

A Novel and Computationally Efficient Approach to Setting Decision Thresholds in an Adaptive Speaker Verification System

Nikki Mirghafori

Joint work with
Matthieu Hebert at Nuance

Friday the 13th, February 2004



Outline

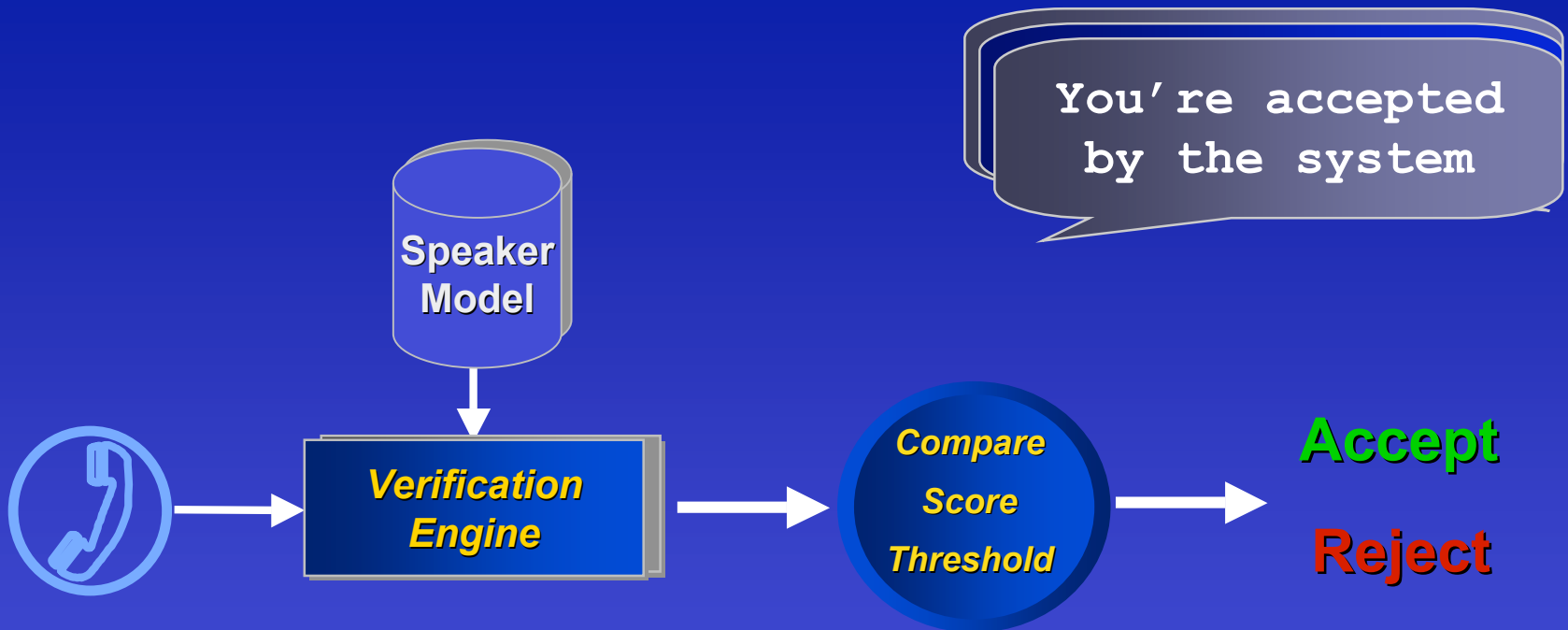
- Beginning
- Middle
- End

Outline

- Motivation
- Approach
- Data Analysis
- Calibration
- Experimental Results
- Conclusions

ICASSP 2004: Mirghafori, N. and Hebert, M. "Parameterization of the Score Threshold for a Text-Dependent Adaptive Speaker Verification System", Montreal, Canada.

