Voice Conversion Based On Maximum-Likelihood Estimation of Speech Parameter Trajectory

Tomoki Toda, Alan W. Black and Keiichi Tokuda

By:-

Mayank Sirotiya (Y8104036)

Vipul Arora (Y5508)

Under Guidance of:-

Dr. R. Hegde

Introduction:

This paper describes a method for Spectral Estimation for Voice Conversion application, based on Maximum Likelihood Estimation technique. A Gaussian mixture model (GMM) of the joint probability density of source and target features is employed for performing spectral conversion between speakers. The conventional method converts spectral parameters frame by frame based on the minimum mean square error. However there are some problems associated with this method. In order to address those problems, we propose a conversion method based on the maximum-likelihood estimation of a spectral parameter trajectory.

Voice conversion technology enables a user to transform one person's speech pattern into another pattern with distinct characteristics .A mapping function is used which consists of utterance pairs of source and target voices

Applications

1. **Speaker conversion -** This technique can modify nonlinguistic information such as voice characteristics while keeping linguistic information unchanged. The resulting mapping function allows the conversion of any sample of the source into that of the target without any linguistic features such as phoneme transcription.

2. **Cross-language speaker conversion -** This conversion framework can straightforwardly be extended to cross-language speaker conversion, to realize target speakers' voices in various languages by applying the mapping function trained in a certain language into the conversion process in another language.

3. **Modeling of speech production -** Models speech as a combination of a sound source, such as the vocal cords, and a linear acoustic filter, the vocal tract (and radiation characteristic).

4. Narrow-band to wide-band speech for telecommunication

5. Speaking aid, etc.

Classical approaches and their limitations

A. **Codebook mapping based on hard clustering and discrete mapping -** The converted feature vector at frame 't'is determined by quantizing the source feature vector to the nearest centroid vector of the source codebook and substituting it with a corresponding centroid vector of the mapping codebook as follows:

$$\hat{\boldsymbol{y}}_t = \boldsymbol{c}_m^{(y)}$$

Where, $\hat{\boldsymbol{y}}_t$ is the converted feature vector, and $\boldsymbol{c}_m^{(y)}$ is the centroid vector.

Since this method is based on hard clustering, therefore, large quantization error occurs, which is removed by fuzzy vector quantization.

B. **Fuzzy vector quantization, for soft clustering -** Continuous weights for individual clusters are determined at each frame according to the source feature vector. The converted feature vector is defined as a weighted sum of the centroid vectors of the mapping codebook as follows:

$$\hat{\boldsymbol{y}}_t = \sum_{m=1}^{M} w_{m,t}^{(x)} \boldsymbol{c}_m^{(y)}$$

where M is the number of centroid vectors

C. More variable representations of the converted feature vector are achieved by modeling a difference vector between the source and target feature vectors as follows:

$$\hat{\boldsymbol{y}}_t = \boldsymbol{x}_t + \sum_{m=1}^{M} w_{m,t}^{(x)} \left(\boldsymbol{c}_m^{(y)} - \boldsymbol{c}_m^{(x)} \right)$$

In this method, a very strong correlation between those two vectors is assumed.

D. **Linear multivariate regression (LMR) -** This is based on continuous mapping based on hard clustering

$$\hat{\boldsymbol{y}}_t = \boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m$$

where $\boldsymbol{A}_m$ and $\boldsymbol{b}_m$ are regression parameters.

E. **Gaussian mixture model -** This realizes continuous mapping based on soft clustering as follows:

$$\hat{\boldsymbol{y}}_t = \sum_{m=1}^{M} w_{m,t}^{(x)} \left(\boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m \right)$$

This mapping method is reasonably effective.

