Department of Electrical Engineering
Technion – Israel Institute of Technology

Artificial Bandwidth Extension of Band Limited Speech Based on Vocal Tract Shape Estimation

Itai Katsir

MSc. Research

Supervised by Prof. David Malah
and Prof. Israel Cohen

30-Nov-11

Outline

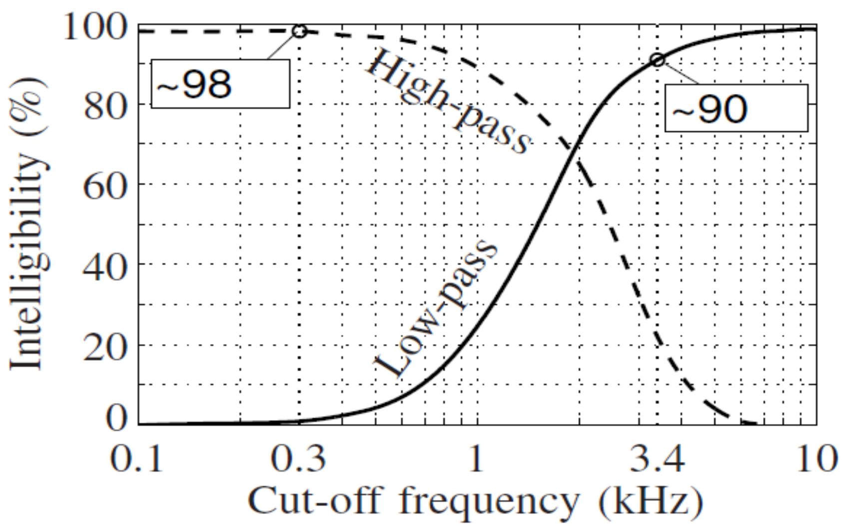
- Introduction
- Methods of BWE
- Proposed BWE Algorithm
- Performance Evaluation
- Conclusion

Outline

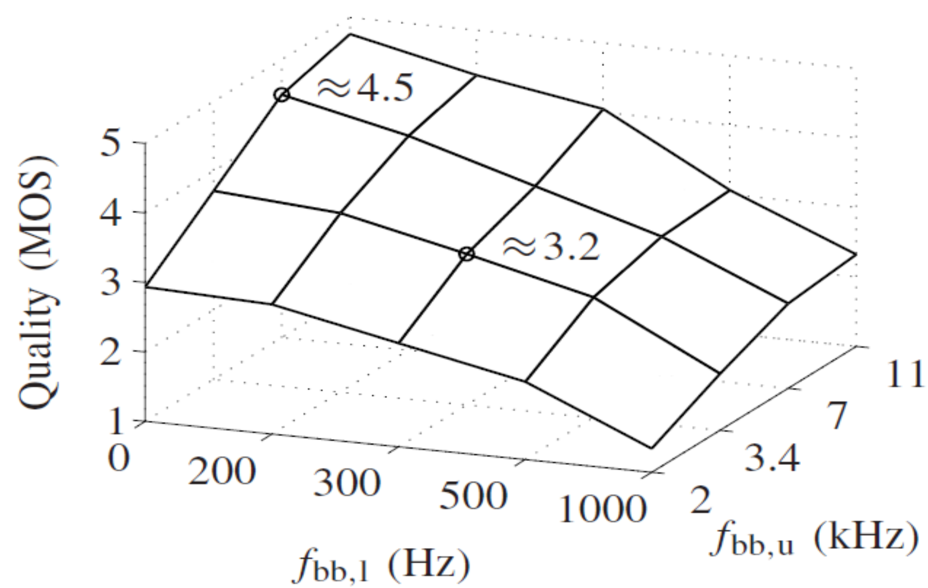
- **Introduction**
- Methods of BWE
- Proposed BWE Algorithm
- Performance Evaluation
- Conclusion

Introduction (1/8)

- **Fact** – Growing consumer demand for HD media –> high quality speech communication.
- **Problem** – Today's analog telephone and PSTN limit the speech to narrowband (NB) frequency range of about 300-3400Hz –> lower speech quality compared to wideband (WB) speech of range 50-7000Hz.



(a) Intellig. of meaningless syllables

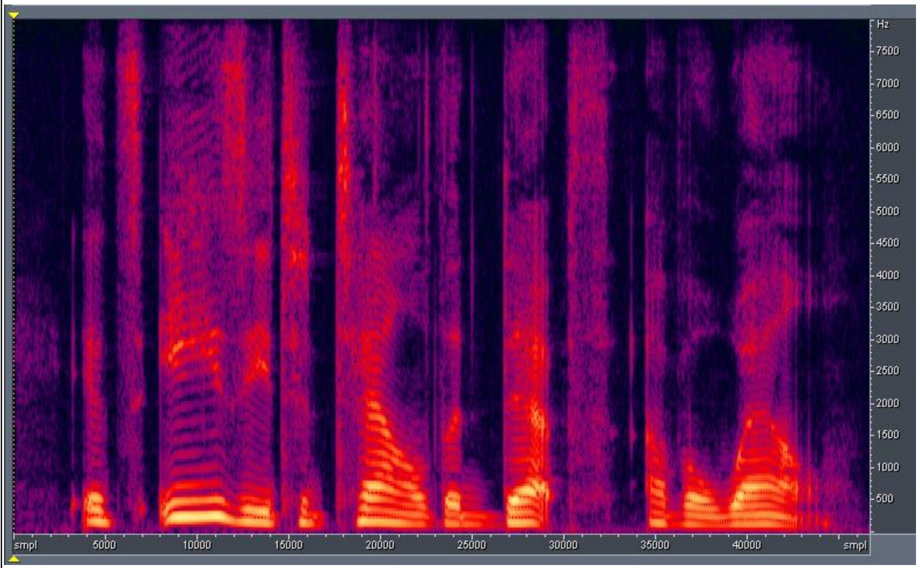


(b) Subjective speech quality

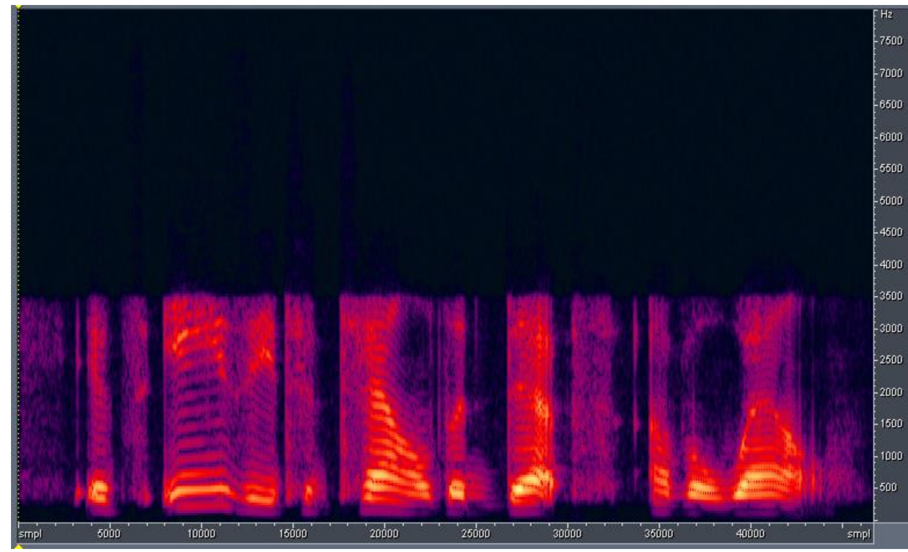
Introduction (2/8)

Wideband vs. Narrowband Speech

Wideband (WB)



Narrowband (NB)



Introduction (3/8)

Wideband vs. Narrowband Speech

- "Seed, Feed" – Spoken at different bandwidths.
- What do they sound like?
  - Reference: G.A. Miller and P.E. Nicely, "An analysis of perceptual confusions among some English consonants" Lincoln Laboratory, MIT, 1955 (*J. Acoust. Soc. Amer.* Vol. 27, pp. 338-352)

300-3400 [kHz] 50-5000 [kHz] 50-7000 [kHz]

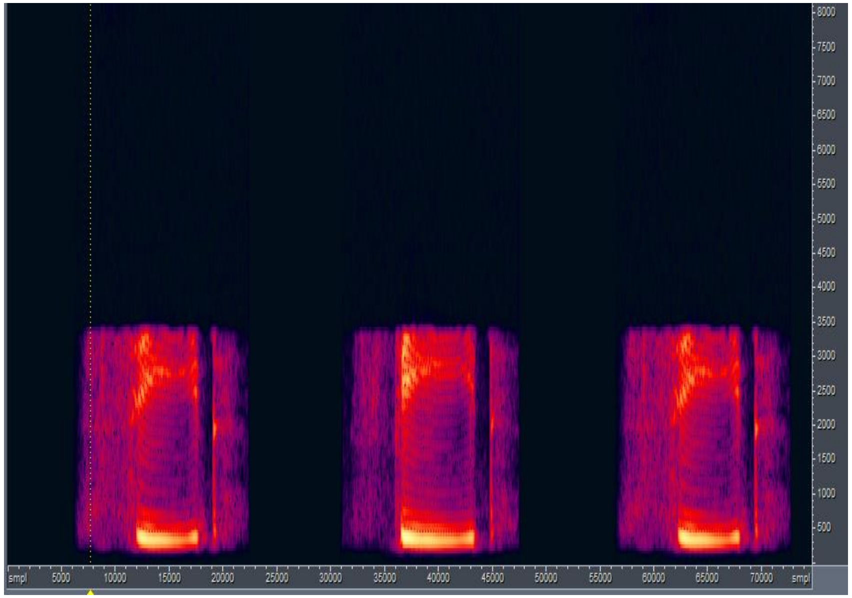
- The same sentence – "Seed, Feed, Seed" in all cases !
- Spectrograms…

Introduction (4/8)

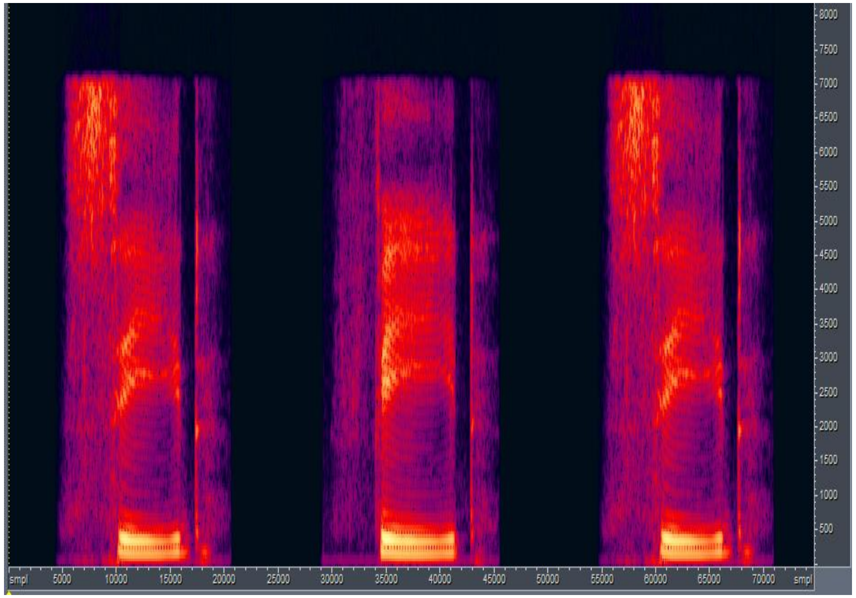
Wideband vs. Narrowband

- Spectrograms

300-3400 [kHz]



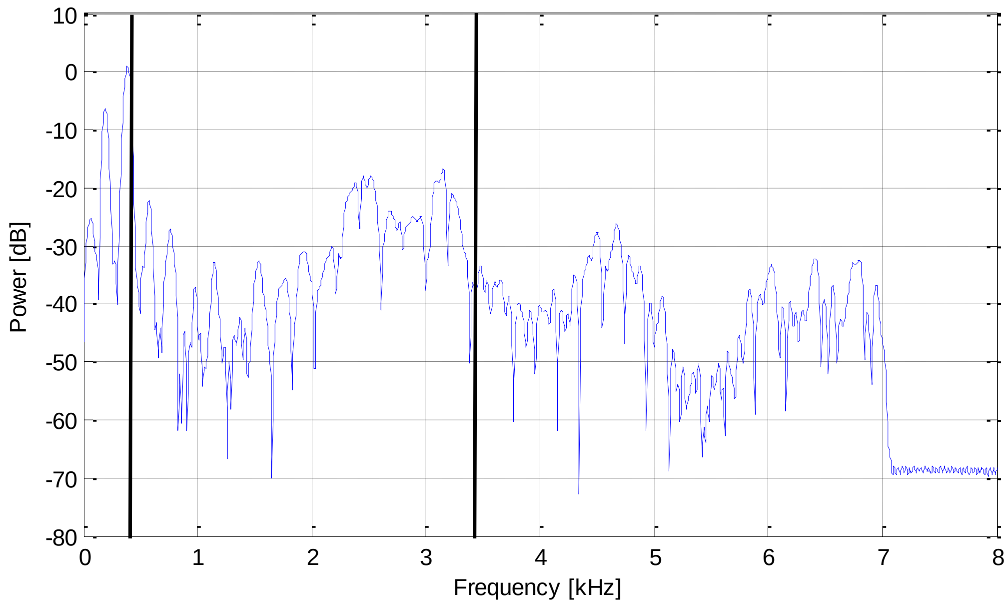
50-7000 [kHz]



Introduction (5/8)

Voiced sounds

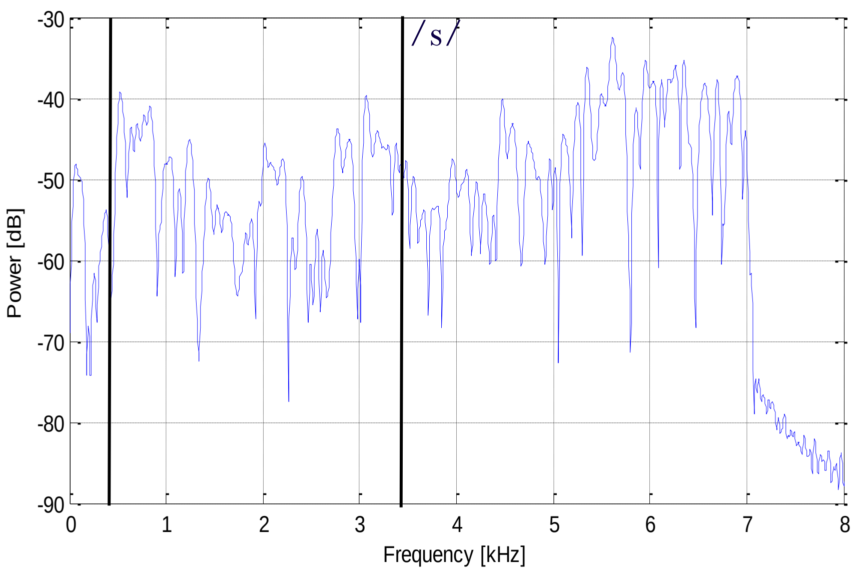
- Most of the energy is present in the low frequencies -> filtering out below 300 Hz affects **speech naturalness.**

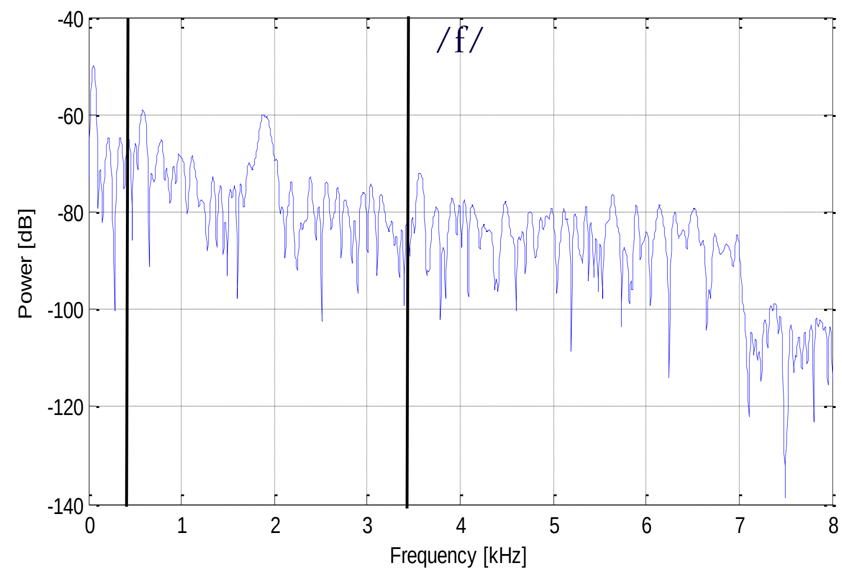


Introduction (6/8)

Unvoiced sounds

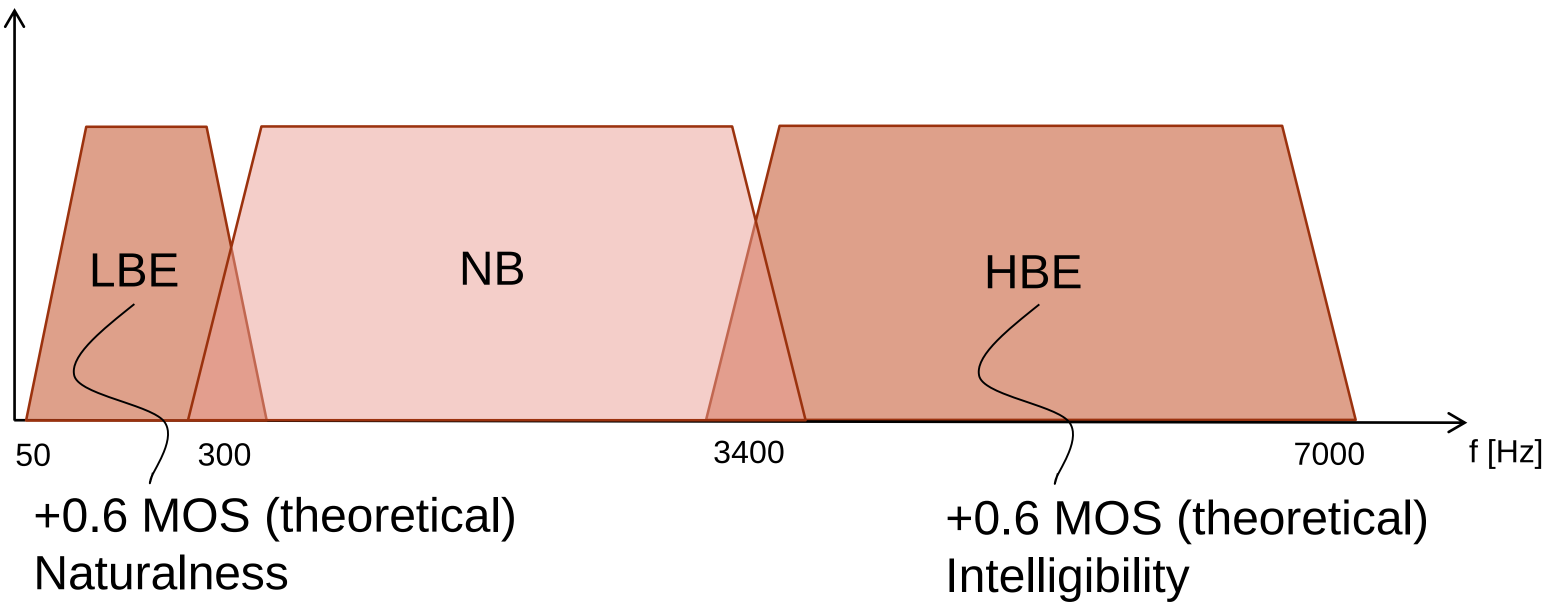
- Important energy above 3.4 kHz -> filtering out affects **speech intelligibility** such as differentiation between “s” and “f”.