Context: Speaker Verification Field Application



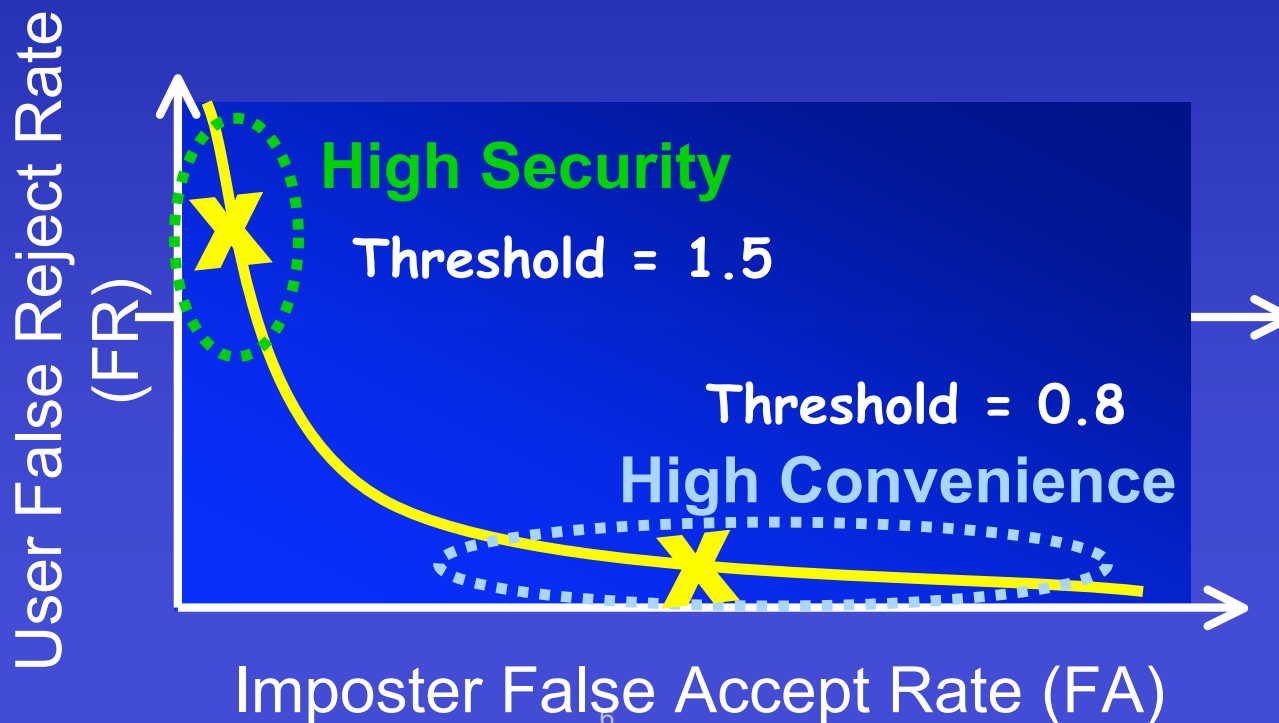
An example of a verification session

Problems with Score Thresholds

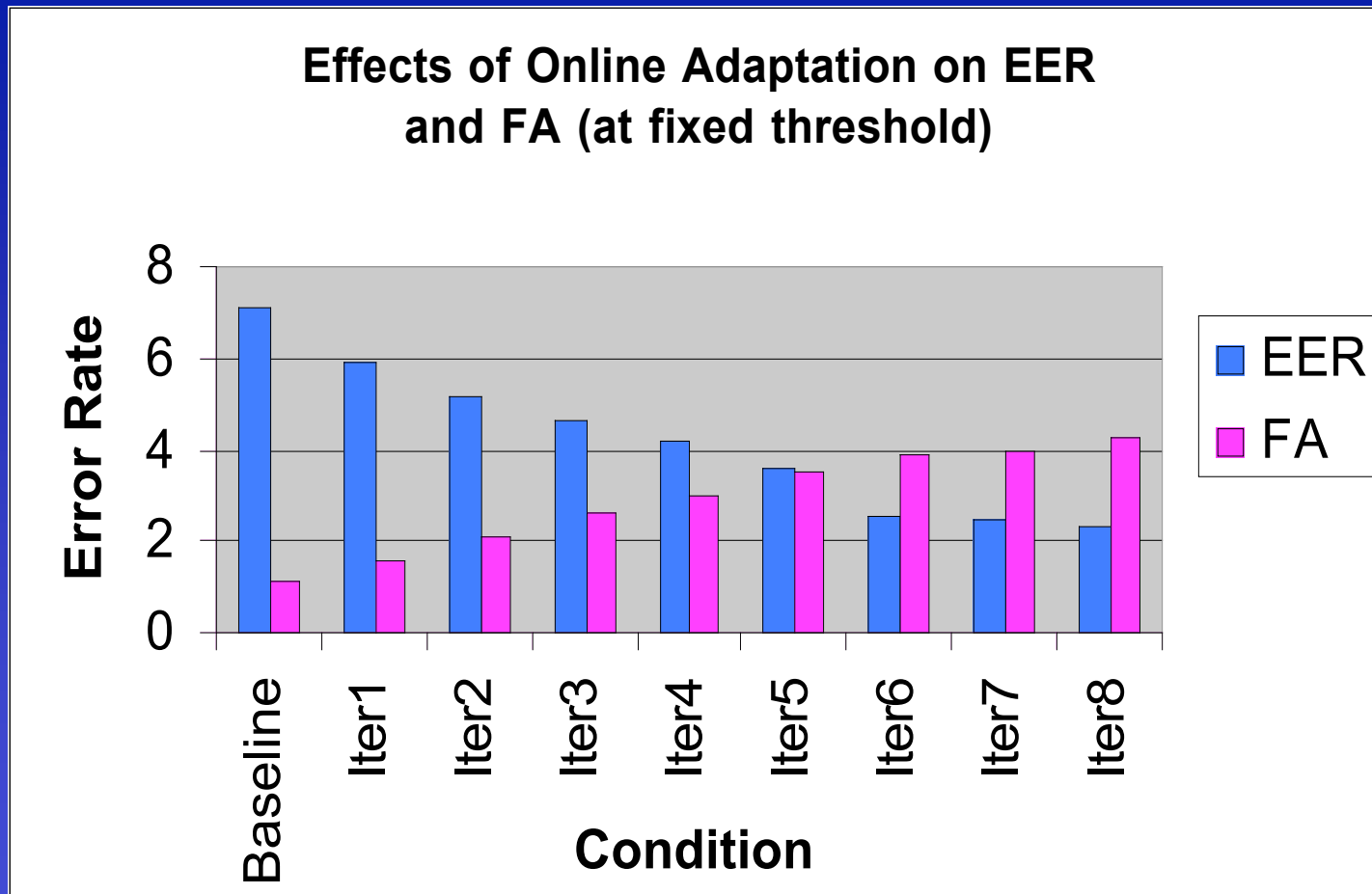
- Overall fixed system thresholds are difficult to set in advance
- If combined with online adaptation, thresholds must be dynamic and adaptive

