Pattern Recognition

Part 6: Bandwidth Extension

Gerhard Schmidt

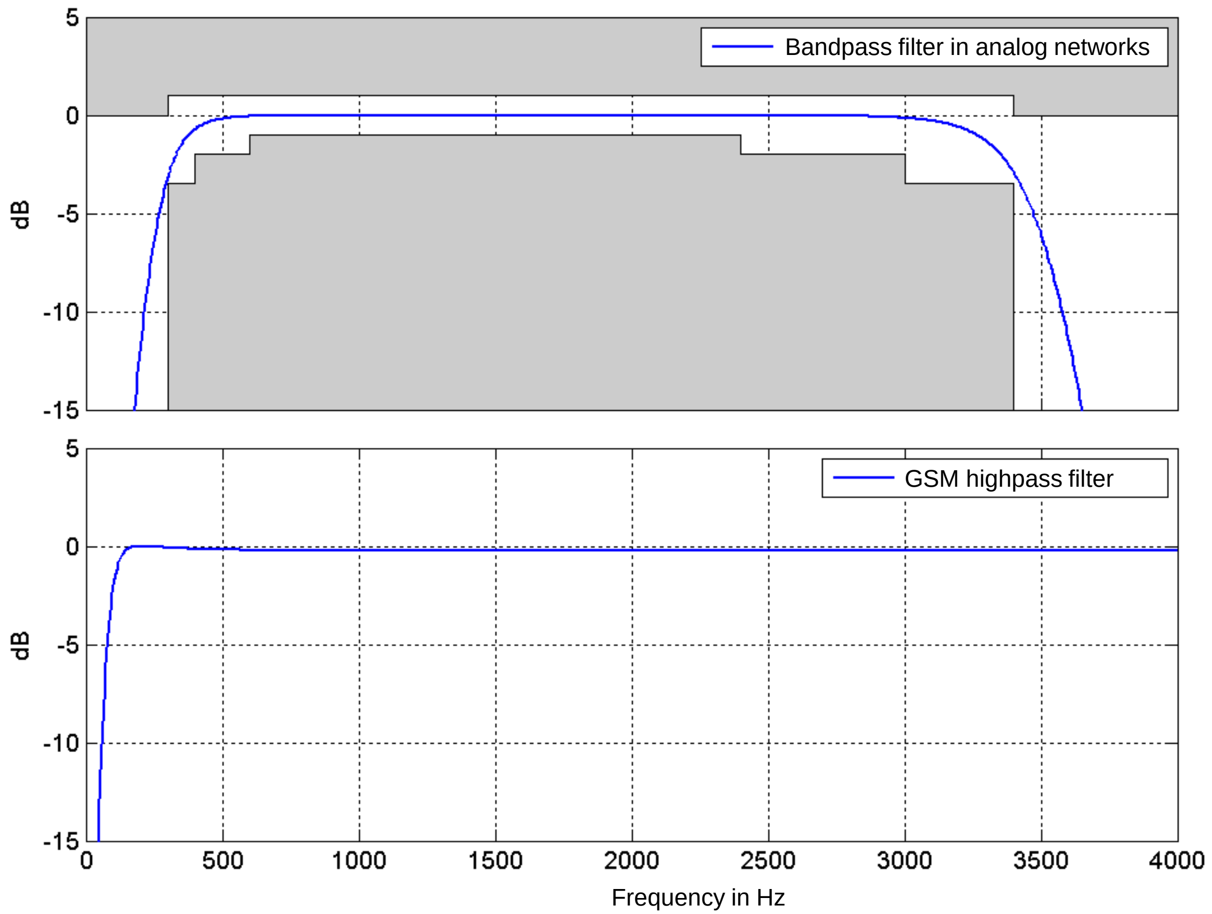
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering
Institute of Electrical and Information Engineering
Digital Signal Processing and System Theory

Contents

Christian-Albrechts-Universität zu Kiel

Motivation – Part 1

Band- or Highpass filter of the analog or of GSM telephone networks:

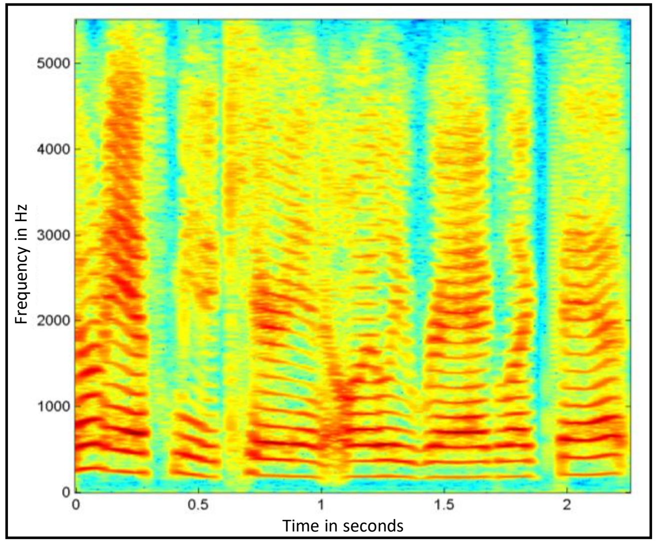


Signal components below 300 Hz and above 3.4 kHz are strongly attenuated (ITU-T Rec. G.712).

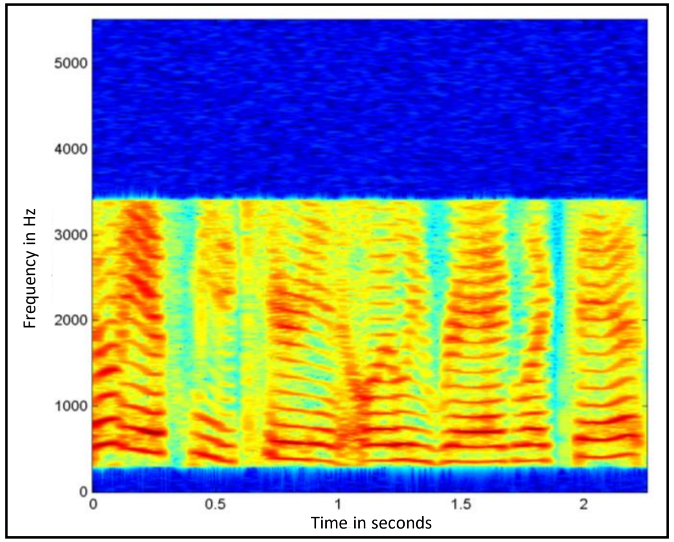
Signal components below 70 Hz are strongly attenuated. The maximum signal frequency is 4 kHz.

Motivation – Part 2

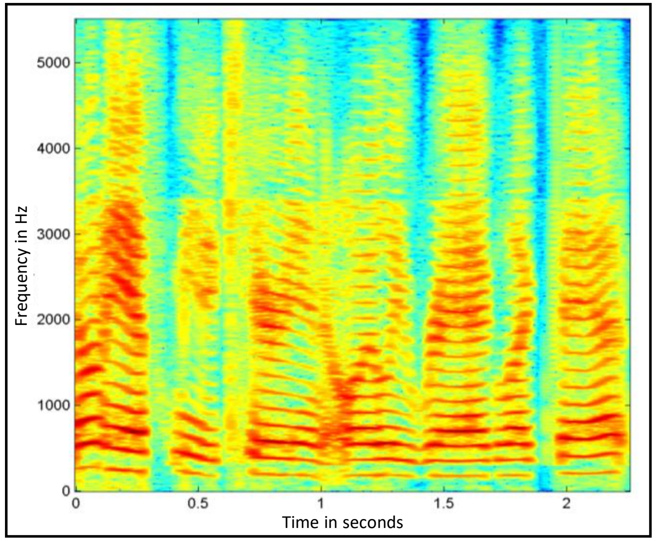
Examples of signals:



Speech signal (Bandwidth: 0 – 5500 Hz)



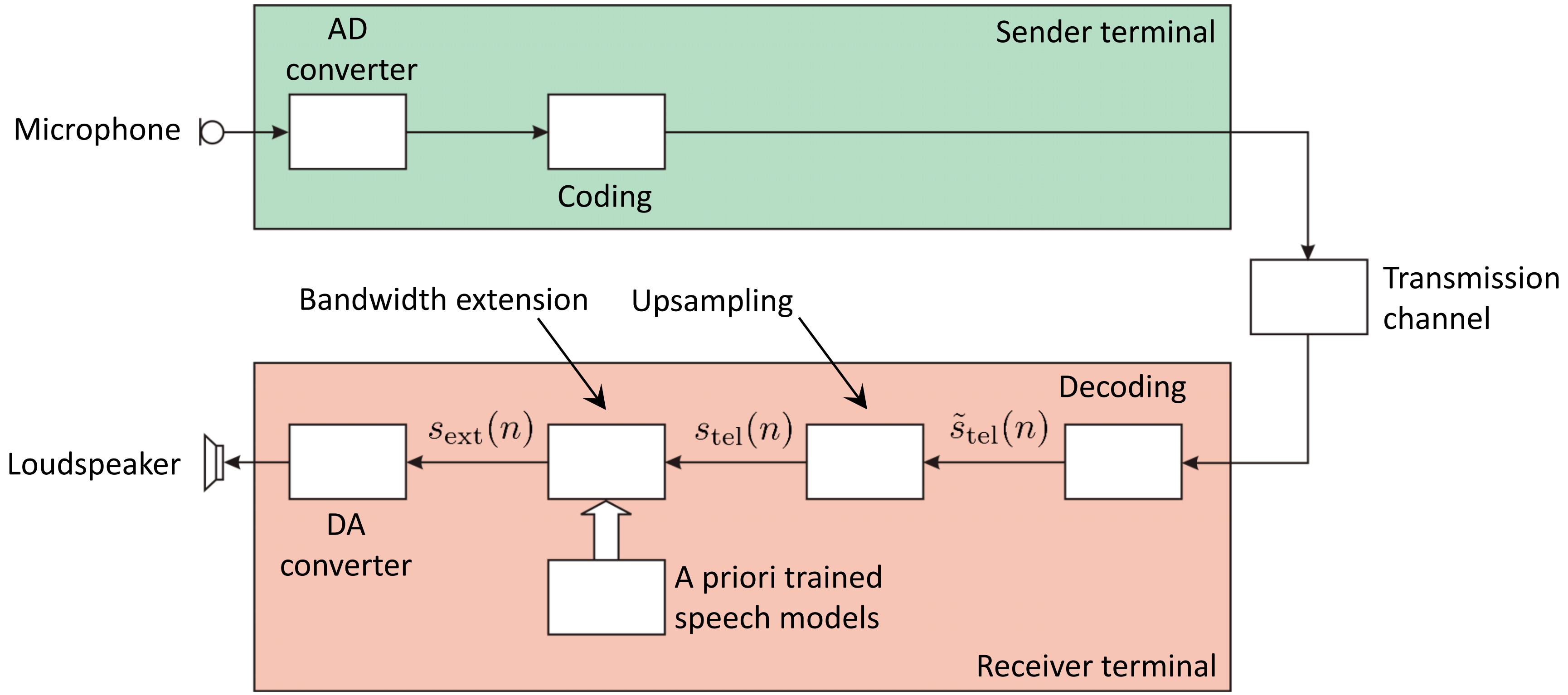
Signal after transmission
over an analog telephone network
(Bandwidth: 300 – 3400 Hz)



