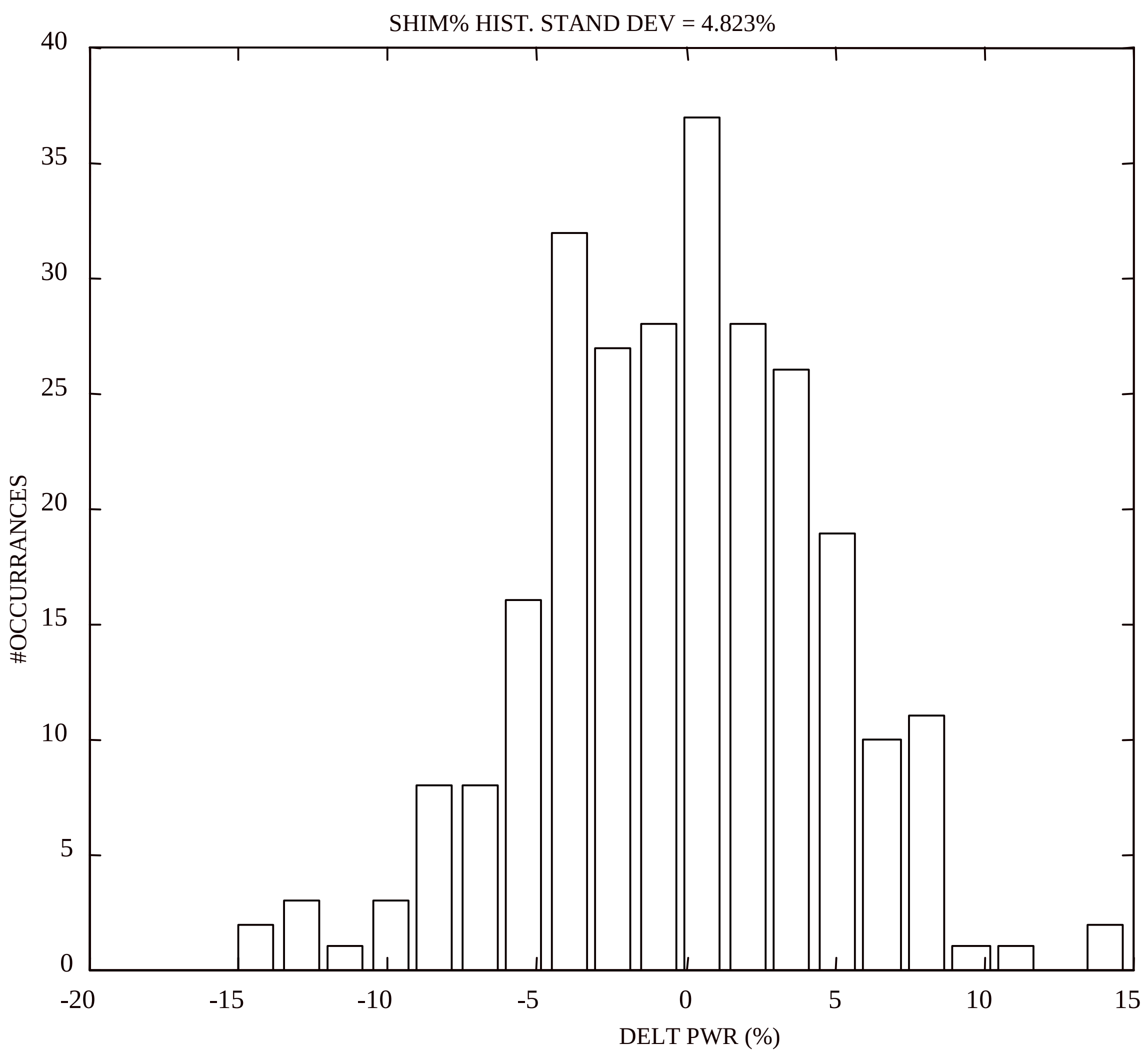
The power track is low/hi pass filtered to yield low frequency power variations and high frequency shimmer.





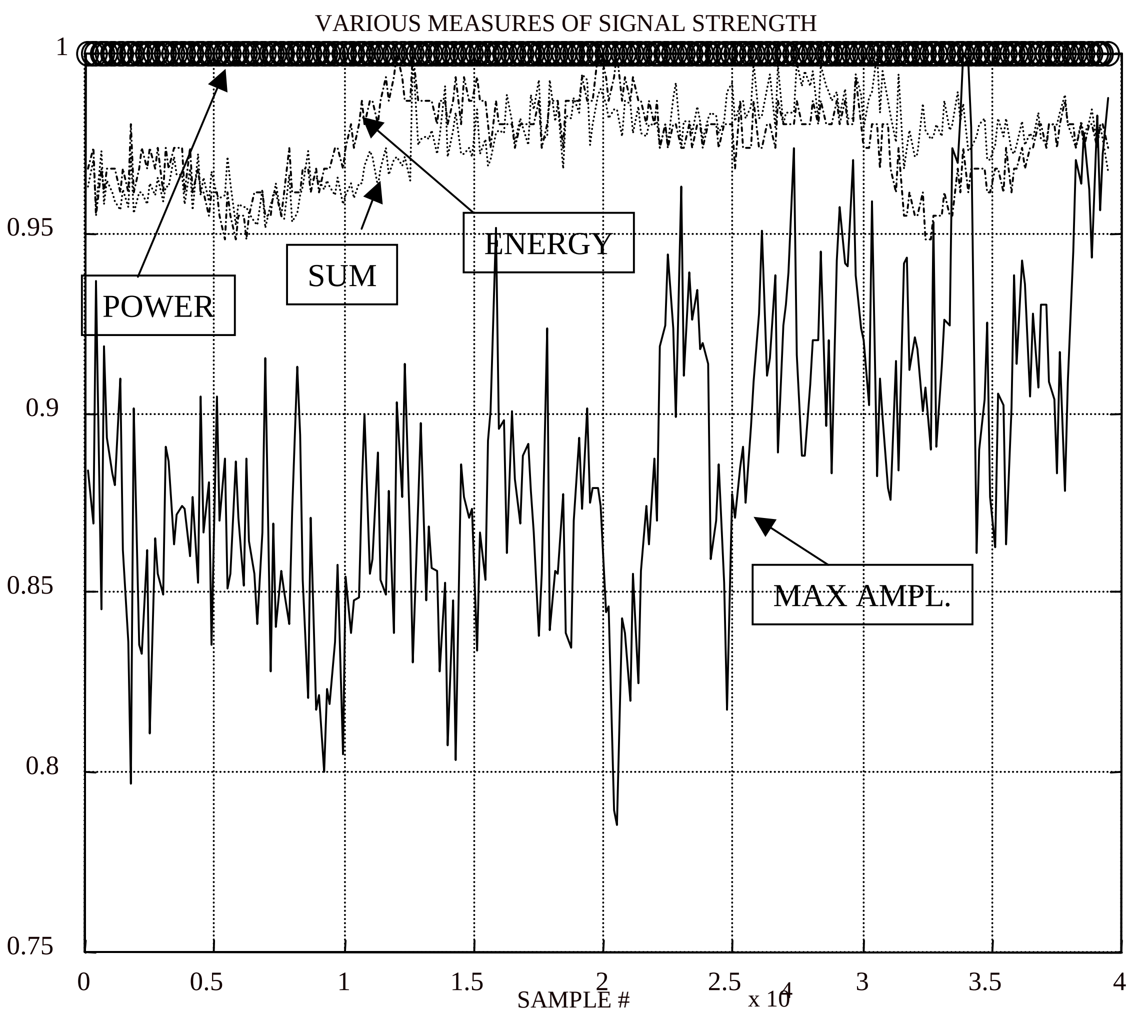
GAUSSIAN POWER VARIATIONS

Shimmer also displays Gaussian variations.