Problem I – Setting Overall Threshold

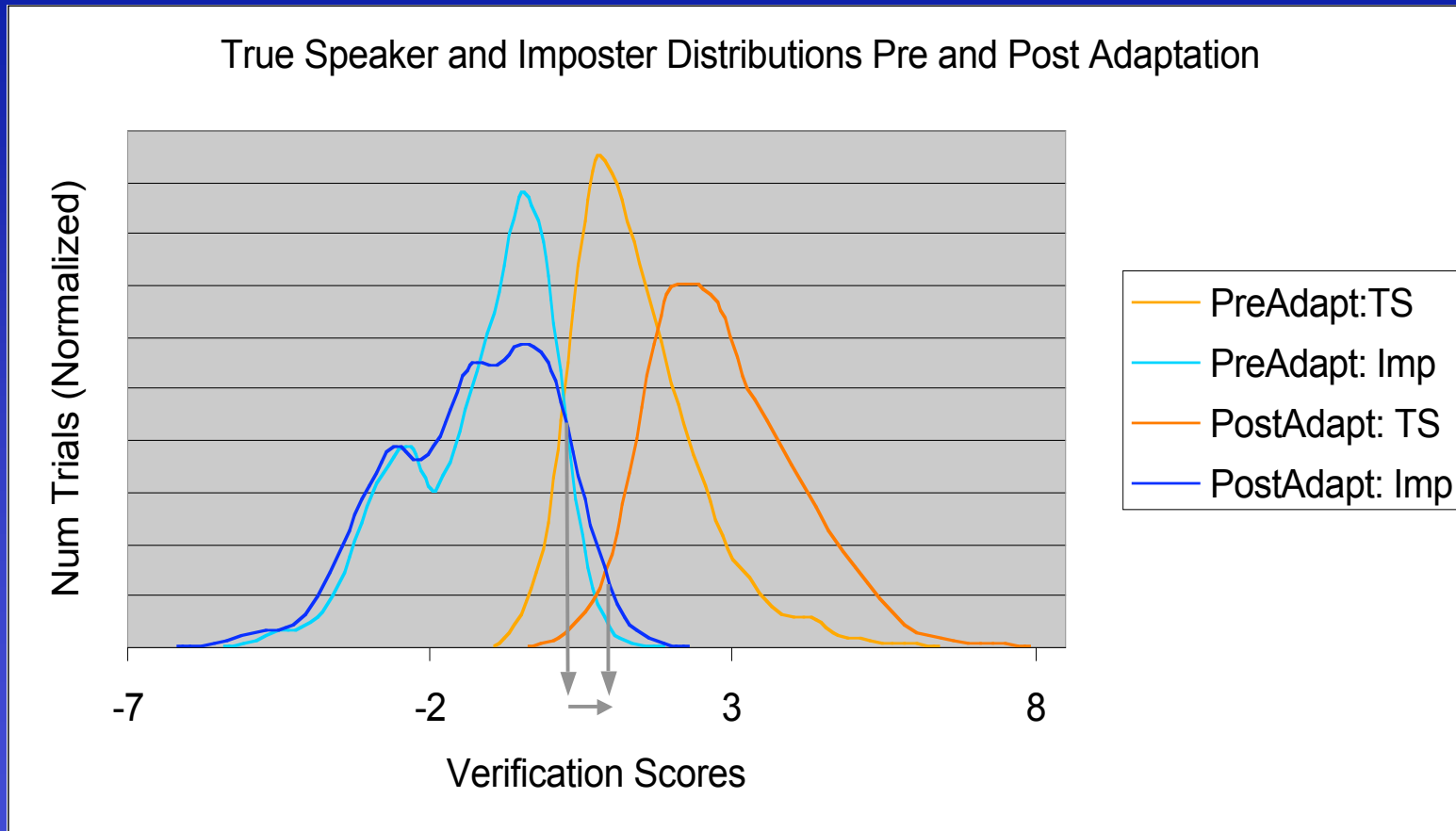
- Setting overall thresholds before-hand is challenging
 - High convenience? High security?
 - Difference between lab and field conditions can cause differences in thresholds
 - Not all speakers are created equal