CONVENTIONAL GMM-BASED MAPPING

Probability Density Function - The joint probability density of the source and target feature vectors is

$$P\left(z_t | \lambda^{(z)}\right) = \sum_{m=1}^{M} w_m N\left(z_t; \mu_m^{(Z)}, \Sigma_m^{(z)}\right)$$

where z_t is a joint vector $\left[\mathbf{x}_t^T, \mathbf{y}_t^T\right]^T$
and the mean vector and covariance matrix are written as

$$\mu_m^{(Z)} = \begin{bmatrix} \mu_m^{(x)} \\ \mu_m^{(y)} \end{bmatrix}, \boldsymbol{\Sigma}_m^{(z)} = \begin{bmatrix} \boldsymbol{\Sigma}_m^{(xx)} & \boldsymbol{\Sigma}_m^{(xy)} \\ \boldsymbol{\Sigma}_m^{(yx)} & \boldsymbol{\Sigma}_m^{(yy)} \end{bmatrix}$$

Mapping function - Conditional probability density can also be represented as

$$P\left(\mathbf{y}_t|\mathbf{x}_t,\lambda^{(z)}\right)=\sum_{m=1}^{M}P\left(m|\mathbf{x}_t,\lambda^{(z)}\right)P\left(\mathbf{y}_t|\mathbf{x}_t,m,\lambda^{(z)}\right)$$

where

$$P\left(m|\mathbf{x}_t,\lambda^{(z)}\right)=\frac{w_m N\left(\mathbf{x}_t;\mu_m^{(x)},\Sigma_m^{(xx)}\right)}{\sum_{m=1}^{M}w_n N\left(\mathbf{x}_t;\mu_n^{(x)},\Sigma_n^{(xx)}\right)}$$

$$P\left(\mathbf{y}_t|\mathbf{x}_t,m,\lambda^{(z)}\right)=N\left(\mathbf{y}_t;\mathrm{E}_{m,t}^{(y)},\mathrm{D}_m^{(y)}\right)$$

The mean vector and the covariance matrix of mth conditional probability distribution are written as

$$\mathrm{E}_{m,t}^{(y)}=\mu_m^{(y)}+\Sigma_m^{(yx)}\Sigma_m^{(xx)-1}\left(\mathbf{x}_t-\mu_m^{(x)}\right)$$

$$\mathrm{D}_m^{(y)}=\Sigma_m^{(yy)}-\Sigma_m^{(yx)}\Sigma_m^{(xx)-1}\Sigma_m^{(xy)}$$

In conventional method the conversion is based on MMSE as follows

$$\begin{aligned}\hat{\boldsymbol{y}}_t&=E[\boldsymbol{y}_t|\boldsymbol{x}_t]\\&=\int P\left(\boldsymbol{y}_t|\boldsymbol{x}_t,\boldsymbol{\lambda}^{(z)}\right)\boldsymbol{y}_t d\boldsymbol{y}_t\\&=\int\sum_{m=1}^{M}P\left(m|\boldsymbol{x}_t,\boldsymbol{\lambda}^{(z)}\right)P\left(\boldsymbol{y}_t|\boldsymbol{x}_t,m,\boldsymbol{\lambda}^{(z)}\right)\boldsymbol{y}_t d\boldsymbol{y}_t\\&=\sum_{m=1}^{M}P\left(m|\boldsymbol{x}_t,\boldsymbol{\lambda}^{(z)}\right)\boldsymbol{E}_{m,t}^{(y)}\end{aligned}$$

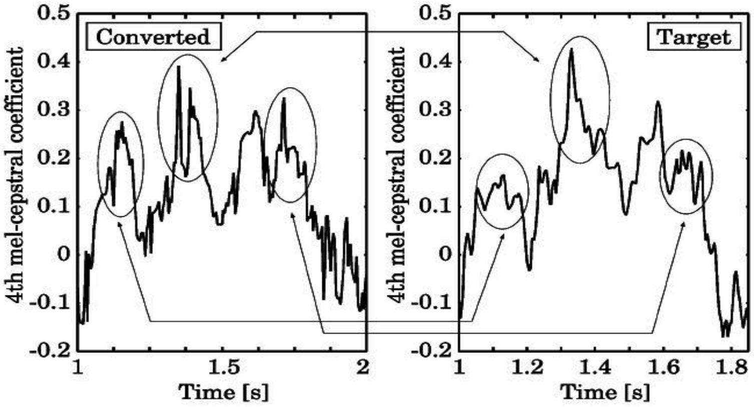
Note that this mapping function has the same form as in our general equation

with $P(m|\boldsymbol{x}_t,\boldsymbol{\lambda}^{(z)})\;=\;w_{m,t}^{(x)},\ \boldsymbol{\Sigma}_m^{(yx)}\boldsymbol{\Sigma}_m^{(xx)^{-1}}\;=\boldsymbol{A}_m$, and $\boldsymbol{\mu}_m^{(y)}-\boldsymbol{\Sigma}_m^{(yx)}\boldsymbol{\Sigma}_m^{(xx)^{-1}}\boldsymbol{\mu}_m^{(x)}=\boldsymbol{b}_m.$

