PLDA with Two Sources of Inter-session Variability

Jesús Villalba

Communications Technology Group (GTC),
Aragon Institute for Engineering Research (I3A),
University of Zaragoza, Spain
villalba@unizar.es

Nov 6, 2012

1 The Model

1.1 PLDA

We take a linear-Gaussian generative model $M$. We suppose that we have i-vectors of the same conversations recorded simultaneously by different channels or different noisy conditions. Then, an i-vector $\phi_{ijl}$ of speaker $i$, session $j$ recorded in a channel $l$ can be written as:

$$\phi_{ijl} = \mu + V y_i + U x_{ij} + \epsilon_{ijl}$$  (1)

where $\mu$ is speaker independent term, $V$ is the eigenvoices matrix, $y_i$ is the speaker factor vector, $U$ is an the eigenchannels matrix, $x_{ij}$ and $\epsilon_{ijl}$ is a channel offset. The term $x_{ij}$ must be the same for all the recordings of the same conversation. The term $\epsilon_{ijl}$ accounts for the channel variability.

We assume the following priors for the variables:

$$y \sim \mathcal{N}(y | 0, I)$$  (2)
$$x \sim \mathcal{N}(x | 0, I)$$  (3)
$$\epsilon \sim \mathcal{N}(\epsilon | 0, W^{-1})$$  (4)

where $\mathcal{N}$ denotes a Gaussian distribution; and $D$ is a full rank precision matrix. $\phi$ is an observable variable and $y$ and $x$ are hidden variables.

1.2 Notation

We are going to introduce some notation:

- Let $\Phi_d$ be the development i-vectors dataset.
- Let $\Phi_t = \{l, r\}$ be the test i-vectors.
- Let $\Phi$ be any of the previous datasets.
- Let $\theta_d$ be the labelling of the development dataset. It partitions the $N_d$ i-vectors into $M_d$ speakers. Each speaker has $H_i$ sessions and each session can be recorded by $L_{ij}$ different channels.
- Let $\theta_t$ be the labelling of the test set, so that $\theta_t \in \{T, N\}$, where $T$ is the hypothesis that $l$ and $r$ belong to the same speaker and $N$ is the hypothesis that they belong to different speakers.
- Let $\theta$ be any of the previous labellings.
- Let $\Phi_i$ be the i-vectors belonging to the speaker $i$. 

• Let $Y_d$ be the speaker identity variables of the development set. We will have as many identity variables as speakers.
• Let $Y_t$ be the speaker identity variables of the test set.
• Let $Y$ be any of the previous speaker identity variables sets.
• Let $X_d$ be the channel variables of the development set.
• Let $X_t$ be the channel variables of the test set.
• Let $X$ be any of the previous channel variables sets.
• Let $X_i = [x_{i1}, x_{i2}, \ldots, x_{iH}]$ be the channel variables of speaker $i$.
• Let $\mathcal{M} = (\mu, V, U, D)$ be the set of all the model parameters.

2 Likelihood calculations

2.1 Definitions

We define the sufficient statistics for speaker $i$. The zero-order statistic is the number of observations of speaker $i$ $N_i$. The first-order and second-order statistics are

$$F_i = \sum_{j=1}^{H_i} \sum_{l=1}^{L_{ij}} \phi_{ijl}$$  \hspace{1cm} (5) \\
$$S_i = \sum_{j=1}^{H_i} \sum_{l=1}^{L_{ij}} (\phi_{ijl} - \mu)(\phi_{ijl} - \mu)^T$$  \hspace{1cm} (6)

We define the centered statistics as

$$\bar{F}_i = F_i - N_i \mu$$  \hspace{1cm} (7) \\
$$\bar{S}_i = S_i - \mu F_i^T F_i - N_i \mu \mu^T$$  \hspace{1cm} (8)

We define the session statistics as

$$F_{ij} = \sum_{l=1}^{L_{ij}} \phi_{ijl}$$  \hspace{1cm} (9) \\
$$\bar{F}_{ij} = F_{ij} - L_{ij} \mu$$  \hspace{1cm} (10)

where $L_{ij}$ is the number of channels for the conversation $ij$.

