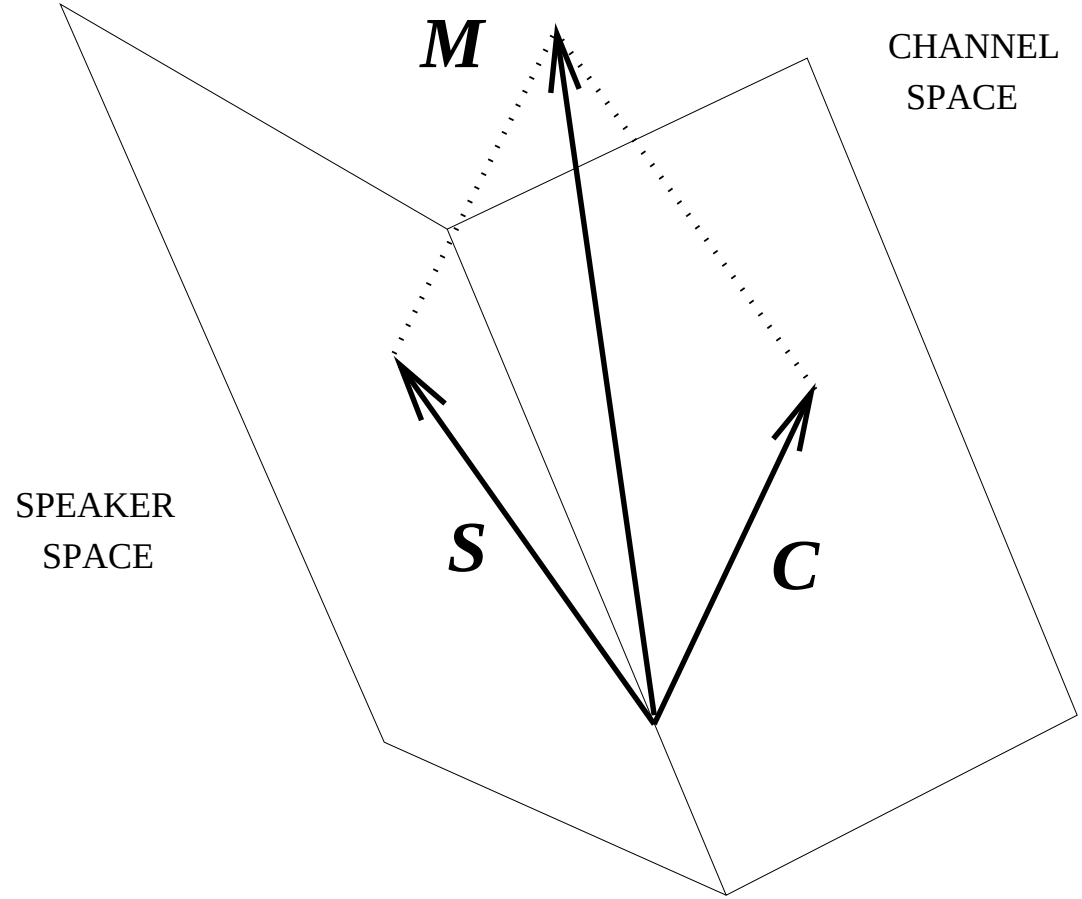
The CRIM System for the 1conv4w-1convmic condition

Patrick Kenny

Centre de recherche informatique de Montreal (CRIM)

Tous droits réservés © 2005 CRIM

First Approach: Rebuild the Channel Space



$$M = S + C$$

Tous droits réservés © 2005 CRIM

The Distribution of Channel Supervectors

- 8 microphone types suggests multimodal (mixture) distribution rather than unimodal
  - Insufficient development data
  - See Odyssey presentation
- Since there are few development speakers not even a UBM can be trained properly
- If there is a mismatch between the UBM and the data
  - UBM likelihood of data is very low
  - Factor analysis with a unimodal channel distribution gives
    - Low likelihoods
    - Slowly decreasing channel eigenvalues

Tous droits réservés © 2005 CRIM

Second Approach: Begin by Mapping into Telephone Domain

$$Y \rightarrow Y - \sum_{c} P(c \mid Y) o_c$$

- SPLICE transformation estimated for each microphone/gender pair
  - Maps microphone data onto telephone speech
  - Parameters consist of a GMM and a collection of offset vectors (one for each mixture component c)
  - GMM trained by Baum-Welch
  - Offset vectors estimated by ML from stereo data

