Electrolaryngeal (EL) speech enhancement aims to improve the naturalness and intelligibility of speech produced by Electrolarynx. In this way, people without larynx could retrieve the capability to produce a normal voice. EL speech does not sound like human-produced speech in three ways:

- The sound quality degrades due to the noise caused by the continuous vibration of the EL;
- EL speech sounds unnatural because it is generated by the mechanical excitation signals;
- The intelligibility is limited since the EL produces monotonous speech.

The intelligibility of the converted EL speech is constrained by fundamental frequency (F0) contour, especially in a tone based language like Mandarin. Hence, our proposed framework aims to predict the F0 contour using additional linguistic information.

**Proposed Conversion Framework**

**Training Phase**
Joint density Gaussian model (JDGMM) is used in our systems for estimating the converted Mel Cepstral Coefficients (CVMCC) and F0.

- **GMM Training:**
  \[ P(X_i|y_i) = \sum_{h=1}^{n} \alpha_h N(x_i; \mu_h, \Sigma_h) \]

- **Acoustic model training:**
  1. Train Hidden Markov Model (HMM) – GMM automatic speech recognition system for phoneme recognition;
  2. Based on the alignment produced by HMM-GMM, train neural network acoustic model.

**Conversion Phase**

The converted speech parameters are estimated by employing maximum likelihood estimation upon the trained GMM.

\[ y = \arg\max P(Y|X, J) \]

subject to \( Y = W y \)

The Mel-frequency cepstral coefficient (MFCC) feature sequences are fed into the acoustic model to obtain the phonetic posterior probabilities (PPP). Then the PCA matrix is applied to reduce the dimension of PPP vectors to get LPPP. The F0 contour is estimated by GMM2 given the input LPPP sequences.

**Objective Evaluations**

- **Mel-Cepstral Distortion**
  \[ MCD[db] = 10 \log_{10} \sum_{t=1}^{T} (p_{true} - p_{est})^2 \]

- **Voicing Decision Error**
  \[ VDE = N_{true} - N_{false} \times 100\% \]

- **Gross Pitch Error**
  \[ GPE = N_{false} \times 100\% \]

- **F0 correlation coefficient**

**Evaluation and Result**

- **Subjective Evaluations**
  22 native mandarin speakers scored the converted speech regarding naturalness, intelligibility and similarity.

- **Results**
  - The LPPP system outperforms significantly in comparison to the CVMCC system regarding the F0 contour coefficient.
  - The intelligibility and the similarity of the speech converted by LPPP system outperform those converted by CVMCC system.
  - The phonetic feature rather than acoustic feature is useful in F0 contour estimation for EL speech enhancement.