5 NORMALIZACIÓN DE PUNTUACIONES

En los capítulos anteriores, en diferentes momentos, se ha enfatizado en el hecho de que la voz de una persona se caracteriza por tener un alto grado de variabilidad. En función de la edad, del estado anímico y de muchos otros factores, dos segmentos de voz pertenecientes a un mismo locutor pueden tener características completamente diferentes (“*Intra-Speaker Variabilities”)*. Además, las condiciones de entrenamiento y de evaluación pueden ser distintas, especialmente en conversaciones telefónicas, donde el tipo de auricular, canal de transmisión o altavoz juegan un papel fundamental. Todos estos inconvenientes se intentaron corregir trabajando a nivel de parámetros (coeficientes cepstrales) mediante CMVN. Incluso a nivel de modelos, en NAP, por ejemplo, se llevaba a cabo una ligera modificación del kernel para compensar la posible distorsión producida por el desajuste entre los segmentos de entrenamiento y evaluación (distorsión de canal). Sin embargo, en cualquier área de reconocimiento de voz, y en general, de reconocimiento biométrico, existe una tercera opción de contrarrestar los efectos negativos mencionados anteriormente, que consiste en normalizar las puntuaciones obtenidas durante la fase de test. Los dos tipos de normalización de puntuación más populares y eficientes son:

- **Znorm (zero normalisation)** [Li and Porter, 1988]. En este caso lo que se hace es enfrentar el modelo del locutor que hemos generado en la fase de entrenamiento a un conjunto de segmentos de evaluación pertenecientes a una cohorte de impostores (“*non-target*”) donde obtendremos un conjunto de puntuaciones asociadas a cada uno de los correspondientes impostores. A estas puntuaciones les calculamos su media y desviación típica, representadas por μ_Z y σ_Z, respectivamente. A la puntuación que había sido obtenida previamente enfrentando el modelo del locutor con su correspondiente fichero de evaluación se le resta la medía y después dividimos por la desviación típica:

$$S_{\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Y}) = \frac{S(\boldsymbol{X},\boldsymbol{Y}) - \mu_Z}{\sigma_Z}, \tag{5.1}$$

  con $S(\boldsymbol{X},\boldsymbol{Y})$ representando la puntuación obtenida entre el fichero de entrenamiento $\boldsymbol{X}$ y de test $\boldsymbol{Y}$. $S_Z(\boldsymbol{X},\boldsymbol{Y})$ representa la puntuación después de haber llevado a cabo el proceso de Znorm.

  Este tipo de normalización tiene la ventaja principal de que no tiene necesariamente que llevarse a cabo en la fase de evaluación (*“offline”*), con lo cual se puede realizar durante la etapa de entrenamiento, convirtiéndose en una normalización ideal para una aplicación real, ya que no ralentiza la etapa del test en la que las exigencias temporales pueden ser relativamente importantes.

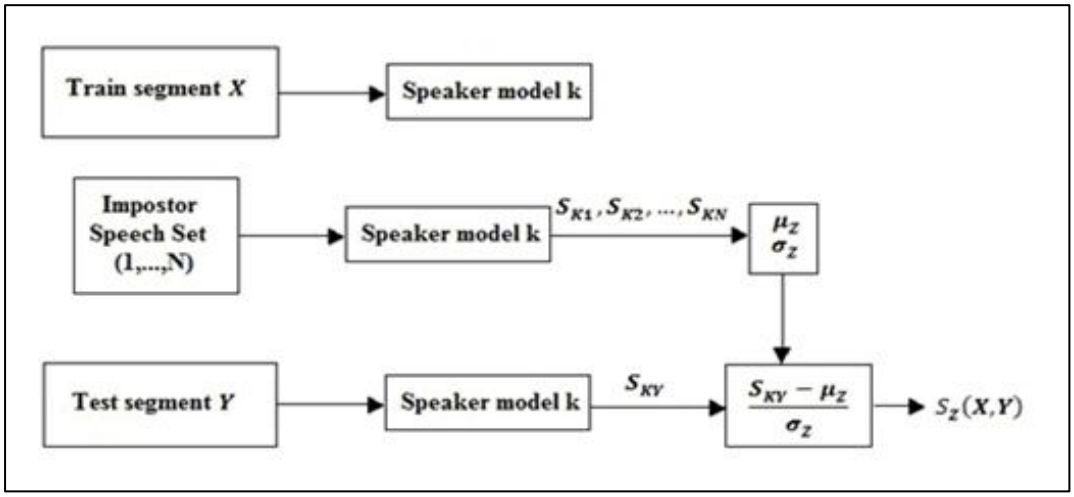

Figura 31. *Estructura de la normalización Znorm.*

- **Tnorm (test normalization)** [Auckenthaler, R., Carey, M., Lloyd-Thomas, H.]. Nuevamente se normaliza el resultado obtenido con un locutor con los resultados obtenidos por una cohorte de impostores. Pero en este caso, para obtener la media (μ_T) y la varianza (σ_T) de los impostores, lo que se hace es enfrentar el mismo segmento de test del interlocutor que estamos evaluando a un conjunto de modelos que han sido creados mediante entrenamiento de impostores, representando la normalización de la siguiente forma:

$$S_T(\boldsymbol{X},\boldsymbol{Y}) = \frac{S(\boldsymbol{X},\boldsymbol{Y}) - \mu_T}{\sigma_T}. \qquad (5.2)$$