AM DEMODULATION

The power track may be used to demodulate the original voice to obtain a version with almost no variation in strength; re-tracking verifies constant power.



ASPIRATION NOISE ANALYSIS

Aspiration noise is segregated via spectral techniques. Peaks in the FFT of the log of the FFT (cepstrum) represent periodic energy, and are filtered out with a comb filter (lifter). Results are used to calculate noise-to-signal ratio (NSR).

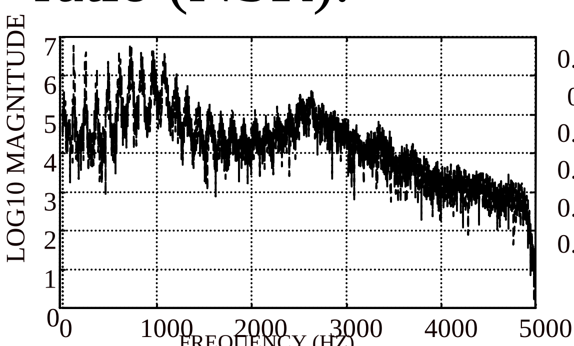


Figure 2.30a. PSD oforiginal voice.

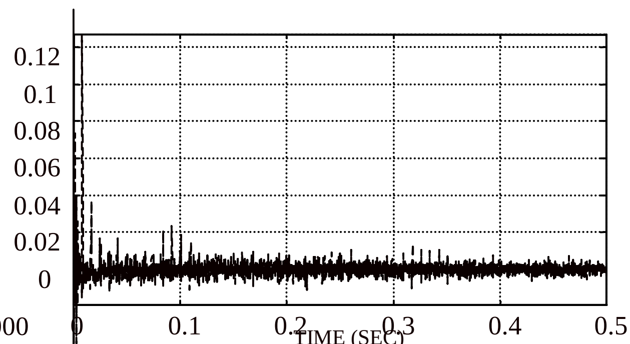


Figure 2.30b. Cepstrum of original voice.

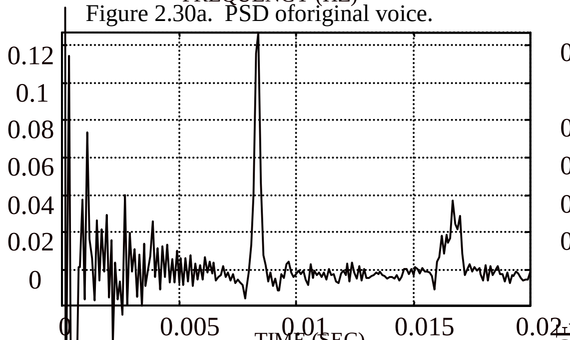


Figure 2.30c. Cepstrum (expanded scale).

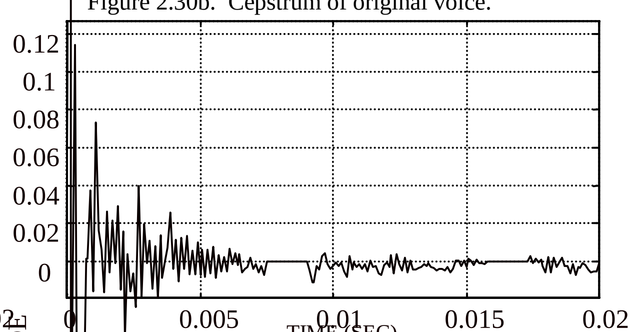


Figure 2.30d. Comb-liftered cepstrum of 14c.

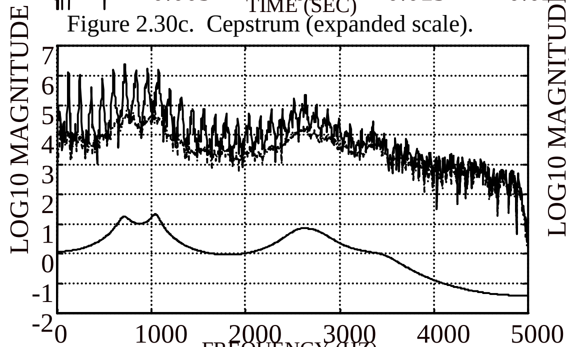


Figure 2.30e. Orig PSD, aspriation PSD, and vocal tract.

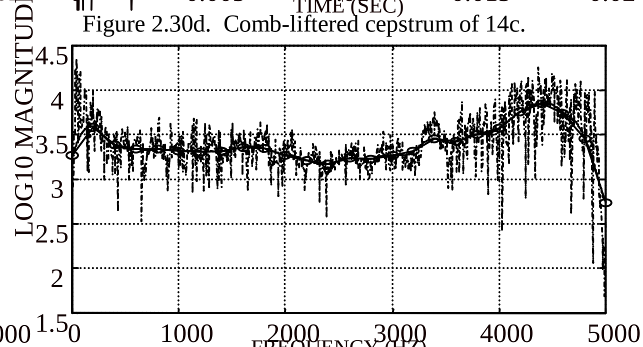
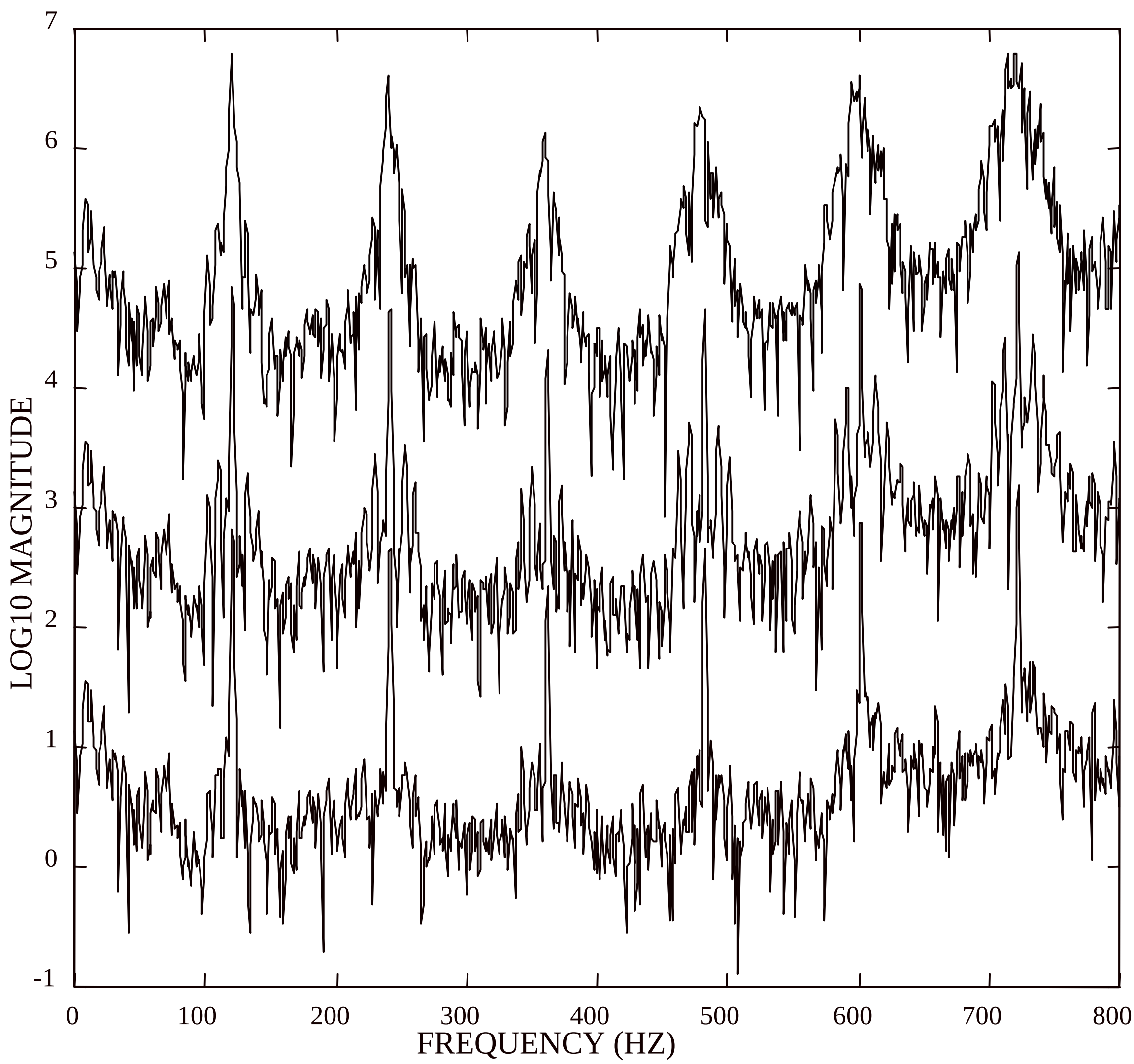