Introduction (7/8)

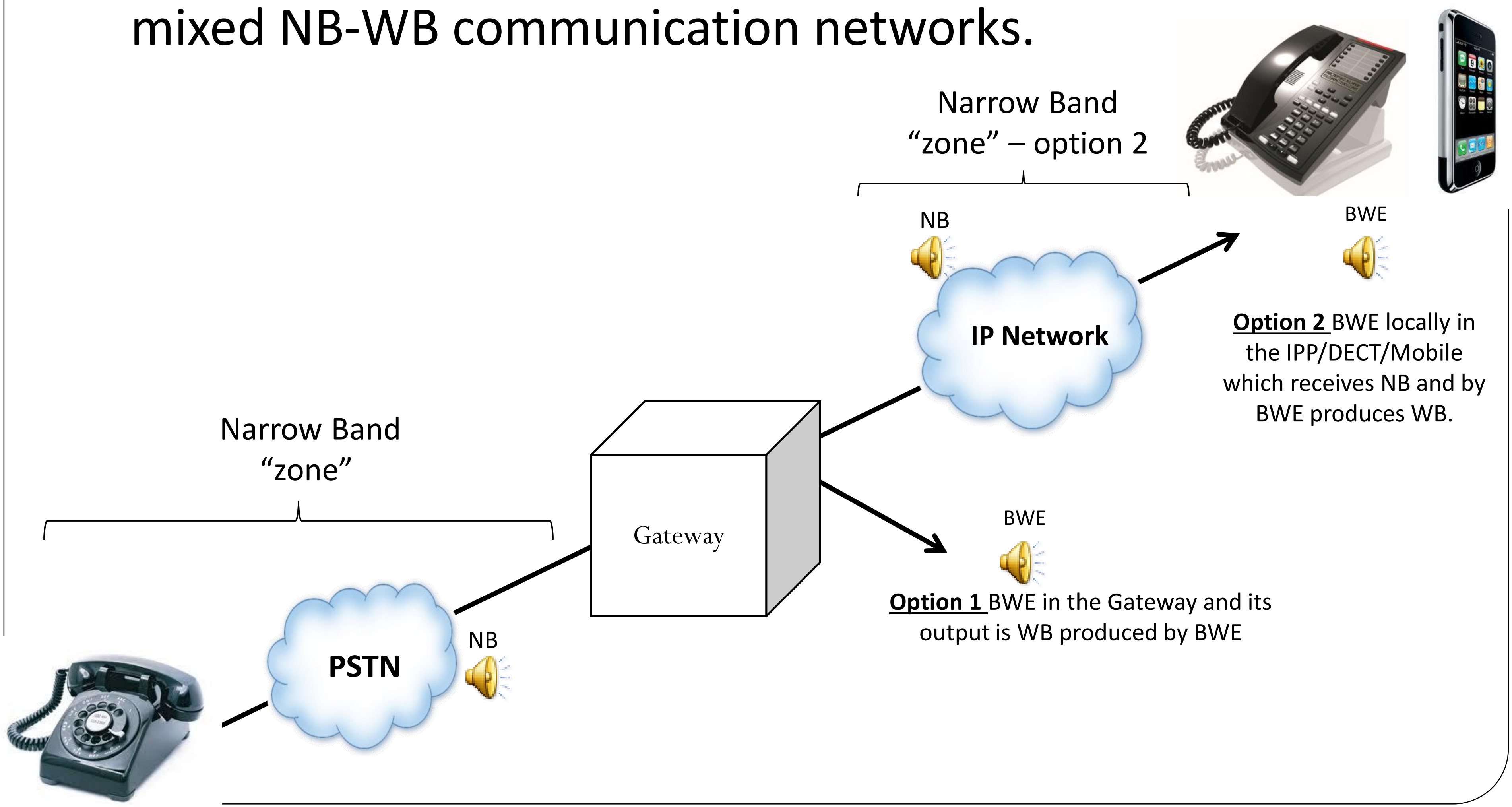
- **Solution –** Artificially extend speech bandwidth to achieve speech quality enhancement and “listening effort” reduction.
  - 3.4-7kHz – Higher intelligibility and quality
  - 0-0.3kHz – Higher naturalness and quality



- +0.3 MOS - state of the art published BWE algorithms.

Introduction (8/8)

- **Application –** In the transition time to full WB communication networks, BWE can be used in mixed NB-WB communication networks.



Outline

- Introduction
- **Methods of BWE**
- Proposed BWE Algorithm
- Performance Evaluation
- Conclusion

Methods of BWE (1/8)

Model-less methods (Non parametric)

- Extend the bandwidth of the input narrowband speech signal directly, i.e., without any signal analysis.
- Utilize rather simple signal processing techniques (filtering and resampling), in time or frequency domain.

Model-based methods (Parametric)

- Estimate WB speech parameters from NB parameters.
- Rely on state-of-the-art signal processing techniques taken from pattern recognition, estimation theory, signal classification, etc.

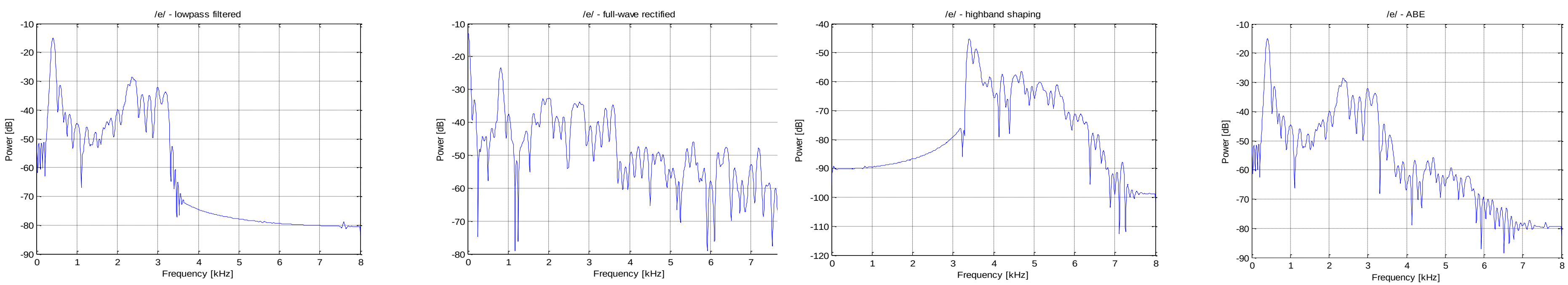
Methods – Model-less BWE (2/8)

General scheme

$s_{\mathrm{NB}}(n)$ → Spectral component generation of LB/HB → Spectral shaping of LB/HB → Gain adjustment and post processing of LB/HB → Synthesis of estimated wideband signal → $\tilde{s}_{\mathrm{WB}}(n)$

Narrowband speech signal

BWE speech signal

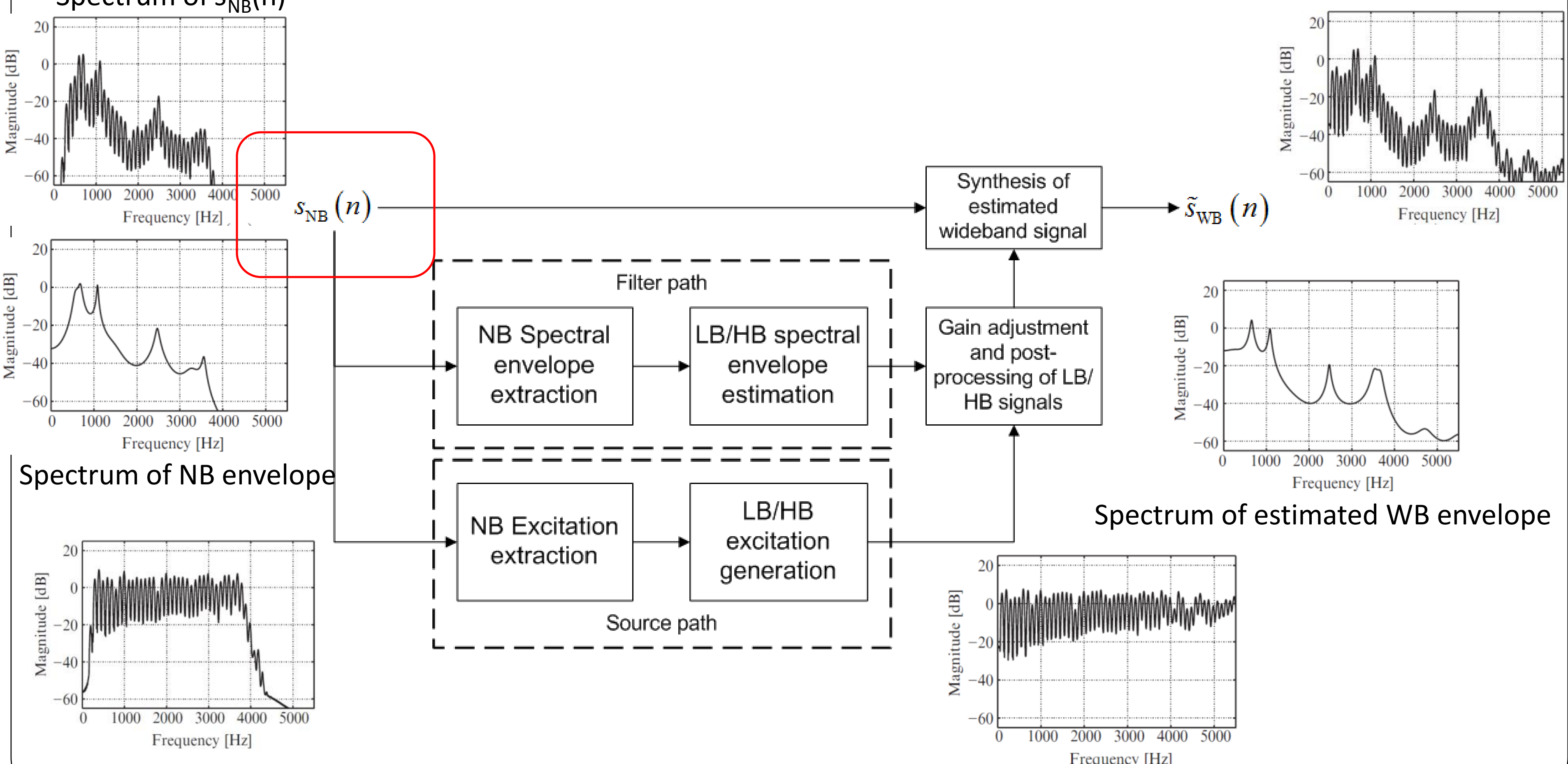


- Advantages
  - ➢ Low complexity as compared to parametric methods.
- Drawback
  - ➢ Lower quality as compared to parametric methods.

Methods – Model Based BWE (3/8)

General scheme

- Allows separate and independent algorithms for excitation extension and spectral envelope extension.



Spectrum of s_{NB}(n)

Spectrum of NB envelope

Spectrum of NB excitation

Spectrum of estimated s_{WB}(n)

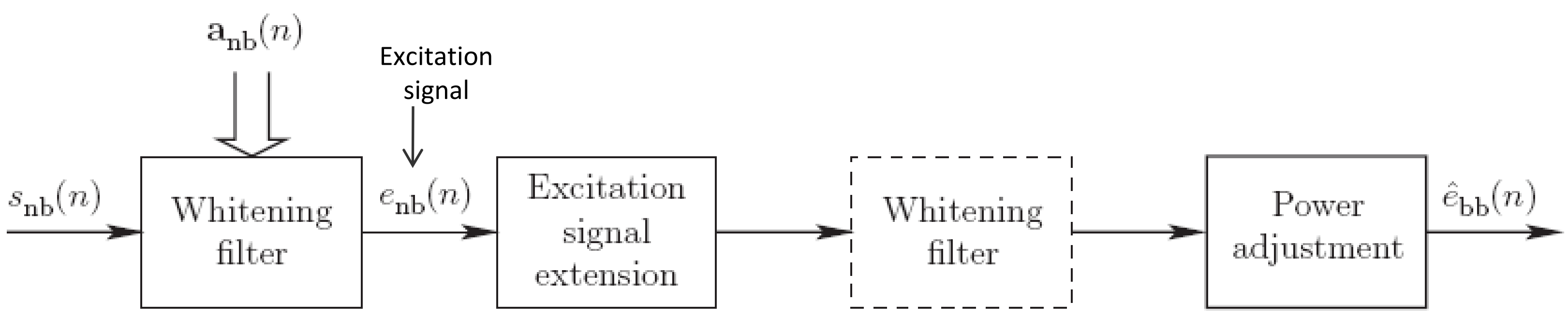
Spectrum of estimated WB envelope

Spectrum of generated WB excitation

Methods – Model Based BWE (4/8)

Excitation BWE - General scheme

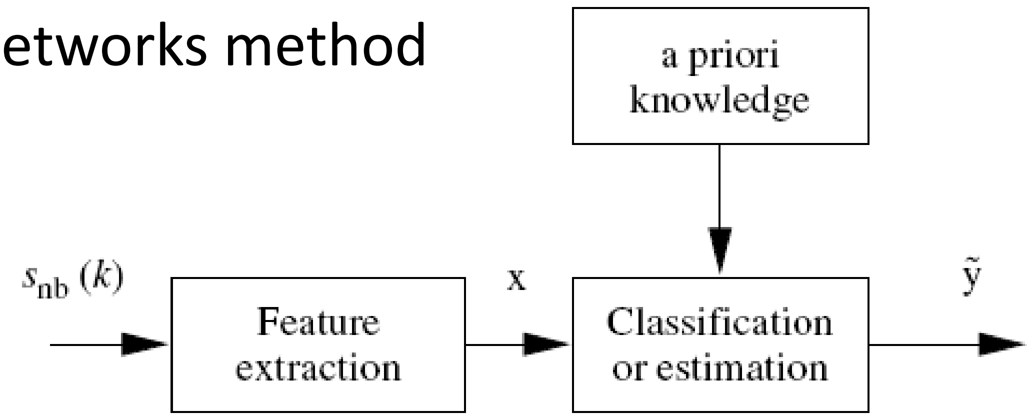
- Generation of WB excitation from NB excitation and NB scalar parameters.
- Various methods for wideband excitation generation.
  - Spectral shifting
  - Non-linear operators
  - Function generation



Methods – Model Based BWE (5/8)

Spectral envelope BWE - General scheme

- Estimate wideband spectral envelope $\tilde{y}$ from narrowband signal features $\mathbf{x}$ (narrowband spectral envelope and voiced/unvoiced scalar features).
- Various methods for wideband spectral envelope estimation.
  - Linear or piece-wise linear mapping
  - Codebook mapping
  - Bayesian estimation based on Gaussian Mixture Models (GMMs) or Hidden Markov Models (HMMs)
  - Neural networks method



Methods – Existing Algorithms (6/8)

	Energy shift	Spectral shaping	Gain adjustment	Results
Nokia, Laaksonen [2008] (model-less)	Spectral folding	Frequency domain filtering using off-line trained control points	Normalized frequency domain filter	0.2 points improvement in MOS score for AMR coded signals
Ericsson, Gustafsson [2006] (model-based)	Spectral copy	Frequency domain filtering using estimated formants peaks	linear mapping based on off-line training	Evaluation of absence of distortion annoyance (using MOS scale)
Motorola, Ramabadran [2008] (model-based)	Full wave rectifier and noise generation	frequency domain filtering using 6, energy based, off-line trained spectral envelopes	mapping using off-line trained control constants	0.25 points improvement in MOS score
NTT – Quasi WB		Extending signal up to 6kHz		

Methods – Existing Algorithms (7/8)

	Energy shift	Spectral shaping	Gain adjustment	Results
Kornagel [2006] (model-based)	Spectral copy	LPC based codebook	narrowband energy equalization of estimated wideband signal to original signal	Used objective cepstral distance
Nilsson and Kleijn [2002] (model-based)	Repeated spectral folding of 2-3kHz excitation band and smooth transition to white noise	Estimate spectral envelope using GMM based mapping	Estimate gain using GMM based mapping	Used subjective degree of artifacts survey
Jax and Vary [2001] (model-based)	Modulation of narrowband excitation	Estimate spectral envelope and gain using HMM based mapping	Estimate gain using HMM based mapping	Used objective RMS LSD

Methods – Existing Algorithms (8/8)

Summary

- The major challenges are to estimate the high-band gain for unvoiced sounds and the spectral envelope for voiced sounds.
- BWE algorithms focus on tuning the extended bandwidth to minimize possible artifacts, by sophisticated spectral envelope and gain estimation techniques.
- Jumpy behavior of spectral envelope estimation, from one frame to another, in time, and from NB to HB shapes, in frequency, causes noticeable artifacts, like high frequency whistling.
- Power adjustment of synthesized signal to match narrow-band signal is crucial for artifact removal.
- Low-band frequencies estimation allows major improvement of BWE speech quality, but it is much more sensitive to erroneous estimation compared to high-band estimation.