Problem II – Effects of Online Adaptation



Problem II – Increase in Impostor Scores



Problem -- II

- Speaker adaptation successfully reduces EER in verification [Heck & Mirghafori, ICSLP '00]
- However, both true speaker and, to a lesser extent, impostor scores may increase
- Increase in impostor false accept rates can cause speaker model corruption
- Similar score increase observed in speech recognition for confidence scores [Sankar & Kannan, ICASSP '02]. Score transformation used as a solution.

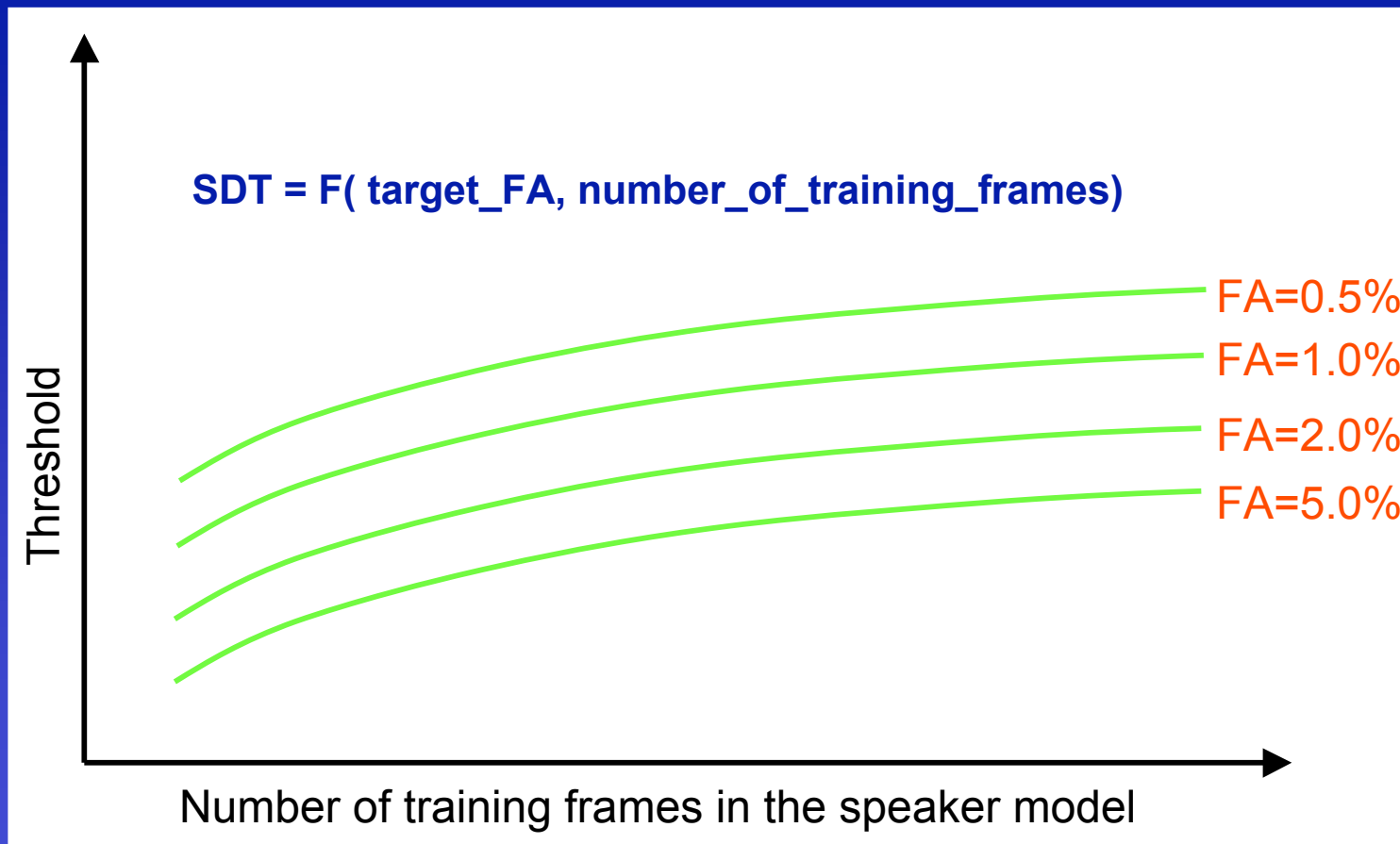
Motivation

1. Eliminate the challenge of setting overall application thresholds in advance
