TD/TP en traitement numérique du signal et applications
Niveau : RT4, INSAT
Chargé de cours/TD/TP : R. Amara Boujemâa

Questions à préparer pour l'implémentation de solutions pour le *speech enhancement* de signaux de parole
Référence : les 2 premières pages de l'article "Monaural Speech Enhancement Using a Multi-Branch Temporal Convolutional Network"
auteurs : Qiquan Zhang, Aaron Nicolson, Mingjiang Wang, Kuldip K. Paliwal, and Chenxu Wang

1. What is the problem addressed by the article ? Identify Prof. Kuldip K. Paliwal and his achievements.
2. What is the kind of NN (Neural Networks) that is proposed to replace RNN ones for sequence modelling ?
3. What is the objective of speech enhancement ? Give some applications for which it is designed.
4. Give the conventional methods for monoraul speech enhancement.
5. On which quantity the statistical model-based speech enhancement methods are depending ?
6. Which are the statistical model-based speech enhancement methods cited in the article ? Identify their authors.
7. What informations MLPs are missing when processing audio/speech data ?
8. What is the temporal convolutional network (TCN) ?
9. Write the noisy speech signal $x[n]$ and specify the assumed hypthesis.
10. What does $X[l,k]$ design in the article ? What do indices l and k refer to. Explain the procedure to obtain $X[l,k]$.
    $|X[l,k]|$ is ....
    $|\phi[l,k]|$ is ....

11. What are the *a priori* and *a posteriori* SNR ? Give the general expression for the estimate of the clean speech spectral magnitude.
    It is a multiplicative correction of what quantity ? On what does the estimate of the clean speech depend ?
12. Give the expressions of the gain functions for SWRF, MMSE-STSA and MMSE-LSA.
    On what does the enhancement performance obtained by these 3 estimators depend ?
13. What is the objective of Deep Xi learning framework ?
14. Write a pseudo-algorithm implementing these 3 estimators (you can also give a block-diagram). The algorithm should have as input : the clean audio signal, the noise, the frame length (the same as the number of FFT coefficients, the length of overlap between consecutive analysis frames) and should construct

    - the noisy speech x and its STFT X
    - ξ and γ matrices
    - the estimators' gains $G_{(.)}(\xi, \gamma)$
    - the corrected spectrograms
    - regenerate the temporal clean speech signal