Outline

- Introduction
- Methods of BWE
- **Proposed BWE Algorithm**
- Performance Evaluation
- Conclusion

Proposed BWE Algorithm (1/26)

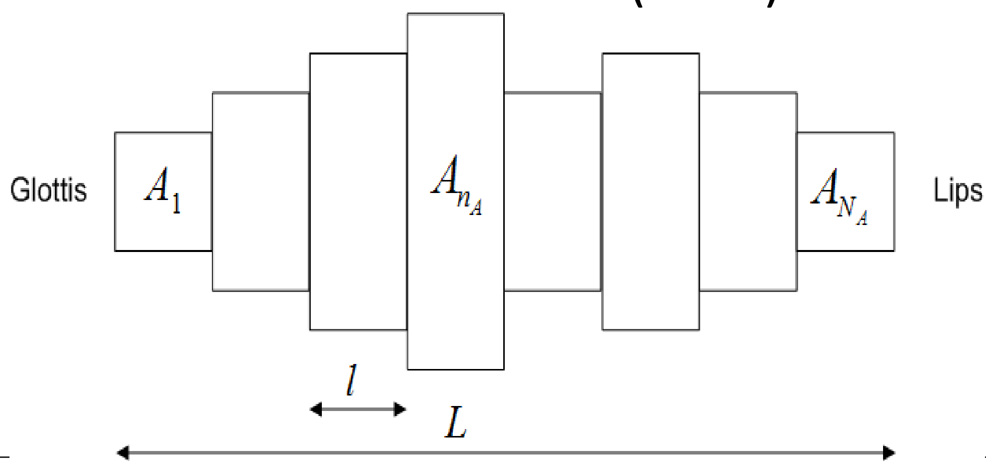
Vocal Tract Modeling

- The proposed algorithm tries to estimate the speaker physical vocal tract shape.
- Atal and Wakita showed the equivalence of acoustic tube model and the linear prediction (LP) model under certain conditions [Atal, 1970; Wakita, 1973].
- The M^{th} order filter transfer function derived through LP is equivalent to the transfer function of an acoustical tube made up of M equal length sections of variable areas.
- This is referred to as Vocal Tract Area Function (VTAF).

$$M = f_s \frac{2L}{c}$$

c - sound velocity

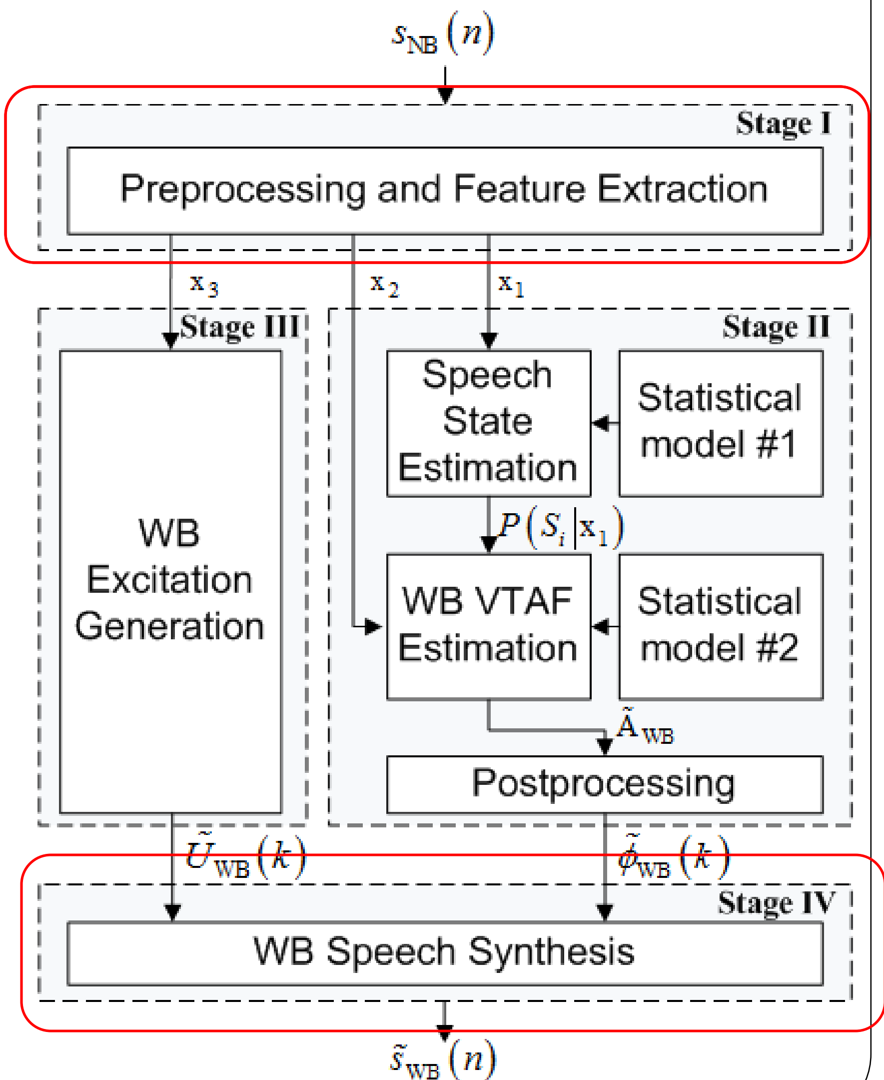
$$A_{n_A} = \frac{1+r_{n_A}}{1-r_{n_A}} A_{n_A+1}$$



Proposed BWE Algorithm (2/26)

Algorithm stages:

I. NB signal preprocessing and features extraction
II. HB spectral envelope estimation and postprocessing
III. WB excitation generation
IV. Wideband signal synthesis



Proposed BWE Algorithm (2/26)

Algorithm stages:

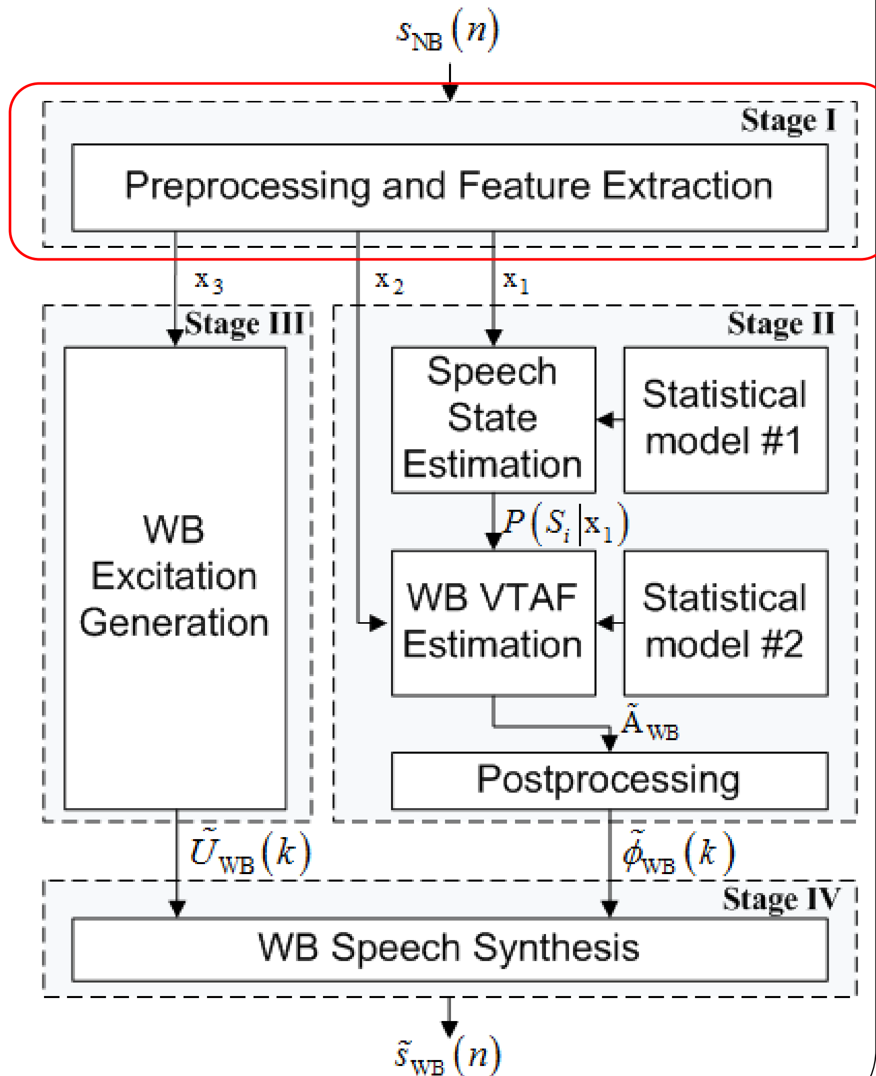
I. **NB signal preprocessing and features extraction**

II. HB spectral envelope estimation and postprocessing

III. WB excitation generation

IV. Wideband signal synthesis

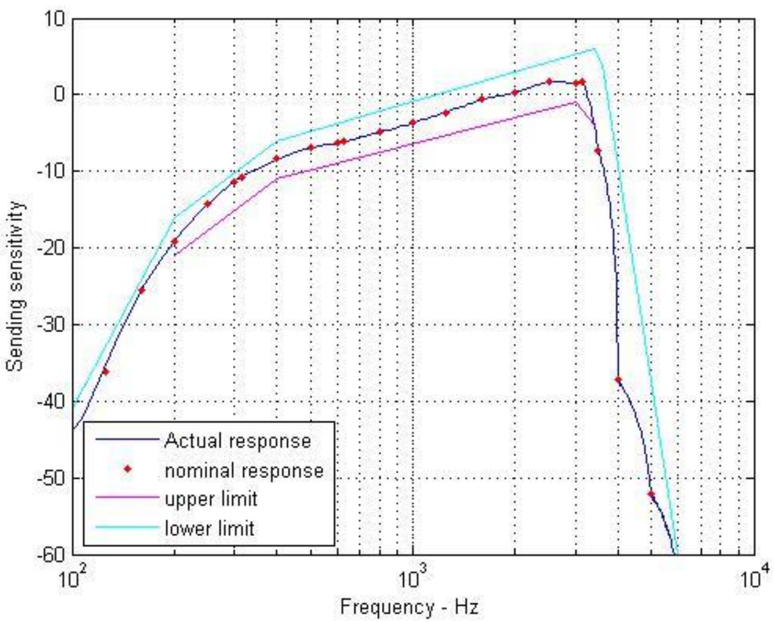
- x_1 - Features for speech state estimation (MFCC, spectral centroid, spectral flatness, spectral slope and normalized energy)
- x_2 – NB VTAF for WB VTAF estimation
- x_3 – NB excitation for WB excitation generation



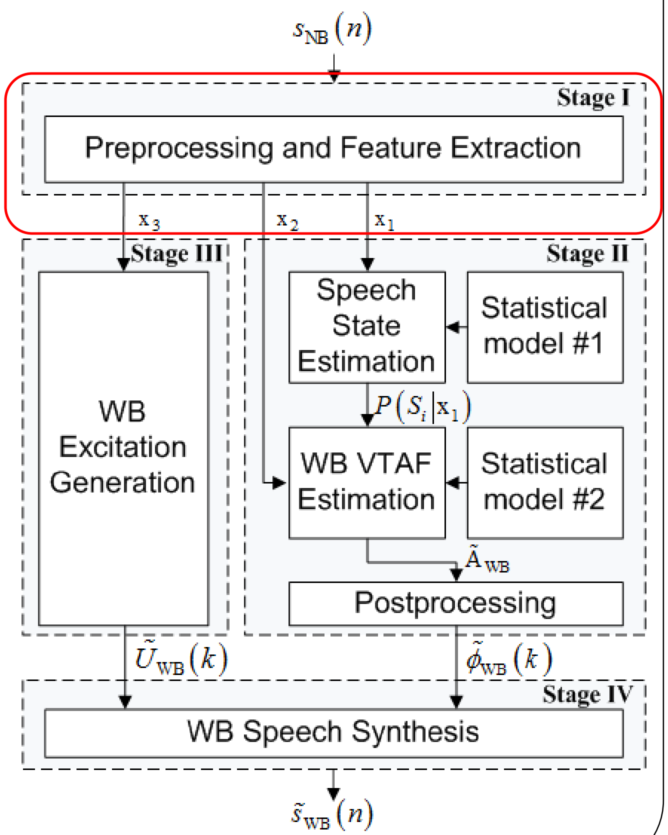
Preprocessing and Features Extraction (3/26)

Incentives

- Compensate for the IRS filter response.
- Extract features that allow good classification and estimation.



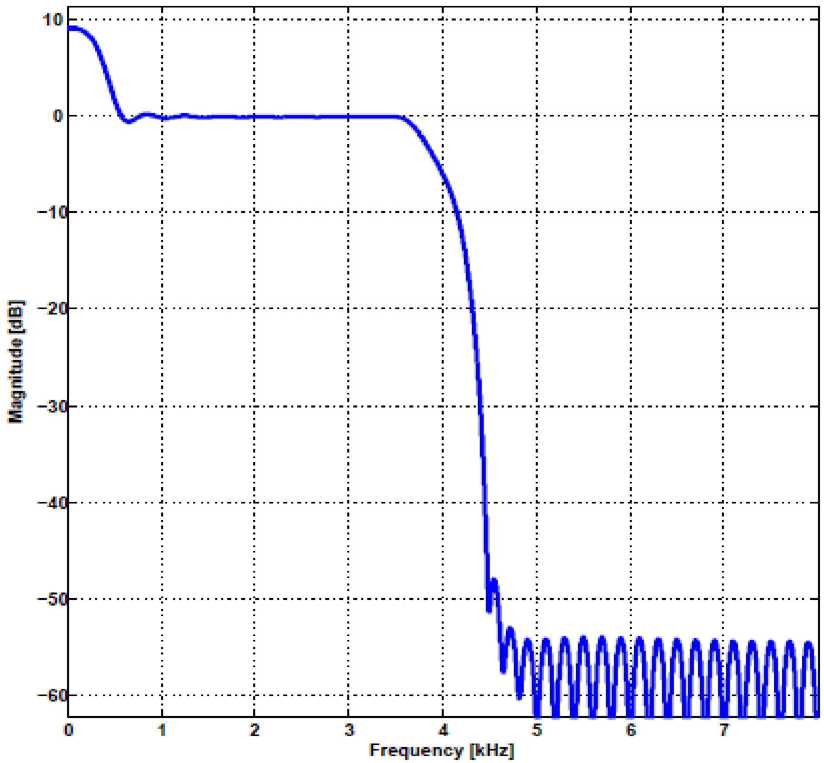
IRS send filter



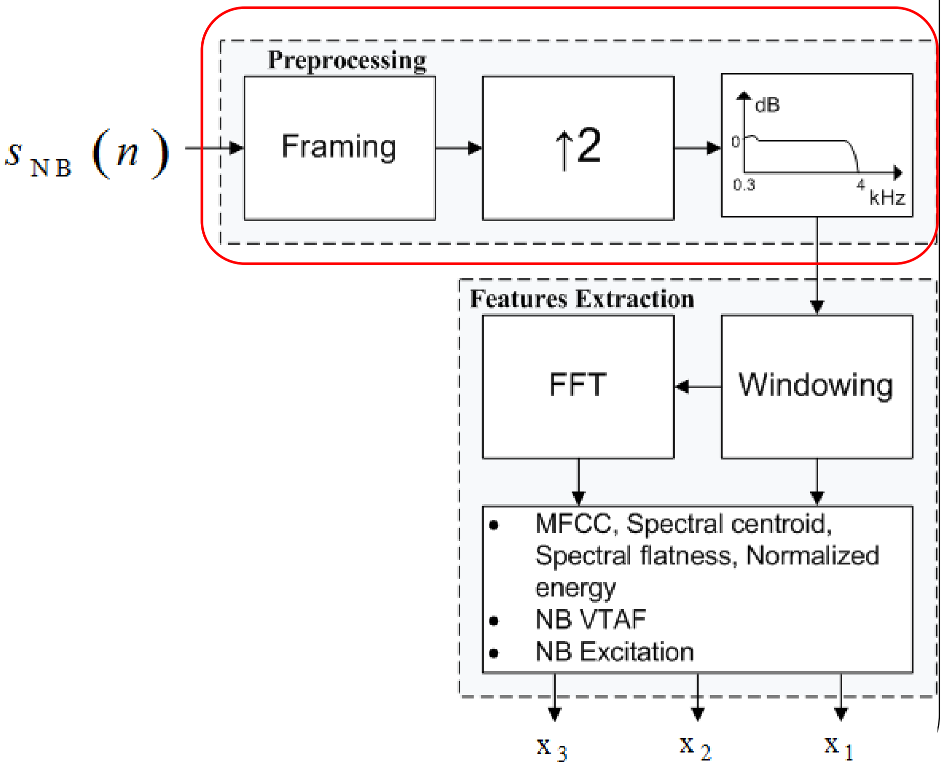
Preprocessing and Features Extraction (4/26)

Preprocessing

- NB signal interpolation.
- Equalization – 10dB boost to compensate for IRS filter at 300 Hz:



Equalizer filter



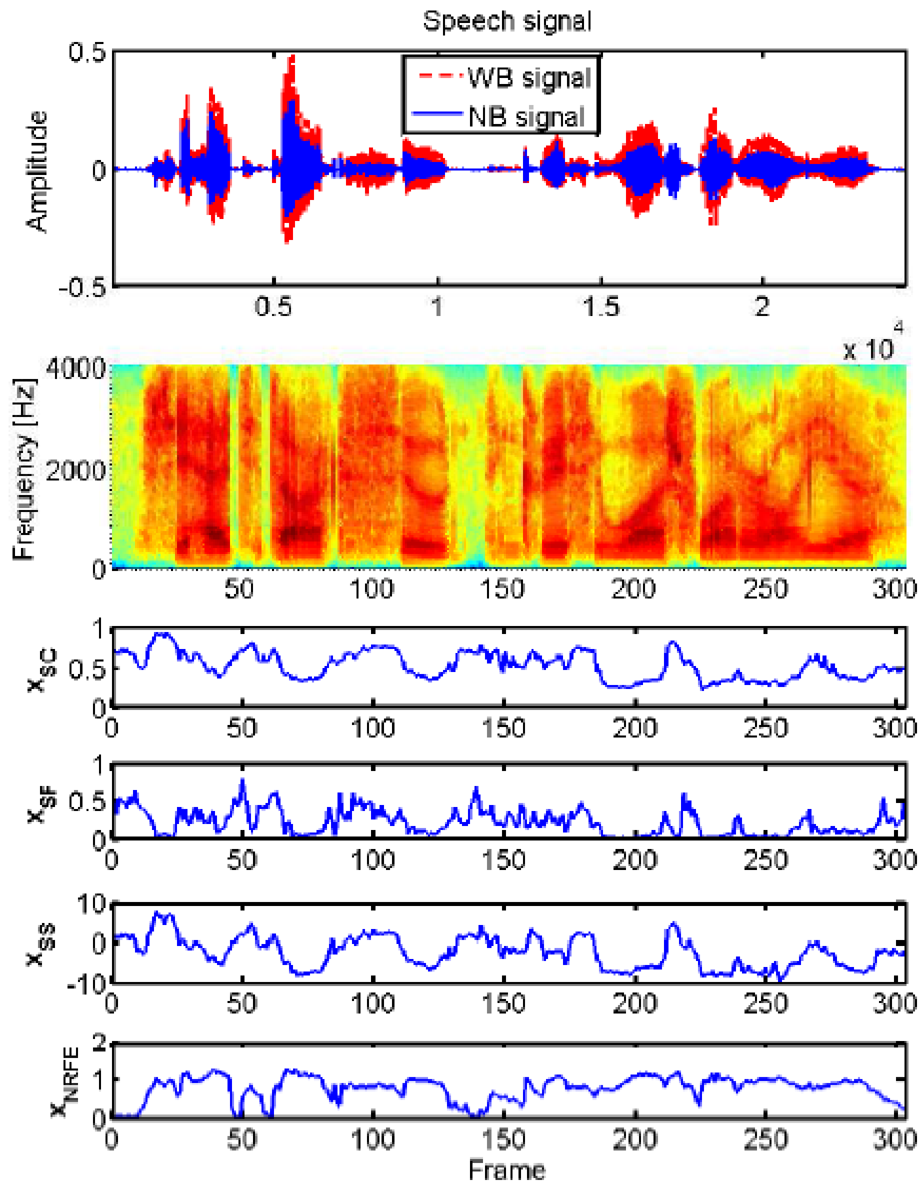
Preprocessing and Features Extraction (5/26)

- Feature selection for BWE algorithm depends on the following considerations:
  - Computational complexity
  - Mutual information, $I(\mathbf{x};\mathbf{y})$, between narrow- and high-band parameters.
  - Separability, $\zeta(\mathbf{x})$, the quality of particular feature set **x**, for a classification problem. A higher separability value indicates a better suitability for classification.