We define the global statistics

$$N = \sum_{i=1}^{M} N_i$$  \hspace{1cm} (11) \\
$$F = \sum_{i=1}^{M} F_i$$  \hspace{1cm} (12) \\
$$\bar{F} = \sum_{i=1}^{M} \bar{F}_i$$  \hspace{1cm} (13) \\
$$S = \sum_{i=1}^{M} S_i$$  \hspace{1cm} (14) \\
$$\bar{S} = \sum_{i=1}^{M} \bar{S}_i$$  \hspace{1cm} (15)
2.2 Data conditional likelihood

The likelihood of the data given the hidden variables for speaker $i$ is

$$\ln P(\Phi_i|y_i, X_i, M) = \sum_{j=1}^{H_i} \sum_{l=1}^{L_{ij}} \ln \mathcal{N}(\phi_{ijl}|\mu + V y_i + U x_{ij}, W^{-1})$$

$$= \frac{N_i}{2} \ln \left| \frac{W}{2\pi} \right| - \frac{1}{2} \sum_{j=1}^{H_i} \sum_{l=1}^{L_{ij}} (\phi_{ijl} - \mu - V y_i - U x_{ij})^T W (\phi_{ijl} - \mu - V y_i - U x_{ij})$$

$$= \frac{N_i}{2} \ln \left| \frac{W}{2\pi} \right| - \frac{1}{2} \text{tr} \left( W S_i \right) + y_i^T V^T W F_i - \frac{N_i}{2} y_i^T V^T W V y_i$$

$$+ \sum_{j=1}^{H_i} x_{ij}^T U^T W F_i y_i - L_{ij} y_i^T V^T W U x_{ij} - \frac{1}{2} L_{ij} x_{ij}^T U^T W U x_{ij}$$

We can write this likelihood in other form if we define:

$$\tilde{y}_{ij} = \begin{bmatrix} y_i \\ x_{ij} \\ 1 \end{bmatrix}, \quad \tilde{V} = [V \ U \ \mu]$$

Then

$$\ln P(\Phi_i|y_i, X_i, M) = \sum_{j=1}^{H_i} \sum_{l=1}^{L_{ij}} \ln \mathcal{N}(\phi_{ijl}|\tilde{V} \tilde{y}_{ij}, W^{-1})$$

$$= \frac{N_i}{2} \ln \left| \frac{W}{2\pi} \right| - \frac{1}{2} \sum_{j=1}^{H_i} \sum_{l=1}^{L_{ij}} (\phi_{ijl} - \tilde{V} \tilde{y}_{ij})^T W (\phi_{ijl} - \tilde{V} \tilde{y}_{ij})$$

$$= \frac{N_i}{2} \ln \left| \frac{W}{2\pi} \right| - \frac{1}{2} \text{tr} \left( W S_i + \sum_{j=1}^{H_i} \tilde{y}_{ij}^T \tilde{V}^T W F_i \tilde{y}_{ij} \right)$$

$$= \frac{N_i}{2} \ln \left| \frac{W}{2\pi} \right| - \frac{1}{2} \text{tr} \left( W \left( S_i + \sum_{j=1}^{H_i} -2F_i \tilde{y}_{ij}^T \tilde{V}^T + L_{ij} \tilde{V} \tilde{y}_{ij} \tilde{V}^T \right) \right)$$