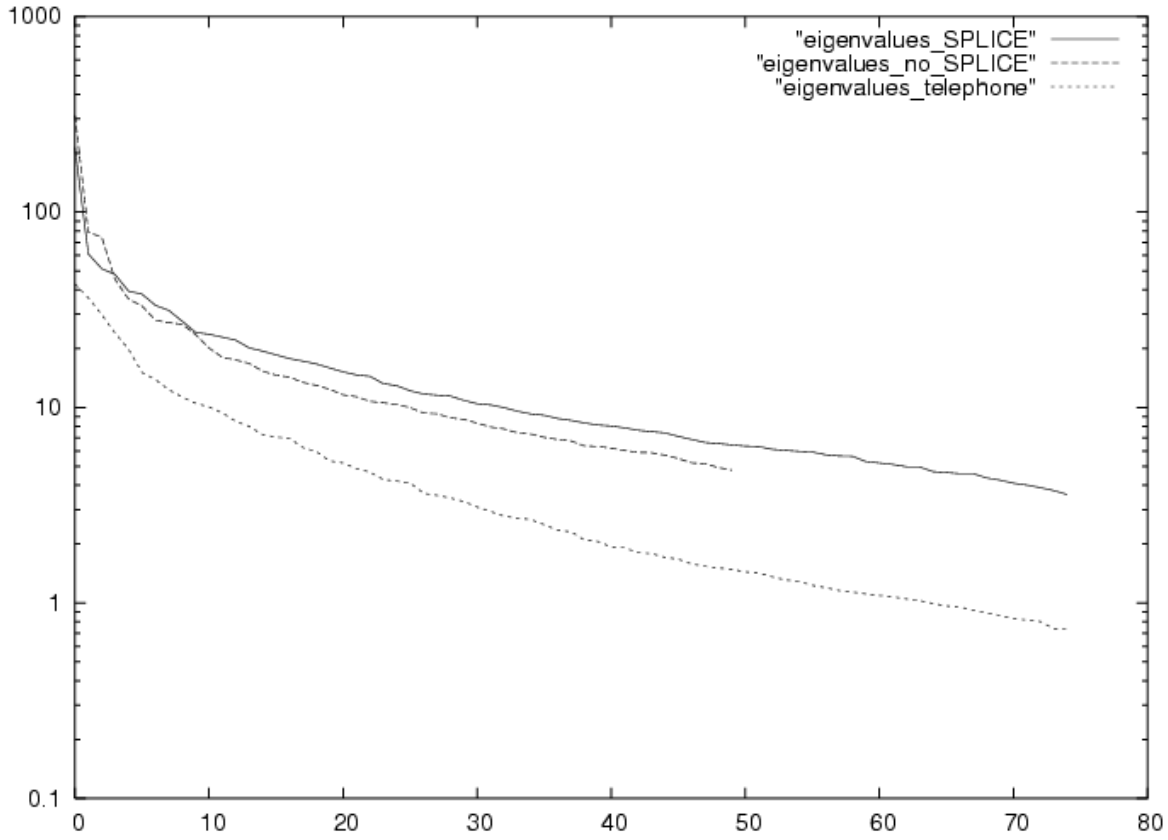
Tous droits réservés © 2005 CRIM

Accurate Microphone Identification is a Pre-Requisite

- Features for speaker verification:
  - Log energy, cepstra, feature warping, first deltas
- Features for microphone identification:
  - Normalized log energy, raw cepstra (no warping, no deltas)
  - 512 component GMM for each microphone/channel condition
- 5% error rate ***on development set*** (8 microphone types)

Tous droits réservés © 2005 CRIM

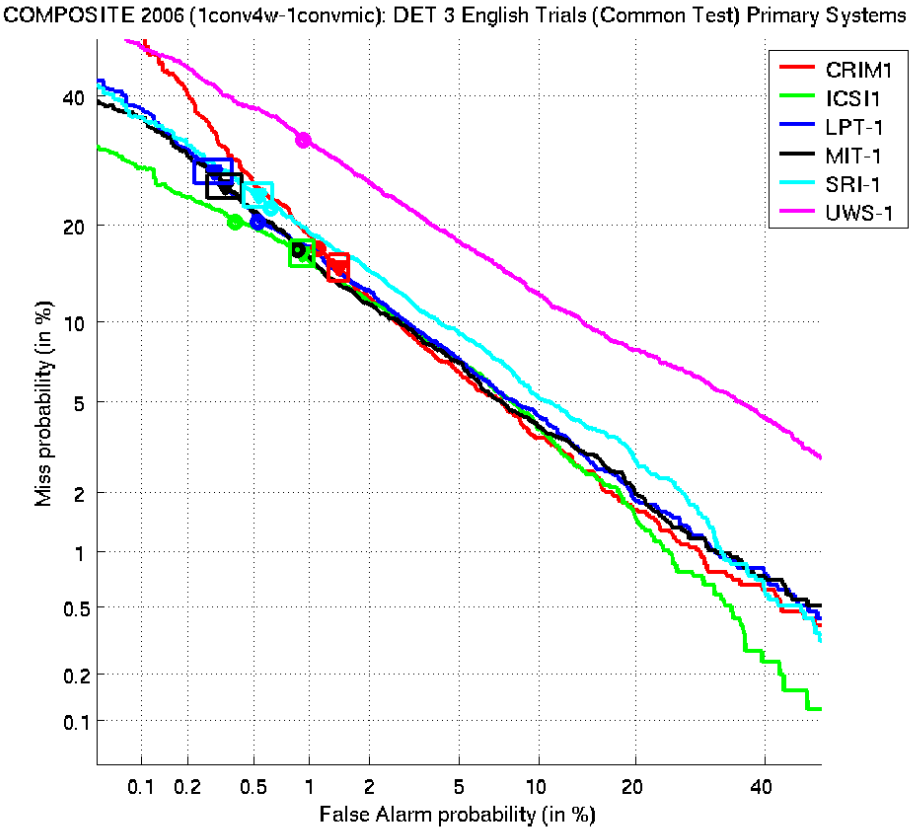
Paradoxical Eigenvalue Profiles with and without SPLICE



Eigenvalues of Cov(C, C) sorted in decreasing order. Female data.

Tous droits réservés © 2005 CRIM

1conv4w-1convmic



Tous droits réservés © 2005 CRIM