Signal after bandwidth extension
(Bandwidth: 0 – 5500 Hz)

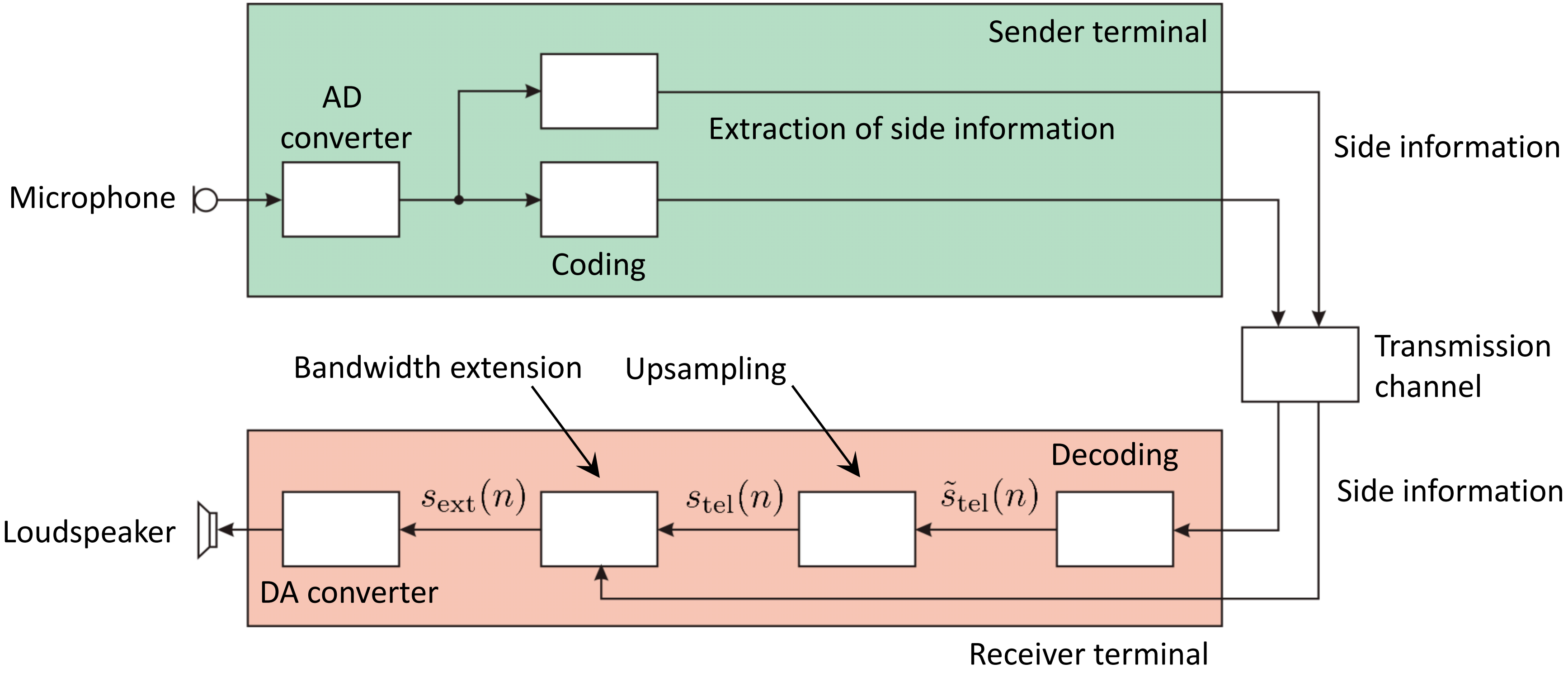
System Concept – Part 1

Approaches without transmission of side information:



System Concept – Part 2

Approaches with transmission of side information:



Literature

Bandwidth extension:

- B. Iser, G. Schmidt: ***Bandwidth Extension of Telephony Speech***, Chapter from E. Hänsler, G. Schmidt (Editor), Speech and Audio Processing in Adverse Environments, Springer, 2008
- P. Jax: ***Bandwidth Extension for Speech***, Chapter fromE. Larsen, R. M. Aarts (Editor), Audio Bandwidth Extension, Wiley, 2004
- P. Vary, R. Martin: ***Digital Speech Transmission***, Wiley, 2006

Neural Networks:

- D. Nauck, F. Klawonn, R. Kruse: ***Neuronale Netze und Fuzzy-Systeme***, Vieweg, 1996 (in German)

Bandwidth Extension – Different Methods

Bandwidth extension

Deterministic approach

- Upsampling with “bad“ anti-imaging filter
- Spectral shifting

Model-based approach

- Separation of excitation signal and filtering
- Nonlinearities, modulation, signal generation for generating the excitation signal
- Neural networks, codebooks, linear mapping for estimating spectral envelopes

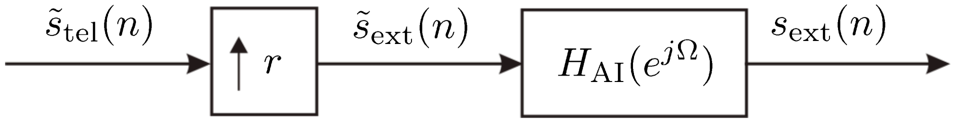
Deterministic Approach

Examples

- Upsampling with “bad“ anti-imaging filters
- Spectral shifting

Approach Without Speech Models – Part 1

Upsampling with images – Basic principle:



- First input the signal with the low sampling rate, ***insert*** $r-1$ ***zeros between the samples***. Although this increases the sampling rate, it also gives rise to ***mirror*** or ***image spectra***.

$$\tilde{s}_{\text{ext}}(n) = \begin{cases} \tilde{s}\left(\frac{n}{r}\right), & \text{if } n \bmod r \equiv 0, \\ 0, & \text{else.} \end{cases}$$

- Normally one would remove the *imaging*-components with anti-imaging filters (a ***lowpass filter*** with appropriate cut-off frequency). For extending the bandwidth the idea is to apply some damping to these components so that bandwidth is extended on average.

$$s_{\text{ext}}(n) = \sum_{i=-\infty}^{\infty} h_{\text{AI},i}\,\tilde{s}_{\text{ext}}(n-i).$$

Christian-Albrechts-Universität zu Kiel

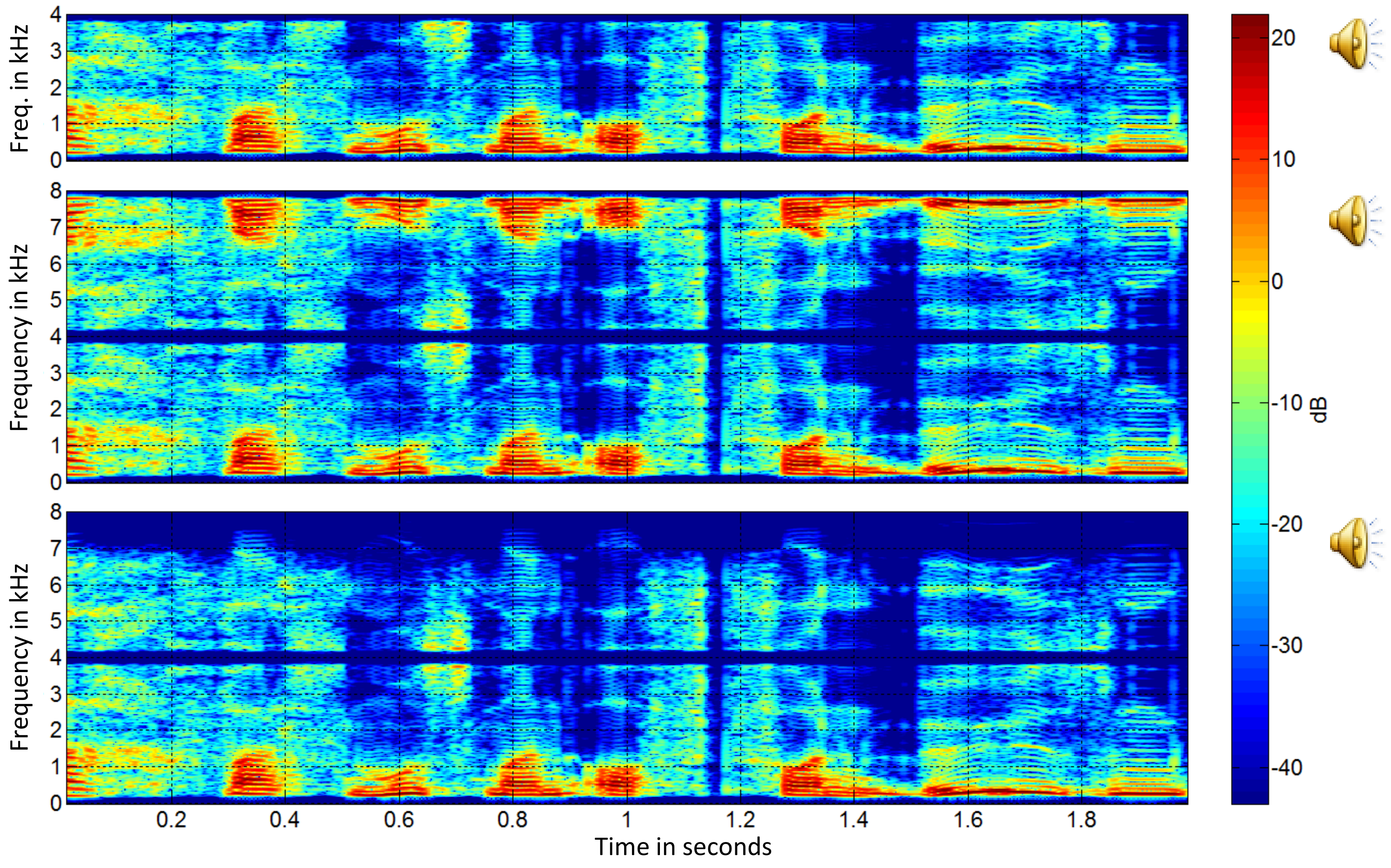
Approach Without Speech Models – Part 2

Upsampling with images – Example:

Input signal

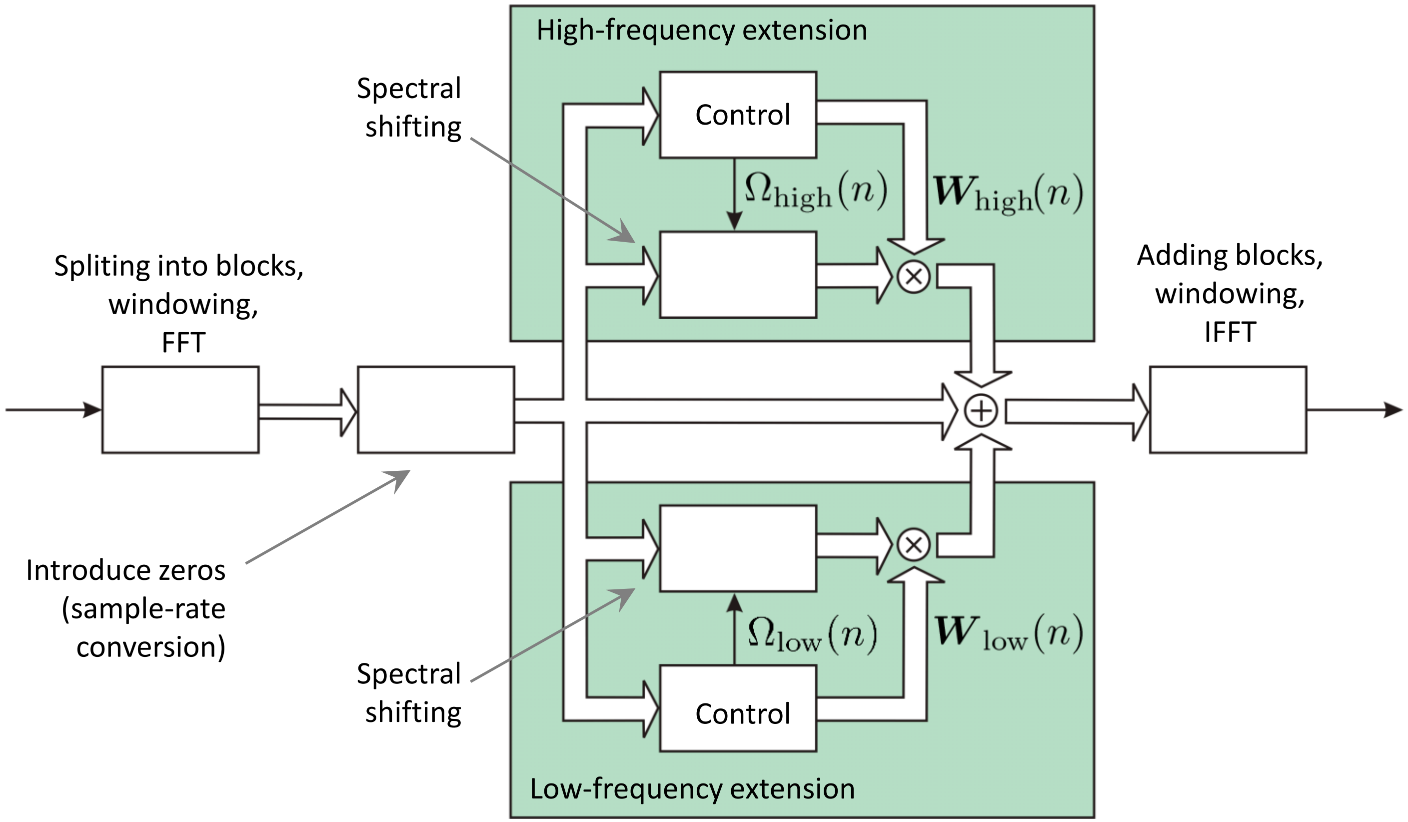
Signal after upsampling

Signal after filtering



Approach Without Speech Models – Part 3

Shifting in the spectral domain – Principle:



Approach Without Speech Models – Part 4

Shifting in the spectral domain – Principle:

- First the sample rate is increased by inserting appropriate number of zeros, which increases the subband vector size.

Input signal sub-band vector:

$$\widetilde{\boldsymbol{S}}_{\text{tel}}(n) = \Big[\widetilde{S}_{\text{tel}}\big(e^{j\frac{2\pi}{\widetilde{N}_{\text{FFT}}}0}, n\big), \ldots, \widetilde{S}_{\text{tel}}\big(e^{j\frac{2\pi}{\widetilde{N}_{\text{FFT}}}(\widetilde{N}_{\text{FFT}}-1)}, n\big)\Big]^{\text{T}}$$

Extended sub-band vector:

$$\boldsymbol{S}_{\text{tel}}(n) = \Big[\widetilde{S}_{\text{tel}}\big(e^{j\frac{2\pi}{\widetilde{N}_{\text{FFT}}}0}, n\big), \ldots, \widetilde{S}_{\text{tel}}\big(e^{j\frac{2\pi}{\widetilde{N}_{\text{FFT}}}(\frac{\widetilde{N}_{\text{FFT}}}{2}-1)}, n\big), \underbrace{0, 0, 0, \ldots 0,}_{N_{\text{FFT}}-\widetilde{N}_{\text{FFT}}+1 \text{ zeros}}$$
$$\widetilde{S}_{\text{tel}}\big(e^{j\frac{2\pi}{\widetilde{N}_{\text{FFT}}}(\frac{\widetilde{N}_{\text{FFT}}}{2}+1)}, n\big), \ldots, \widetilde{S}_{\text{tel}}\big(e^{j\frac{2\pi}{\widetilde{N}_{\text{FFT}}}(\widetilde{N}_{\text{FFT}}-1)}, n\big)\Big]^{\text{T}}$$

- This vector will subsequently be up or down shifted such that both the high and the low frequency range is extended. The resulting sub-band vector is then weighted in such a way that the extended bands are on average the same as the telephone bands.

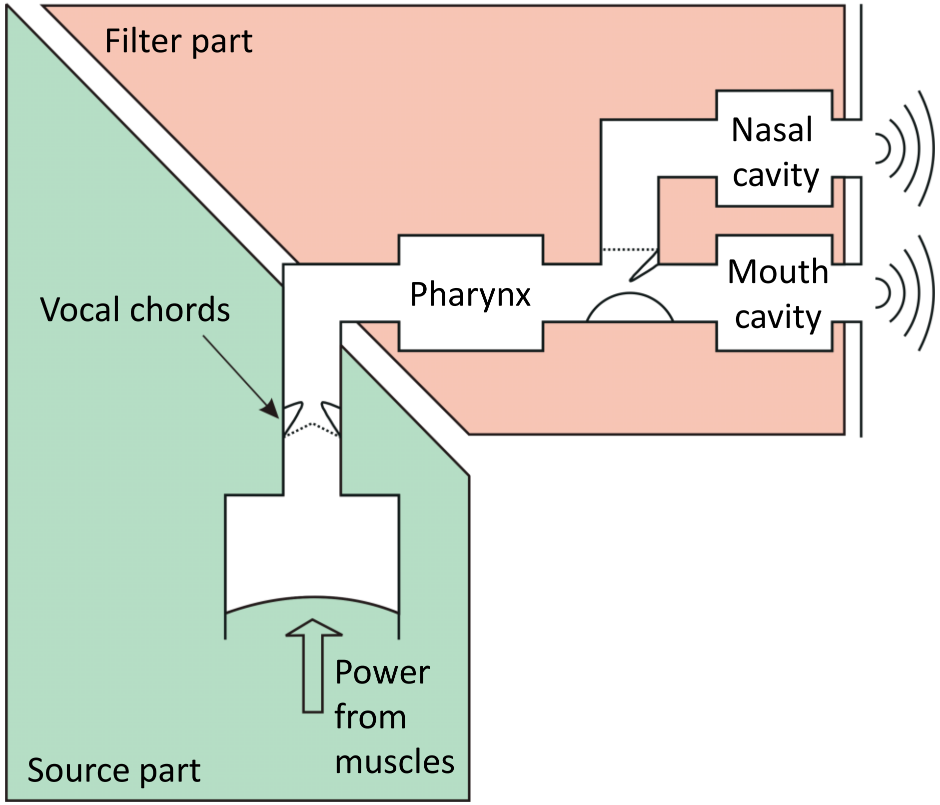
Model-Based Approaches

Examples

- ***Separation*** of the excitation signal and filtering
- Nonlinearities and Modulation approaches to ***extend the excitation signal***
- Neural Networks, codebooks, and linear mapping to ***estimate the spectral envelope***

Modeling Speech Generation – Part 1 (Repetition)

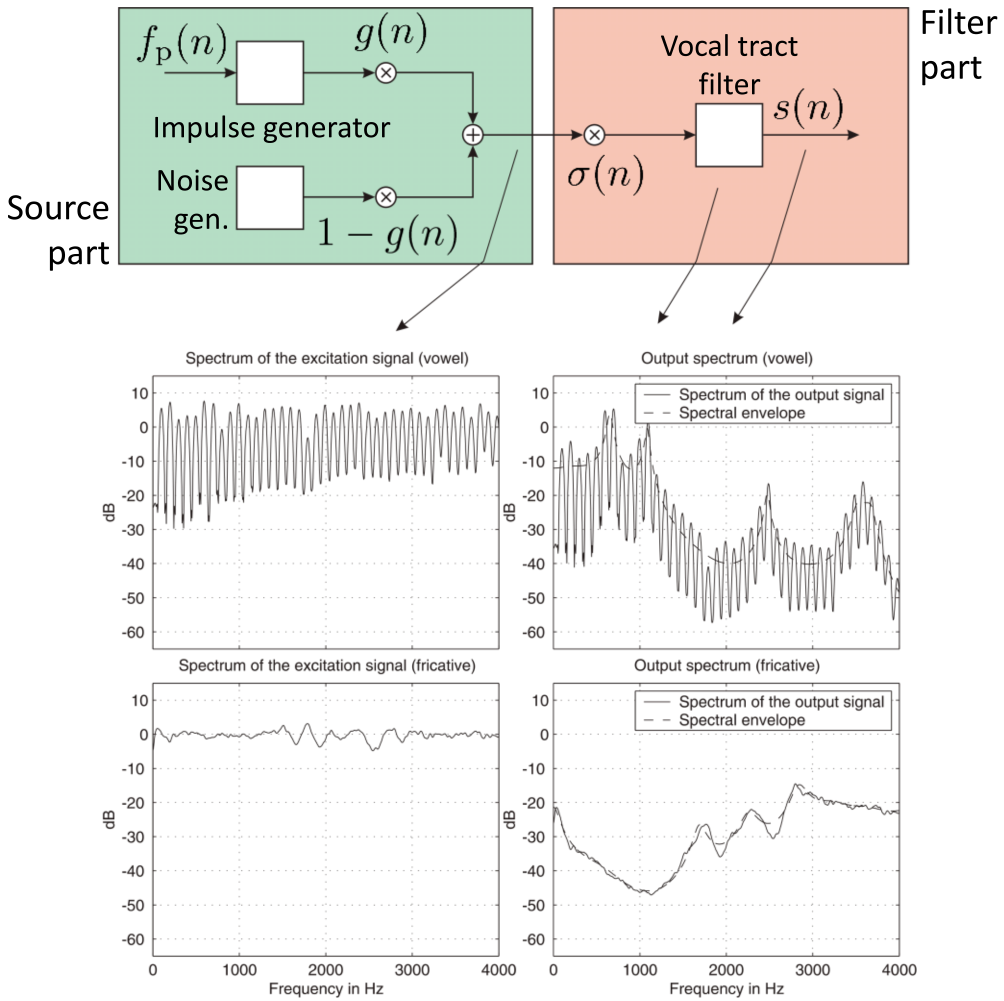
Speech production in humans:



Modeling Speech Generation – Part 2 (Repetition)