2.3 Posterior of the hidden variables

The posterior of the hidden variables can be decomposed into two factors:

$$P(y_i, X_i|\Phi_i, M) = P(X_i|y_i, \Phi_i, M) P(y_i|\Phi_i, M)$$

2.3.1 Conditional posterior of $X_i$

The conditional posterior of $X_i$ is

$$P(X_i|y_i, \Phi_i, M) = \frac{P(\Phi_i|y_i, X_i, M) P(X_i)}{P(\Phi_i|y_i, M)}$$
Using equations (3) and (18),

\[
\ln P(X_i|y_i, \Phi_i, M) = \ln P(\Phi_i|y_i, X_i, M) + \ln P(X_i|M) + \text{const}
\]

(26)

\[
= \sum_{j=1}^{H_i} x_{ij}^T U^T W F_{ij} - L_{ij} y^T V^T W U x_{ij} - \frac{1}{2} L_{ij} x_{ij}^T U^T W U x_{ij} - \frac{1}{2} x_{ij}^T x_{ij} + \text{const}
\]

(27)

\[
= \sum_{j=1}^{H_i} x_{ij}^T U^T W (F_{ij} - L_{ij} V y_i) - \frac{1}{2} x_{ij}^T L_{ij} x_{ij} + \text{const}
\]

(28)

\[
= \sum_{j=1}^{H_i} x_{ij}^T \zeta_{ij} - \frac{1}{2} x_{ij}^T L_{x_{ij}} x_{ij} + \text{const}
\]

(29)

where

\[
\zeta_{ij} = U^T W (F_{ij} - L_{ij} V y_i) = \tilde{\zeta}_{ij} - L_{ij} y_i
\]

(30)

\[
\tilde{\zeta}_{ij} = U^T W F_{ij}
\]

(31)

\[
J = U^T W V
\]

(32)

\[
L_{x_{ij}} = I + L_{ij} U^T W U
\]

(33)

Equation (29) has the form of a product of Gaussian distributions. Therefore

\[
P(X_i|y_i, \Phi_i, M) = \prod_{j=1}^{H_i} N \left( x_{ij} | \overline{x}_{ij}, L_{x_{ij}}^{-1} \right)
\]

(34)

where

\[
\overline{x}_{ij} = L_{x_{ij}}^{-1} \zeta_{ij}
\]

(35)

2.3.2 Posterior of \(y_i\)

The marginal posterior of \(y\) is

\[
P(y|\Phi_i, M) = \frac{P(\Phi_i|y_i, M) P(y) P(\Phi_i|M)}{P(\Phi_i|y_i, M) P(X_i|\Phi_i, y_i, M) P(\Phi_i|y_i, M)}
\]

(36)

We can use Bayes Theorem to write

\[
P(\Phi_i, X_i|y_i, M) = P(\Phi_i|y_i, X_i, M) P(X_i|y_i, M) = P(X_i|\Phi_i, y_i, M) P(\Phi_i|y_i, M)
\]

(37)

Simplifying

\[
P(\Phi_i|y_i, X_i, M) P(X_i) = P(X_i|\Phi_i, y_i, M) P(\Phi_i|y_i, M)
\]

(38)

Then

\[
P(y|\Phi_i, M) = \frac{P(\Phi_i|y_i, X_i, M) P(X_i) P(y)}{P(X_i|\Phi_i, y_i, M) P(\Phi_i|M)} \bigg|_{X_i=0}
\]

(39)

