ARTIFICIAL BANDWIDTH EXTENSION USING THE CONSTANT Q TRANSFORM

Pramod Bachhav, Massimiliano Todisco, Nicholas Evans

EURECOM, France

lastname@eurecom.fr

Moctar Mossi, Christophe Beaugeant

INTEL, Sophia Antipolis, France

firstname.lastname@intel.fr

Introduction

- public telephone networks limit the bandwidth of speech signals to 300-3400Hz
- intelligibility for unvoiced phonemes is generally lower than that for voiced phonemes because their spectra extends beyond 3400Hz
- wider bandwidths generally correspond to higher quality speech [1]
- artificial bandwidth extension (ABE) methods estimate missing frequency components to compensate for the consequential loss speech quality and intelligibility
- most ABE algorithms are based either on the classical source-filter model OR employ short time Fourier transform (STFT) for spectral analysis
- the STFT offers a fixed frequency resolution, and is equivalent to a bank of filters with variable Q factors
- however, the human auditory system exhibits constant Q characteristics between 500Hz to 20kHz

Contributions

- Application of the constant Q transform (CQT), a more perceptually motivated approach to spectral analysis, to ABE.

The constant Q transform (CQT)

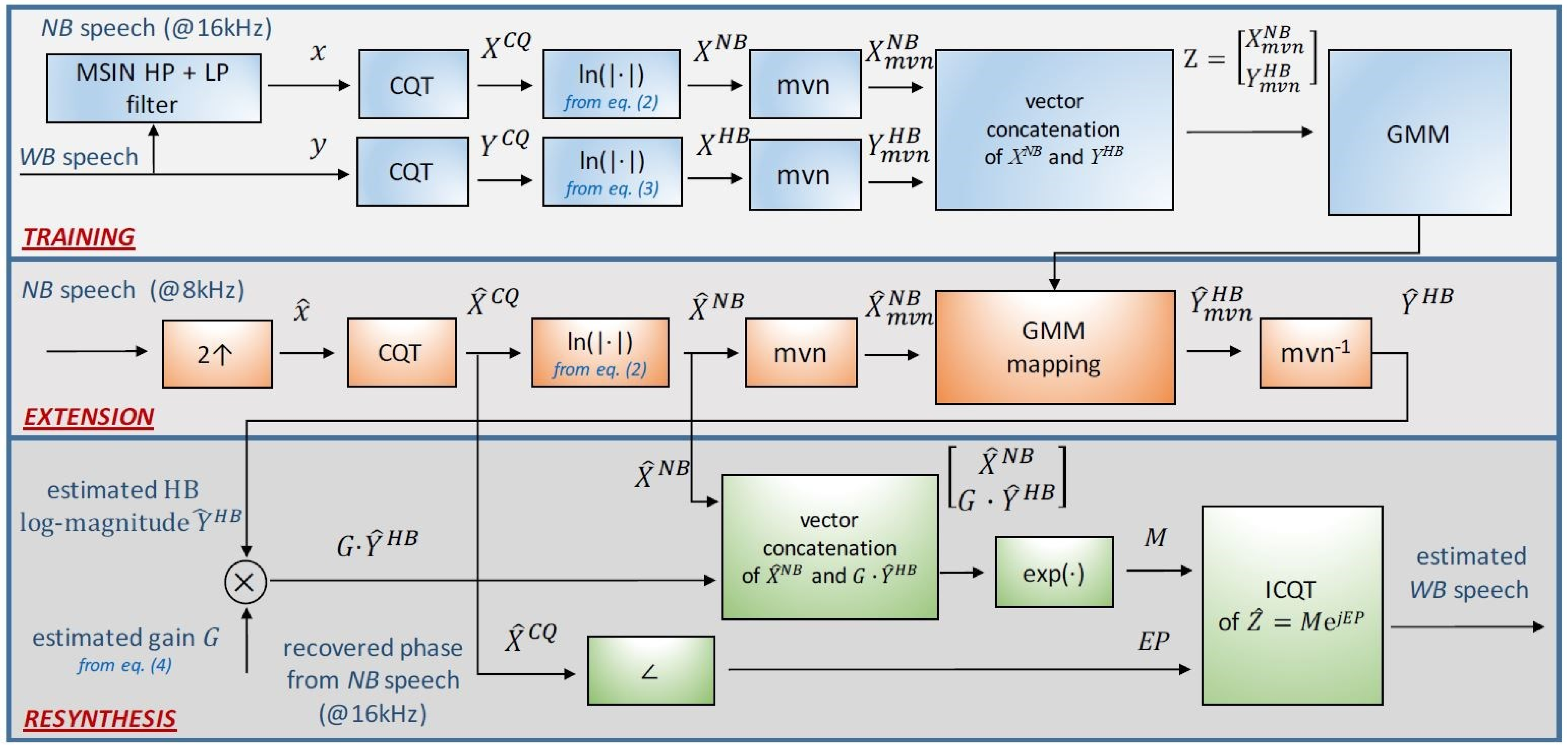
- Uncertainty principle: time and frequency content cannot be measured precisely at the same time
- Q factor is defined as $Q = \frac{f_k}{\delta f}$
- the bandwidth of each STFT filter is constant, whereas the Q factor increases from low to high frequencies
- however, human perception approximates a constant Q transform between 500 Hz and 20 kHz
- CQT: introduced by Youngberg and Boll [2], and refined over the years [3];