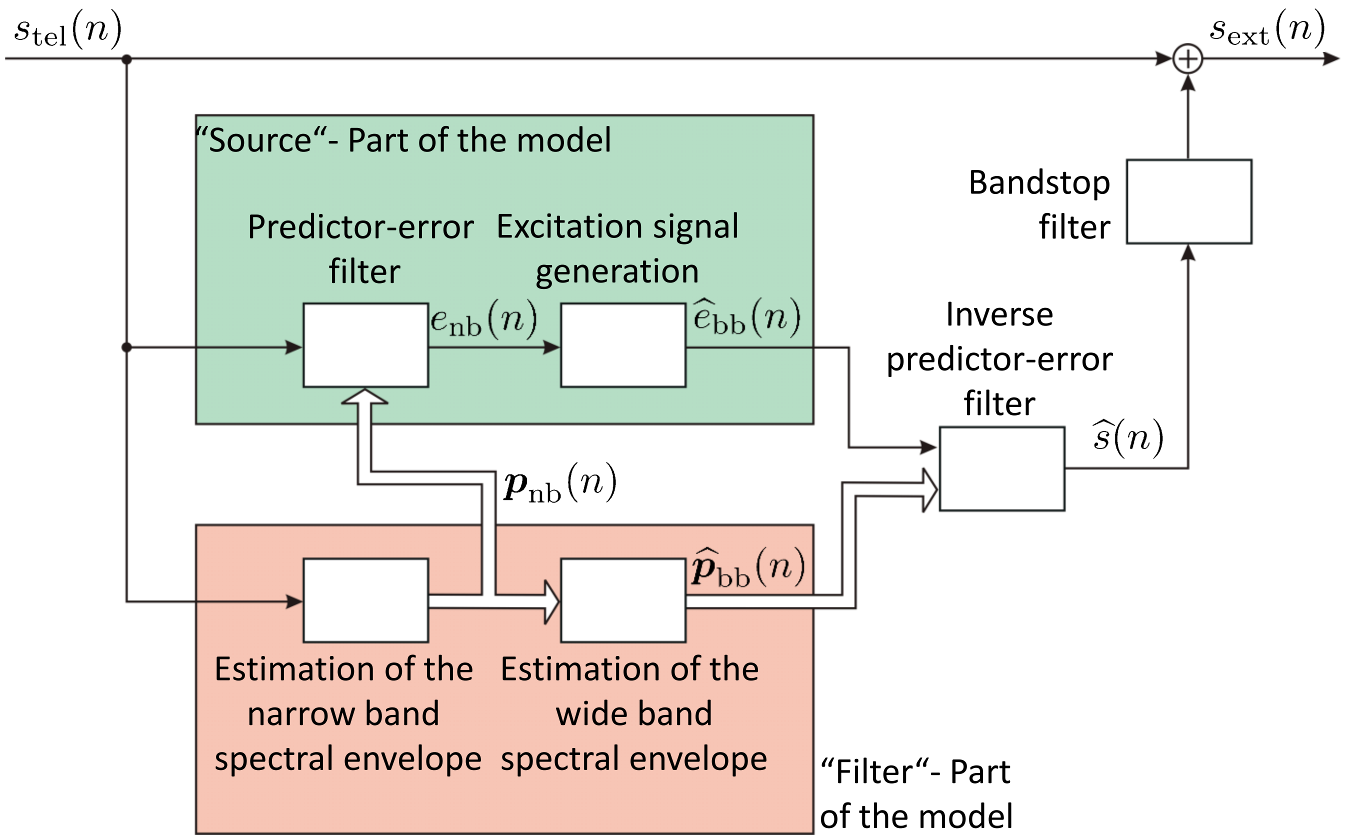
Source-filter model:

- In model-based approaches for bandwidth extension, the ***source-filter model*** is applied.
- That is, there are two separate producing parts, one is the ***excitation signal*** (wide band white signal directly behind the vocal chords) and the other is the broadband spectral envelope.
- The ***envelope estimation*** is done with the a priori trained model (based on a large database).



Model-Based Approaches for Bandwidth Extensions

Time-domain structure:



Prediction in Bandwidth Extension

Removal of the narrow-band spectral envelopes:

$$e_{\mathrm{nb}}(n) \quad = \quad s_{\mathrm{tel}}(n) - \sum_{i=1}^{N_{\mathrm{pre,nb}}} p_{\mathrm{nb},i}(n)\, s_{\mathrm{tel}}(n-i)$$

Predictor-error filter (FIR structure)

Impose the wide-band spectral envelope:

$$\widehat{s}(n) \quad = \quad \widehat{e}_{\mathrm{bb}}(n) + \sum_{i=1}^{N_{\mathrm{pre,bb}}} \widehat{p}_{\mathrm{bb},i}(n)\, \widehat{s}(n-i)$$

Inverse predictor-error filter (IIR structure)

Extension of the Excitation Signal – Part 1

Modulation or Spectral Shifting – Principle:

- With a ***multiplication*** of one (or more) ***cosine carrier***

$$\widehat{e}_{\mathrm{bb}}(n) \quad = \quad e_{\mathrm{nb}}(n) \cdot 2\,\cos(\Omega_0 n)$$

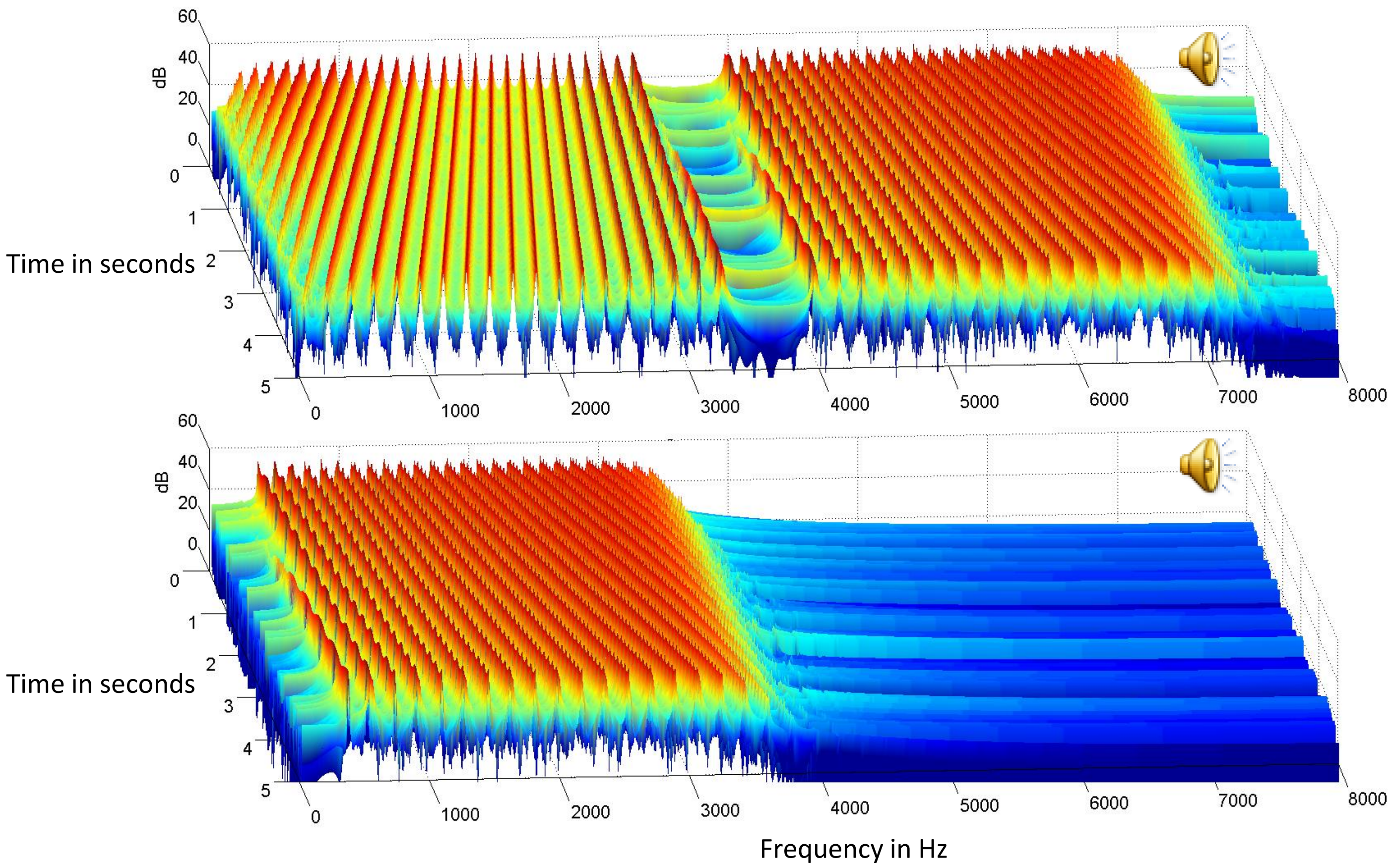
  we can generate one (or more) copies of the original spectrum:

$$\begin{aligned}\widehat{E}_{\mathrm{bb}}\left(e^{j\Omega}\right) &= E_{\mathrm{nb}}\left(e^{j\Omega}\right) * \Big[\delta\big(\Omega - \Omega_0\big) + \delta\big(\Omega + \Omega_0\big)\Big] \\ &= E_{\mathrm{nb}}\big(e^{j(\Omega-\Omega_0)}\big) + E_{\mathrm{nb}}\big(e^{j(\Omega+\Omega_0)}\big)\,.\end{aligned}$$

- Some of the resulting ***spectral components are inverted on the frequency axis*** and have to be removed by using appropriate filtering (preferably by the final bandstop filter).

Extension of the Excitation Signal – Part 2

Modulation or spectral shifting – Example:



Output signal (after multiplication with a 4-kHz-cosine carrier)

Input signal (after Predictor-error filtering)

Extension of the Excitation Signal – Part 3

Modulation or spectral shifting – Remark:

- The ***spectral gap*** in the mid-band of the extended spectra can be avoided by choosing ***an adaptive modulation frequency of the cosine-carrier***, i.e. the modulation frequency is determined by looking from which or up to which frequency the input signal power is present.
- Alternatively the ***modulation can be realized by directly using a spectral shift***. For this then an analysis-synthesis system is necessary and a delay is added to the overall system.

Extension of the Excitation Signal – Part 4

Non-linearities – Principle:

- One problem with the previous approach using modulation is that the ***fundamental frequency of the speech signal*** has to be determined if the lower frequency range has to be extended.
- An ***inexpensive alternative*** to this problem is to introduce some nonlinearities so that the signal characteristics in terms of pitch continuity are maintained. An example is the quadratic characteristic

$$\widehat{e}_{\mathrm{bb}}(n) \quad = \quad e_{\mathrm{nb}}^2(n).$$

In the spectral domain the nonlinearity is obtained with a convolution with itself

$$\widehat{E}_{\mathrm{bb}}\left(e^{j\Omega}\right) \quad = \quad E_{\mathrm{nb}}\left(e^{j\Omega}\right) * E_{\mathrm{nb}}\left(e^{j\Omega}\right).$$

With a line spectrum the pitch properties remain and new pitch lines are created at the correct distance.

Extension of the Excitation Signal – Part 5

Non-linearities – Principle:

- In case of nonlinearities the output power of the signal on the input has to be adjusted. This depends mainly on the type of nonlinearity.