Using equations (2), (18) and (33)
\[ \begin{aligned}
\ln P(y_i|\Phi_i, M) &= \ln P(\Phi_i|y_i, X_i, M) + \ln P(y_i) - \ln P(X_i|\Phi_i, y_i, M) + \text{const} \\
&= y^T V^T WF_i - \frac{N_i}{2} y_i^T V^T W y_i - \frac{1}{2} y_i^T y_i + \frac{1}{2} \sum_{i=1}^{H_i} \bar{\bar{X}}_{x_i}^T L_{x_i} \bar{\bar{x}}_{x_i} + \text{const} \\
&= y^T V^T WF_i - \frac{1}{2} y_i^T (I + N_i V^T W V) y_i \\
&\quad + \frac{1}{2} \sum_{i=1}^{H_i} (F_{ij} - L_{ij} y_i) L_{x_{ij}}^{-1} U^T W (F_{ij} - L_{ij} y_i) + \text{const} \\
&= y^T V^T WF_i - \frac{1}{2} y_i^T (I + N_i V^T W V) y_i \\
&\quad + \frac{1}{2} \sum_{i=1}^{H_i} F_{ij}^T W L_{x_{ij}}^{-1} U^T W F_{ij} \\
&\quad - 2 L_{ij} y_i^T V^T U L_{x_{ij}}^{-1} U^T W F_{ij} \\
&\quad + L_{ij}^2 y_i^T V^T W L_{x_{ij}}^{-1} U^T W y_i + \text{const} \\
&= y^T V^T \left( WF_i - \sum_{i=1}^{H_i} L_{ij} W L_{x_{ij}}^{-1} U^T W F_{ij} \right) \\
&\quad - \frac{1}{2} y_i^T \left( I + V^T \left( N_i W - \sum_{i=1}^{H_i} L_{ij} W L_{x_{ij}}^{-1} U^T W \right) V \right) y_i + \text{const}
\end{aligned} \]

Then

\[ P(y_i|\Phi_i, M) = N(y_i|\bar{\bar{y}}_i, L_{y_i}^{-1}) \]

where

\[
L_{y_i} = I + V^T \left( N_i W - \sum_{i=1}^{H_i} L_{ij} W L_{x_{ij}}^{-1} U^T W \right) V
\]

\[ = I + N_i V^T W W - \sum_{i=1}^{H_i} L_{ij}^T J^T L_{x_{ij}}^{-1} J \]

\[ \gamma_i = V^T \left( WF_i - \sum_{i=1}^{H_i} L_{ij} W L_{x_{ij}}^{-1} U^T W F_{ij} \right) = \tilde{\gamma}_i - \sum_{i=1}^{H_i} L_{ij} J^T L_{x_{ij}}^{-1} \tilde{\xi}_{ij} \]

\[ \bar{\gamma}_i = V^T W F_i \]

\[ \bar{\gamma}_i = L_{y_i}^{-1} \gamma_i \]

### 2.4 Marginal likelihood of the data

The marginal likelihood of the data is

\[ P(\Phi_i|M) = \frac{P(\Phi_i|y_i, X_i, M) P(y_i) P(X_i)}{P(X_i|\Phi_i, y_i, M) P(y_i|\Phi_i, M)} \bigg|_{y_i=0, X_i=0} \]

Taking equations (18), (2), (3), (45) and (52)

\[
\ln P(\Phi_i|M) = \frac{N_i}{2} \ln \left| \frac{W}{2\pi} \right| - \frac{1}{2} \text{tr} (W \bar{\bar{S}}_i) - \frac{1}{2} \sum_{i=1}^{H_i} \ln |L_{x_{ij}}| + \frac{1}{2} \sum_{i=1}^{H_i} \tilde{\xi}_{ij}^T L_{x_{ij}}^{-1} \tilde{\xi}_{ij} - \frac{1}{2} \ln |L_{y_i}| + \frac{1}{2} \gamma_i^T L_{y_i}^{-1} \gamma_i
\]
3 EM algorithm

3.1 E-step
In the E-step we calculate the posterior of $y$ and $X$ with equation (24)

$$Q(M) = \sum_{i=1}^{M} E_{Y,X} [\ln P(\Phi_i, y_i, X_i | M)]$$

Taking equation (23)

$$Q(M) = \frac{N}{2} \ln |W| - \frac{1}{2} \text{tr} \left( W \left( S + \sum_{i=1}^{M} \sum_{j=1}^{H_i} -2F_{ij} E_{Y,X} [\tilde{y}_{ij}]^T \tilde{V}^T + L_{ij} \tilde{V} E_{Y,X} [\tilde{y}_{ij}]^T \tilde{V}^T \right) \right)$$