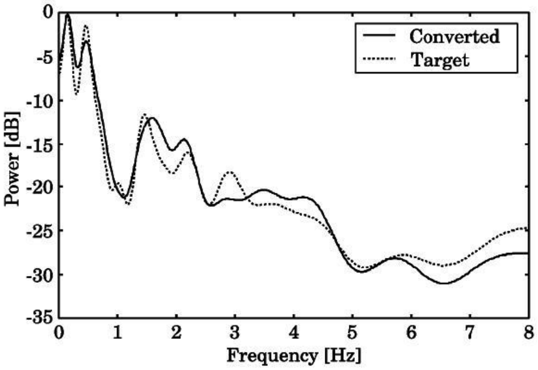
In each mixture component, the conditional target mean vector for the given source feature vector is calculated by a simple linear conversion based on the correlation between the source and target feature vectors problems

Limitations

1. **Time-independent mapping -** Fig. 2 shows an example of the parameter trajectory converted using the GMM-based mapping function and the natural target trajectory. Although these two trajectories seem similar, they sometimes have different local patterns. Such differences are often observed because the correlation of the target feature vectors between frames is ignored in the conventional mapping.



2. **Oversmoothing -** Fig. 3 shows an example of the converted and natural target spectra. We can see that the converted spectrum is excessively smoothed compared with the natural one. This smoothing undoubtedly causes error reduction of the spectral conversion. However, it also causes quality degradation of the converted speech because the removed structures are still necessary for synthesizing high-quality speech.



PROPOSED SPECTRAL CONVERSION:

The paper proposes a trajectory based spectral conversion, instead of conventional frame based one. The estimated static feature vector trajectory is given as,

$$\hat{\boldsymbol{y}} = f(\boldsymbol{x})$$

where,

$$\boldsymbol{x} = \left[\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top, \ldots, \boldsymbol{x}_t^\top, \ldots, \boldsymbol{x}_T^\top\right]^\top$$
$$\boldsymbol{y} = \left[\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top, \ldots, \boldsymbol{y}_t^\top, \ldots, \boldsymbol{y}_T^\top\right]^\top$$

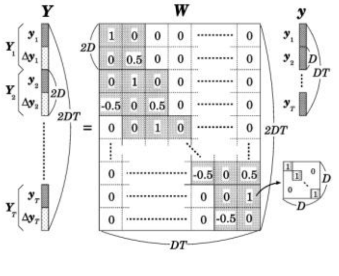
Conversion Considering Dynamic Features:

This method takes into account the feature correlation between frames.

The source and target feature vectors are 2D dimensional

$$\boldsymbol{X} = \left[\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top, \ldots, \boldsymbol{X}_t^\top, \ldots, \boldsymbol{X}_T^\top\right]^\top$$
$$\boldsymbol{Y} = \left[\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top, \ldots, \boldsymbol{Y}_t^\top, \ldots, \boldsymbol{Y}_T^\top\right]^\top$$

where,



$$\boldsymbol{X}_t = \left[\boldsymbol{x}_t^\top, \Delta\boldsymbol{x}_t^\top\right]^\top \text{ and } \boldsymbol{Y}_t = \left[\boldsymbol{y}_t^\top, \Delta\boldsymbol{y}_t^\top\right]^\top$$

The static feature vector sequence and static & dynamic feature vector sequence relate as,

$$\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{y}$$

Maximum Likelihood Estimation of Parameter Trajectory:

The joint vector is written as,

$$Z_t = \left[\boldsymbol{X}_t^\top, \boldsymbol{Y}_t^\top\right]^\top$$

The Gaussian Mixture Model (GMM) for joint probability density is trained using conventional training framework.