- ***Typical nonlinearities***:

Half-way rectification
$$f(x) = \begin{cases} x, & \text{if } x > 0\,, \\ 0 & \text{else.} \end{cases}$$

Full-way rectification
$$f(x) = |x|$$

Saturation characteristic
$$f(x) = \begin{cases} a + (x-a)\,b, & \text{if } x > a, \\ a - (x+a)\,b, & \text{if } x < a, \\ x, & \text{else.} \end{cases}$$

Quadratic function
$$f(x) = x^2$$
$$f(x) = x\,|x|$$

Extension of the Excitation Signal – Part 6

Nonlinearities – Principle:

- ***Typical nonlinearities*** (continued):

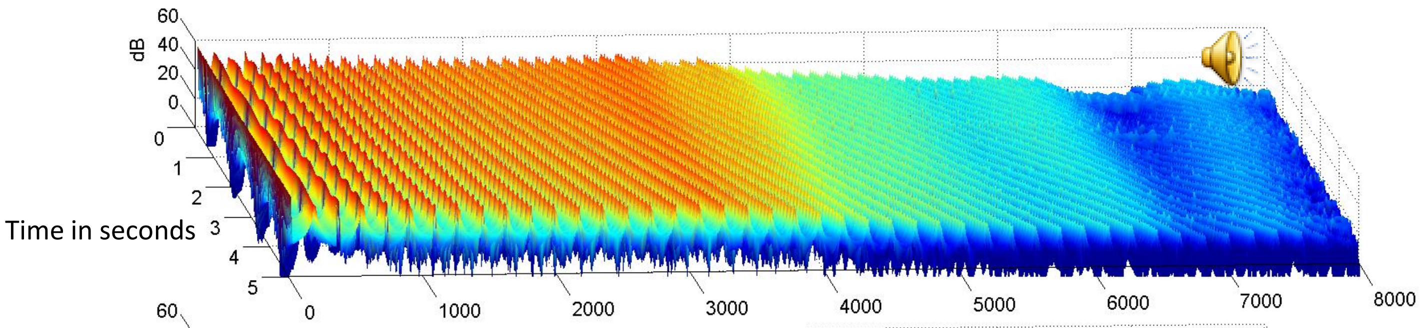
  Cubic function $f(x) = x^3$

  Tanh characteristic $f(x) = \tanh(\mu x)$

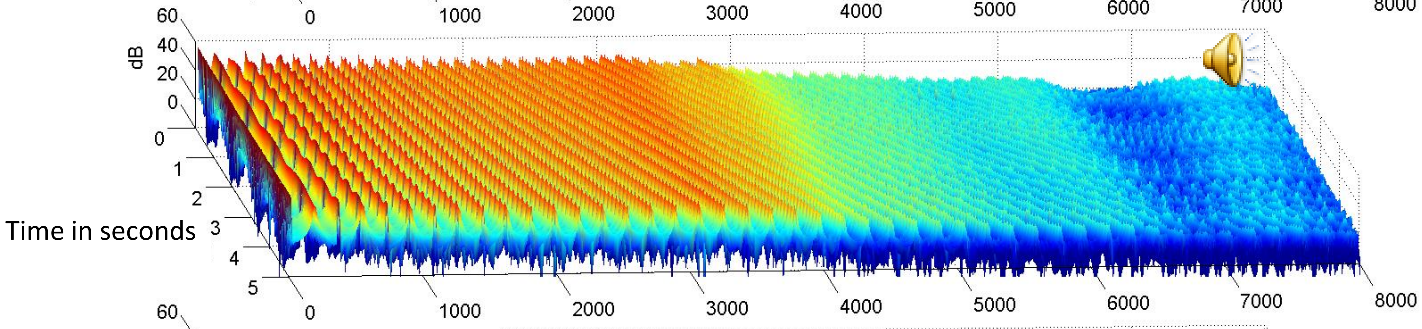
- With these curves it is important to insist that any ***DC components*** produced as a result of the non-linearity (e.g. $f(x) = x^2$ or $f(x) = |x|$) ***should be removed again***.
- Next, care must be taken that the excessive harmonics of the sampling frequency “mirror“and may hurt the pitch properties. In these cases upsampling (and again downsampling) must be applied before the application of a nonlinearity.

Extension of the Excitation Signal – Part 7

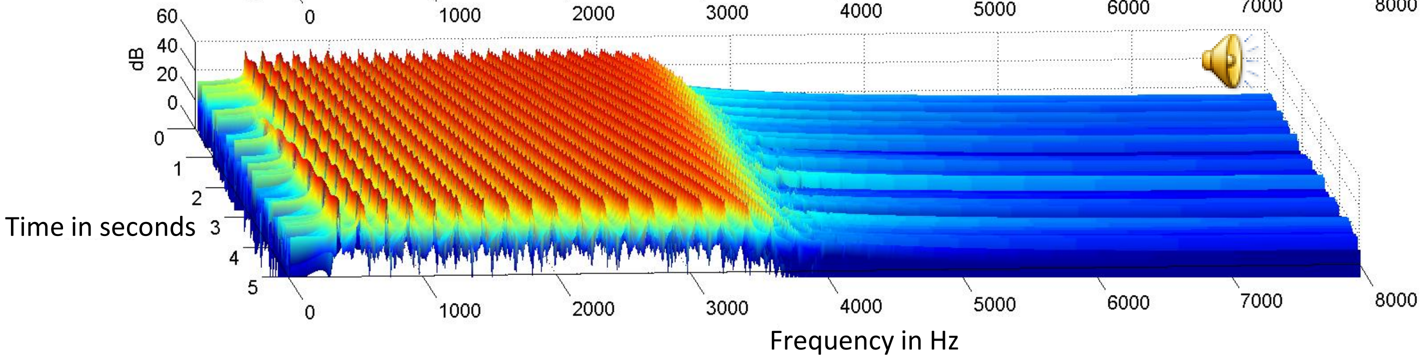
Nonlinearities – Example:



Output signal (after cubic characteristic, power normalization, and up- und down-sampling).

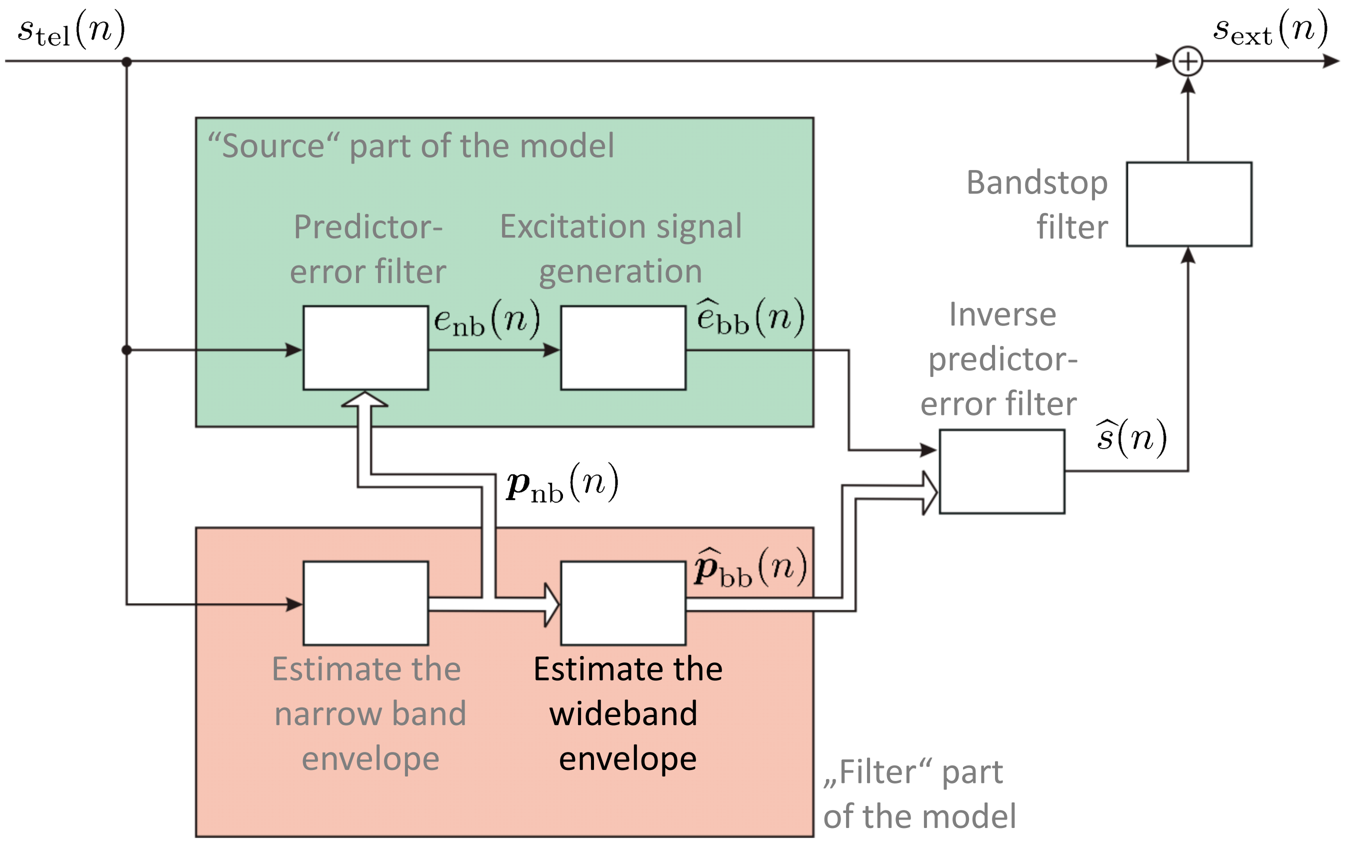


Output signal (after cubic characteristic, power normalization)



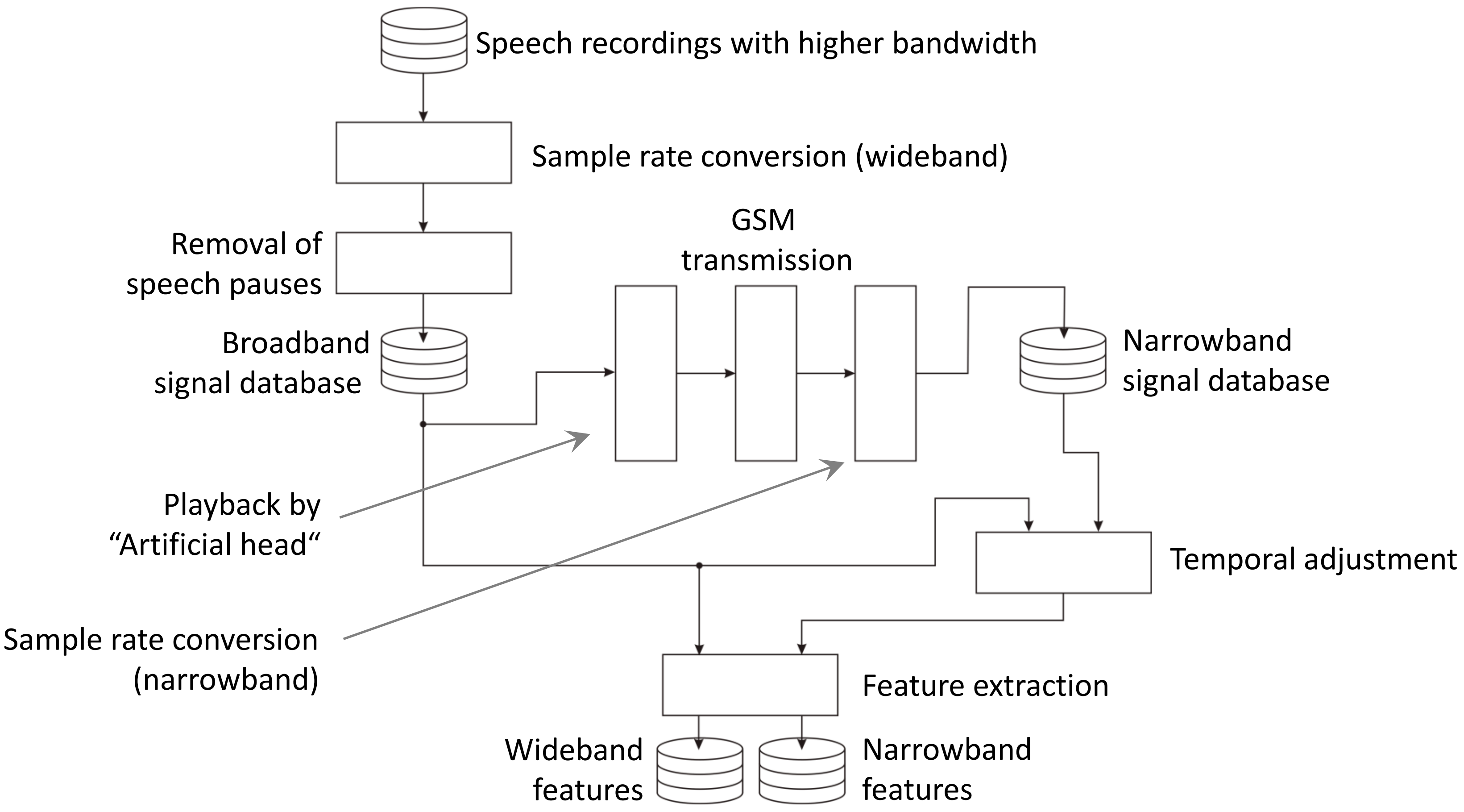
Model-based Approach for Bandwidth Extension:

Time-domain structure:



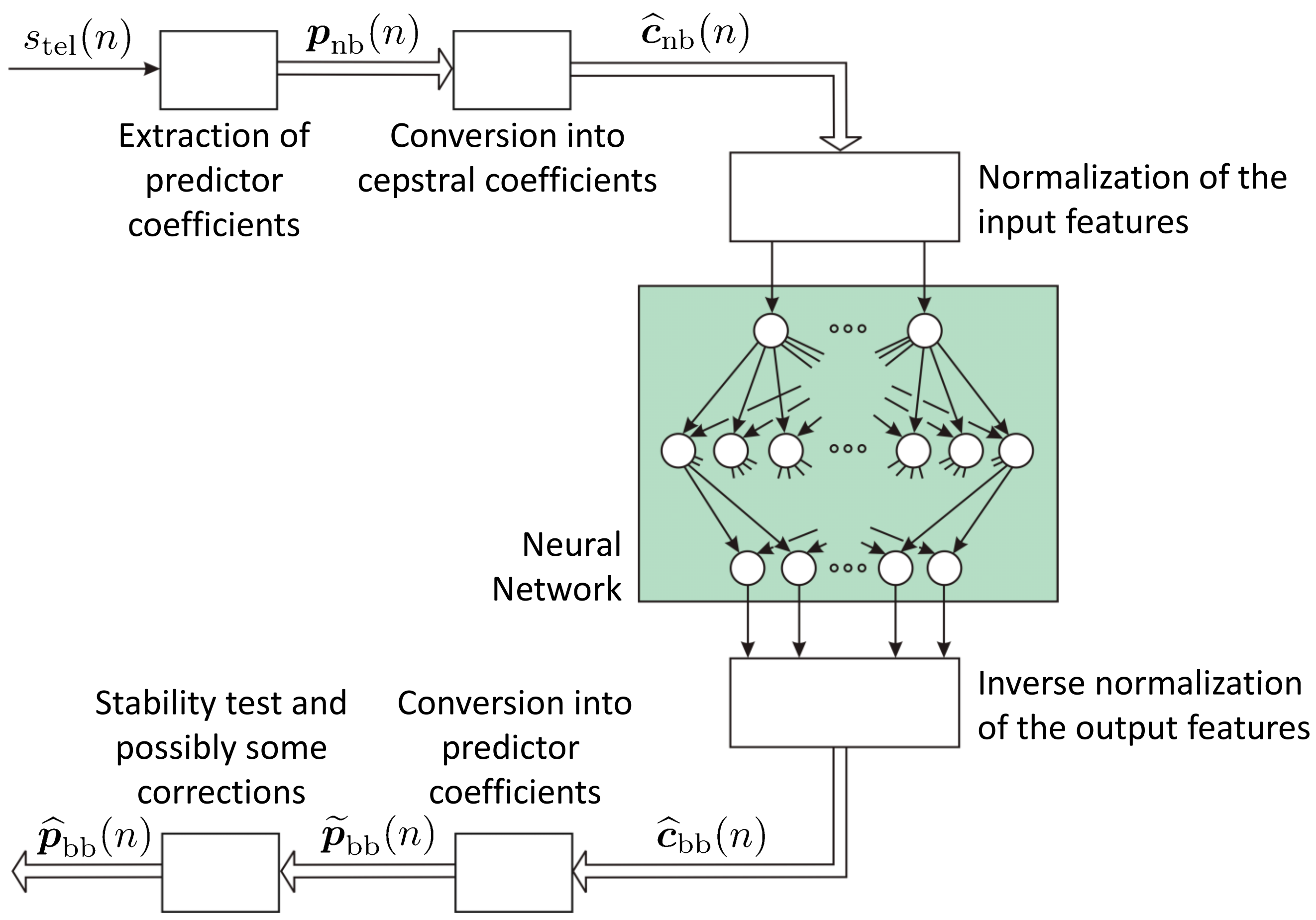
Extension of the spectral Envelope – Database for the Model Generation

Creation of the database:



Extension of the Spectral Envelope – Approaches with Neural Networks (Part 1)

Basic structure:



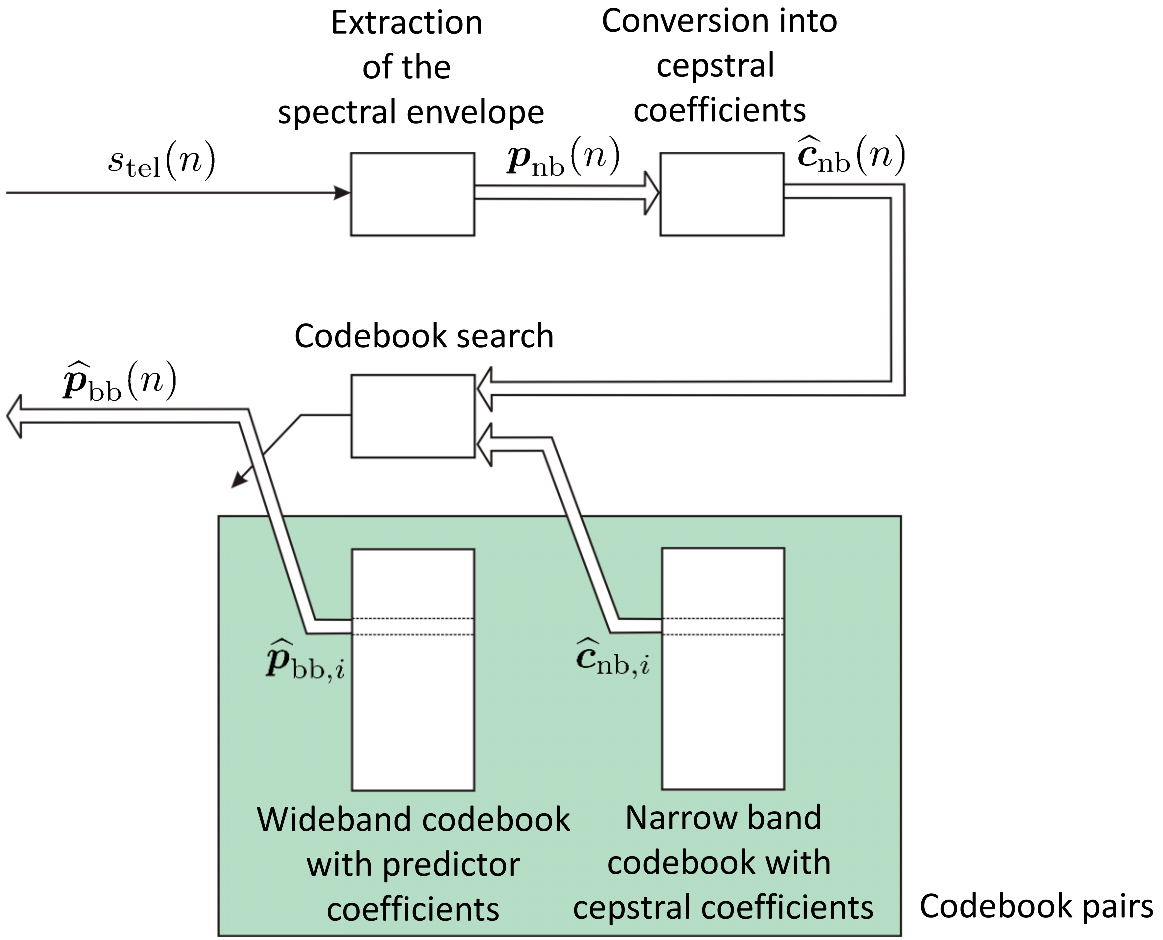
Extension of the Spectral Envelope – Approaches with Neural Networks (Part 2)

Properties:

- Neural networks can essentially learn any arbitrary correlations – it is ***not limited*** to a linear approach.
- Network structures are often ***multilayer perceptrons***, but networks with ***radial basis functions*** are also used.
- But creating the neural network cannot be fully defined. It is used very often and good quality is achieved but ***artifacts*** may occur ***temporarily***.
- Just to avoid such artifacts a ***stability test*** must be implemented at the ***end*** of the processing chain.

Extension of the Spectral Envelope – Approaches with Codebook Pairs (Part 1)

Basic structure:



Extension of the Spectral Envelope – Approaches with Codebook Pairs (Part 2)

Properties:

- When generating the wideband codebook a ***conversion into an appropriate form*** (e.g. predictor coefficients) can be added. This ***saves computation complexity*** during real-time operation.
- Beside the best codebook entry also a ***weighted sum of the best N entries*** can be utilized for the wideband estimation. The weights should be chosen such, that they are, e.g., ***inversely proportional to the corresponding distance functions and that they sum up to one***.
- Beside the distances between the individual codebook entries and the current narrowband envelope also ***the distance with the previous narrowband entry*** is sometimes taken into account. This ***avoids temporal switching effects*** among only a few codebook entries.

„Intermezzo“

Partner exercise:

- Please answer (in groups of two people) the questions that you will get during the lecture!

Evaluation of the Envelope Estimation Methods – Part 1

Subjective evaluation – Boundary Conditions:

- For the ***evaluation*** a number band-limited telephone signals were available. The excitation signal is generated by a nonlinear characteristic. For the estimation of the spectral envelope on one hand the codebook approach was chosen and on the other hand an approach based on neural networks.
- The resulting signals were presented to 10 experienced subjects. First they decide on the two variants as compared to the narrow band signals and give a rating based on the ***seven-point scale*** given below:
  - The extended version sounds ***much worse*** than the reference.
  - The extended version sounds ***worse*** than the reference.
  - The extended version sounds ***slightly worse*** than the reference.
  - The extended version and the reference sound the ***same***.
  - The extended version sounds ***slightly better*** than the reference.
  - The extended version sounds ***better*** than the reference.
  - The extended version sounds ***much better*** than the reference.