we define

$$R_{\tilde{y}} = \sum_{i=1}^{M} \sum_{j=1}^{H_i} L_{ij} E_{Y} [\tilde{y}_{ij}]$$

$$C = \sum_{i=1}^{M} \sum_{j=1}^{H_i} F_{ij} E_{Y} [\tilde{y}_{ij}]^T$$

then

$$Q(M) = \frac{N}{2} \ln |W| - \frac{1}{2} \text{tr} \left( W \left( S - 2C\tilde{V}^T + \tilde{V}R_{\tilde{y}} \tilde{V}^T + \tilde{V} \tilde{V}^T \right) \right) + \text{const}$$

$$\frac{\partial Q(M)}{\partial \tilde{V}} = C - \tilde{V}R_{\tilde{y}} = 0 \implies \tilde{V} = CR_{\tilde{y}}^{-1}$$

$$\frac{\partial Q(M)}{\partial W} = \frac{N}{2} \left( 2W^{-1} - \text{diag}(W^{-1}) \right) - \frac{1}{2} (K + K^T - \text{diag}(K)) = 0$$

where $K = S - 2C\tilde{V}^T + \tilde{V}R_{\tilde{y}} \tilde{V}^T$, so

$$W^{-1} = \frac{1}{N} \left( \frac{K + K^T}{2} \right)$$

$$= \frac{1}{N} \left( S - \tilde{V}C\tilde{V}^T + \tilde{V}R_{\tilde{y}} \tilde{V}^T \right)$$

Finally, we need to evaluate the expectations $E_{Y} [\tilde{y}_{ij}]$ and $E_{Y} [\tilde{y}_{ij}\tilde{y}_{ij}^T]$ and compute $R_{\tilde{y}}$ and $C$.

$$C = \sum_{i=1}^{M} \sum_{j=1}^{H_i} F_{ij} E_{Y,X} [\tilde{y}_{ij}]^T = \sum_{i=1}^{M} \sum_{j=1}^{H_i} F_{ij} \left[ \frac{E_{Y,X}[y_i]}{1} \right]^T = [C_y \ C_x \ F]$$
We assume a more general prior for the hidden variables:

\[ P(y_i) = \mathcal{N}(y_i | \mu_y, \Lambda_y^{-1}) \] (79)

\[ P(x_{ij}|y_i) = \mathcal{N}(x_{ij}|H y_i + \mu_x, \Lambda_{xy}^{-1}) \] (80)

3.3 M-step MD

We assume a more general prior for the hidden variables:
To minimize the divergence we maximize

$$Q(\mu_y, \Lambda_y, H, \mu_x, \Lambda_x) = \sum_{i=1}^{M} E_Y \left[ \ln \mathcal{N} \left( y_i | \mu_y, \Lambda_y^{-1} \right) \right] + \sum_{i=1}^{H_i} E_{Y,X} \left[ \ln \mathcal{N} \left( x_{ij} | H y_i + \mu_x, \Lambda_x^{-1} \right) \right]$$

(81)

$$= -\frac{M}{2} \ln |\Lambda_y| - \frac{1}{2} \text{tr} \left( \Lambda_y \sum_{i=1}^{M} E_Y \left[ (y_i - \mu_y) (y_i - \mu_y)^T \right] \right)$$

$$+ \frac{H}{2} \ln |\Lambda_x| - \frac{1}{2} \text{tr} \left( \Lambda_x \sum_{i=1}^{H_i} \sum_{j=1}^{H_i} E_{Y,X} \left[ (x_{ij} - H y_i - \mu_x) (x_{ij} - H y_i - \mu_x)^T \right] \right)$$

+ const

(82)

$$\frac{\partial Q(\mu_y, \Lambda_y, H, \mu_x, \Lambda_x)}{\partial \mu_y} = \frac{1}{2} \sum_{i=1}^{M} \Lambda_y E_Y [y_i - \mu_y] = 0 \implies \mu_y = \frac{1}{M} \sum_{i=1}^{M} E_Y [y_i]$$

