

Bandwidth Extension

1 Questions

1. What can (not) be improved using bandwidth extension?
2. What are the principles of bandwidth extension based on imaging? What is the reason for sound distortions?
3. What are the principles of bandwidth extension based on linear prediction? What is the benefit of the band-stop filter? Why don't we use the completely synthesized signal $\hat{s}(n)$ as output signal?
4. What problems do appear when bandwidth extension using modulation is applied?
5. Which properties that are relevant for bandwidth extension does the convolution of a line spectrum with itself have? What has to be taken care of when applying bandwidth extension using nonlinearities?
6. Compare the approach based on neural networks with the system concepts on slides 5 and 6. Which system concept is being used? Which block in the system concept corresponds to the neural network? What effects complicate the application of neural networks?
7. In the basic structure, there are two codebooks. For what reasons do these two codebooks contain different types of coefficients? What happens to the selected wide-band codebook vectors $\hat{\mathbf{a}}_{bb}(n)$?

2 Answers

1. An improvement of the sound quality at the receiver can be reached. But usually no improvement of speech intelligibility.
2. Imaging will generate imaging effects by oversampling and filtering (slide 10). Sound distortions are created by the non-natural continuation of the spectrum. For example harmonics and the spectral envelope are not continued correctly.
3. At first remove the narrow band envelope using a predictor error filter. After that upsample and generate an excitation signal. At last apply the wide band envelope using the inverse predictor error filter (slide 17).
In the frequency range of the narrow band signal, only the original signal should be used, because the resynthesized signal $\hat{s}(n)$ does not reach the quality of the original (narrow band) signal.
4. Some of the resulting spectral components are inverted on the frequency axis and have to be removed by using appropriate filtering. Also the spectral gap in the mid-band can lead to distortions (slides 19-21).
5. The harmonics are conserved and are continued in a natural way.
Removal of the unintentionally inserted DC components, adjustment of the output power and reduction of aliasing effects have to be considered (slides 23, 24).
6. Usually, the approach without transmission of side information is used. The neural network corresponds to the block "a priori trained speech models" as well as partly to the block "bandwidth extension". The predictor coefficients that are output by the neural network do not guarantee a stable IIR filter. Furthermore, the reaction of neural networks to unlearned input is uncertain (slide 29).
7. For the recognition of spectral envelopes, cepstral coefficients are suitable, because a distance function can easily be defined (compare part 4, feature extraction). But in order to recreate a spectral envelope, predictor coefficients are necessary. The wide-band codebook vectors feed the inverse prediction error filter (slide 17).