Evaluation of the Envelope Estimation Methods – Part 2

Subjective Evaluation – Boundary Conditions:

- After the tests the listeners were asked which of the two extension variants they prefer.
  Here they had to decide on a variant– no grades.
    - Variant 1 sounds ***worse*** than variant 2.
    - Variant 1 sounds ***better*** than variant 2.
- The order and the assignment of variant 1 and 2 was randomly chosen.
- Before the test, the listeners were made to listen to some test examples that are not tested, to make them familiar.

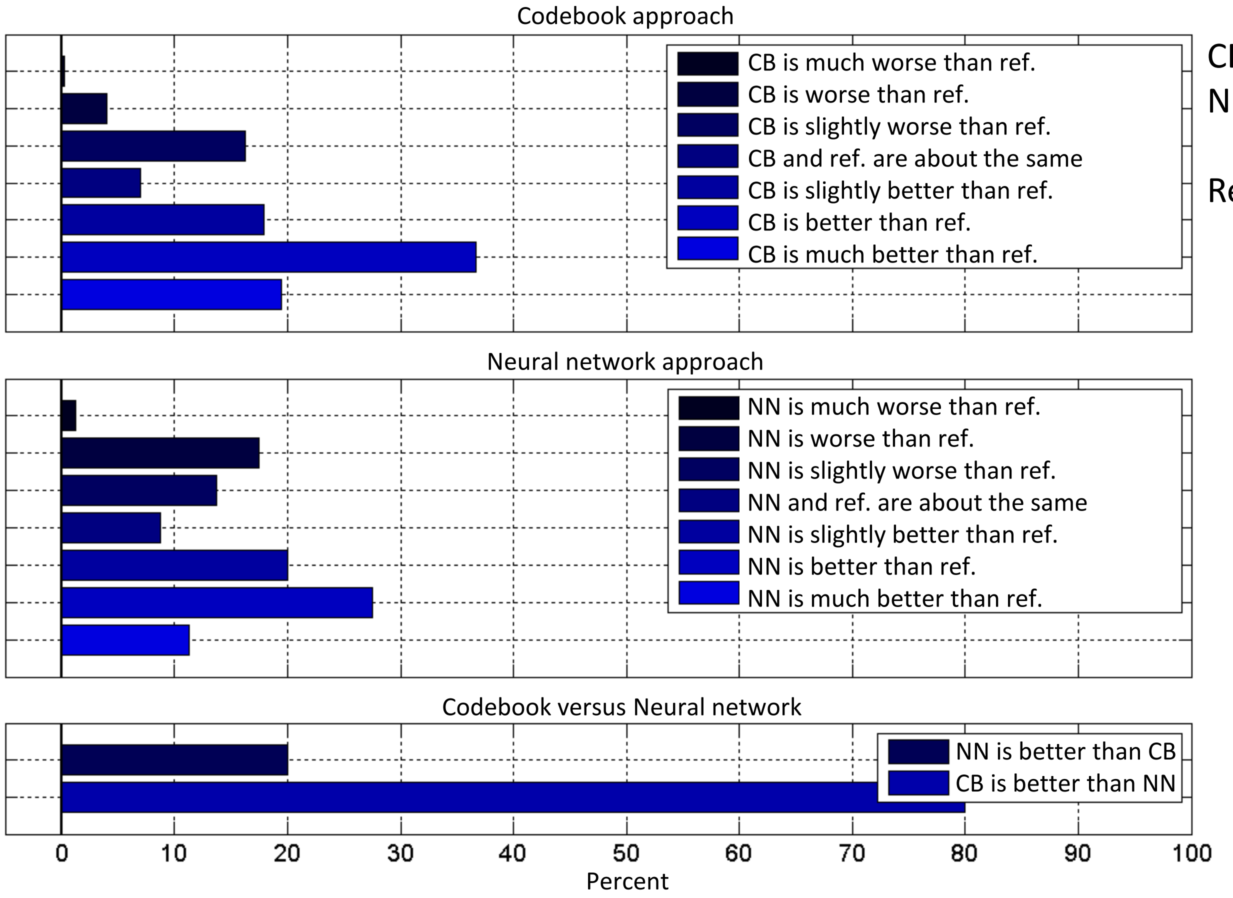
Evaluation of the Envelope Estimation Methods – Part 3

Subjective Evaluation – Results:

Comparison between extended signal with codebook and narrow band signal

Comparison between extended signal with neural network and narrow band signal

Comparison between codebook and neural network



CB = Codebook
NN = Neural network
Ref = Reference

Extension of the Spectral Envelopes – Linear Mapping Approach (Part 1)

Principle:

- Linear approach:

$$\hat{\boldsymbol{y}}(n) \quad = \quad \boldsymbol{W}\left(\boldsymbol{x}(n) - \boldsymbol{m}_x\right) + \boldsymbol{m}_y.$$

- Cost function:

$$F(\boldsymbol{W}) = \sum_{n=0}^{N-1} \left\| \boldsymbol{y}(n) - \hat{\boldsymbol{y}}(n) \right\|^2 \rightarrow \min \ .$$

- Determination of the mean vectors:

$$\boldsymbol{m}_x \quad = \quad \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{x}(n)$$

$$\boldsymbol{m}_y \quad = \quad \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{y}(n)$$

Extension of the Spectral Envelopes – Linear Mapping Approach (Part 2)

Principle (continued):

- Linear approach:

$$\hat{\boldsymbol{y}}(n) \quad = \quad \boldsymbol{W}\left(\boldsymbol{x}(n) - \boldsymbol{m}_x\right) + \boldsymbol{m}_y.$$

- Determination of the matrix:

$$\boldsymbol{W}_{\text{opt}} = \boldsymbol{Y}\,\boldsymbol{X}^T\left(\boldsymbol{X}\,\boldsymbol{X}^T\right)^{-1}$$

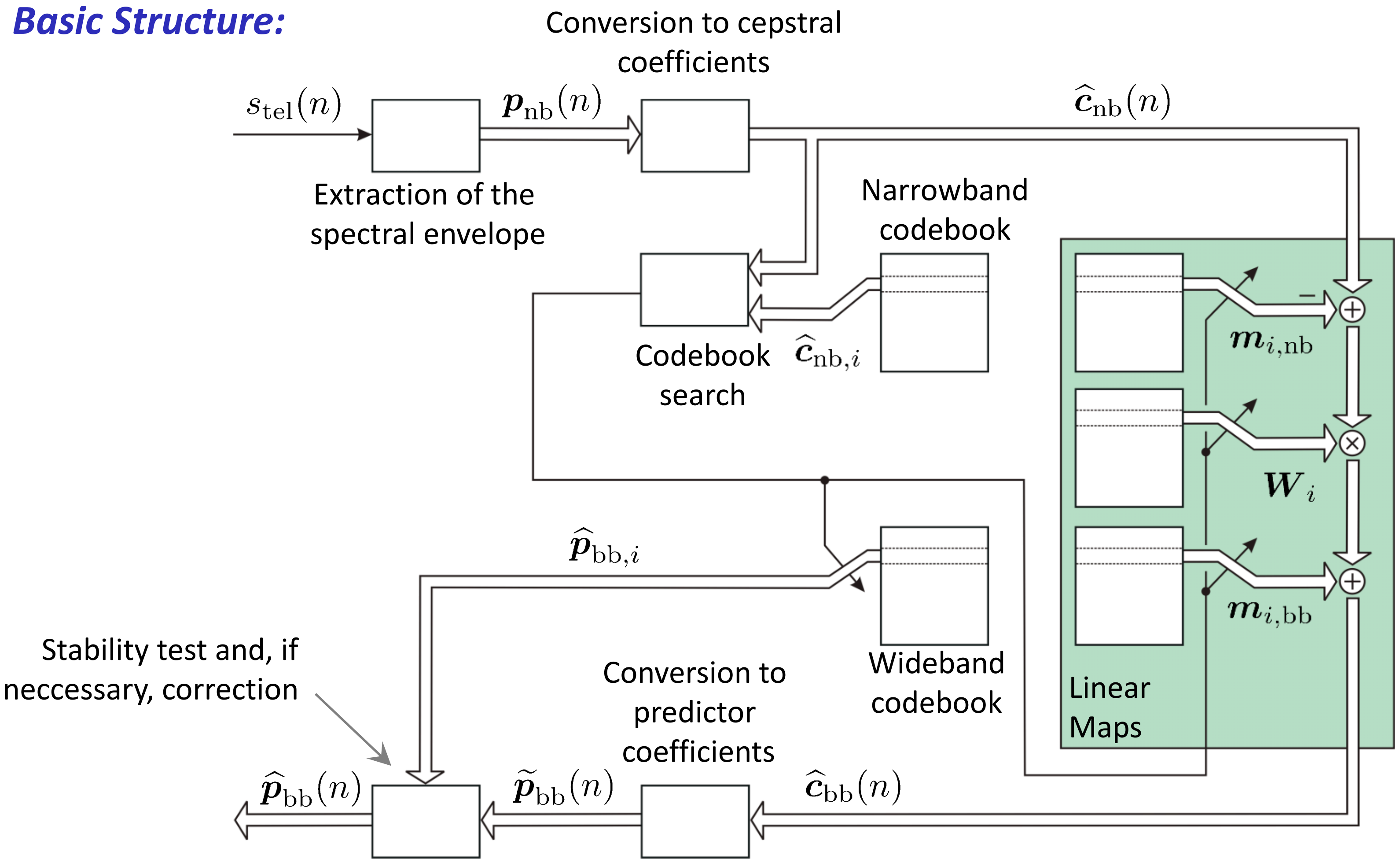
with

$$\boldsymbol{X} \quad = \quad \left[\boldsymbol{x}(0) - \boldsymbol{m}_x,\, \boldsymbol{x}(1) - \boldsymbol{m}_x,\, \ldots,\, \boldsymbol{x}(N-1) - \boldsymbol{m}_x\right],$$

$$\boldsymbol{Y} \quad = \quad \left[\boldsymbol{y}(0) - \boldsymbol{m}_y,\, \boldsymbol{y}(1) - \boldsymbol{m}_y,\, \ldots,\, \boldsymbol{y}(N-1) - \boldsymbol{m}_y\right].$$

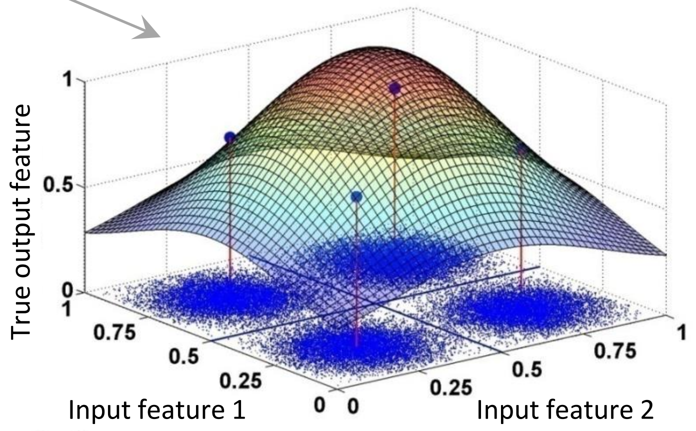
Extension of the Spectral Envelope – Approaches with Codebooks and Linear Mapping