(83)

$$\mu_x = \left( \frac{1}{H} \sum_{i=1}^{H_i} \sum_{j=1}^{H_i} E_X [x_{ij}] - H \sum_{i=1}^{M} E_Y [y_i] \right) \frac{1}{H} (P_{x1} - H P_{y1})$$

(92)

where

$$S = \sum_{i=1}^{M} E_Y \left[ (y_i - \mu_y) (y_i - \mu_y)^T \right],$$

so

$$\Sigma_y = \Lambda_y^{-1}$$

(86)

$$= \frac{1}{M} \sum_{i=1}^{M} E_Y \left[ (y_i - \mu_y) (y_i - \mu_y)^T \right]$$

(87)

$$= \frac{1}{M} \sum_{i=1}^{M} E_Y [y_i y_i^T] - \mu_y E_Y [y_i] y_i^T - E_Y [y_i] \mu_y^T + \mu_y \mu_y^T$$

(88)

$$= \frac{1}{M} \sum_{i=1}^{M} E_Y [y_i y_i^T] - \mu_y \mu_y^T$$

(89)

$$\frac{\partial Q(\mu_y, \Lambda_y, H, \mu_x, \Lambda_x)}{\partial \mu_x} = \Lambda_x \sum_{i=1}^{M} \sum_{j=1}^{H_i} E_{Y,X} [x_{ij} - H y_i - \mu_x] = 0 \implies \mu_x = \left( \frac{1}{H} \sum_{i=1}^{H_i} \sum_{j=1}^{H_i} E_X [x_{ij}] - H \sum_{i=1}^{M} E_Y [y_i] \right) \frac{1}{H} (P_{x1} - H P_{y1})$$

(90)

$$= \left( \frac{1}{H} \sum_{i=1}^{H_i} \sum_{j=1}^{H_i} E_X [x_{ij}] - H \sum_{i=1}^{M} E_Y [y_i] \right) \frac{1}{H} (P_{x1} - H P_{y1})$$

(91)

where

$$P_{x1} = \sum_{i=1}^{M} \sum_{j=1}^{H_i} E_X [x_{ij}]$$

(93)

$$P_{y1} = \sum_{i=1}^{M} E_Y [y_i]$$

(94)
\[
\frac{\partial Q(\mu_y, \Lambda_y, H, \mu_x, \Lambda_x)}{\partial H} = \Lambda_y \sum_{i=1}^{M} \sum_{j=1}^{H_i} E_{Y, X} \left[ (x_{ij} - Hy_i - \mu_x) y_i^T \right] = 0
\] (95)

\[
\Rightarrow P_{xy} - HP_y - \mu_x P_{y1} = 0
\] (96)

\[
\Rightarrow P_{xy} - HP_y - \frac{1}{H} (P_{x1} - HP_{y1}) P_{y1} = 0
\] (97)

\[
\Rightarrow P_{xy} - \frac{1}{H} P_{x1} P_{y1} - H \left( P_y - \frac{1}{H} P_{y1} P_{y1} \right) = 0 \quad \Rightarrow
\] (98)

\[
H = \left( P_{xy} - \frac{1}{H} P_{x1} P_{y1} \right) \left( P_y - \frac{1}{H} P_{y1} P_{y1} \right)^{-1}
\] (99)

where

\[
P_{xy} = \sum_{i=1}^{M} \sum_{j=1}^{H_i} E_{Y, X} \left[ x_{ij} y_i^T \right]
\] (100)

\[
P_y = \sum_{i=1}^{M} H_i E_Y \left[ y_i y_i^T \right]
\] (101)

\[
P_x = \sum_{i=1}^{M} \sum_{j=1}^{H_i} E_X \left[ x_{ij} x_{ij}^T \right]
\] (102)

