to $\boldsymbol{\theta}_n$, only it is with respect to the channel, where $\boldsymbol{\theta}_n$ pertains to the speaker. $\mathbf{x}_{n,c}$ has two indices since it is dependent on the recording from speaker n with channel c. JFA basically solves the combination of Equations 16.2 and 16.3 simultaneously, trying to separate speaker and channel effects.

16.4.3 Total Factors (Total Variability)

In Section 12.5 we presented a full treatment of *factor analysis*. In that section, Equation 12.54 basically states that the common factors and the specific factors are uncorrelated. However, [16] postulates that the *mean vector* associated with the channel variability, $\boldsymbol{\mu}_c$, still contains some speaker-related information, by showing, empirically, that speakers may still be somewhat identified using this information.

For this reason, *Dehak, et al.* [16] propose recombining the *speaker variability* and the *channel variability spaces* back into one, essentially ending with,

$$\mathbf{y}_n = \boldsymbol{\mu} + \mathbf{V}\boldsymbol{\theta}_n \tag{16.4}$$

They call $\boldsymbol{\theta}_n$, *total variability*, and the space associated with it, the *total variability space* [14]. However, as we discussed in detail, in Section 12.5, they are basically reverting back to *PCA*, since they are removing the residual term which is one of the main factors that differentiates factor analysis from PCA. Nevertheless, aside from the problem with the terminology, it is perfectly fine to use PCA techniques for performing speaker recognition. In later incarnations of their work [15, 45, 18], *Dehak et al.* have called these vectors *i-vectors*. They have used these vectors in conjunction with support vector machines, employing the *Cosine kernel* (Section 15.4.4) as well as others.

Speaker factor coefficients are related to the speaker coordinates, in which each speaker is represented as a point. This space is defined by the *Eigenvoice matrix*. These speaker factor vectors are relatively short, having in the order of about 300 elements [19], which makes them desirable for use with *support vector machines*, as the observed vector in the observation space ($\mathbf{x}$).

16.5 Audio Segmentation

Audio segmentation is one of the challenges faced in processing telephone or recorded speech. Consider telephone speech for the moment. Majority of telephone conversations take place between two individuals at the different ends of a channel. It is conceivable to record a two-party conversation into two separate channels us-