2. Counter the verification score shifts resulting from online adaptation

Previous Work in Setting Thresholds in Speaker Verification

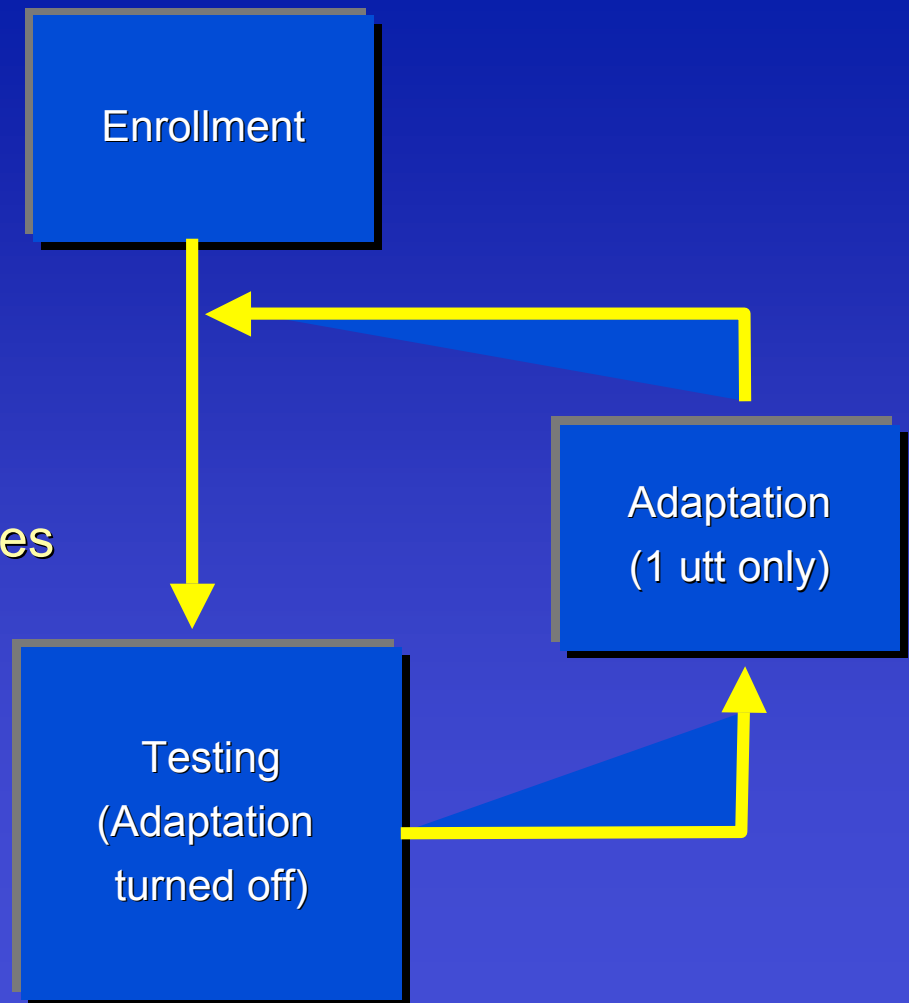
- Mirghafori, N. and Heck, L.P., "An Adaptive Speaker Verification System With Speaker Dependent A Priori Decision Thresholds", Proc. International Conference on Spoken Language Processing, Denver, Colorado, 2002.
- A.C. Surendran and C.-H Lee. "A priori threshold selection in fixed vocabulary speaker verification system." ICSLP 2000
- J. Lindberg, J.W. Koolwaaij, H.-P. Hutter, D. Genoud, M. Blomberg, J.-B. Pierrot, and F. Bimbot. "Normalization and selection of speech segments for speaker recognition scoring." RLA2C 1998.