Preprocessing and Features Extraction (6/26)

Feature vector $\mathbf{x}$	dim $\mathbf{x}$	Towards high frequencies		Towards low frequencies	
		$I(\mathbf{x};\mathbf{y})$ [bit/frame]	$\zeta(\mathbf{x})$ (16 classes)	$I(\mathbf{x};\mathbf{y})$ [bit/frame]	$\zeta(\mathbf{x})$ (16 classes)
ACF	10	2.6089	1.6349	2.7530	2.3977
LPC	10	2.3054	1.5295	2.1100	1.7901
LSF	10	2.3597	1.5596	2.2125	2.5817
LPC-cepstrum	10	2.2401	1.4282	2.1778	2.3879
Cepstrum	10	2.3075	1.5483	1.9398	2.5473
MFCC	10	2.3325	2.2659	3.0771	6.6142
ACF (1)	1	0.7514	1.1237	0.7324	1.1065
ACF (pitch period)	1	0.4450	0.4058	0.5441	0.6745
Frame energy	1	0.9285	1.0756	1.3968	4.2328
Gradient index	1	0.8011	1.2520	0.5403	0.6983
Zero-crossing rate	1	0.7453	1.0795	0.7456	1.1685
Pitch period	1	0.2451	0.0530	0.4823	0.1122
Local kurtosis	1	0.2037	0.0225	0.2979	0.0809
Spectral centroid	1	0.7913	1.0179	0.6630	0.9276
Spectral flatness	1	0.4387	0.3538	0.4201	0.4648

Preprocessing and Features Extraction (7/26)



Speech signal

Spectrogram

Spectral centroid

Spectral flatness

Spectral slope

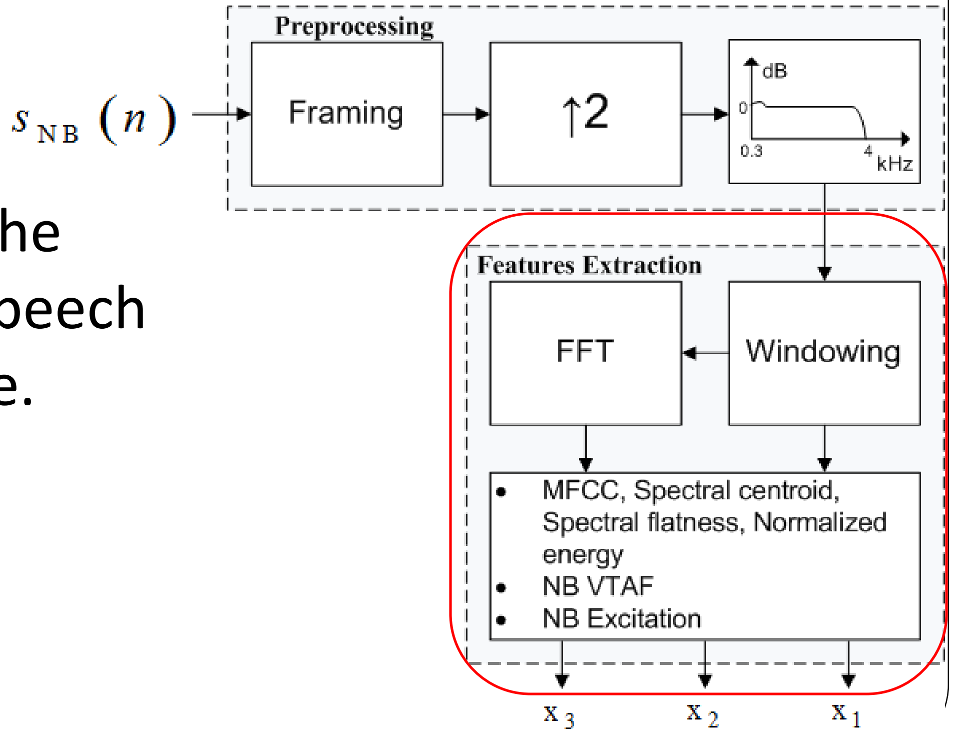
Normalized relative frame energy

Preprocessing and Features Extraction (8/26)

Feature extraction

- x_1 – Frequency based features for phoneme estimation (MFCC, spectral centroid, spectral flatness, spectral slope and normalized energy.
- x_2 – NB VTAF calculated from the reflection coefficients

$$A_{n_A} = \frac{1+r_{n_A}}{1-r_{n_A}} A_{n_A+1}$$

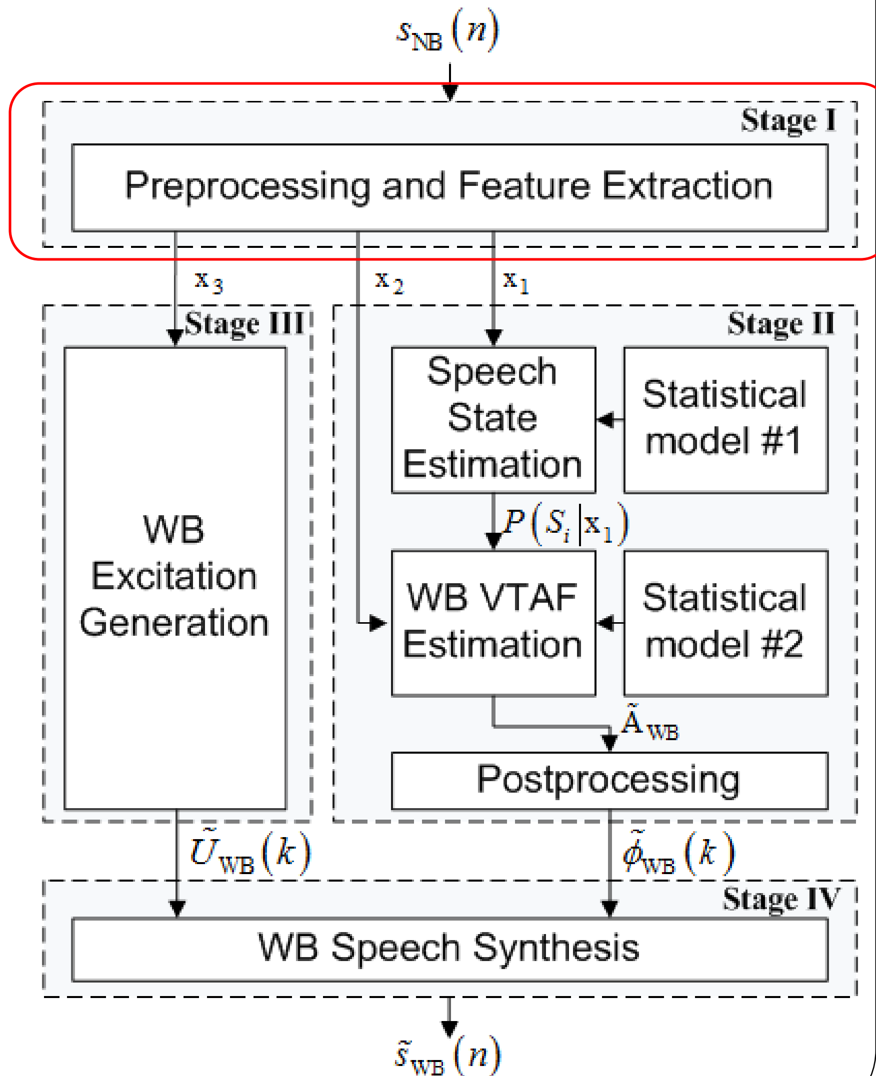


- x_3 – NB excitation calculated in the frequency domain by dividing speech signal with the spectral envelope.

Proposed BWE Algorithm (9/26)

Algorithm stages:

I. **NB signal preprocessing and features extraction**
II. HB spectral envelope estimation and postprocessing
III. WB excitation generation
IV. Wideband signal synthesis



Proposed BWE Algorithm (9/26)

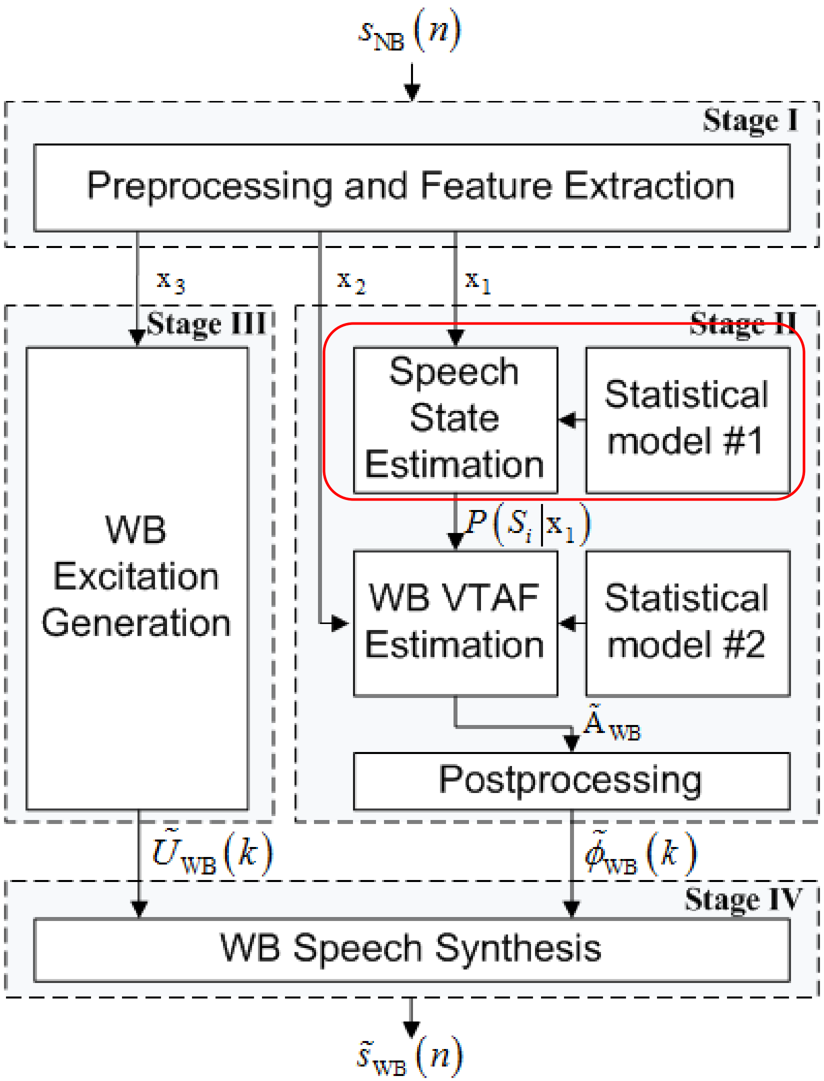
Algorithm stages:

I. NB signal preprocessing and features extraction

II. HB spectral envelope estimation and postprocessing

III. WB excitation generation

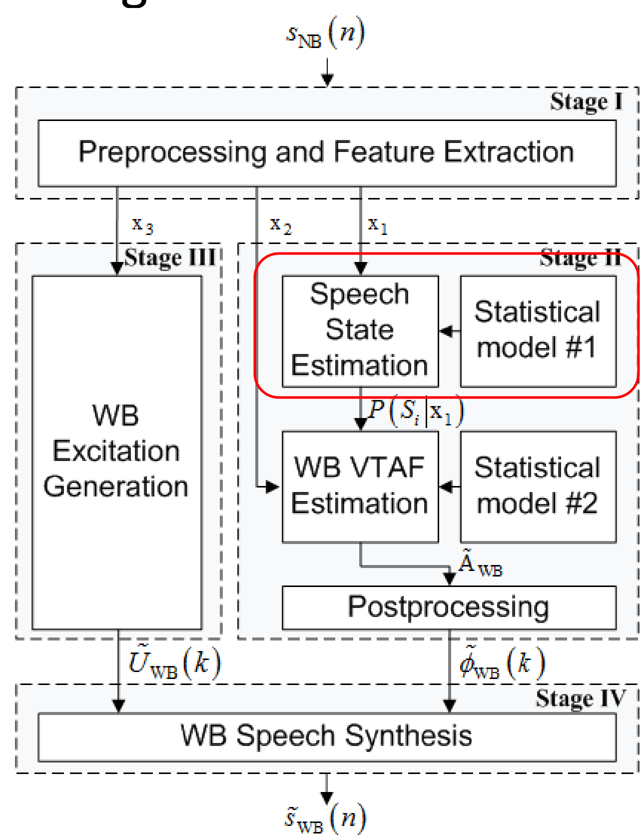
IV. Wideband signal synthesis



WB Spectral Envelope Estimation (10/26)

Phoneme estimation: using HMM to estimate each speech frame linguistic state.

- <u>Off-line process:</u> using TIMIT transcription to build HMM statistical model using phoneme based states. Calculating the following PDFs:
  - $p(S_i)$ - Initial probability of each state.
  - $p(S_i(m)|S_j(m-1))$ - Transition probability of the Markov chain from state *j* to state *I.*
  - $p(\mathrm{x}_1|S_i)$ - Observation probability for each state. Approximated by GMM parameters using the EM algorithm.



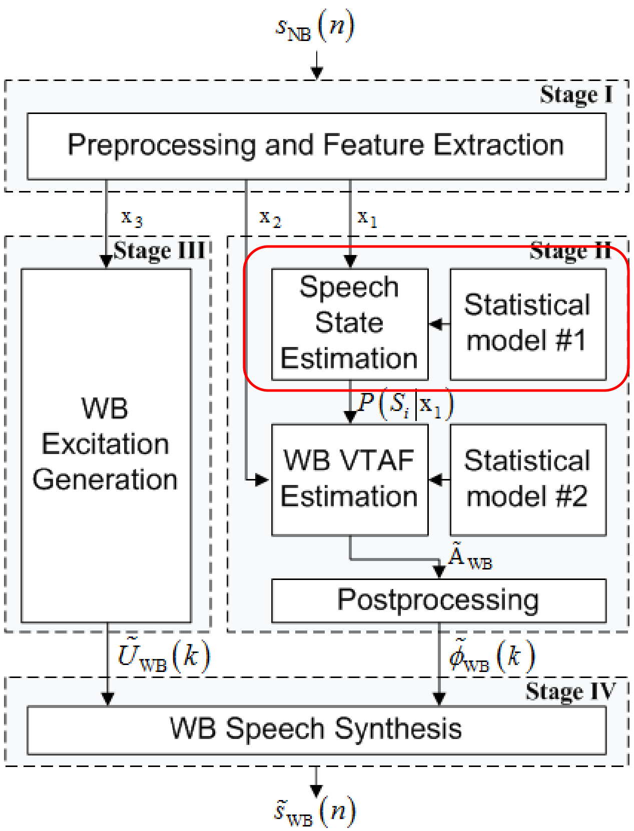
WB Spectral Envelope Estimation (11/26)

Phoneme estimation: using HMM to estimate each speech frame linguistic state.