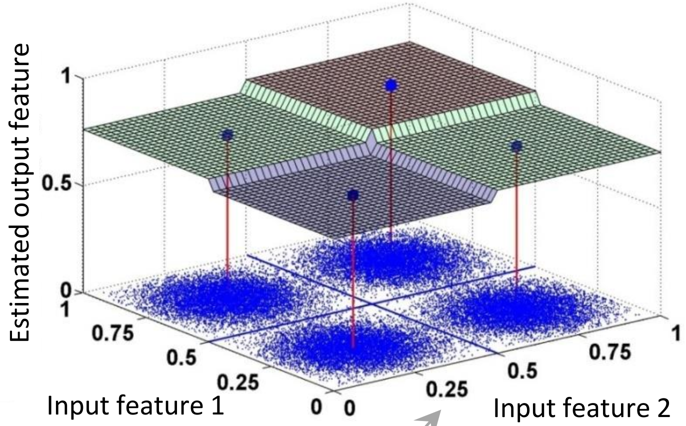
Basic Structure:



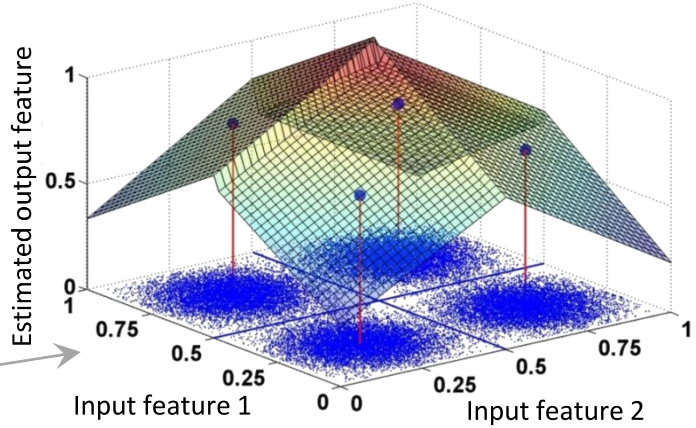
Extension of the Spectral Envelope – Approaches with Codebooks and Linear Mapping

Example for the relation between input and output features





Approximation by codebook pairs



Example for a locally optimized linear mapping

Distance Measure for the Evaluation of the Envelope Estimation Methods – Part 1

Definition of the distance measure:

- First the logarithmic distance between two sampling points of the true (only available in simulations) and the estimated spectral envelope is determined:

$$\Delta\left(e^{j\Omega},n\right) \;=\; 20\log_{10}\left\{\frac{\left|\widehat{P}_{\rm ext}\left(e^{j\Omega},n\right)\right|}{\left|P_{\rm bb}\left(e^{j\Omega},n\right)\right|+\epsilon}\right\}.$$

The positive constant in the denominator prevents division by zero.

- The distance is now weighted (in a nonlinear manner). Taking into account the frequency resolution of the human hear, the ***lower frequencies are weighted larger than the higher frequencies***:

$$\xi\left(e^{j\Omega},n\right) \;=\; \begin{cases} \Delta\left(e^{j\Omega},n\right)\cdot e^{\alpha\Delta\left(e^{j\Omega},n\right)-\beta\Omega}, & \text{if } \Delta\left(e^{j\Omega},n\right)\leq 0\,, \\ \ln\left(-\Delta\left(e^{j\Omega},n\right)+1\right)\cdot e^{-\beta\Omega}, & \text{else.} \end{cases}$$

Distance Measure for the Evaluation of the Envelope Estimation Methods – Part 2

Definition of the distance measure:

- The parameter can be adjusted to user preferences. Typical values are:

$$\begin{aligned}\alpha &= 0.1,\\ \beta &= 0.0005.\end{aligned}$$

- The modified distances are now integrated with the entire frequency range:

$$d_{\mathrm{SDM}}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \xi\left(e^{j\Omega}, n\right) d\Omega$$

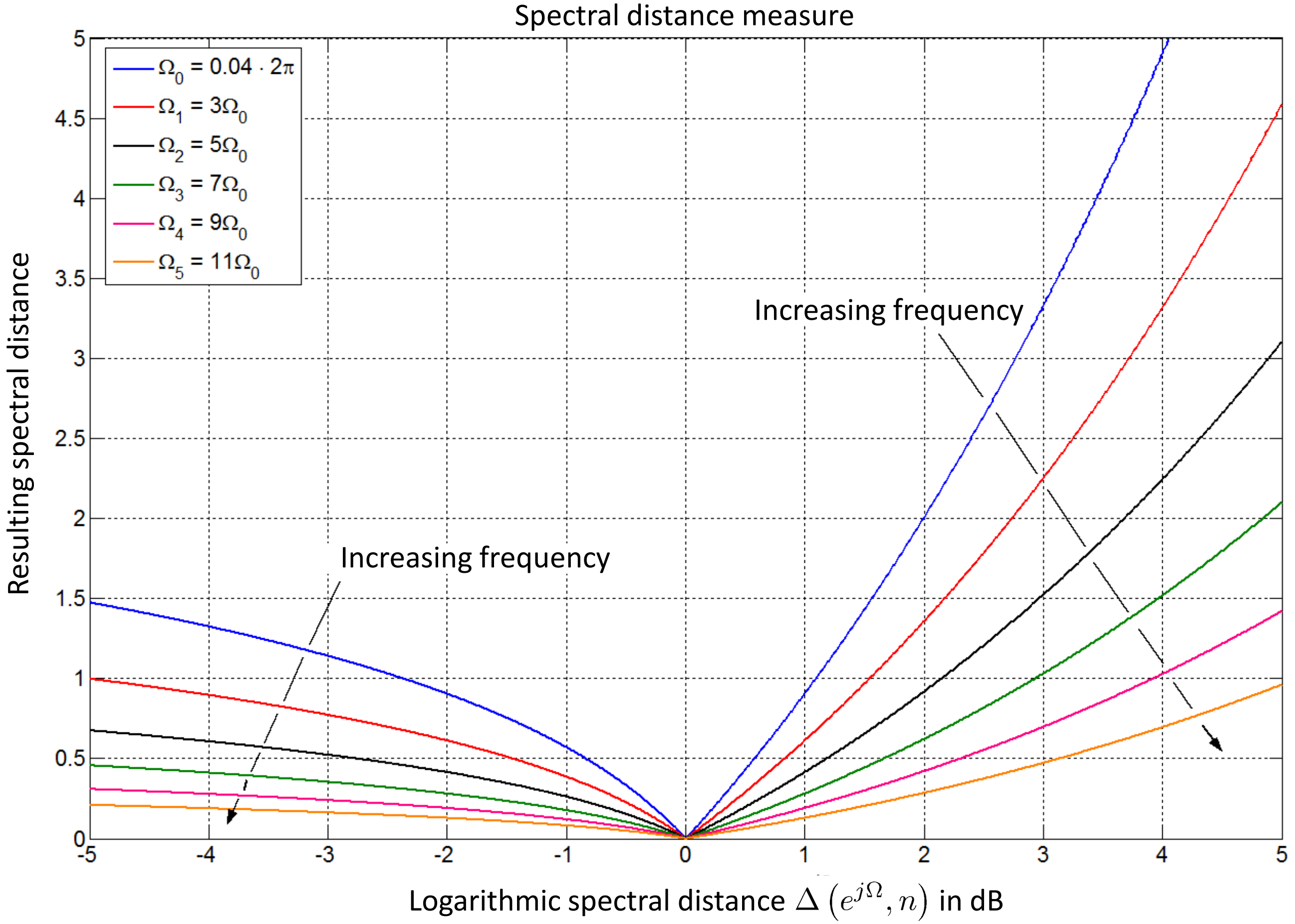
  or as an approximation, summation over a sufficient number of support points be carried out.

- For evaluation, the individual mean distance measure per frame are averaged over all frames:

$$\overline{d}_{\mathrm{SDM}} = \frac{1}{N_{\mathrm{eval}}} \sum_{n=0}^{N_{\mathrm{eval}}-1} d_{\mathrm{SDM}}(n)\,.$$

Distance Measure for the Evaluation of the Envelope Estimation Methods – Part 3

Definition of the distance measure:



Distance Measure for the Evaluation of the Envelope Estimation Methods – Part 4

Measured distance measure:

Codebook size	Only Codebook	Codebook followed by linear mapping
2	38.47	15.36
4	23.66	11.54
8	17.12	9.21
16	14.64	8.71
32	13.30	8.10
64	12.44	7.64
128	11.89	7.38
256	11.41	7.23

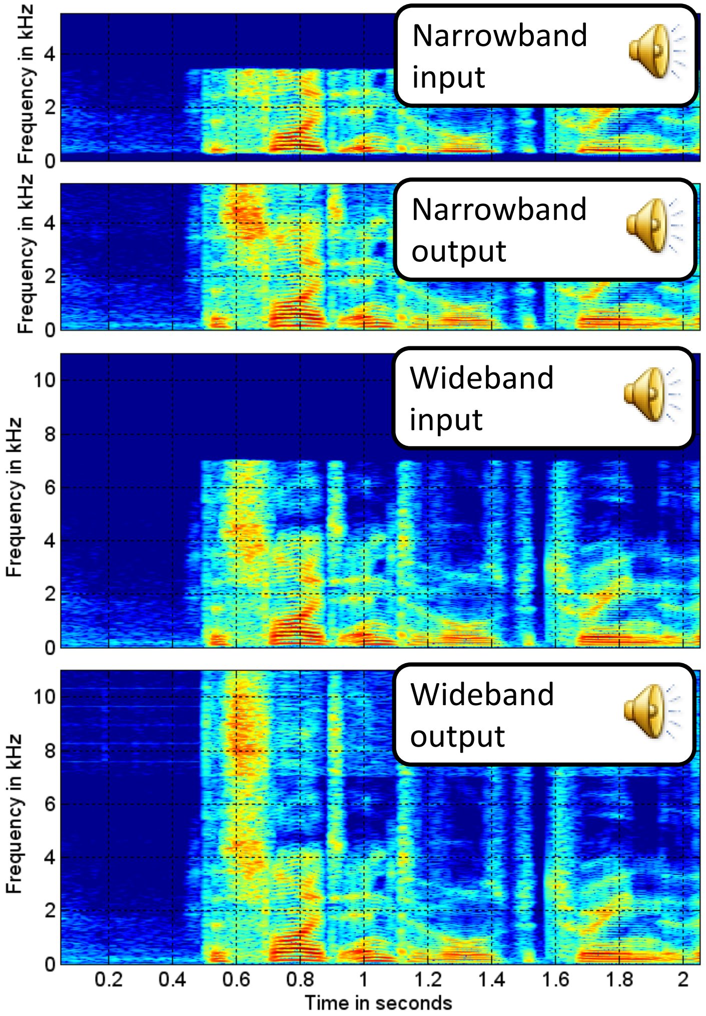
Examples

Narrow band connection:

Bandwidth extension for narrowband telephony (bandwidth 3.4 … 3.8 kHz) – extension of the lower frequencies and higher frequencies up to 5.5 … 8 kHz.

Wideband connection:

Bandwidth extension for wideband telephony (bandwidth 7 kHz, e.g. with the AMR wideband codec G.722.2) – extension of the higher frequency signal portions up to 11 kHz.



Summary and Outlook

Summary:

- Motivation
- System overview
- Extension of the excitation signal
  - Spectral shifting / modulation
  - Non-linear characteristics
- Extension of the spectral envelope
  - Schemes based on neural networks
  - Schemes based on codebooks
  - Schemes based on linear mapping
- Examples

Next week:

- Gaussian mixture models (GMMs)