Our Approach

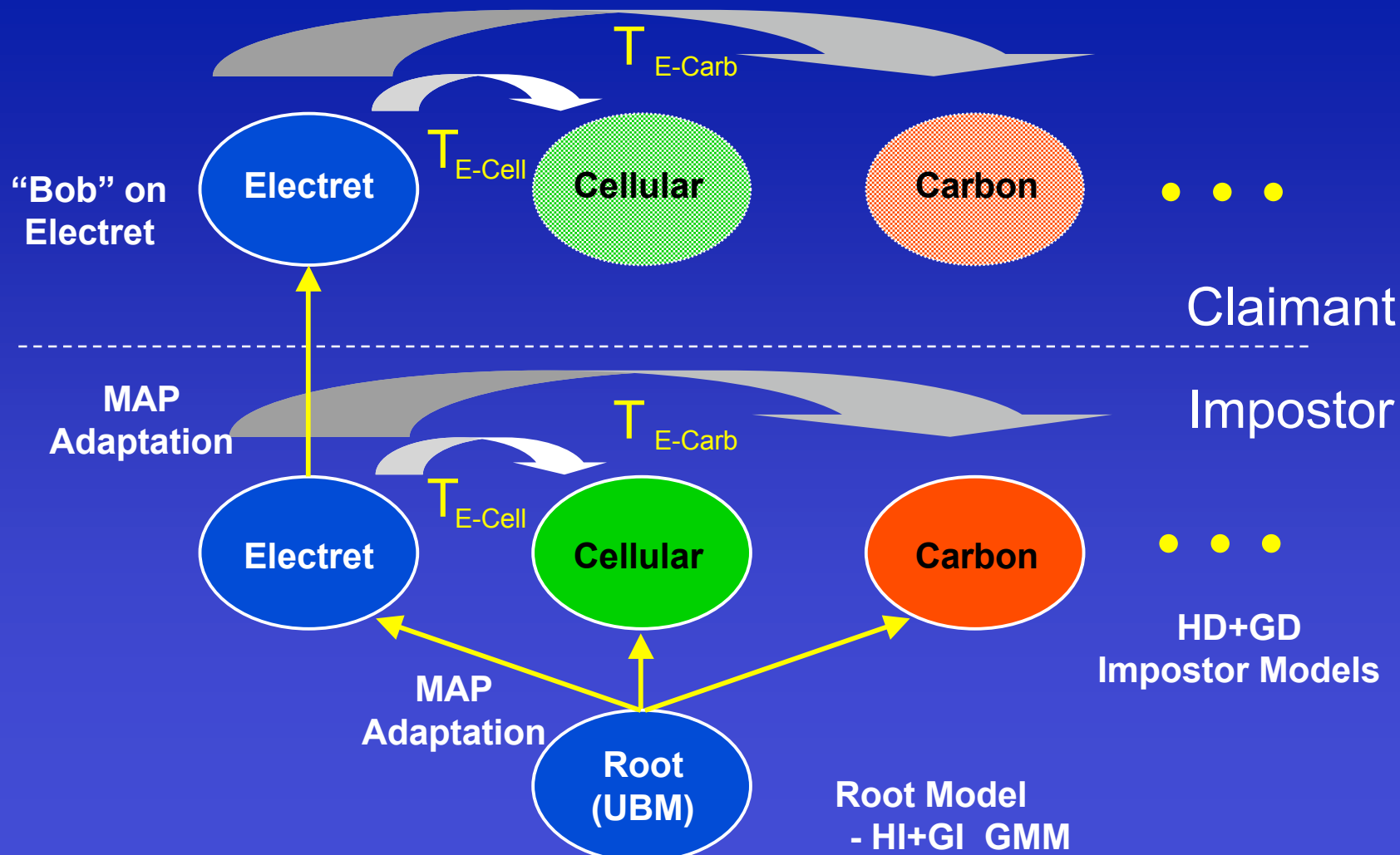


Test Setup

- Japanese Digits
- 6,477 speaker models
- Enrollment: 8-digit string (3x)
- Held out test set:
 - 8-digit string (1x)
- Adaptation:
 - 8-digit string (1x)
- Gather all test scores
- Divide into bins w.r.t. to num frames
- Calculate threshold at target FA