$$X^{CQ}(k) = \frac{1}{N_k}\sum_{n<N_k} x(n) w_{N_k}(n) e^{-2\pi i n \frac{f_k}{f_s}}$$

$$N_k = Q\frac{f_s}{f_k} \qquad Q = (2^{\frac{1}{B}} - 1)^{-1}$$

where B is the number of bins per octave

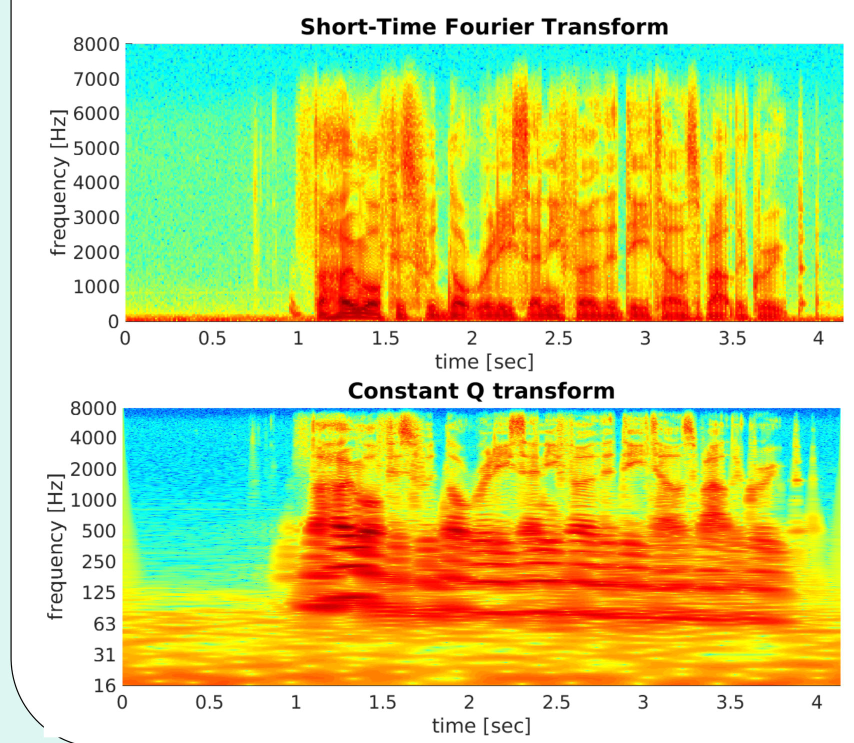
- filter center frequencies are geometrically spaced
- greater frequency resolution for lower frequencies and a greater time resolution for higher frequencies



Block diagram of the CQT-based ABE system.

Experimental Setup

- Database: TSP speech database [4] consisting of 1378 utterances spoken by 12 male and 12 female speakers.
- CQT Parameters: B=48 bins/octave, f_{max} = 8000 Hz, f_1 = 250 Hz
- Mapping: GMM regression (using 512 components)
- input NB (250-3.4kHz) features – 187D, Output HB (3.4-8kHz) features- 52D
- during resynthesis, gain ***G*** corrects the energy of estimated HB which is learned through polynomial regression of order 4.
- whereas phase is copied from HB of upsampled NB CQT signal



Spectrograms of the utterance 'the woman is a star who has grown to love the limelight' for a male speaker in the ASVspoof database. Spectrograms computed with the short-time Fourier Transform (top) and with the constant Q transform (bottom)

Experimental Results

Gain	Phase	Train Mean (σ)	Test Mean (σ)
-	OP	3.01 (0.72)	5.28 (1.51)
-	EP	3.21 (0.71)	5.39 (1.49)
OG	OP	1.89 (0.37)	3.13 (0.67)
OG	EP	2.16 (0.38)	3.30 (0.67)
EG	OP	2.46 (0.40)	4.64 (1.06)
EG	**EP**	2.66 (0.42)	4.77 (1.05)

***RMS-LSD** results (in dB) with and without gain normalization and different phase extensions. OG - oracle gain, EG – estimated gain, OP - oracle phase, EP - estimated phase. EG-EP is the proposed method.*