Figure 2.30f. Source aspiration PSD with vocal tract removed and fitted to 25-point piecewise-linear model.

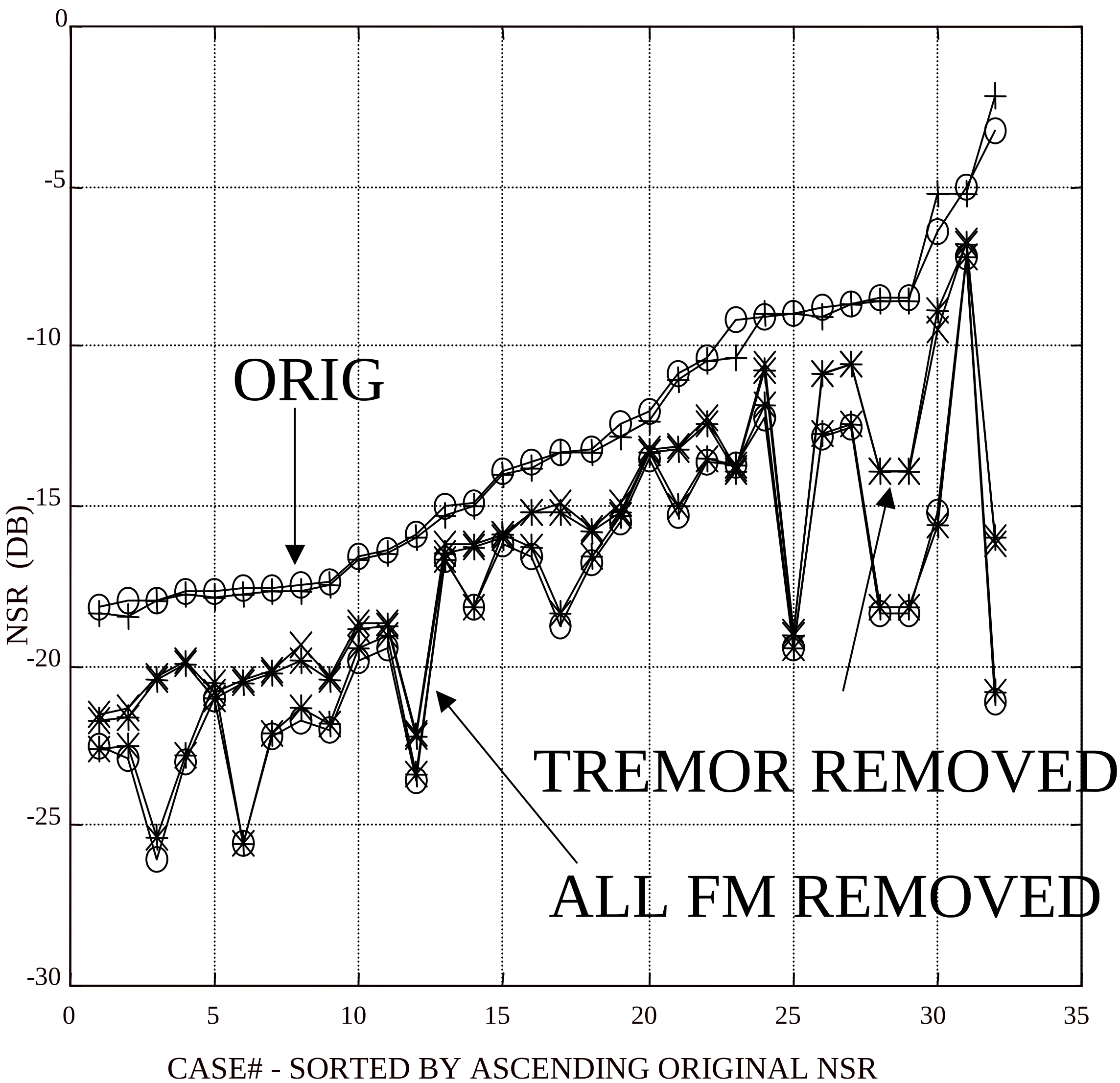
FM DEMODULATION IMPROVES NSR ACCURACY

Using FM demodulation improves resolution of spectral peaks of periodic components, thus allowing longer FFT windows and more accurate NSR determination.