- <u>On-line process:</u> making a decision on current frame state (phoneme) by maximizing the a-posteriori PDF:

$$p\left(S_i(m)\middle|\mathrm{X}_1(m)\right)=p\left(\mathrm{x}_1(m)\middle|S_i(m)\right)\cdot$$

$$\sum_{j=1}^{N_s} p\left(S_i(m)\middle|S_j(m-1)\right)p\left(S_j(m-1)\middle|\mathrm{X}_1(m-1)\right)$$



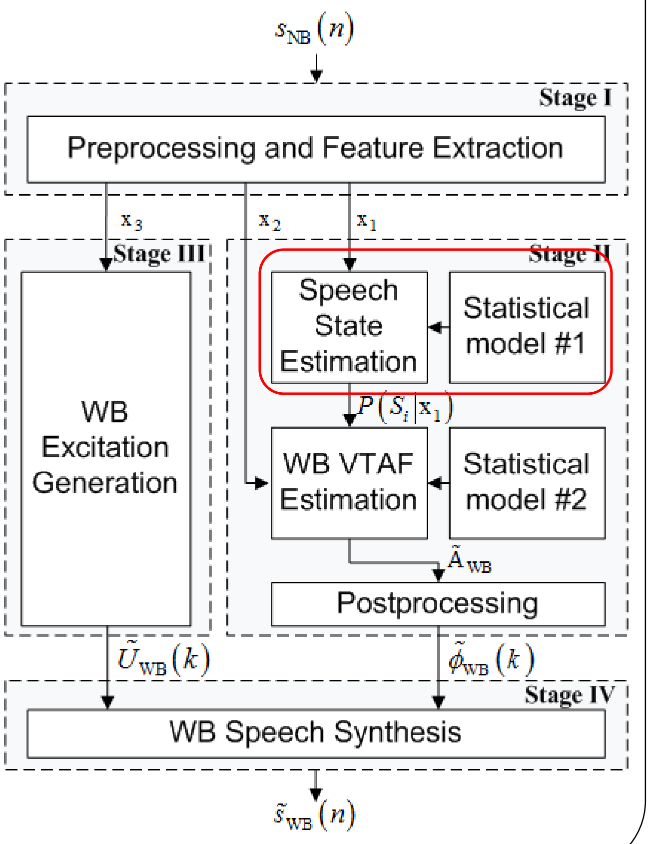
WB Spectral Envelope Estimation (12/26)

Estimate WB VTAF: using codebook matching of calculated NB VTAF to WB VTAF.

- <u>Off-line process:</u> for each speech state, clustering of N_{CB} WB VTAF using vector quantization of speech frames training set.
- <u>On-line process:</u> finding closest WB VTAF codeword to extracted NB VTAF using Euclidean distance.

$$\tilde{\mathrm{A}}_{\mathrm{WB}}^{S_i} = \mathrm{A}_{\mathrm{WB}}^{S_i}\left(j^{opt}\right)$$

$$j^{opt} = \arg\min_{j=1}^{N_{\mathrm{CB}}} \left\| \log\left(\mathrm{A}_{\mathrm{NB}}\right) - \log\left(\mathrm{A}_{\mathrm{WB}}^{S_i}\left(j\right)\right) \right\|_2^2$$

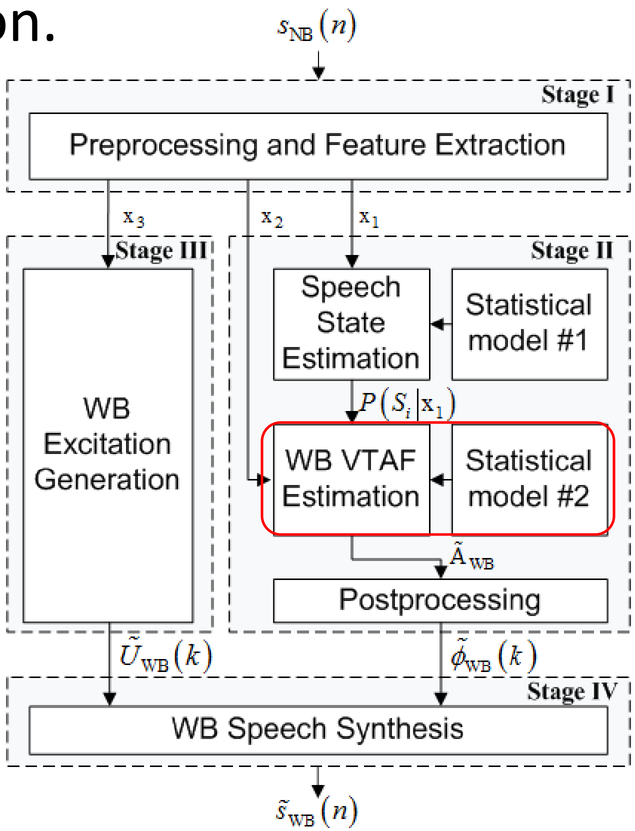


WB Spectral Envelope Estimation (13/26)

Postprocessing

- Reduce artifacts due to erroneous estimation in first two steps.
- Reduce artifacts due to erroneous state estimation by using N_{best} highest probability states for VTAF estimation.

$$\tilde{\mathrm{A}}_{\mathrm{WB}} = C \cdot \left(p_1 \cdot \tilde{\mathrm{A}}_{\mathrm{WB}}^{S_{i_1}} + \ldots + p_{N_{\text{best}}} \cdot \tilde{\mathrm{A}}_{\mathrm{WB}}^{S_{i_{N_{\text{best}}}}} \right)$$

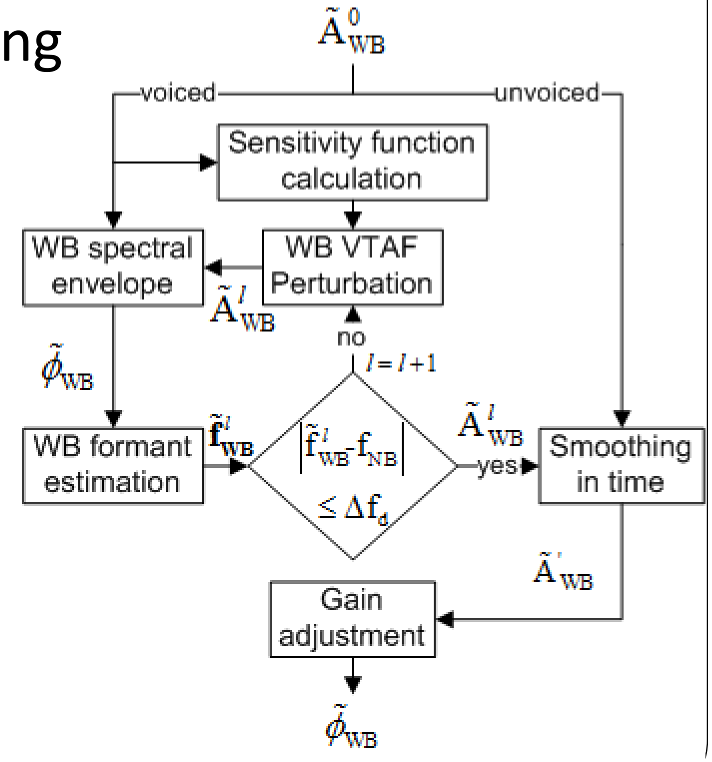


WB Spectral Envelope Estimation (14/26)

Postprocessing

- Reduce artifacts due to erroneous WB VTAF estimation.
- Estimated WB envelope fit to NB envelope by tuning formant frequencies of estimated WB VTAF to allow better gain adjustment to NB envelope. Iterative tuning by VTAF perturbation.
- Iterative VTAF perturbation based on the sensitivity function:

$$\frac{\Delta f_{n_f}}{f_{n_f}} = \sum_{n_A}^{N_A} S_{n_f, n_A} \frac{\Delta A_{n_A}}{A_{n_A}}$$



WB Spectral Envelope Estimation (15/26)

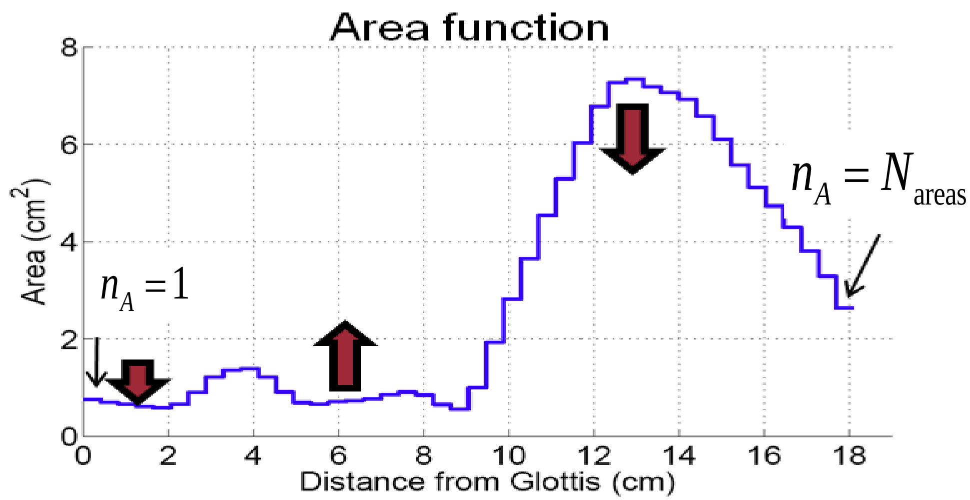
Vocal Tract Sensitivity Function

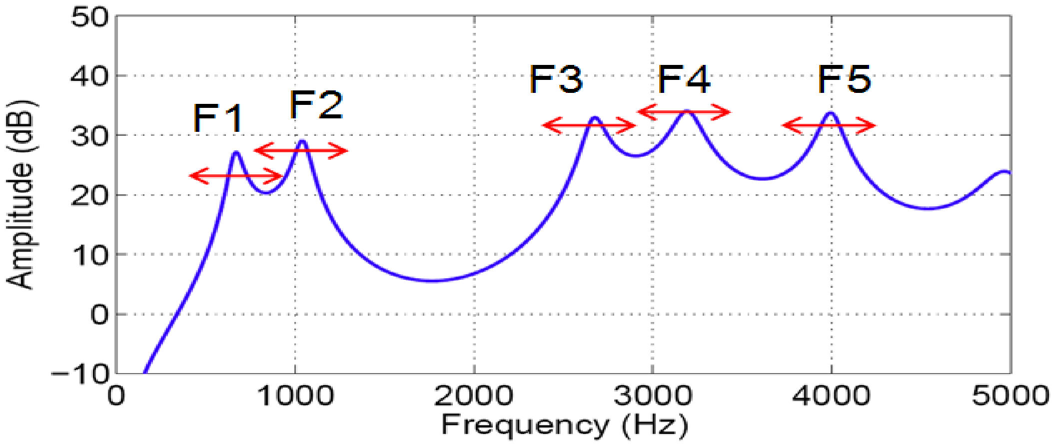
Relate changes in area to changes in formant frequencies

$$\frac{\Delta f_{n_f}}{f_{n_f}} \Leftrightarrow \frac{\Delta A_{n_A}}{A_{n_A}}$$

Sensitivity function

$$\frac{\Delta f_{n_f}}{f_{n_f}} = \sum_{n_A}^{N_A} S_{n_f,n_A} \frac{\Delta A_{n_A}}{A_{n_A}}$$



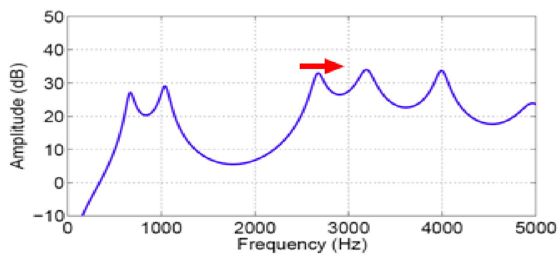


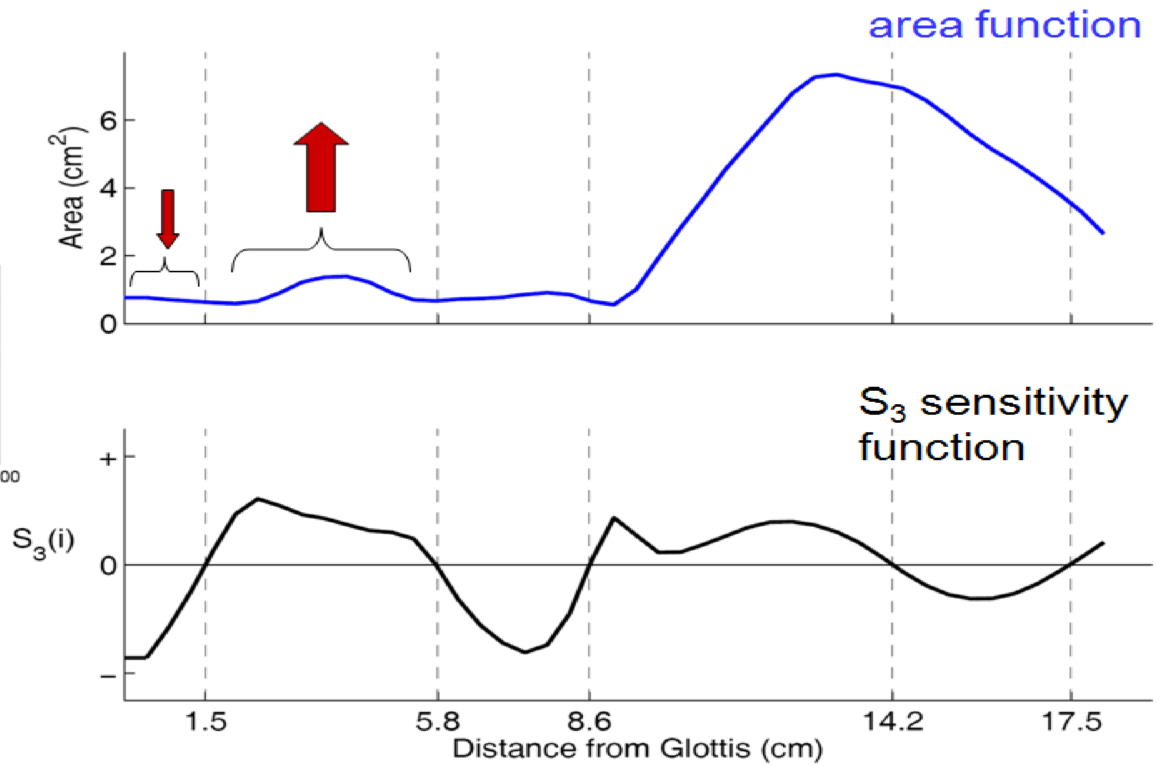
WB Spectral Envelope Estimation (16/26)

Vocal Tract Sensitivity Function

$$\frac{\Delta f_3}{f_3} = \sum_{n_A}^{N_A} S_{3,n_A} \frac{\Delta A_{n_A}}{A_{n_A}}$$

$\Delta a(i)$	$S_3(i)$	ΔF_3
+	+	+
+	−	−
−	+	−
−	−	+





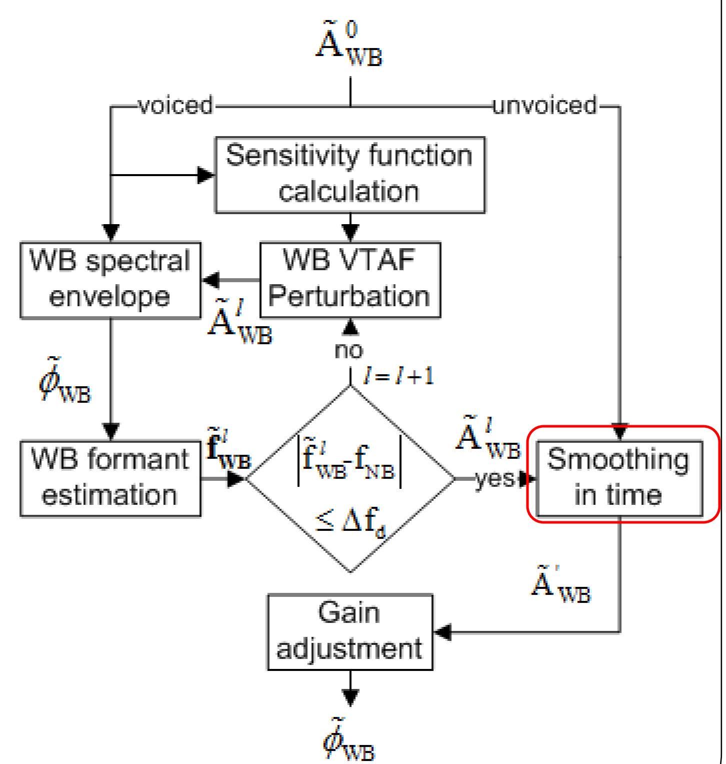
WB Spectral Envelope Estimation (17/26)

Postprocessing

- Stopping condition for iterative process is formant frequencies difference.
- Smoothing in time of tuned estimated VTAF.

$$\tilde{A}'_{WB}(m) = \beta \cdot \tilde{A}'_{WB}(m-1) + (1-\beta) \cdot \tilde{A}_{WB}(m)$$

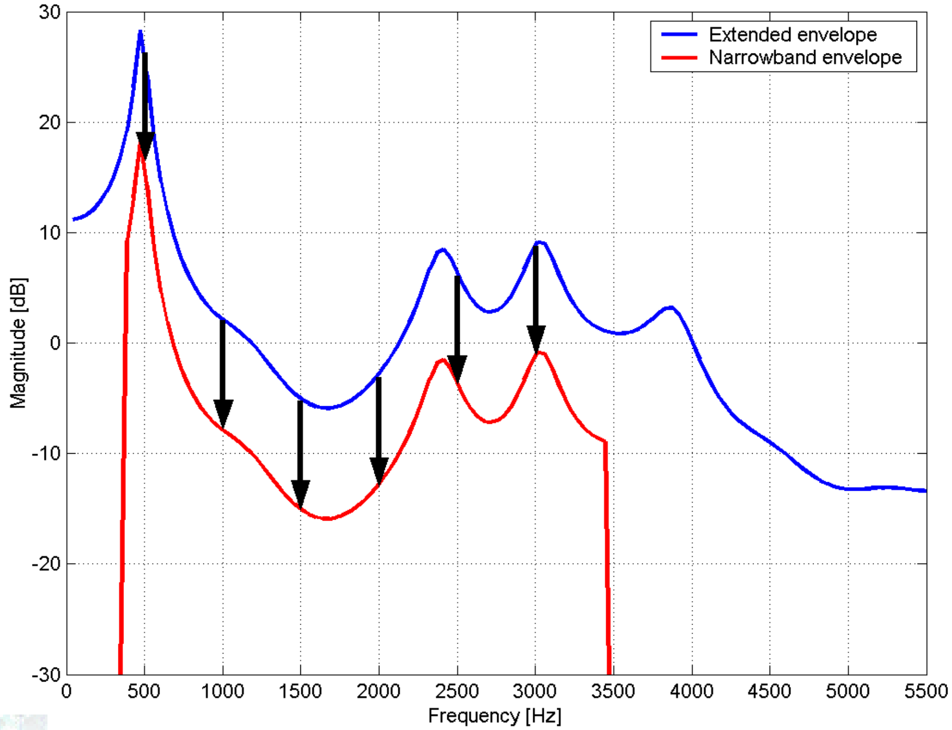
- Converting WB VTAF to WB spectral envelope.