\[
\frac{\partial Q(\mu_y, \Lambda_y, H, \mu_x, \Lambda_x)}{\partial \Lambda_x} = \frac{H}{2} \left( 2\Lambda_x^{-1} - \text{diag}(\Lambda_x^{-1}) \right) - \frac{1}{2} \left( 2S - \text{diag}(S) \right) = 0
\] (103)

where \( S = \sum_{i=1}^{M} \sum_{j=1}^{H_i} E_{Y, X} \left[ (x_{ij} - Hy_i - \mu_x) (x_{ij} - Hy_i - \mu_x)^T \right] \), so

\[
\Sigma_x = \Lambda_x^{-1}
\] (104)

\[
\frac{1}{H} \left( P_x - P_{xy} H^T - HP_{x1}^T - P_{x1} \mu_x^T - \mu_x P_{x1}^T + HP_y H^T \right)
\]

\[
+ HP_y \mu_x^T + \mu_x P_{y1}^T + H \mu_x \mu_x^T
\] (105)

\[
\frac{1}{H} \left( P_x - P_{xy} H^T - HP_{x1}^T - HP_y H^T \right)
\]

\[- (P_{x1} - HP_{y1}) \mu_x^T - \mu_x (P_{x1} - HP_{y1})^T + H \mu_x \mu_x^T
\] (106)

\[
\frac{1}{H} \left( P_x - P_{xy} H^T - HP_{x1}^T - HP_y H^T - (P_{x1} - HP_{y1}) \mu_x^T \right)
\] (107)

The transform \((y, x) = \phi(y', x')\) such as \(y'\) and \(x'\) has a standard prior is

\[
y = \mu_y + \left( \Sigma_y^{1/2} \right) y'
\] (108)

\[
x = \mu_x + Hy + \left( \Sigma_x^{1/2} \right)^T x'
\] (109)

\[
= \mu_x + H \mu_y + H \left( \Sigma_y^{1/2} \right)^T y' + \left( \Sigma_x^{1/2} \right)^T x'
\] (110)

We can transform \(\mu, V\) and \(U\) using that transform

\[
U' = U \left( \Sigma_x^{-1/2} \right)
\] (111)

\[
V' = (V + UH) \left( \Sigma_y^{-1/2} \right)
\] (112)

\[
\mu' = \mu + (V + UH) \mu_y + U \mu_x
\] (113)
3.4 Objective function

The EM objective function is equation (52) summed for all speakers

\[
\ln P(\Phi|M) = \frac{N}{2} \ln \left| \frac{W}{2\pi} \right| - \frac{1}{2} \text{tr} \left( WS \right) - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{H_i} \ln |L_{x_{ij}}| + \frac{1}{2} \sum_{i=1}^{M} \frac{H_i}{2} \sum_{j=1}^{H_i} \xi_{ij}^T L_{x_{ij}}^{-1} \xi_{ij} \\
- \frac{1}{2} \sum_{i=1}^{M} \ln |L_{y_i}| + \frac{1}{2} \sum_{i=1}^{M} \gamma_i^T L_{y_i}^{-1} \gamma_i
\]  

(114)

4 Likelihood ratio

Given a model $M$ we can calculate the ratio of the posterior probabilities of target and non target as shown in [1]:

\[
\frac{P(T|\Phi_t, M, \pi)}{P(N|\Phi_t, M, \pi)} = \frac{P_T P(\Phi_t|T, M)}{P_N P(\Phi_t|N, M)} = \frac{P_T}{P_N} R(\Phi_t, M)
\]  

(115)

where we have defined the plug-in likelihood ratio $R(\Phi_t, M)$. To get this ratio we need to calculate $P(\Phi|\theta, M)$. Given a model $M$, the $y_1, y_2, \ldots, y_M \in Y$ are sampled independently from $P(y|M)$. Besides, given the $M$ and a speaker $i$ the set $\Phi_i$ of i-vectors produced by that speaker is drawn independently from $P(\Phi_i|y_i, M)$. Using these independence assumptions we can write:

\[
P(\Phi|\theta, M) = \prod_{i=1}^{M} P(\Phi_i|M)
\]  