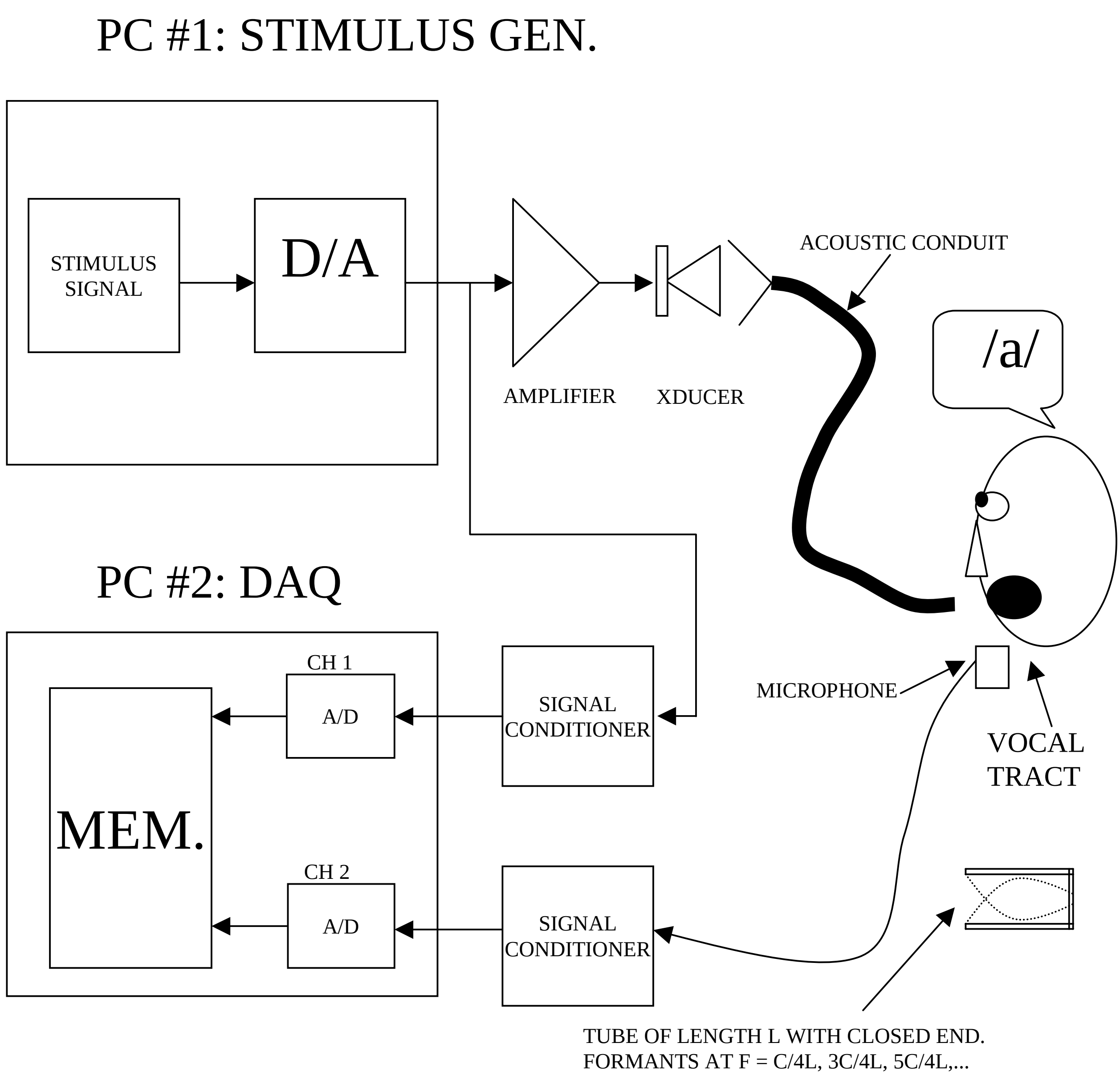
CHANGES IN NSR AFTER FM AND AM DEMODULATION

FM demodulation reduces NSR measures by up to 20 dB, yielding results closer to perceived levels. AM demodulation has much less effect.



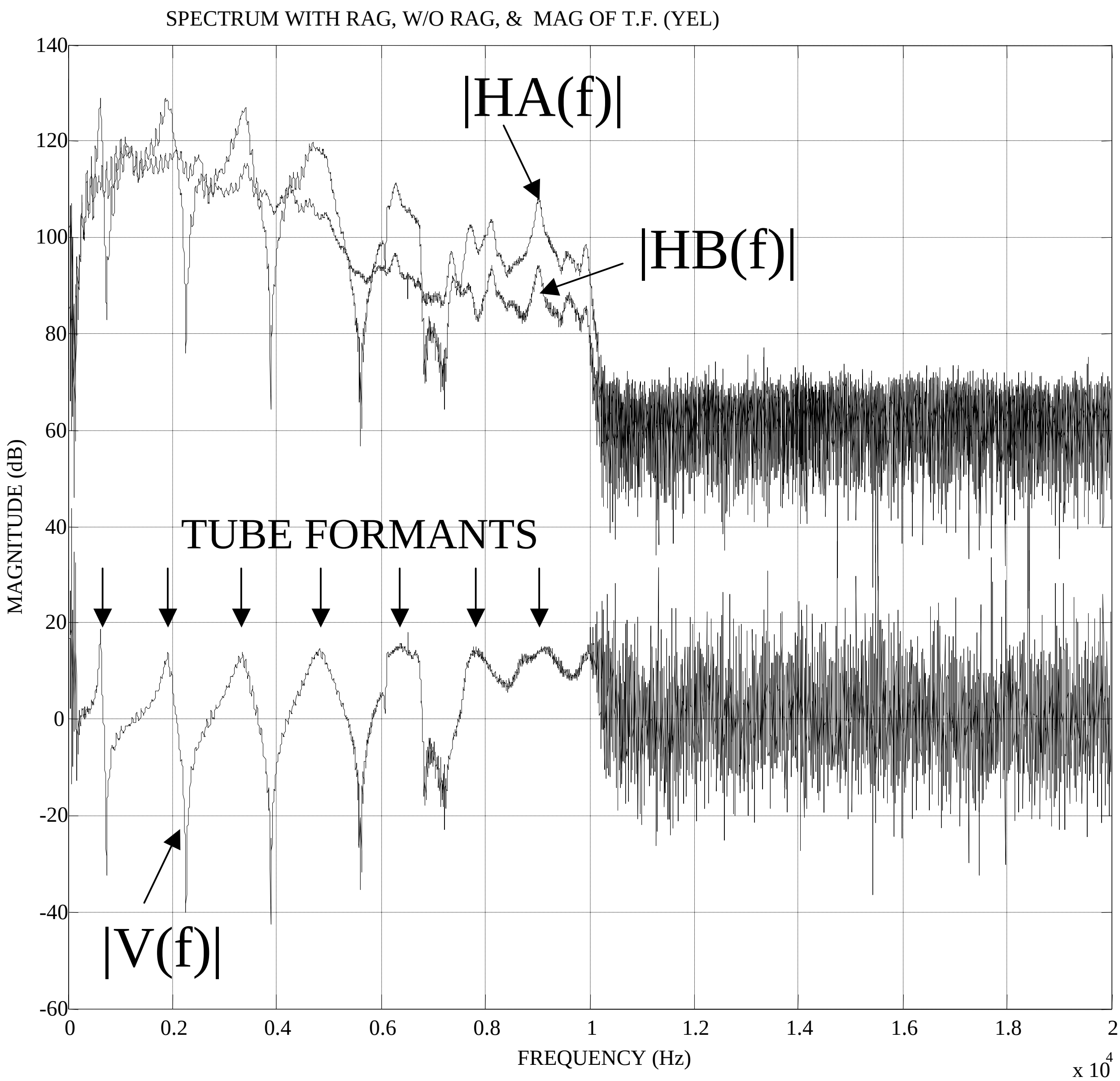
EXTERNAL (ES) SOURCE ANALYSIS

Source-filter ambiguity may be resolved by augmenting the glottal source with an known external stimulus. [Epps].