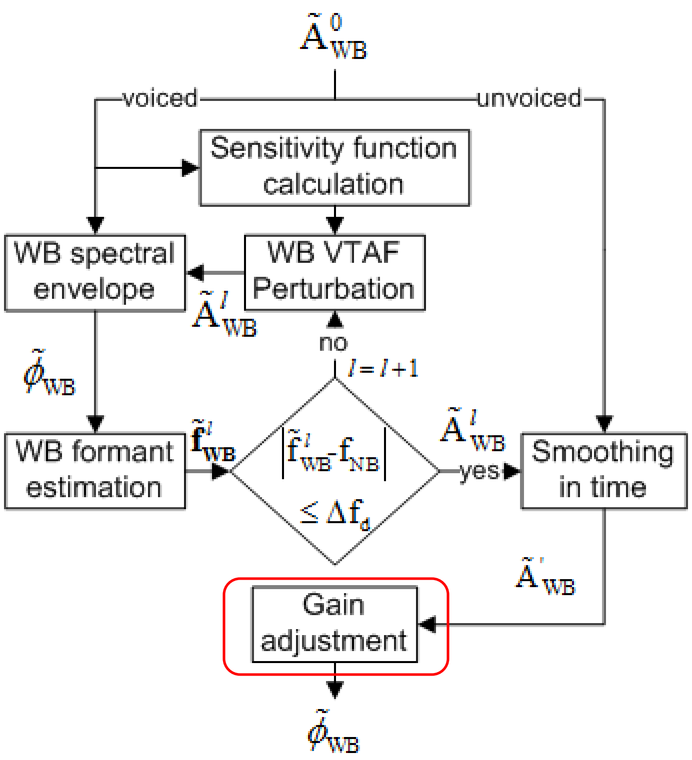


WB Spectral Envelope Estimation (18/26)

Postprocessing

- Gain adjustment of estimated WB spectral envelope to match the input NB spectral envelope.

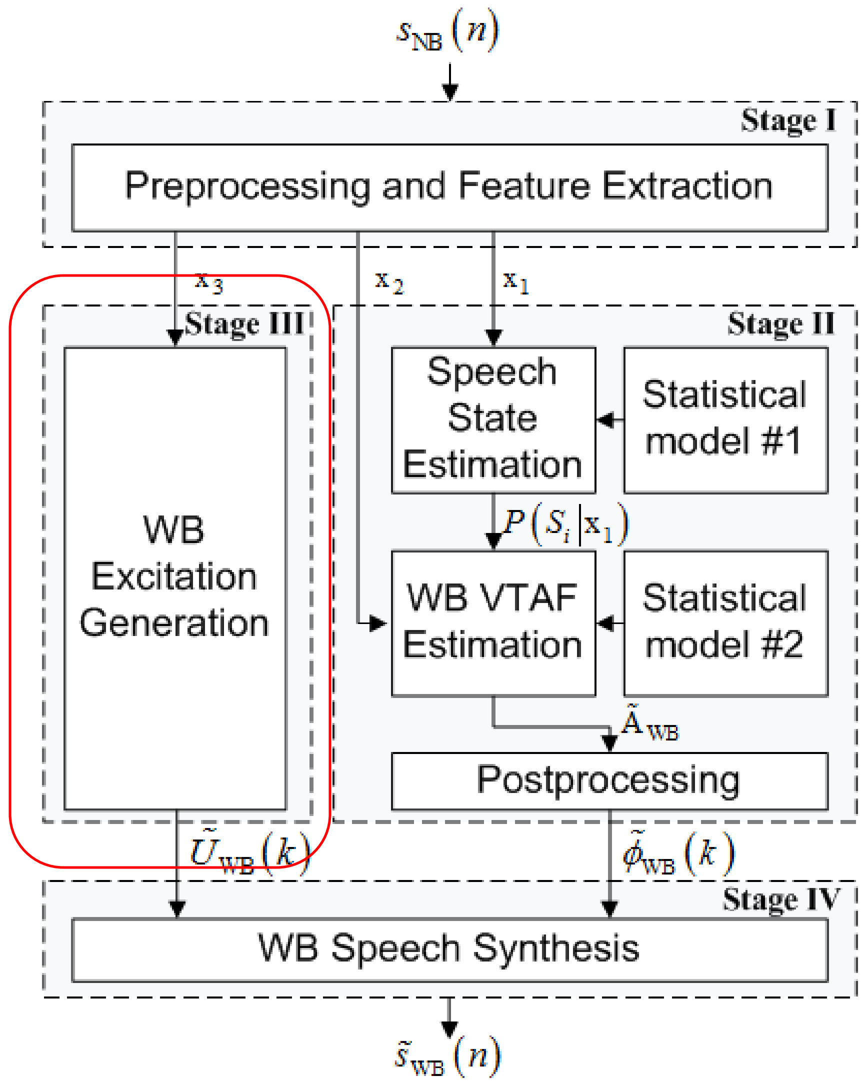




Proposed BWE Algorithm (19/26)

Algorithm stages:

I. NB signal preprocessing and features extraction
II. HB spectral envelope estimation and postprocessing
III. WB excitation generation
IV. Wideband signal synthesis



WB Excitation Generation (20/26)

Excitation generation - Shifting / modulation approaches:

- Fixed spectral translation:

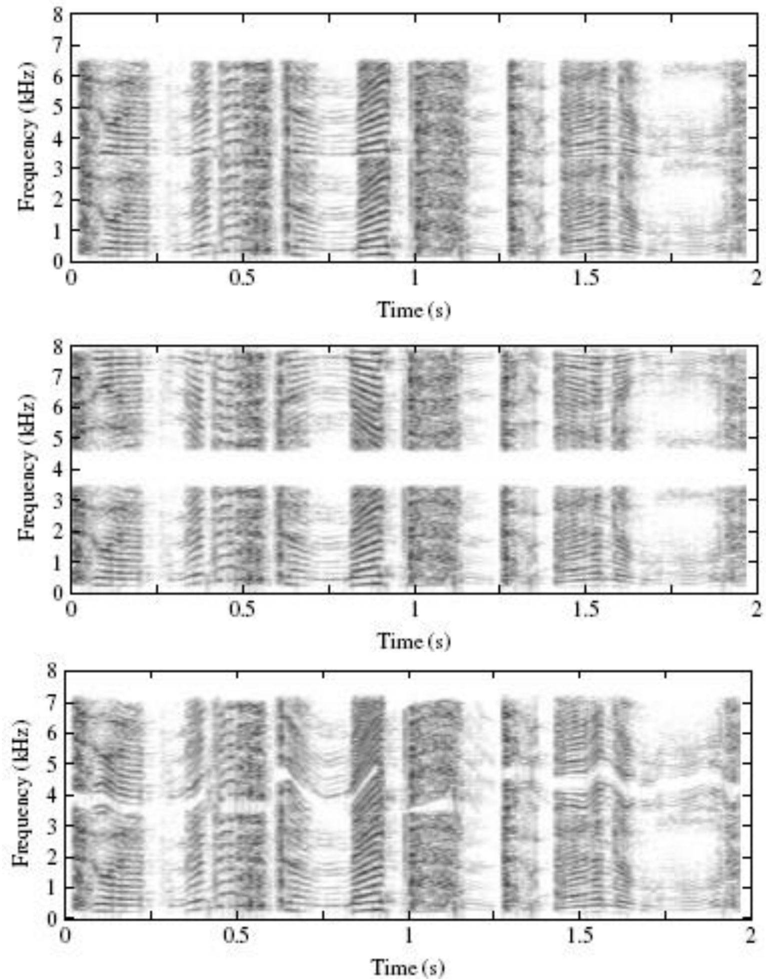
$$\Omega_m = 2\pi \frac{3.4kHz}{f_s}$$

- Spectral folding:

$$\Omega_m = 2\pi \frac{8kHz}{f_s}$$

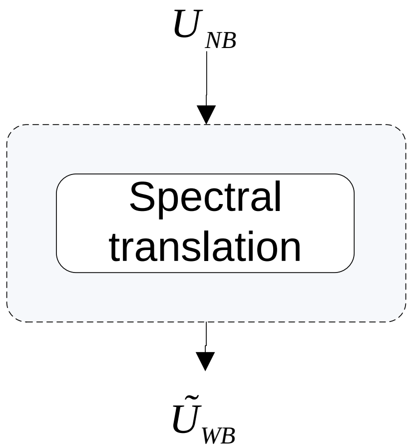
- Pitch adaptive modulation:

$$\Omega_m = 2\pi \frac{(nF_0)kHz}{f_s}, n = \max_n nF_0 \leq 3.4kHz$$



WB Excitation Generation (21/26)

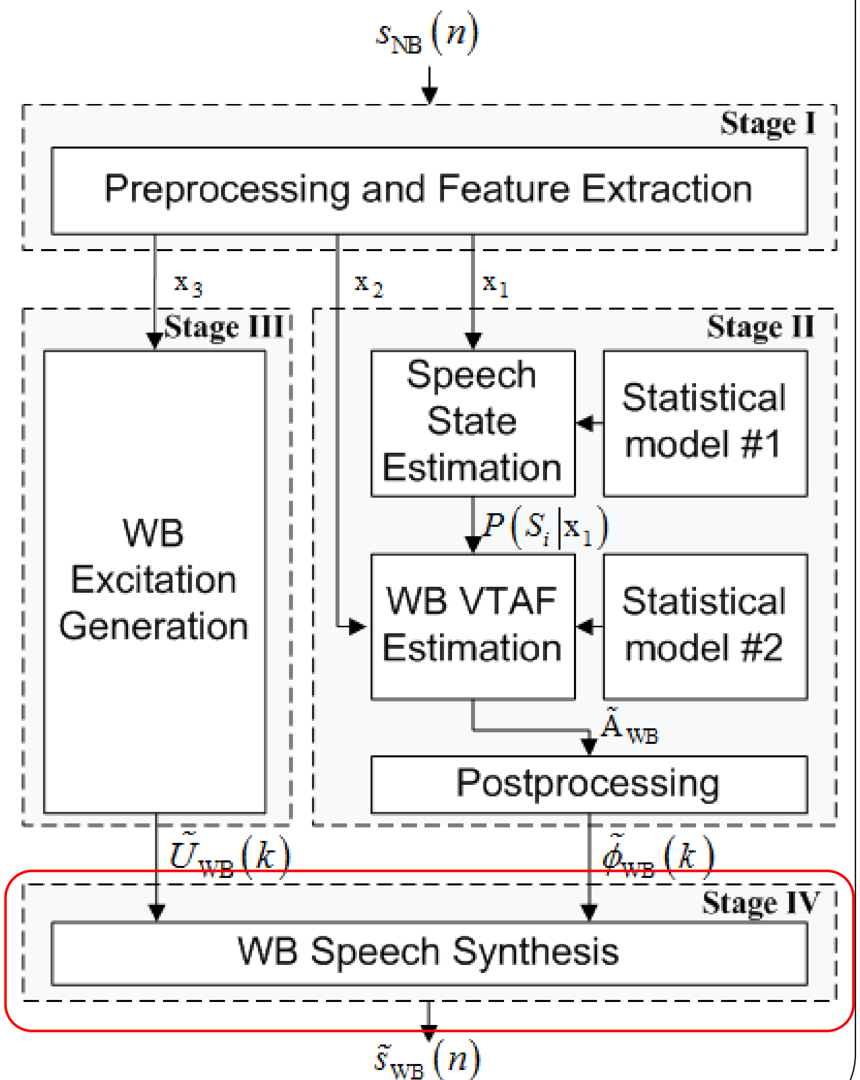
- Extract narrowband (NB) excitation using analyzed LPC.
- Using fixed spectral translation method.
- Advantages
  - Fill all high-band frequencies.
  - Does not change narrowband excitation.
  - Low computational complexity.
- Disadvantages
  - Does not keep harmonic structure of voiced excitation.



Proposed BWE Algorithm (22/26)

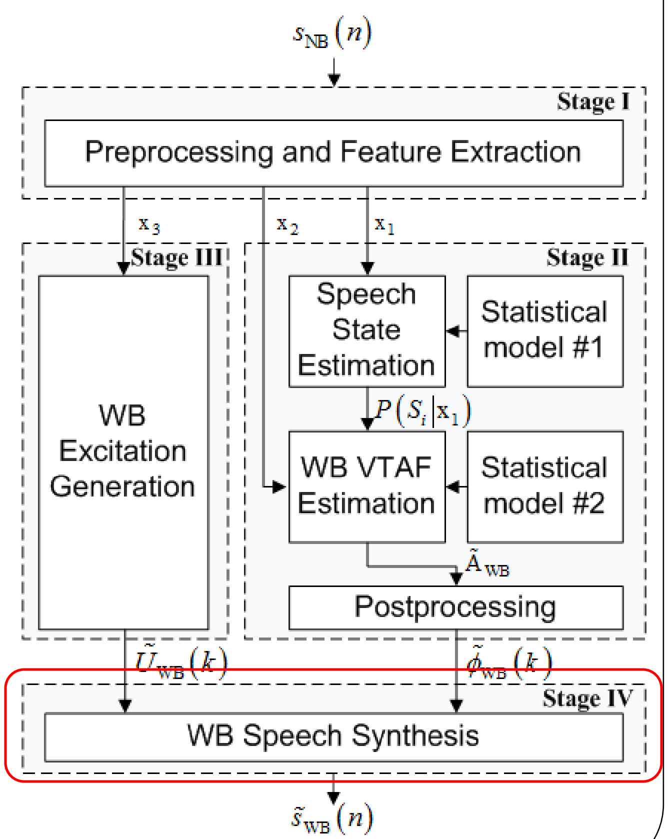
Algorithm stages:

I. NB signal preprocessing and features extraction
II. HB spectral envelope estimation, and post processing
III. WB excitation generation
IV. **Wideband signal synthesis**



Wideband Speech Synthesis (23/26)

- Make sure not to change received narrow-band signal.
- Combine the estimated high-band signal and the narrow band signal.
  - Summation in time domain.
  - Concatenation in frequency domain.



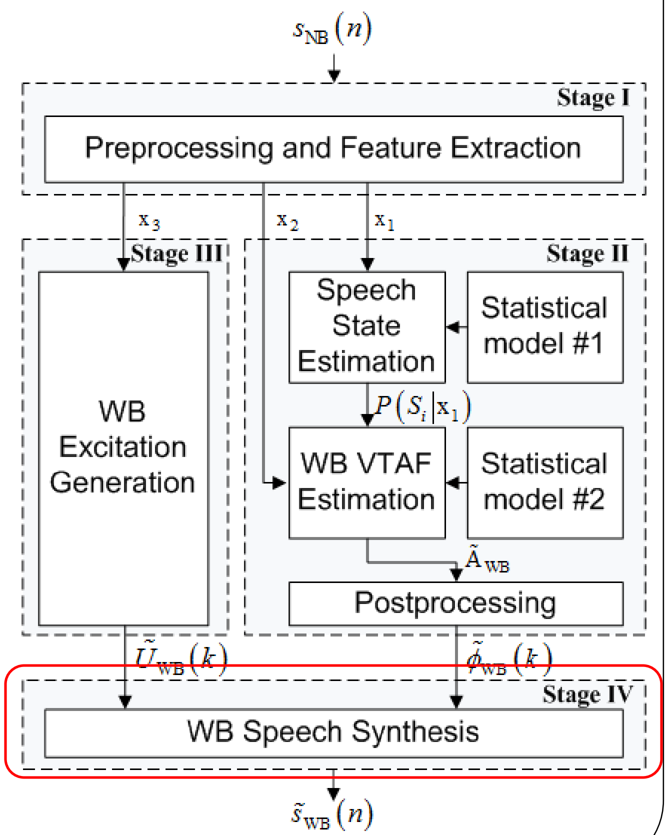
Wideband Speech Synthesis (24/26)

- Frequency domain synthesis.
- Calculate high-band estimated signal using:

$$\tilde{S}_{hb}(k) = \tilde{U}_{hb}(k) \cdot \tilde{\phi}_{hb}(k)$$

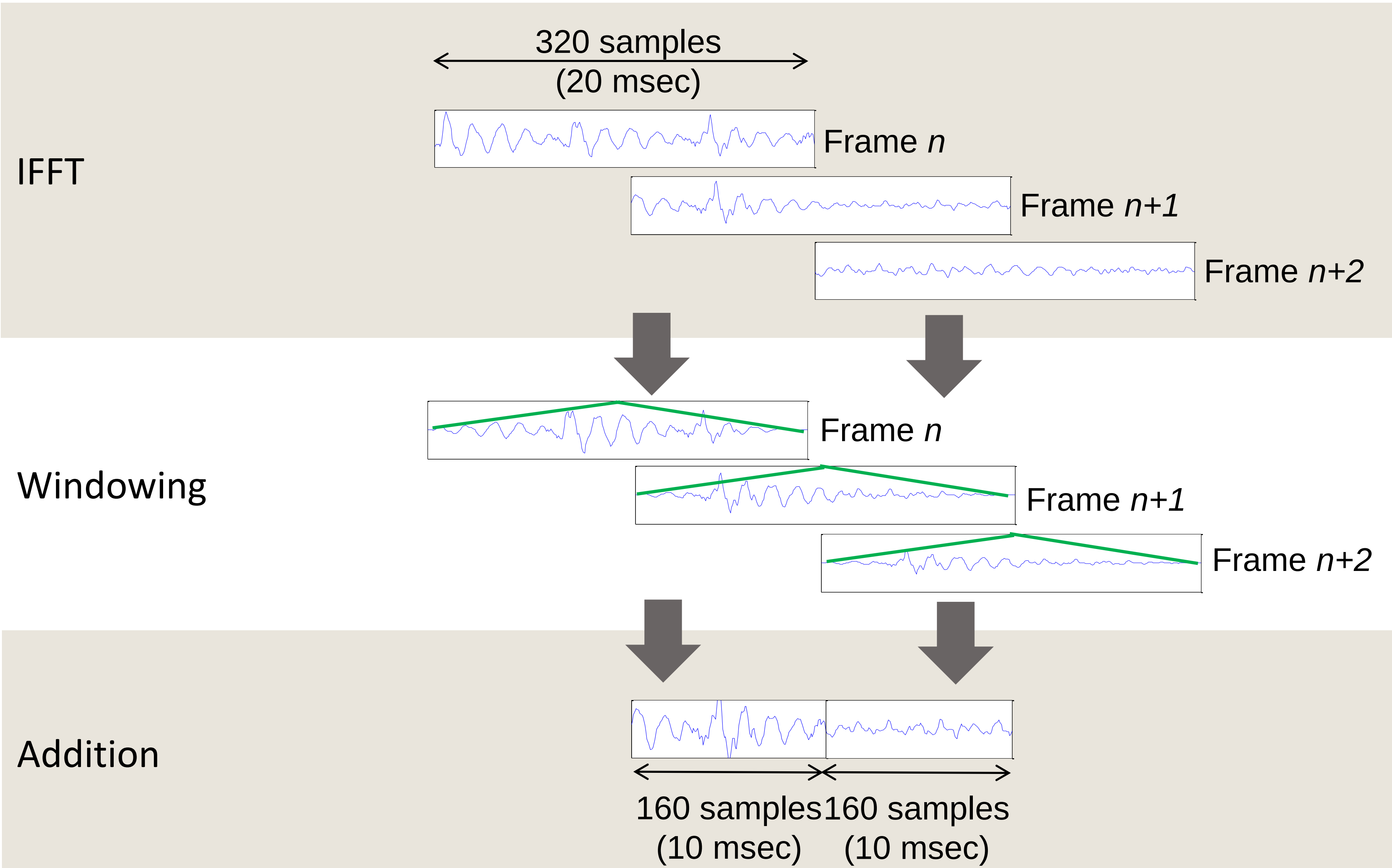
- Concatenating narrowband and high band signal in frequency.