$$P(\boldsymbol{Z}_t|\boldsymbol{\lambda}^{(Z)})$$

The likelihood function to be maximized is:

$$P\left(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\lambda}^{(Z)}\right) = \prod_{t=1}^{T}\sum_{m=1}^{M} P\left(m|\boldsymbol{X}_t,\boldsymbol{\lambda}^{(Z)}\right) \times P\left(\boldsymbol{Y}_t|\boldsymbol{X}_t, m, \boldsymbol{\lambda}^{(Z)}\right)$$

where,

$$P\left(m|\boldsymbol{X}_t,\boldsymbol{\lambda}^{(Z)}\right) = \frac{w_m\mathcal{N}\left(\boldsymbol{X}_t;\boldsymbol{\mu}_m^{(X)},\boldsymbol{\Sigma}_m^{(XX)}\right)}{\sum_{n=1}^{M} w_n\mathcal{N}\left(\boldsymbol{X}_t;\boldsymbol{\mu}_n^{(X)},\boldsymbol{\Sigma}_n^{(XX)}\right)}$$

$$P\left(\boldsymbol{Y}_t|\boldsymbol{X}_t, m, \boldsymbol{\lambda}^{(Z)}\right) = \mathcal{N}\left(\boldsymbol{Y}_t;\boldsymbol{E}_{m,t}^{(Y)},\boldsymbol{D}_m^{(Y)}\right)$$

here,

$$\boldsymbol{E}_{m,t}^{(Y)} = \boldsymbol{\mu}_m^{(Y)} + \boldsymbol{\Sigma}_m^{(YX)}\boldsymbol{\Sigma}_m^{(XX)-1}\left(\boldsymbol{X}_t - \boldsymbol{\mu}_m^{(X)}\right)$$

$$\boldsymbol{D}_m^{(Y)} = \boldsymbol{\Sigma}_m^{(YY)} - \boldsymbol{\Sigma}_m^{(YX)}\boldsymbol{\Sigma}_m^{(XX)-1}\boldsymbol{\Sigma}_m^{(XY)}$$

This relation is derived as follows [2].

Derivation of Conditional Probability:

The joint vector with Gaussian distribution $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ is written as,

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}$$

with mean and covariance matrices as,

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$$

For convenience, inverse of covariance matrix, known as Precision matrix, is used,

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix}$$

As inverse of a symmetric matrix is also symmetric, hence, $\boldsymbol{\Lambda}_{aa}$ and $\boldsymbol{\Lambda}_{bb}$ are symmetric, and $\boldsymbol{\Lambda}_{aa} = \boldsymbol{\Lambda}_{bb}^{T}$.

From the product rule of probability, we see that the conditional distribution can be evaluated from the joint distribution **p(x) = p(x$_a$, x$_b$)** simply by fixing **x$_b$** to the observed value and normalizing the resulting expression to obtain a valid probability distribution over **x$_a$**. Instead of performing this normalization explicitly, we can obtain the solution more efficiently by considering the quadratic form,

$$\begin{aligned}
-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \\
-\frac{1}{2}(\mathbf{x}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{\Lambda}_{aa}(\mathbf{x}_a-\boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{x}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b-\boldsymbol{\mu}_b) \\
-\frac{1}{2}(\mathbf{x}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}\boldsymbol{\Lambda}_{ba}(\mathbf{x}_a-\boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{x}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}(\mathbf{x}_b-\boldsymbol{\mu}_b)
\end{aligned}$$

Comparing this with the general Gaussian exponent form

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathrm{const}$$

we obtain, that the second order term is,

$$-\frac{1}{2}\mathbf{x}_a^{\mathrm{T}}\boldsymbol{\Lambda}_{aa}\mathbf{x}_a$$

and the first order term is,

$$\mathbf{x}_a^{\mathrm{T}}\left\{\boldsymbol{\Lambda}_{aa}\boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{ab}(\mathbf{x}_b-\boldsymbol{\mu}_b)\right\}$$

Hence, we obtain the conditional mean and variance as,

$$\begin{aligned}
\boldsymbol{\mu}_{a|b} &= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b-\boldsymbol{\mu}_b) \\
\boldsymbol{\Sigma}_{a|b} &= \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}.
\end{aligned}$$

EM Algorithm:

$\hat{\boldsymbol{y}} = \arg\max P\left(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\lambda}^{(Z)}\right)$