(116)

\[
P(\Phi_i|y, M) = \prod_{\phi \in \Phi_i} P(\phi|y, M)
\]  

(117)

Then, the likelihood of $\Phi$ is

\[
P(\Phi|\theta, M) = \prod_{i=1}^{M} P(\Phi_i|y_0, M) P(y_0|M) = K(\Phi)L(\theta|\Phi)
\]  

(118)

where $K(\Phi) = \prod_{i=1}^{N} P(\phi_j|y_0, M)$ is a term that only dependent on the dataset, not $\theta$, so it vanishes when doing the ratio and we do not need to calculate it. What we need to calculate is:

\[
L(\theta|\Phi) = \prod_{i=1}^{M} Q(\Phi_i)
\]  

(119)

\[
Q(\Phi_i) = \frac{P(y_0|M)}{P(y_0|\Phi_i, M)}
\]  

(120)

and the likelihood ratio is:

\[
R(\Phi_t, M) = \frac{Q(\{1, r\})}{Q(\{r\}) Q(\{\ell\})}
\]  

(121)

Making $y_0 = 0$ we can use [39], [2] to calculate $Q(\Phi)$

\[
\ln Q(\Phi_i) = \frac{1}{2} \left( - \ln |L_{y_i}| + \gamma_i^T L_{y_i}^{-1} \gamma_i \right)
\]  

(122)

Given a set of training observations $\Phi_1$ of a speaker 1 with statistics $N_1$ and $\tilde{F}_1$; and a set of test observations $\Phi_2$ of a speaker 2 with statistics $N_2$ and $\tilde{F}_2$. To test if the speakers 1 and 2 are the same speaker the log-likelihood ratio is

\[
\ln R(\Phi_t, M) = \frac{1}{2} \left( - \ln |L_3| + \gamma_3^T L_3^{-1} \gamma_3 + \ln |L_1| - \gamma_1^T L_1^{-1} \gamma_1 + \ln |L_2| - \gamma_2^T L_2^{-1} \gamma_2 \right)
\]  

(123)
where

\( P(\mathbf{y}|\Phi_1, M) = \mathcal{N}(\mathbf{y}|\gamma_1\mathbf{L}_1^{-1}, \mathbf{L}_1^{-1}) \) \hspace{1cm} (124)

\( P(\mathbf{y}|\Phi_2, M) = \mathcal{N}(\mathbf{y}|\gamma_2\mathbf{L}_2^{-1}, \mathbf{L}_2^{-1}) \) \hspace{1cm} (125)

\( P(\mathbf{y}|\Phi_1, \Phi_2, M) = \mathcal{N}(\mathbf{y}|\gamma_3\mathbf{L}_3^{-1}, \mathbf{L}_3^{-1}) \) \hspace{1cm} (126)

\( \gamma_3 = \gamma_1 + \gamma_2 \):

\[
\ln R(\Phi_t, M) = \frac{1}{2} \left( \ln |\mathbf{L}_1| + \ln |\mathbf{L}_2| - \ln |\mathbf{L}_3| + 2\gamma_1^T\mathbf{L}_3^{-1}\gamma_2 + \gamma_1^T(\mathbf{L}_3^{-1} - \mathbf{L}_1^{-1})\gamma_1 + \gamma_2^T(\mathbf{L}_3^{-1} - \mathbf{L}_2^{-1})\gamma_2 \right) \hspace{1cm} (128)
\]

**References**

[1] Niko Brummer and Edward De Villiers, “The Speaker Partitioning Problem,” in *Oddyssey Speaker and Language Recognition Workshop*, Brno, Czech Republic, 2010.