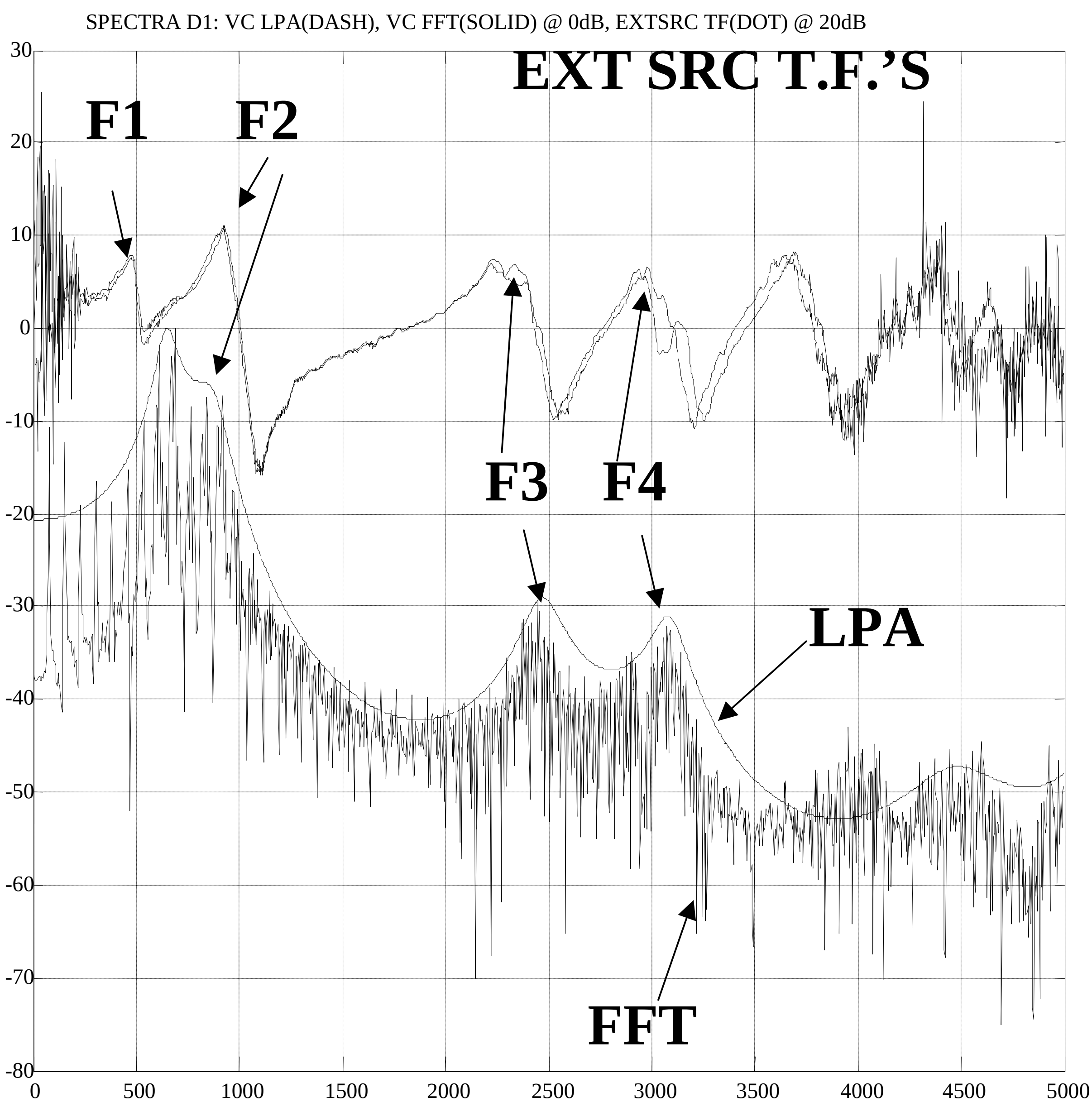
ES VERIFICATION

A simple plastic tube model verified the ES experimental setup. Resonances occur at expected frequencies.



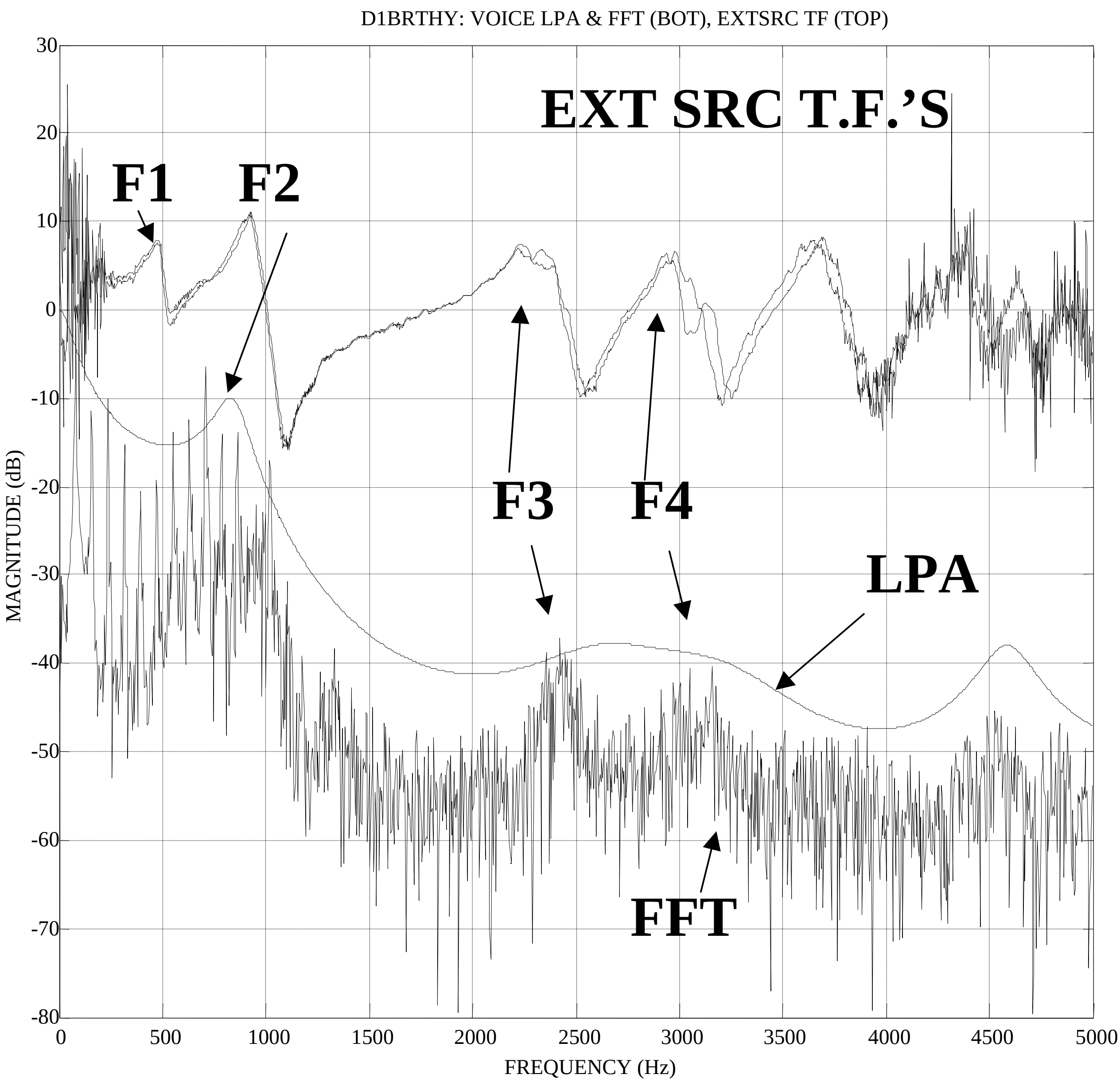
ES: NORMAL /a/

LP & FFT analysis show consistent results with ES analysis for a normal vowel.



ES: SIMULATED BREATHY /a/

LP & FFT analysis show poor resolution for F3 and F4 for a breathy /a/, while the ES resolution for F3 and F4 remains good.



SYNTHESIS OF PATHOLOGICAL VOWELS

Synthesis is a critical step in the study of pathological vowels. It provides evidence of the success of analysis and modeling steps via immediate comparisons of original and synthetic voice.