This direct maximization is very complex, hence EM algorithm is used, in which the following auxiliary function is iteratively maximized with respect to $\hat{\mathbf{y}}$,

$$\begin{aligned}
Q(\boldsymbol{Y},\hat{\boldsymbol{Y}}) &= \sum_{\text{all }\boldsymbol{m}} P\left(\boldsymbol{m}|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\lambda}^{(Z)}\right)\log P\left(\hat{\boldsymbol{Y}},\boldsymbol{m}|\boldsymbol{X},\boldsymbol{\lambda}^{(Z)}\right) \\
&= \sum_{t=1}^{T}\sum_{m=1}^{M} P\left(m|\boldsymbol{X}_t,\boldsymbol{Y}_t,\boldsymbol{\lambda}^{(Z)}\right)\log P\left(\hat{\boldsymbol{Y}}_t,m|\boldsymbol{X}_t,\boldsymbol{\lambda}^{(Z)}\right) \\
&= \sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{m,t}\left(-\frac{1}{2}\hat{\boldsymbol{Y}}_t^{\top}\boldsymbol{D}_m^{(Y)^{-1}}\hat{\boldsymbol{Y}}_t + \hat{\boldsymbol{Y}}_t^{\top}\boldsymbol{D}_m^{(Y)^{-1}}\boldsymbol{E}_{m,t}^{(Y)}\right) + \overline{K} \\
&= \sum_{t=1}^{T} -\frac{1}{2}\hat{\boldsymbol{Y}}_t^{\top}\overline{\boldsymbol{D}_t^{(Y)^{-1}}}\hat{\boldsymbol{Y}}_t + \hat{\boldsymbol{Y}}_t^{\top}\overline{\boldsymbol{D}_t^{(Y)^{-1}}\boldsymbol{E}_t^{(Y)}} + \overline{K} \\
&= -\frac{1}{2}\hat{\boldsymbol{Y}}^{\top}\overline{\boldsymbol{D}^{(Y)^{-1}}}\hat{\boldsymbol{Y}} + \hat{\boldsymbol{Y}}^{\top}\overline{\boldsymbol{D}^{(Y)^{-1}}\boldsymbol{E}^{(Y)}} + \overline{K} \\
&= -\frac{1}{2}\hat{\boldsymbol{y}}^{\top}\boldsymbol{W}^{\top}\overline{\boldsymbol{D}^{(Y)^{-1}}}\boldsymbol{W}\hat{\boldsymbol{y}} + \hat{\boldsymbol{y}}^{\top}\boldsymbol{W}^{\top}\overline{\boldsymbol{D}^{(Y)^{-1}}\boldsymbol{E}^{(Y)}} + \overline{K}
\end{aligned}$$

where,

$$\begin{aligned}
\overline{\boldsymbol{D}^{(Y)^{-1}}} &= \mathrm{diag}\left[\overline{\boldsymbol{D}_1^{(Y)^{-1}}},\overline{\boldsymbol{D}_2^{(Y)^{-1}}},\ldots,\right. \\
&\qquad \left.\overline{\boldsymbol{D}_t^{(Y)^{-1}}},\ldots\overline{\boldsymbol{D}_T^{(Y)^{-1}}}\right] \\
\overline{\boldsymbol{D}^{(Y)^{-1}}\boldsymbol{E}^{(Y)}} &= \left[\overline{\boldsymbol{D}_1^{(Y)^{-1}}\boldsymbol{E}_1^{(Y)}}^{\top},\overline{\boldsymbol{D}_2^{(Y)^{-1}}\boldsymbol{E}_2^{(Y)}}^{\top},\ldots,\right. \\
&\qquad \left.\overline{\boldsymbol{D}_t^{(Y)^{-1}}\boldsymbol{E}_t^{(Y)}}^{\top},\ldots,\overline{\boldsymbol{D}_T^{(Y)^{-1}}\boldsymbol{E}_T^{(Y)}}^{\top}\right]^{\top}
\end{aligned}$$

$$\begin{aligned}
\overline{\boldsymbol{D}_t^{(Y)^{-1}}} &= \sum_{m=1}^{M}\gamma_{m,t}\boldsymbol{D}_m^{(Y)^{-1}} \\
\overline{\boldsymbol{D}_t^{(Y)^{-1}}\boldsymbol{E}_t^{(Y)}} &= \sum_{m=1}^{M}\gamma_{m,t}\boldsymbol{D}_m^{(Y)^{-1}}\boldsymbol{E}_{m,t}^{(Y)} \\
\gamma_{m,t} &= P(m|\boldsymbol{X}_t,\boldsymbol{Y}_t,\boldsymbol{\lambda}^{(Z)}).
\end{aligned}$$