$$\tilde{S}_{wb}(k) = \begin{cases} S_{nb}(k), 0 < k < 3.5/4kHz \\ \tilde{S}_{hb}(k), 3.5/4 < k < 8kHz \end{cases}$$



Wideband Speech Synthesis (25/26)

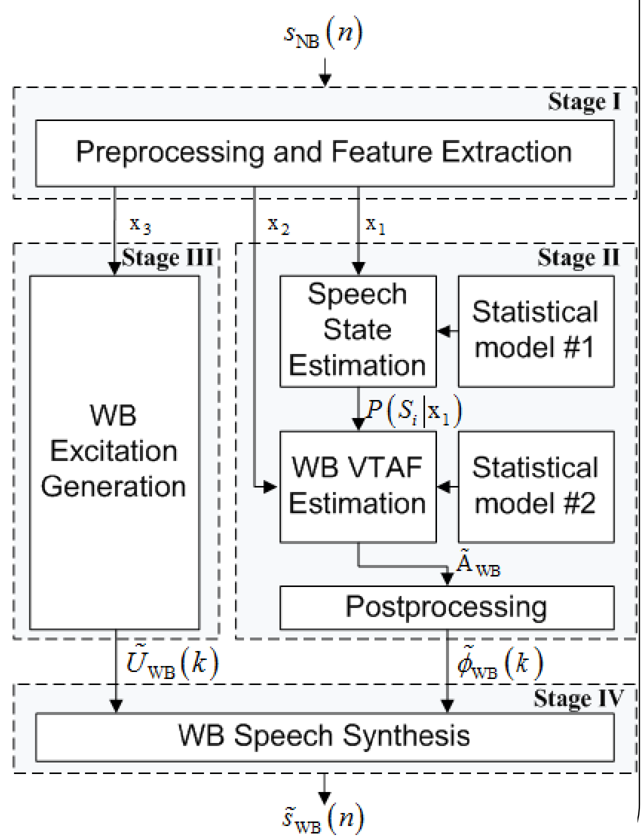
Overlap add - Synthesize 20msec frames, 10msc overlap



Proposed BWE Algorithm (26/26)

Summary

- Separate extension of excitation and spectral envelope.
- signal equalization of NB attenuated band – 50-300 Hz.
- Spectral envelope estimation
  - **phoneme estimation** and **speaker vocal tract area function (VTAF) estimation.**
  - **Iterative tuning of estimated VTAF** to reduce possible artifacts.
- HB Excitation estimation by spectral shifting.
- Overlap-add synthesis for better time transition.



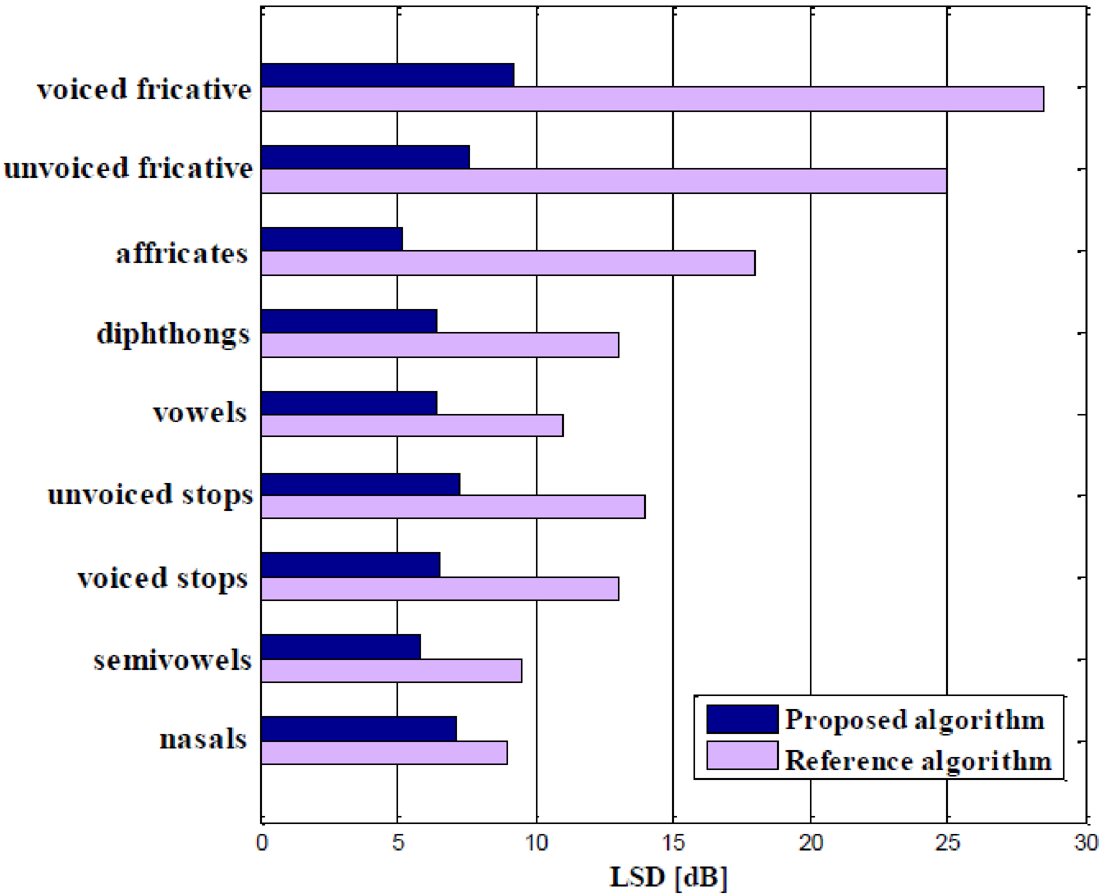
Outline

- Introduction
- Methods of BWE
- Proposed BWE Algorithm
- **Performance Evaluation**
- Conclusion

BWE Performance Evaluation (1/8)

Objective quality evaluation - Log Spectral Distance for phone category

- Reference algorithm: *Evaluation of an Artificial Speech Bandwidth Extension Method in Three Languages* [Pulakka et al., 2008].



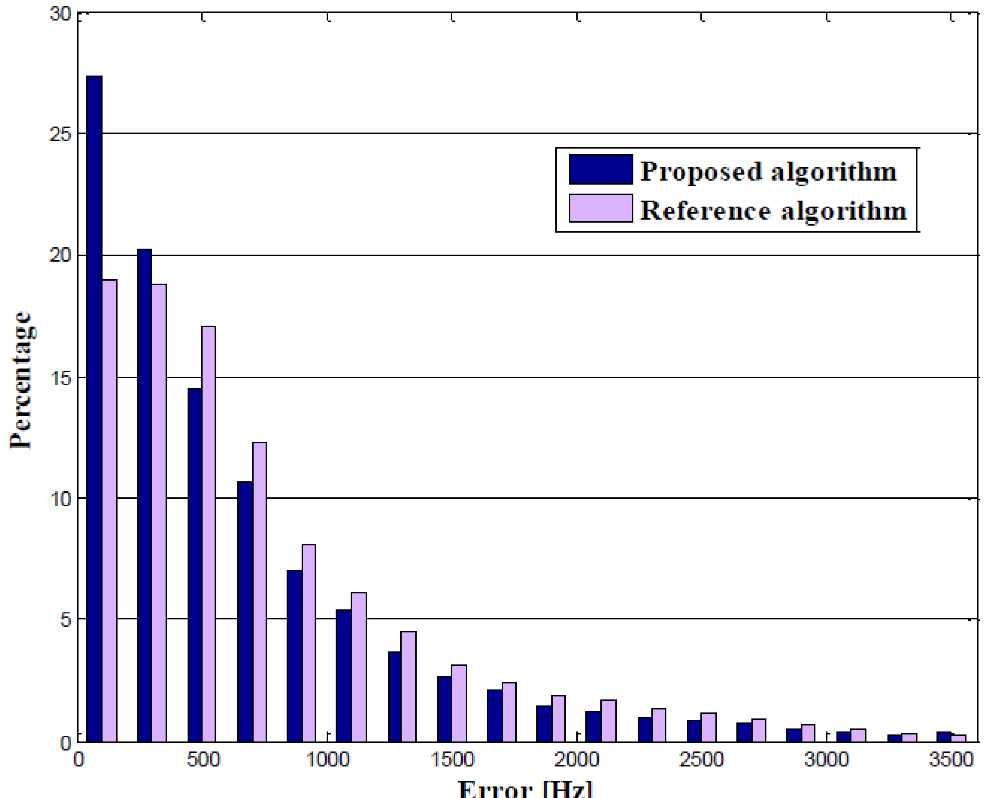
$$\mathrm{LSD} = \sqrt{\frac{1}{k_{\mathrm{high}} - k_{\mathrm{low}} + 1} \sum_{k=k_{\mathrm{low}}}^{k_{\mathrm{high}}} \left[10 \log_{10} \frac{P(k)}{\tilde{P}(k)} \right]^2}$$

BWE Performance Evaluation (2/8)

Objective quality evaluation - Histogram of estimated formant frequencies error

- Reference algorithm: *Low-Complexity Feature-Mapped Speech Bandwidth Extension* [Gustafsson et al., 2006].

	Proposed	Reference
Mean Error	620 Hz	740 Hz

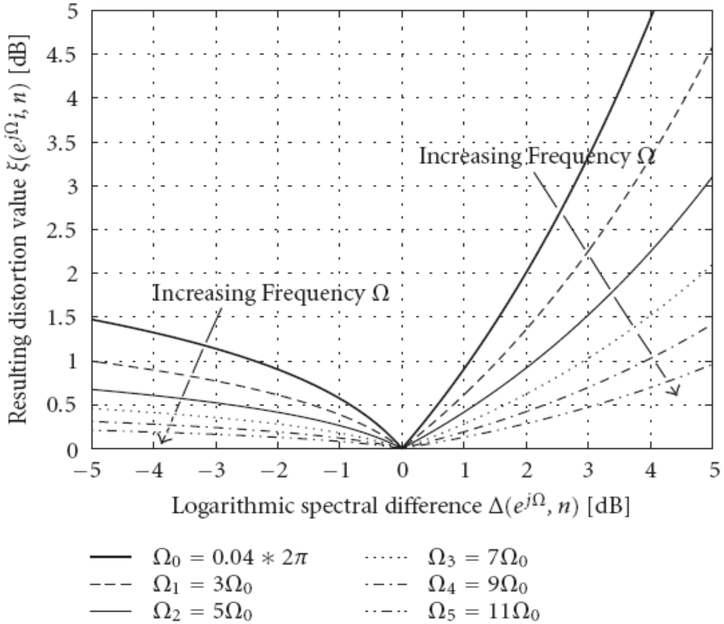


$$Error = \left| \tilde{f}_{\mathrm{HB}} - f_{\mathrm{HB}} \right|$$

BWE Performance Evaluation (3/8)

Objective quality evaluation - Spectral distortion measure of estimated spectral envelope with and without the iterative process

- Spectral Distortion Measure (SDM)
  - Non-symmetric distortion measure.
  - Takes the human auditory system into account.
  - penalizes spectral overestimation higher than underestimation.



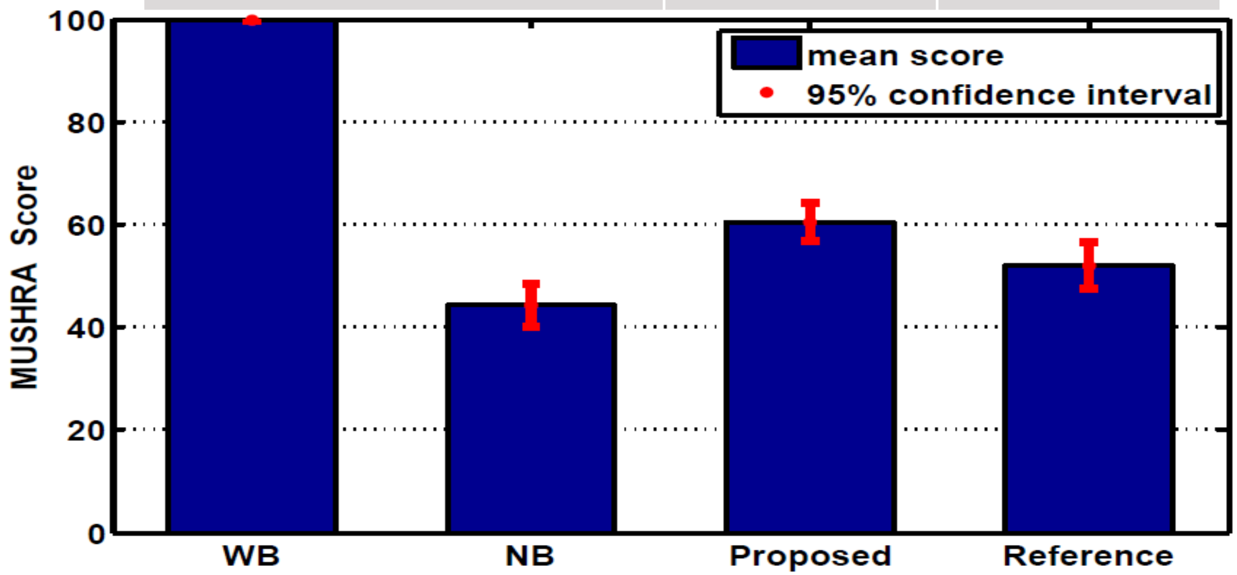
Measured	SDM [dB]	LSD [dB]
Without iterative process	13.6380	9.9759
With iterative process	9.8889	9.9057

BWE Performance Evaluation (4/8)

Subjective quality evaluation

- **MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor)** – ranks several speech samples for score between 0-100.
- Recommendation ITU-R BS.1116-1.
- Test setup: 11 listeners, 6 different sentences (English, 3 male, 3 female) – each with WB reference signal, NB anchor signal, proposed BWE signal and a reference BWE signal from Geiser [Based on Jax et al., 2003. modified in 2010].

	Proposed	Reference
Mean MUSHRA score	60	55

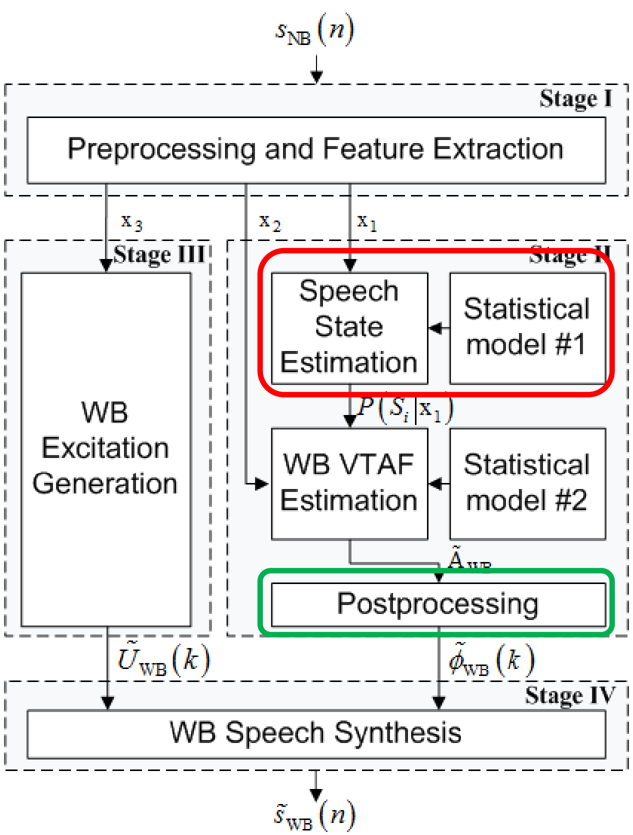


BWE Performance Evaluation (5/8)

Complexity evaluation

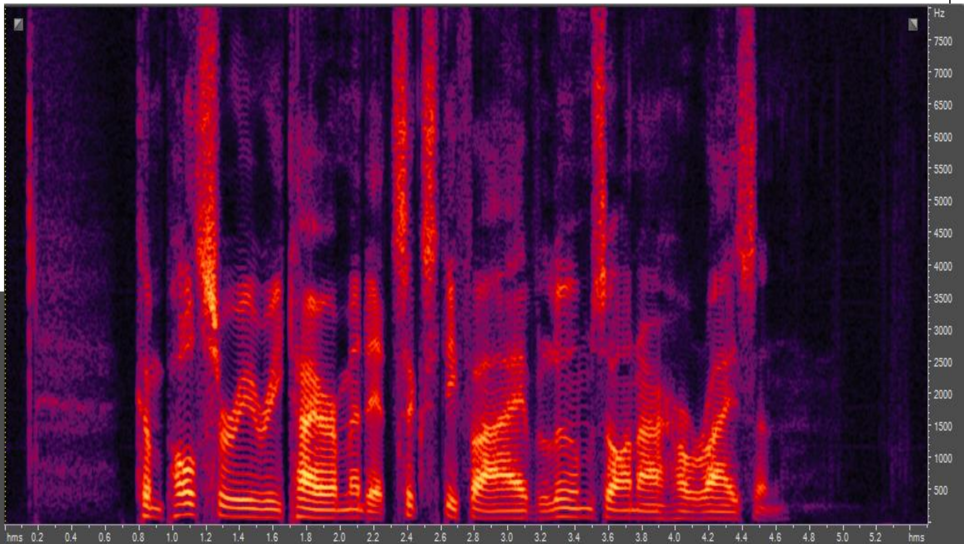
- Calculating Matlab processing time of each major processing block.
- The table indicate the average processing time of a 20 msec speech frame.

