Bandwidth Extension for Speech Enhancement

F. Mustiere, M. Bouchard, M. Bolic

University of Ottawa

Tuesday, May 4th 2010

CCECE 2010: Signal and Multimedia Processing

Outline

1. Context and objectives
2. Proposed solution
3. Results & demos
4. Conclusion

Current Topic

1. Context and objectives
2. Proposed solution
3. Results & demos
4. Conclusion

Context

- Speech enhancement or de-noising now found in several applications (speech transmission, recognition, hearing aids, etc.)
- Noisy speech has frequency dependent SNR
- Higher SNR in lowband (0-5 kHz in this work), lower SNR in highband (5-10 kHz in this work)

Context

- **Speech enhancement in highband**: because of lower SNR, **higher risk of damaging speech** (i.e. distortion) when attempting to remove noise

- Moreover, total complexity of lowband + highband enhancement can be significantly **more costly** than lowband enhancement only

Objectives

To illustrate that a simple **Bandwidth Extension** scheme (BWE, details in next slide) can be both:

- a **competitive** speech enhancement or de-noising tool in the highband (as good as fairly advanced schemes)
- a way to **reduce the complexity** of advanced enhancement schemes, by computing enhancement only in the lowband (less bins or lower model order)

Using BWE could also **allow to use a more complex lowband enhancement scheme**, using the computations freed by the BWE scheme

Bandwidth extension

Some background on classic bandwidth extension:

- production of missing frequency bands with or without additional information
- generic audio bandwidth extension versus source-filter model-based speech bandwidth extension

Main techniques in classical BWE

Excitation signal extension:

- using non-linearities on time sequence
- using spectral shifting or modulation techniques
- using artificial function generators (e.g. harmonic sines)

Spectral envelope extension:

- using codebooks from parameters (LPC, cepstral coeffs.)
- using neural network mapping
- using linear mapping (sometimes combined with codebooks)
- using Bayesian estimation methods (GMMs, HMMs)

Bandwidth extension

- BWE and spectral band replication (SBR) techniques are **found in several speech codecs** (GSM full-rate, AMR-WB, AMR-WB+, G.729EV/G.729.1) **and audio codecs** (MP3pro, Enhanced AACplus, HE-AAC)
- **Different frequency bands** present **different challenges**, e.g. bandwidth extension 300Hz-3.4kHz to 0Hz-5.5kHz is different from 0Hz-11kHz to 0Hz-22kHz
- BWE has received **little attention** in the literature so far as an approach **for speech enhancement or denoising**

Current Topic

1 Context and objectives

2 Proposed solution

3 Results & demos

4 Conclusion

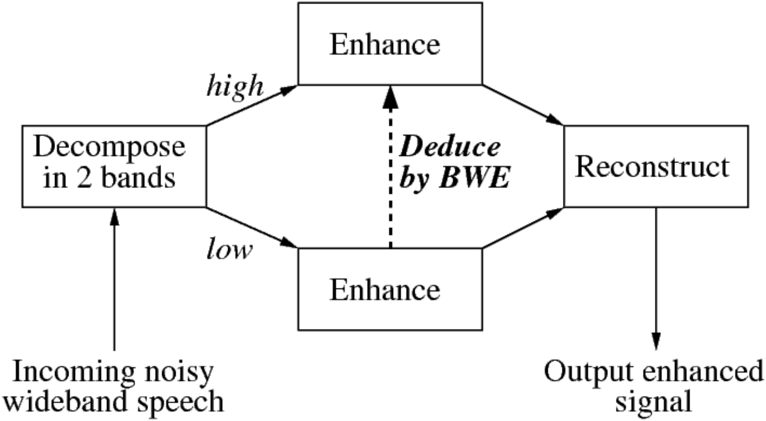
Application of BWE to Speech Enhancement

- In contrast with classical model-based BWE, here **we have access to a coarse envelope estimate**: the noisy signal envelope.
- **In our particular context** (0-10 kHz speech enhancement; SNR > -10 dB; non-synthetic recorded noise) it was found that if a "good" narrowband excitation signal can be obtained and extended, then the **spectral envelope plays a fairly minor role** in the resulting quality.

Application of BWE to Speech Enhancement

- Thus, for **simplicity/efficiency**, in this work **LPC coefficients of noisy fullband spectral envelope are used**:
  - for predicting the enhanced lowband excitation
  - for synthesizing the fullband enhanced signal.
- For the **excitation signal extension**, simple **spectral shifting** is used (spectral band replication, spectral folding).

Application of BWE to Speech Enhancement



Summary of method

1. Obtain analysis/synthesis filter by LPC analysis of *wb* noisy signal $z(k)$
2. From $z(k)$, downsample to *nb* signal $z_n(k)$
3. Enhance downsampled $z_n(k)$, upsample to $\hat{x}_n(k)$
4. Filter $\hat{x}_n(k)$ with analysis filter to get *nb* enhanced excitation $\hat{e}_n(k)$
5. Bandwidth extend $\hat{e}_n(k)$ by modulation to get $\hat{e}_w(k)$
6. Filter $\hat{e}_w(k)$ with synthesis filter to obtain *wb* enhanced speech

Current Topic

1. Context and objectives
2. Proposed solution
3. Results & demos
4. Conclusion

Experimental setup

- **Speech** content from **TIMIT database** (several male and female speakers), upsampled to 20 kHz
- **Noise** from the **NOISEX-92 database** (babble, factory, tank, car), at different levels i.e. SNRs.
- Assessment using a **mixture** of SNR, speech quality and speech intelligibility **objective measures** (SNR, ASNR, CSII, WPESQ, Csig, Cbak, Covl)

Experimental setup

- Assessment of subjective quality using **informal listening tests**
- **Three different fairly advanced speech enhancement algorithms** were used, each **in fullband and BWE modes**: Kalman + EM, multi-band spectral subtractive algorithm, generalized subspace approach.

Results

- In large majority of cases, **objectives measures** results using the **BWE approach** were **better** that those using fullband enhancement, for either low or high input SNR.

- To fully quantify perceptual improvement would require more formal listening tests, but this is not the point here.

- **Informal listening tests** easily confirm that the **BWE approach** is at least **perceptually similar** to the fullband enhancement case, but **at lower cost or complexity**.

Sound demos, for Kalman + EM algorithm, 5 dB input SNR

Noisy	Enhanced *wb*	Enhanced *nb*	Enhanced *nb* + BWE
Factory	$\triangleright$	$\triangleright$	$\triangleright$

Stop

Current Topic

1. Context and objectives
2. Proposed solution
3. Results & demos
4. Conclusion

Conclusion

- Simple BWE-based speech enhancement can reduce complexity of fairly advanced enhancement algorithms, with equivalent quality
- Further quality improvements could likely be obtained by allocating the freed resources on improved narrowband enhancement
- If reduction of complexity is not the main factor, an alternative would be to seek an even better enhancement performance by using a more complex BWE scheme than the one used here

Thank you. Questions?

Frederic Mustiere, Martin Bouchard, Miodrag Bolic
{mustiere,bouchard,mbolic}@site.uottawa.ca