Two synthesizers were implemented:

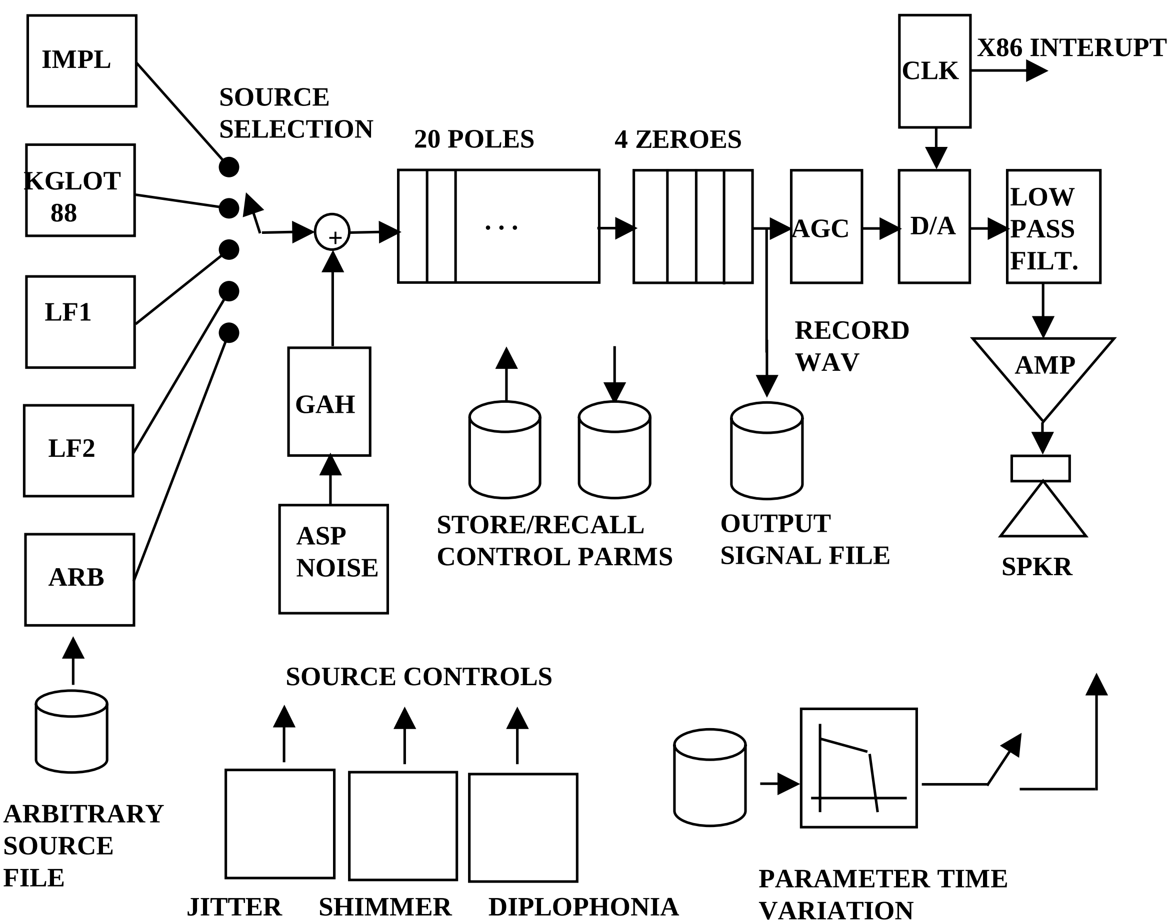
1. A realtime hardware-based synthesizer capable of providing instant response to changes in model parameters.

2. A software synthesizer implemented in MATLAB with extended features, convenient graphical interface, and ease of modification.

REALTIME SYNTHESIZER

- Implemented in native X86 assembly language
- Executes all code within 100us cycle
- Overrides PC OS to achieve determinancy
- Employs dedicated clock, I/O, and control hardware implemented in a wire-wrap PCB adapter card

CURRENT REALTIME SYNTHESIZER FUNCTIONAL OVERVIEW



SYNTHESIS VALIDATION

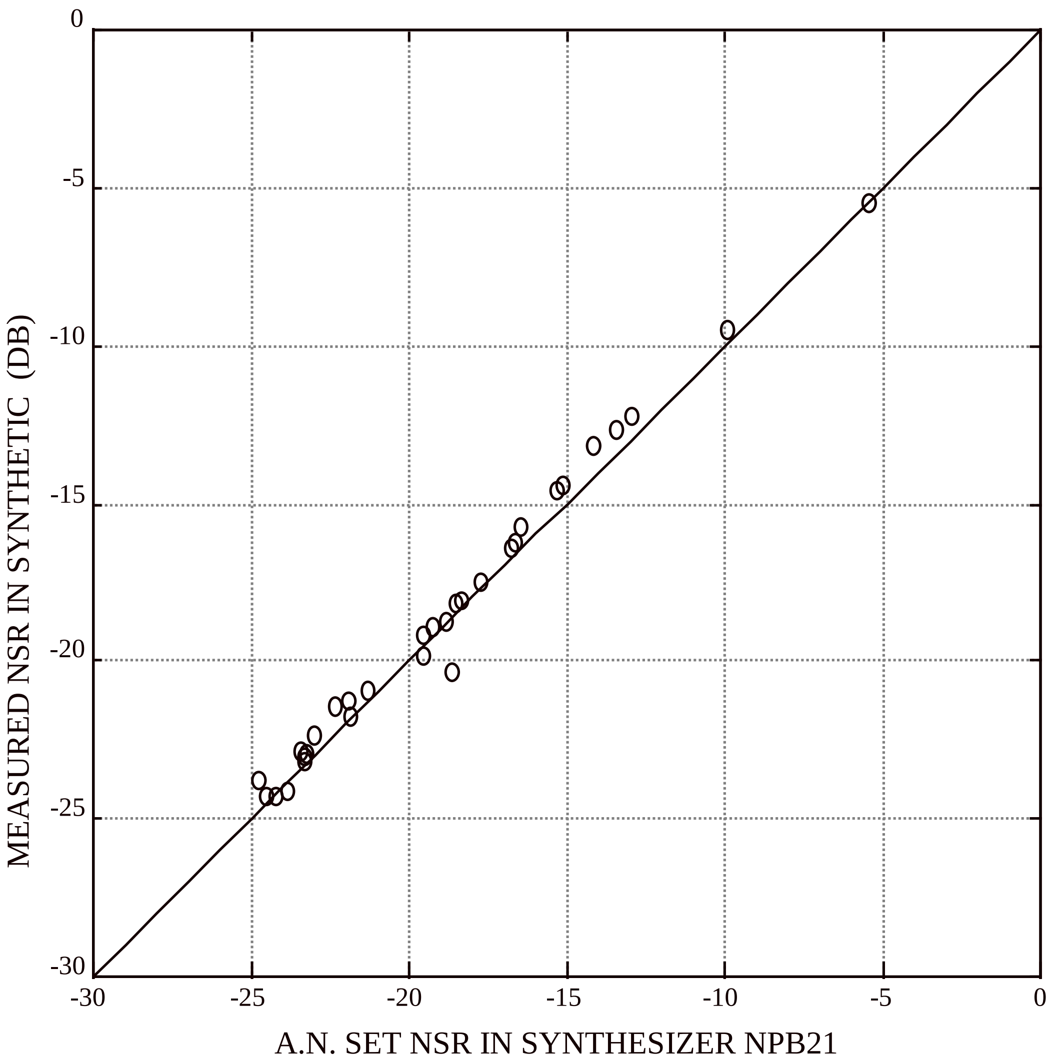
The current model, analysis tools, and synthesizers yield a high level of fidelity in generation of synthetic pathological vowels. The system is currently employed at the UCLA Voicelab for NIH funded perceptual studies.

In order to objectively validate the analysis/synthesis process, the loop is closed by re-analyzing the synthetic time series to confirm parameter values.

Re-analysis also provides opportunity to observe interactions of nonperiodic components.