Tal y como podemos observar en la siguiente figura, Tnorm tiene el inconveniente del alto coste computacional (debido a que tiene que crear previamente N modelos para impostores) y que la obtención de las puntuaciones para la cohorte de impostores debe realizarse paralelamente en la fase de evaluación ("*online*") convirtiéndola en una normalización no muy aconsejable para una aplicación real en la que la verificación deba hacerse en pocos segundos.

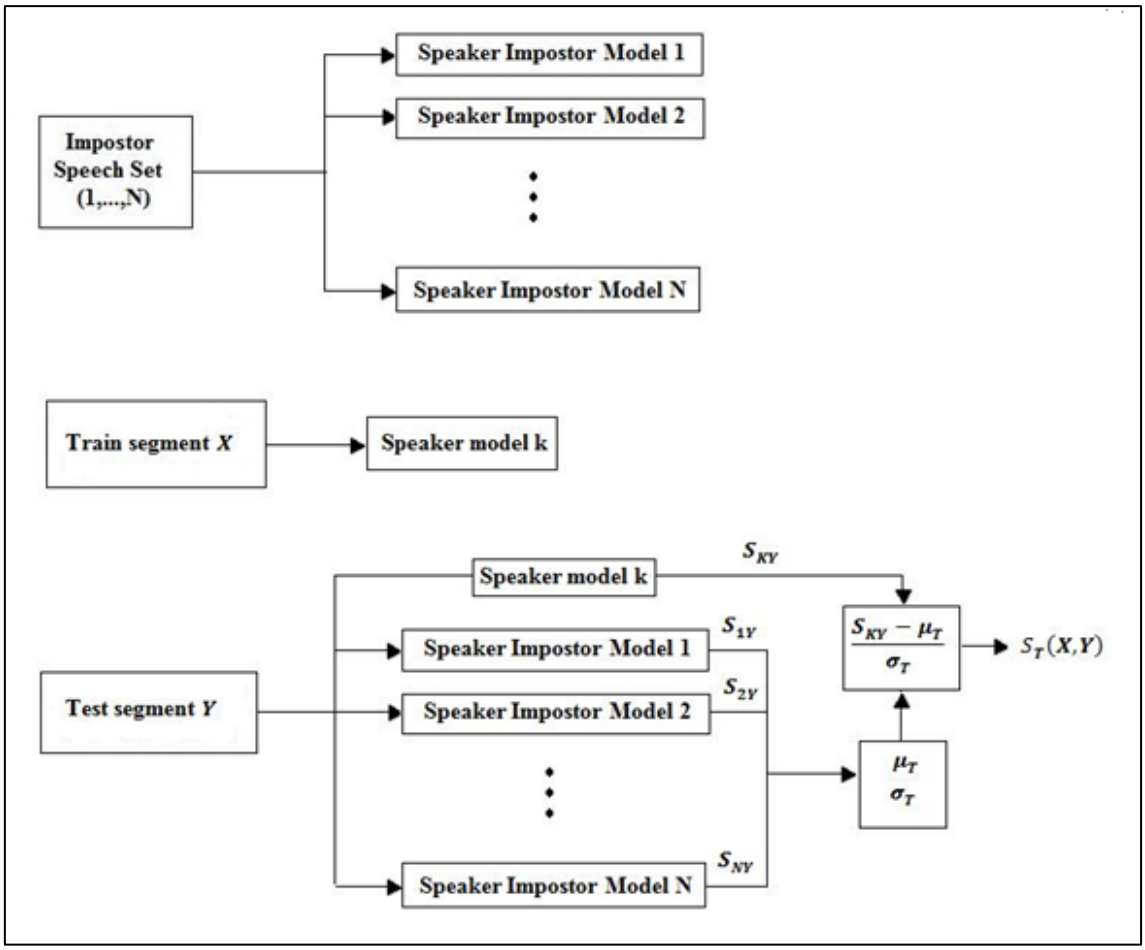

Figura 32. *Estructura de la normalización Tnorm.*

Por otra parte, hay otra forma de mejorar los resultados de las puntuaciones que es desarrollar en serie los dos tipos de normalizaciones anteriores, dando lugar a dos nuevas posibilidades, que en el proyecto las denominaremos como ZTnorm y TZnorm, donde la diferencia entre ambos es si primero se hace la normalización T y después la Z o viceversa. En los resultados de las simulaciones de este proyecto podremos verificar como dependiendo del orden en el que fusionemos las dos técnicas anteriores, obtendremos unos resultados diferentes. Al fusionarlas, intentamos que las posibles carencias que pueda tener una técnica puedan ser subsanadas con otra y viceversa.