Algorithm Processing Block	Computation Time [msec]
Preprocessing and feature extraction	1.27
State estimation	19.39
WB VTAF estimation	0.59
Postprocessing (iterative process)	7.69
Postprocessing (gain adjustment)	0.36
WB excitation generation	0.04
WB speech synthesis	0.57
Total	29.91

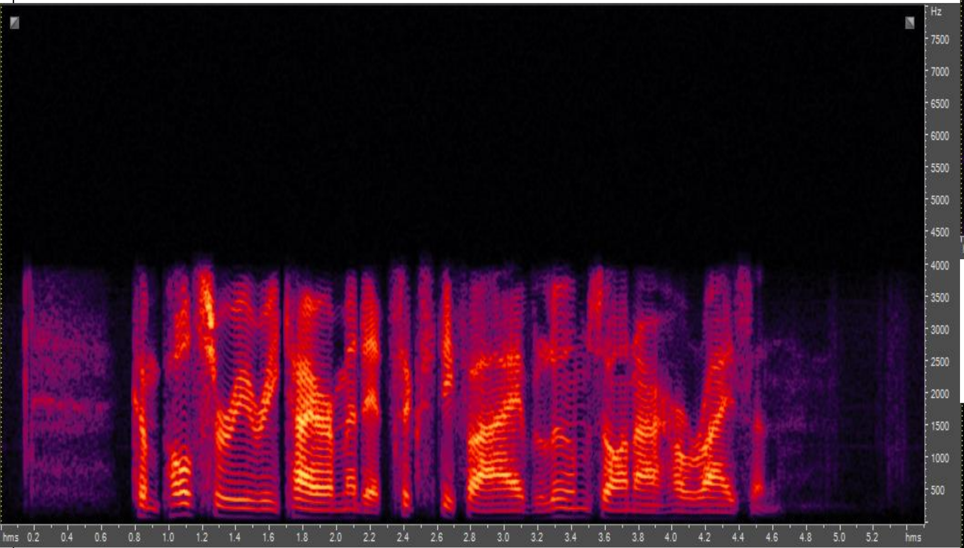


BWE Performance Evaluation (6/8)

Male spectograms

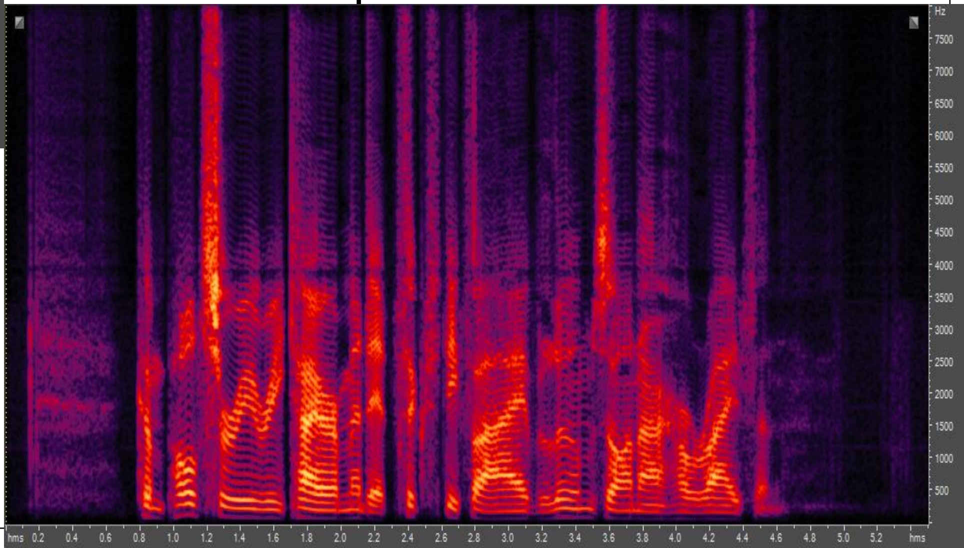


Wideband



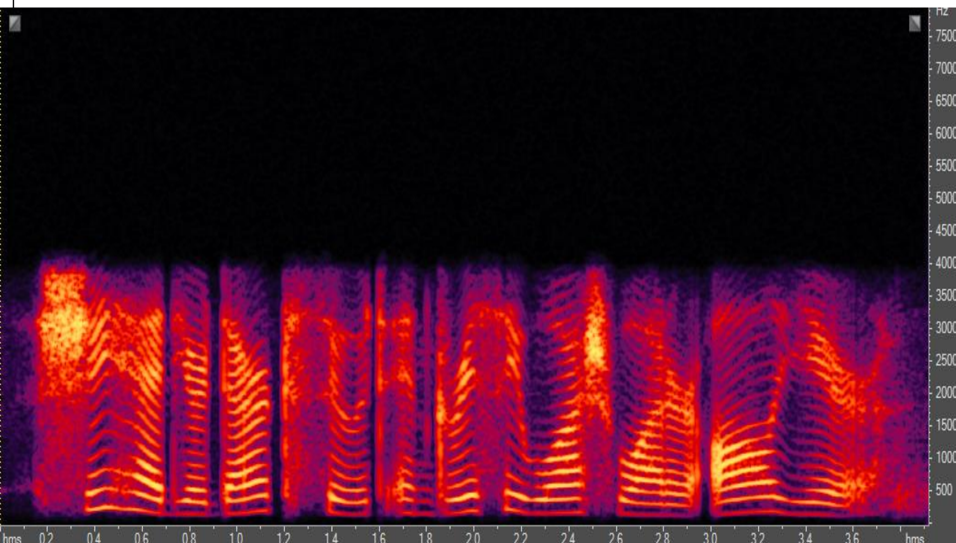
Narrowband

Proposed BWE

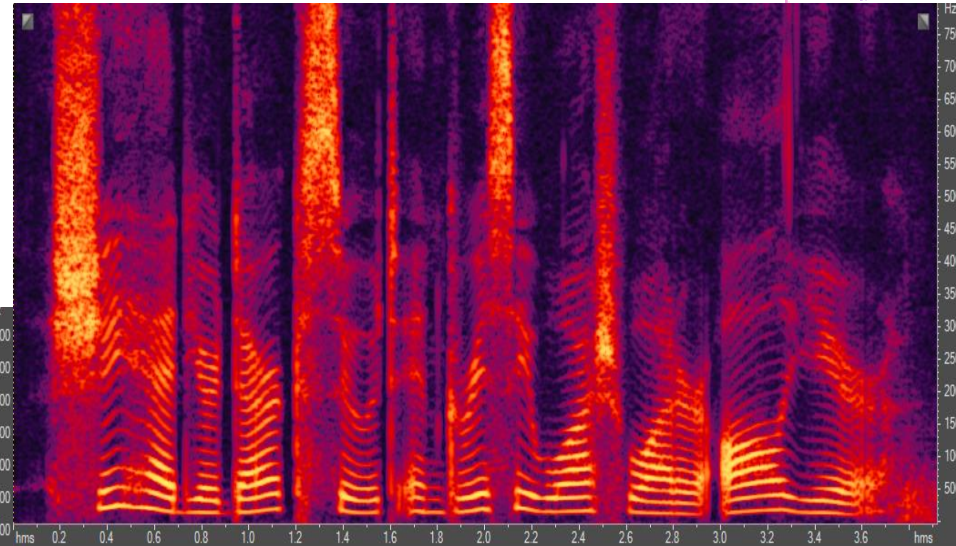


BWE Performance Evaluation (7/8)

Female spectograms

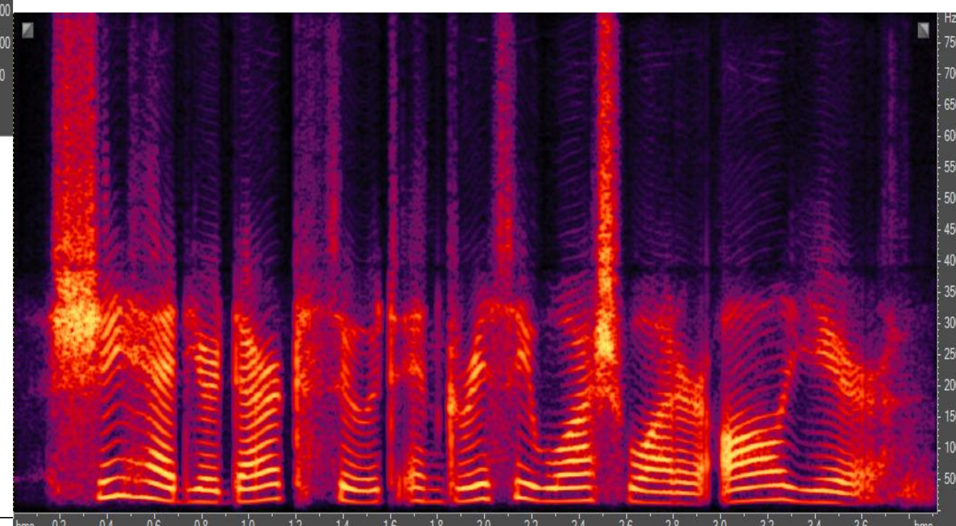


Narrowband



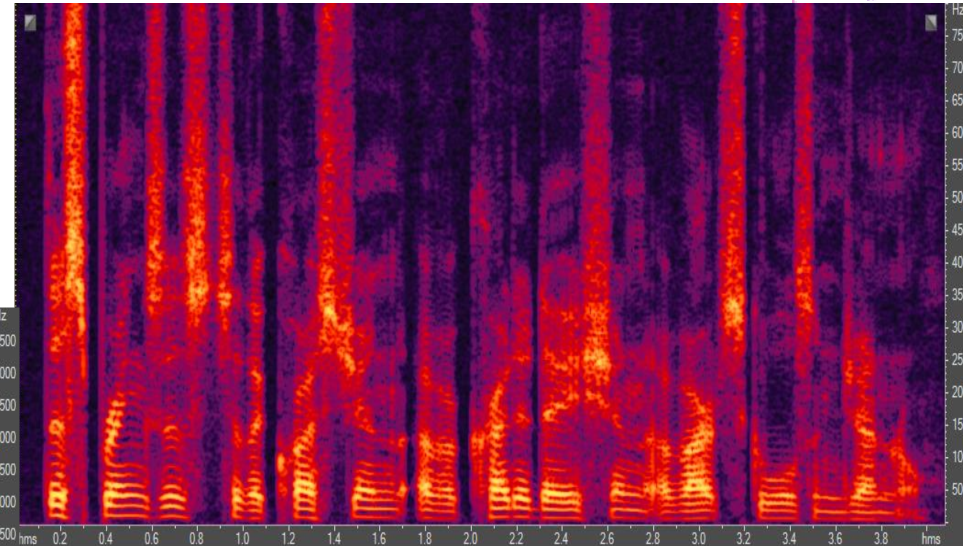
Wideband

Proposed BWE



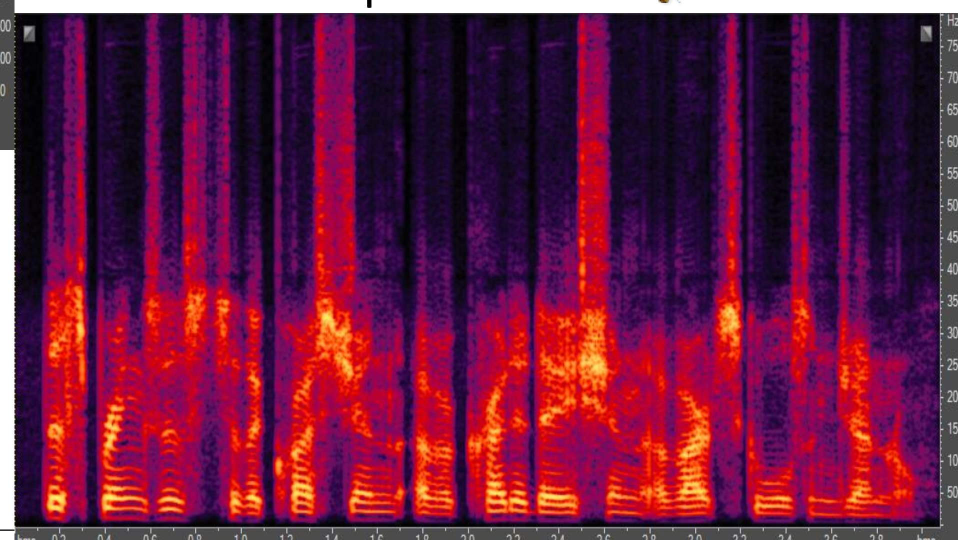
BWE Performance Evaluation (8/8)

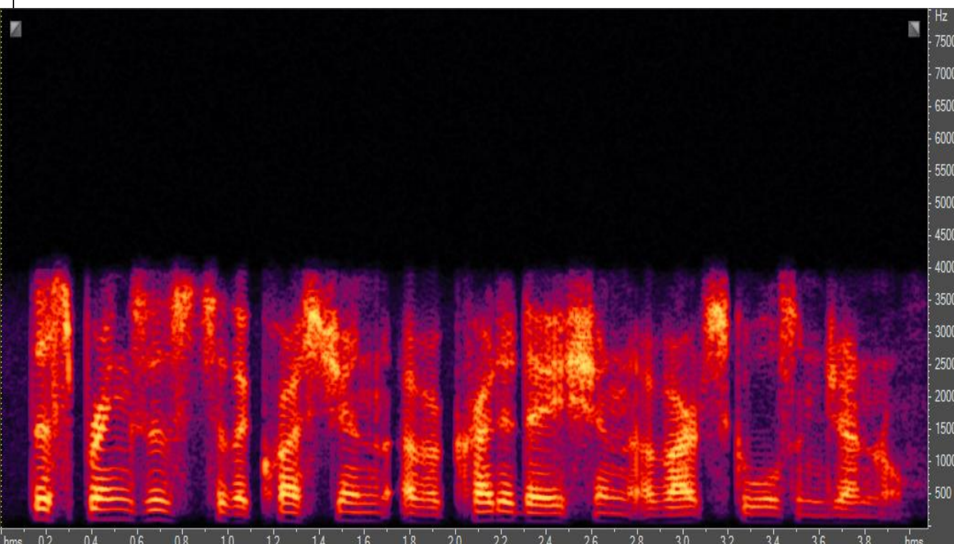
Male spectograms



Wideband

Proposed BWE





Narrowband

Outline

- Introduction
- Methods of BWE
- Proposed BWE Algorithm
- Performance Evaluation
- **Conclusion**

Conclusion (1/2)

- Proposed BWE algorithm innovations
  - Phonetic content estimation for each speech frame.
  - WB VTAF estimation for specific speaker using codebook search.
  - Iterative tuning of estimated WB VTAF for better gain adjustment.
- Algorithm advantages
  - Phonetic-based estimation reduces estimation error for unvoiced frames.
  - Iterative tuning at postprocessing reduces estimation error artifacts.
- Algorithm shortcomings
  - Using VTAF is lacking in modeling nasal and unvoiced sounds.
  - HMM-based phoneme estimation and online sensitivity function calculation at the postprocessing step, result in high algorithm complexity.

Conclusion (2/2)

- Future Work
  - Reduce algorithm complexity by using sensitivity functions tables for each WB VTAF codeword.
  - Use the postprocessing iterative procedure for better refinement and control of estimated spectral envelope by high-band formants tuning to past estimated high-band formants.
  - Change spectral envelope estimation model for unvoiced and nasal sounds.
  - Check algorithm robustness to background noise and to different languages.

Thank You

References - General

- “Bandwidth Extension of Speech Signals”, Bernd Iser, Wolfgang Minker, Gerhard Schmidt, Springer 2008.
- “Audio Bandwidth Extension”, Erik Larsen, Ronald M. Aarts, wiley 2004.
- “Artificial Bandwidth Extension of Narrowband Speech”, Master’s Thesis, Nels Rohde, Svend Aage Vedstesen, Aalborg University 2007.
- “Toward Wideband Speech by Narrowband Speech Bandwidth Extension: Magic Effect or Wideband Recovery?”, Doctor’s Thesis, Gilles MIET, Maine University 2001.
- “Bandwidth Extension of Speech Signals: A Catalyst for the Introduction of Wideband Speech Coding?”, Peter Jax and Peter Vary, RWTH Aachen University, IEEE Communications Magazine, May 2006.

References - Papers

- "Quality Enhancement of Band Limited Speech by Filtering and Multirate Techniques", H. Yasukawa, International Conference on Spoken Language Processing, ICSLP '94,27.7, pp. 1607-1610, Sept. 1994.
- “Adaptive filtering for broad band signal reconstruction using spectrum extrapolation", H. Yasukawa, The 7th IEEE 1996 Digital Signal Processing Workshop, 1996.
- “Bandwidth expansion of speech based on vector quantization of the mel-frequency cepstral coefficients”, N. Enbom, W. B. Kleijn, in Proc. IEEE Workshop on Speech Coding, Porvoo, Finland, pp. 171–173, 1999.
- “Techniques for artificial bandwidth extension of telephone speech”, Ulrich Kornagel, Signal Processing , Vol. 86 , Nr. 6, p. 1296—1306, 2006.
- “Low-complexity feature-mapped speech bandwidth extension”, H. Gustafsson, U. A. Lindgren, I. Claesson, IEEE Trans. On Audio, Speech & Language Processing, vol. 14, pp. 577-588, 2006.
- “On artificial bandwidth extension of telephone speech”, P. Jax and P. Vary, Signal Processing, vol. 83, no. 8, pp. 1707–1719, 2003.
- “Evaluation of an artificial speech bandwidth extension method in three languages”, H. Pulakka, L. Laaksonen, M. Vainio, J. Pohjalainen, P. Alku, IEEE Trans. Audio Speech Lang. Process. 16 (6) (August 2008) 1124–1137.
- “Estimation of vocal-tract shapes from acoustical analysis of the speech wave. The state of the art”, H. Wakita. IEEE Transactions on Acoustics, Speech and Signal Processing. ASSP-27(3), June 1979.
- “Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms”, H. Wakita. IEEE Transactions on Audio and Electroacoustics, AU-21(5)L417. October 1973.