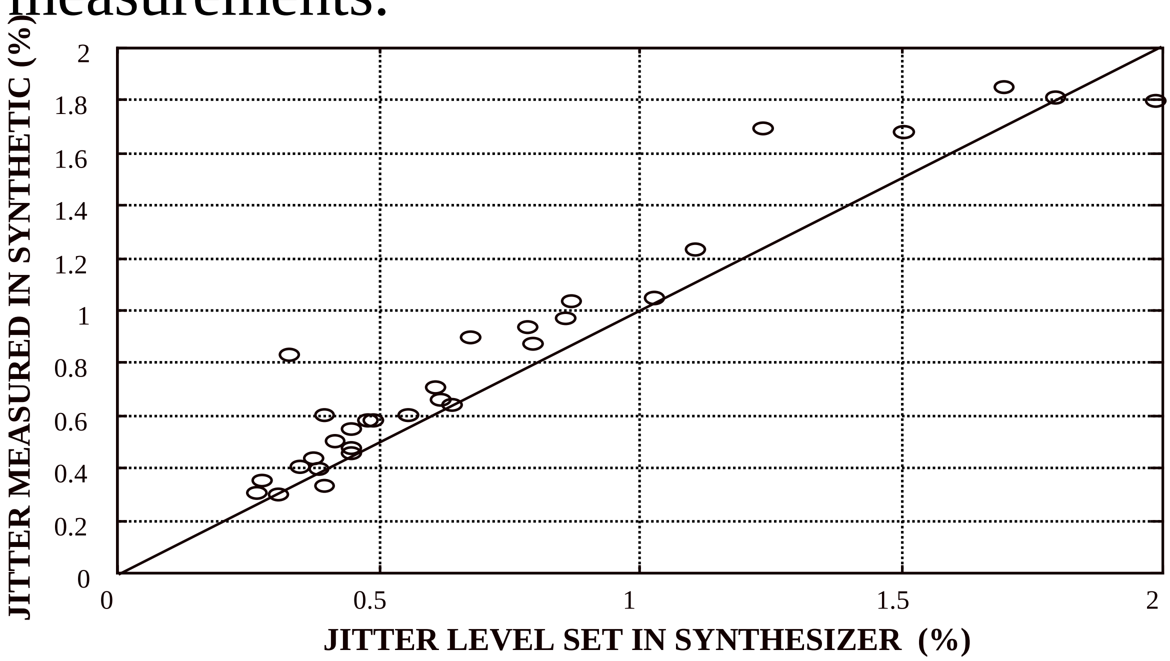
SYNTHESIS VALIDATION

Re-measured synthetic aspiration noise level agrees with level set in synthesizer.

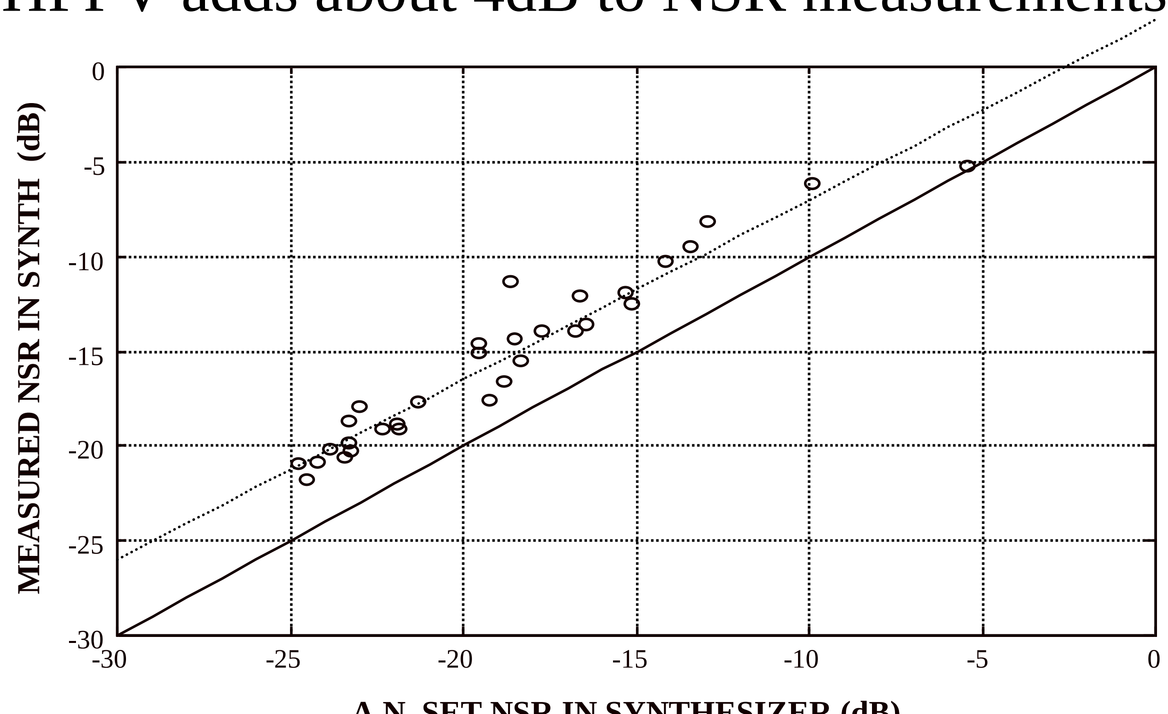


SYNTHESIS VALIDATION

Aspiration noise adds about %0.2 to HFPV measurements.



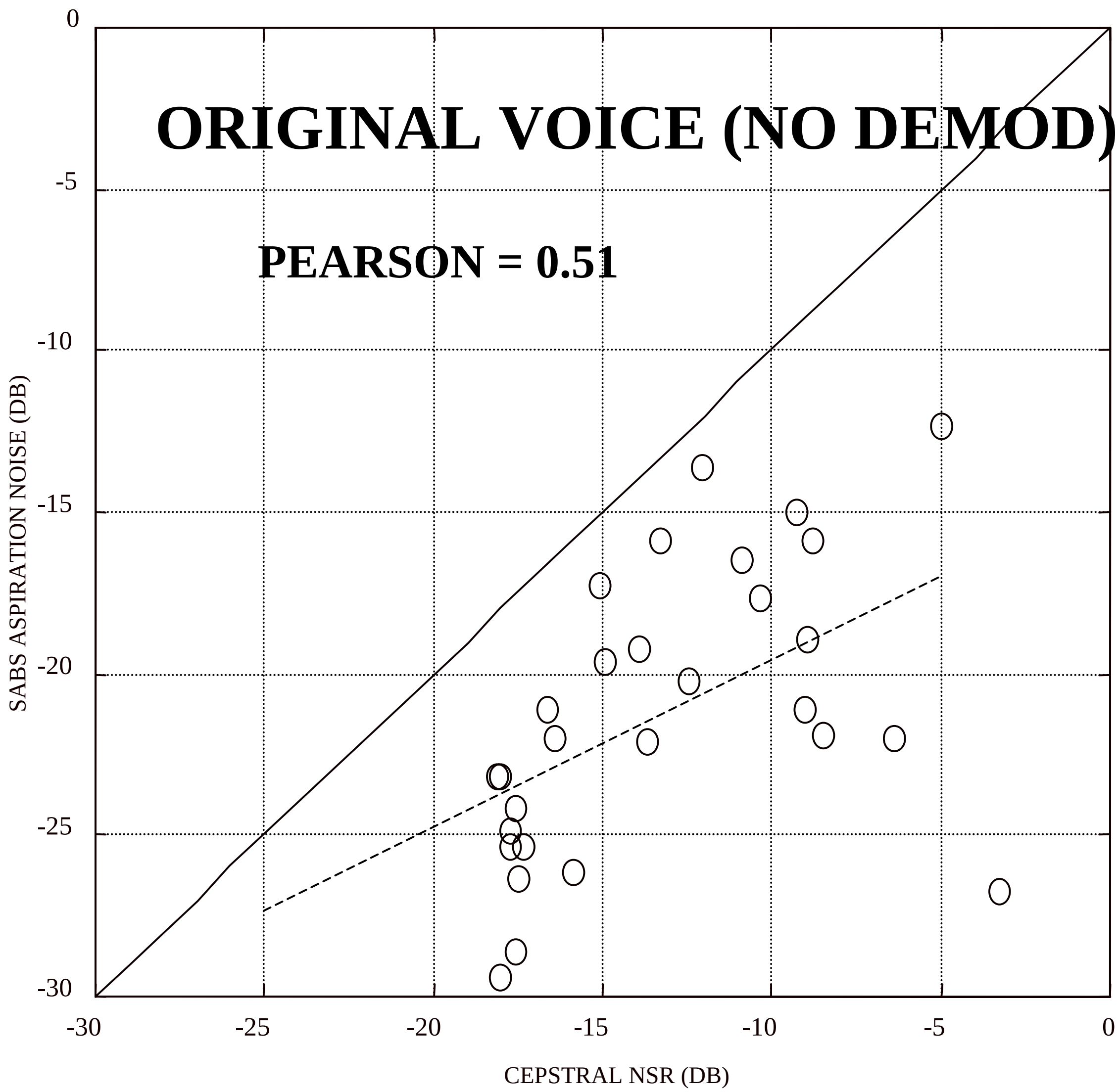
HFPV adds about 4dB to NSR measurements.