The derivative of auxiliary function with respect to $\hat{\mathbf{y}}$ is,

$$\frac{\partial Q(\boldsymbol{Y},\hat{\boldsymbol{Y}})}{\partial\boldsymbol{y}} = -\boldsymbol{W}^{\top}\overline{\boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}}}\boldsymbol{W}\boldsymbol{y} + \boldsymbol{W}^{\top}\overline{\boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}}\boldsymbol{E}_{\hat{\boldsymbol{m}}}^{(Y)}}$$

Equating it to zero, we get,

$$\hat{\boldsymbol{y}} = \left(\boldsymbol{W}^\top \overline{\boldsymbol{D}^{(Y)^{-1}}} \boldsymbol{W}\right)^{-1} \boldsymbol{W}^\top \overline{\boldsymbol{D}^{(Y)^{-1}} \boldsymbol{E}^{(Y)}}$$

In the expectation step, we calculate $\gamma_{m,t}$ and in the maximization step, we estimate $\hat{\mathbf{y}}$.

Approximation with suboptimum mixture sequence:

The log likelihood function is approximated with a single mixture component sequence as

$$\mathcal{L} = \log P\left(\hat{\boldsymbol{m}} | \boldsymbol{X}, \boldsymbol{\lambda}^{(Z)}\right) P\left(\boldsymbol{Y} | \boldsymbol{X}, \hat{\boldsymbol{m}}, \boldsymbol{\lambda}^{(Z)}\right)$$

where,

$$\hat{\boldsymbol{m}} = \operatorname{argmax} P\left(\boldsymbol{m} | \boldsymbol{X}, \boldsymbol{\lambda}^{(Z)}\right)$$

This log likelihood function is easy to maximize by direct differentiation. Hence, we obtain,

$$\hat{\boldsymbol{y}} = \left(\boldsymbol{W}^\top \boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} \boldsymbol{W}\right)^{-1} \boldsymbol{W}^\top \boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} \boldsymbol{E}_{\hat{\boldsymbol{m}}}^{(Y)}$$

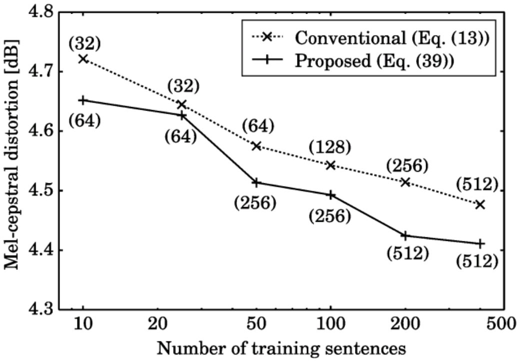
where,

$$\begin{aligned} \boldsymbol{E}_{\hat{\boldsymbol{m}}}^{(Y)} &= \Big[\boldsymbol{E}_{\hat{m}_1,1}^{(Y)}, \boldsymbol{E}_{\hat{m}_2,2}^{(Y)}, \ldots, \\ &\qquad \boldsymbol{E}_{\hat{m}_t,t}^{(Y)}, \ldots, \boldsymbol{E}_{\hat{m}_T,T}^{(Y)}\Big] \\ \boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} &= \operatorname{diag}\Big[\boldsymbol{D}_{\hat{m}_1}^{(Y)^{-1}}, \boldsymbol{D}_{\hat{m}_2}^{(Y)^{-1}}, \ldots, \\ &\qquad \boldsymbol{D}_{\hat{m}_t}^{(Y)^{-1}}, \ldots, \boldsymbol{D}_{\hat{m}_T}^{(Y)^{-1}}\Big] \end{aligned}$$

The approximated solution reduces the computational cost. The preliminary experiments demonstrated that there were no large differences in the results obtained by EM algorithm and the approximated method.

Results:

The evaluation using mel-cepstral distortions show that the proposed method is quite better than the conventional method. This is because the proposed method takes into account the interframe correlation for parameter trajectory estimation, whereas conventional method ignores it.



Summary:

In this paper, we mainly studied following concepts:

- GMM based Feature mapping
- Conditional Gaussian Distributions
- Maximum Likelihood technique for GMMs
- Expectation Maximization Algorithm