Comparison B → A	MOS
EG-EP → NB	1.12
OG-EP → NB	1.14
EG-EP → WB	-1.42
OG-EP → WB	-1.03

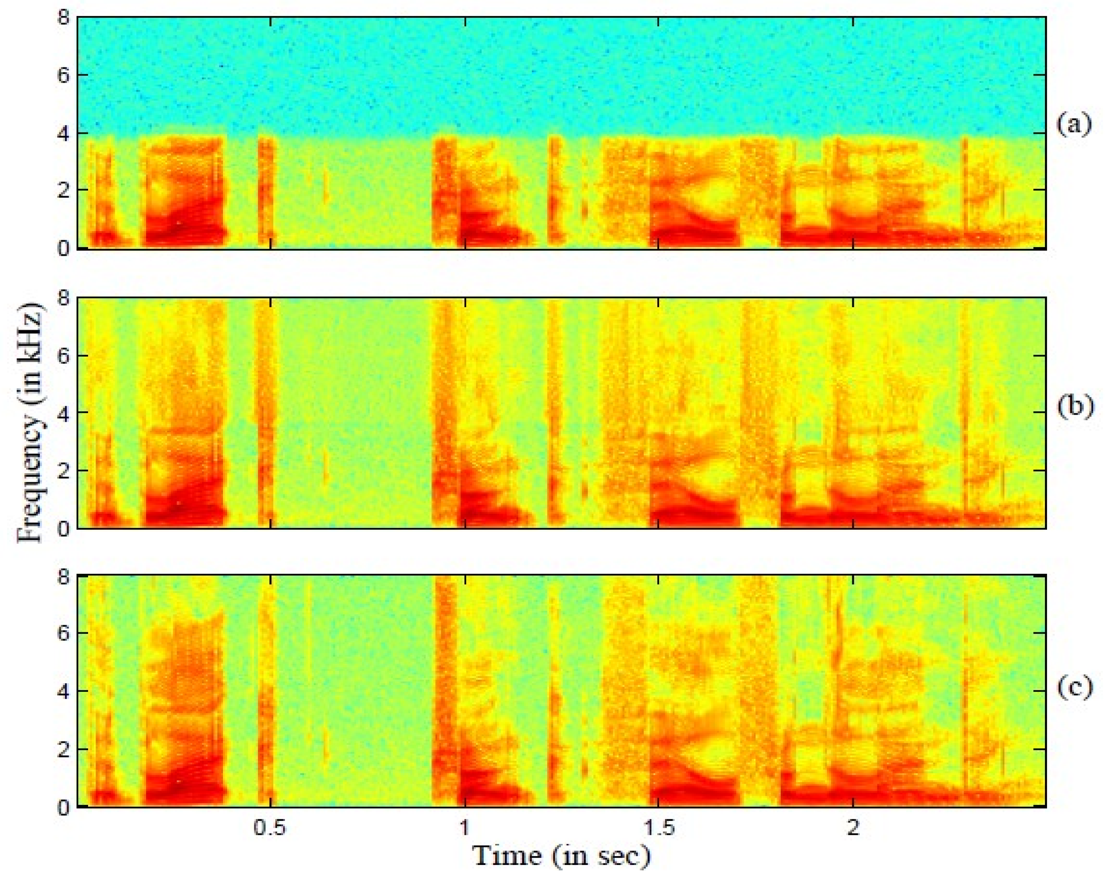
*Comparison based **MOS** for EP with EG and OG. EG-EP is the proposed method (files used for the subjective evaluation are available at **http://audio.eurecom.fr/content/media**)*

Conclusions and Future work

- the CQT is a perceptually motivated approach to time-frequency analysis
- ABE using the CQT produces higher quality, higher bandwidth speech signals using the CQT
- the accurate estimation of the spectral magnitude and gain is critical

Future Work

- analysis and optimisation of the CQT over traditional STFT
- future work will investigate the application of ABE to music signals



Spectrograms of an upsampled NB speech (a), artificially extended WB speech (b) and original WB speech (c).

Acknowledgements

This work was supported with funding from **Intel**.

Selected References

[1] P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech", *Signal Processing*, vol .83 , no. 8, pp. 1707-1719, 2003

[2] J. Youngberg and S. Boll, "Constant-q signal analysis and synthesis," in *Proc. of ICASSP*,1978.

[3] J. Brown, "Calculation of a constant Q spectral transform," *Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, January 1991.

[4] P. Kabal, "TSP Speech Database", McGill University. [Online] Available: http://www-mmsp.ece.mcgill.ca/Documents/Data/.

Presented by Pramod Bachhav at 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2017, New Orleans, USA.