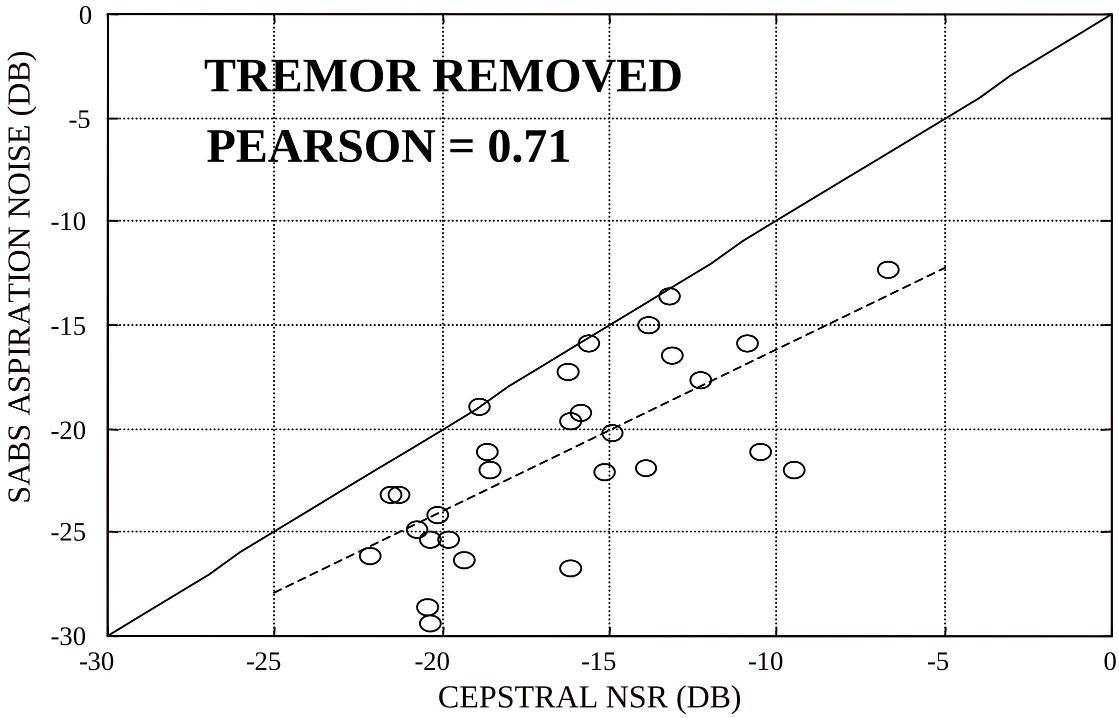
SYNTHESIS VALIDATION

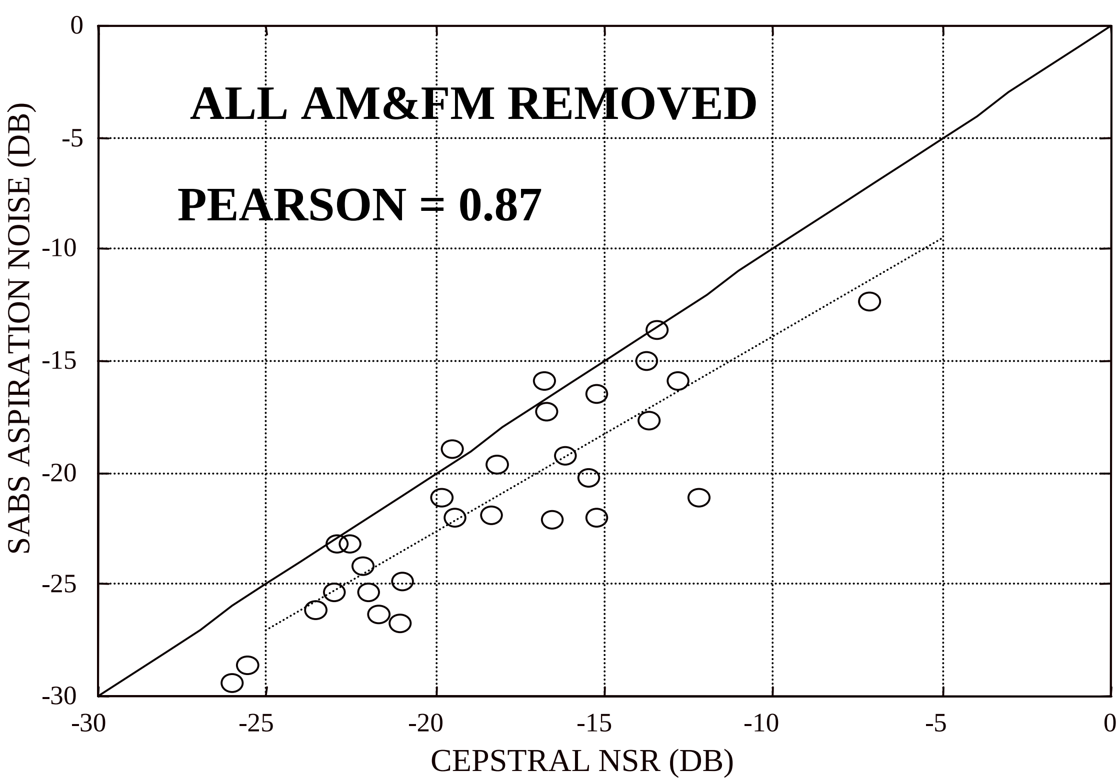
Subjective analysis by synthesis experiments demonstrate the success of AM and FM demodulation in achieving accurate modeling of nonperiodic features. Listeners adjust synthetic aspiration noise to match original.

Match improves with demodulation



SYNTHESIS VALIDATION





SUMMARY

This study has achieved improved automatic, objective analysis and synthesis of speech within the specialization of pathological vowels. Specific accomplishments include:

- A unique, symmetric model for nonperiodic components as AM, FM and spectrally-shaped aspiration noise
- Improved accuracy of noise analysis via AM & FM demodulation
- Application of ES formant identification for pathological vowels.
- Implementation of realtime and offline specialized high fidelity vowel synthesizers