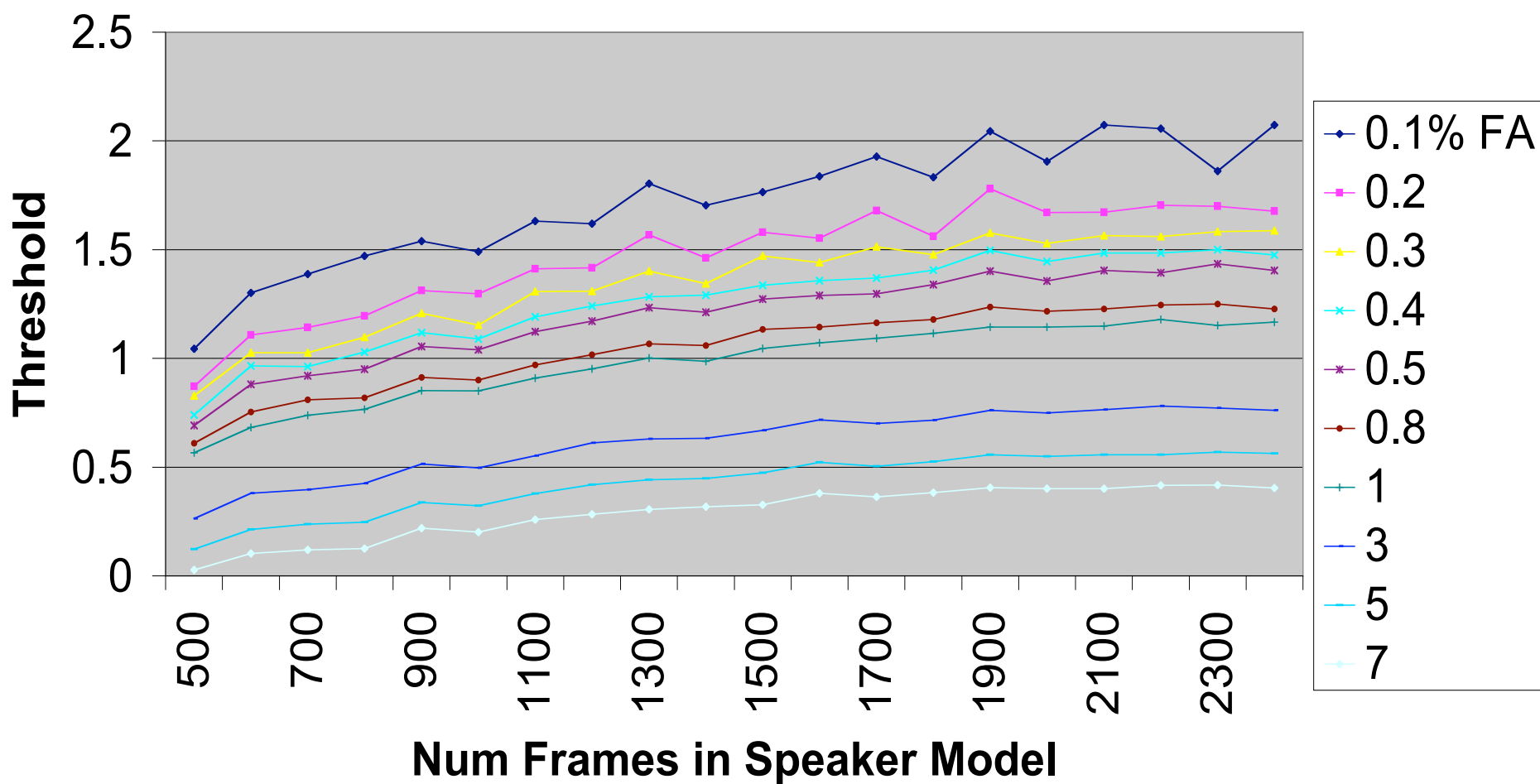


Speaker Verification System



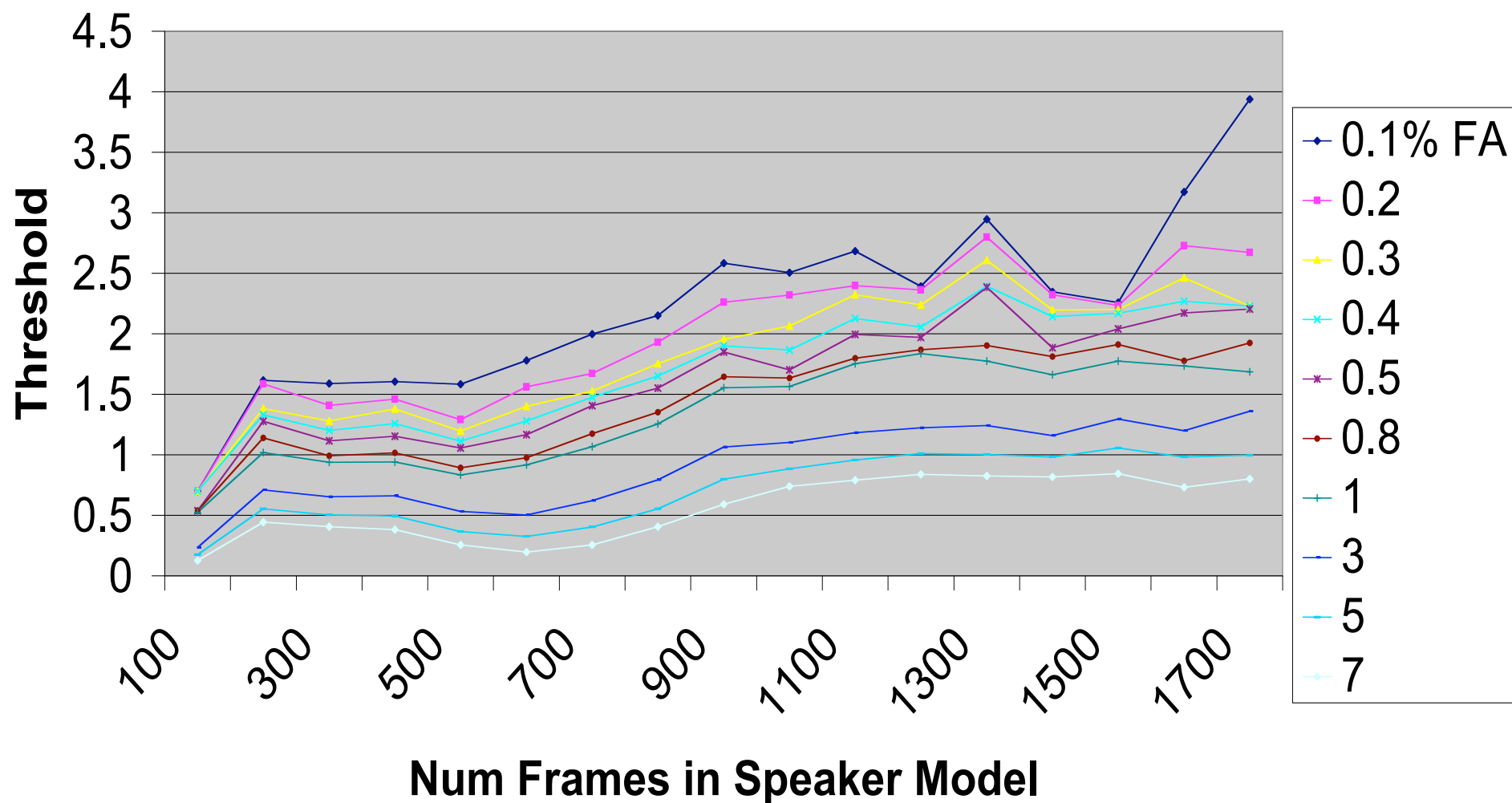
The trend looks good: Threshold = (Target FA, Num Frames)

Japanese Digits



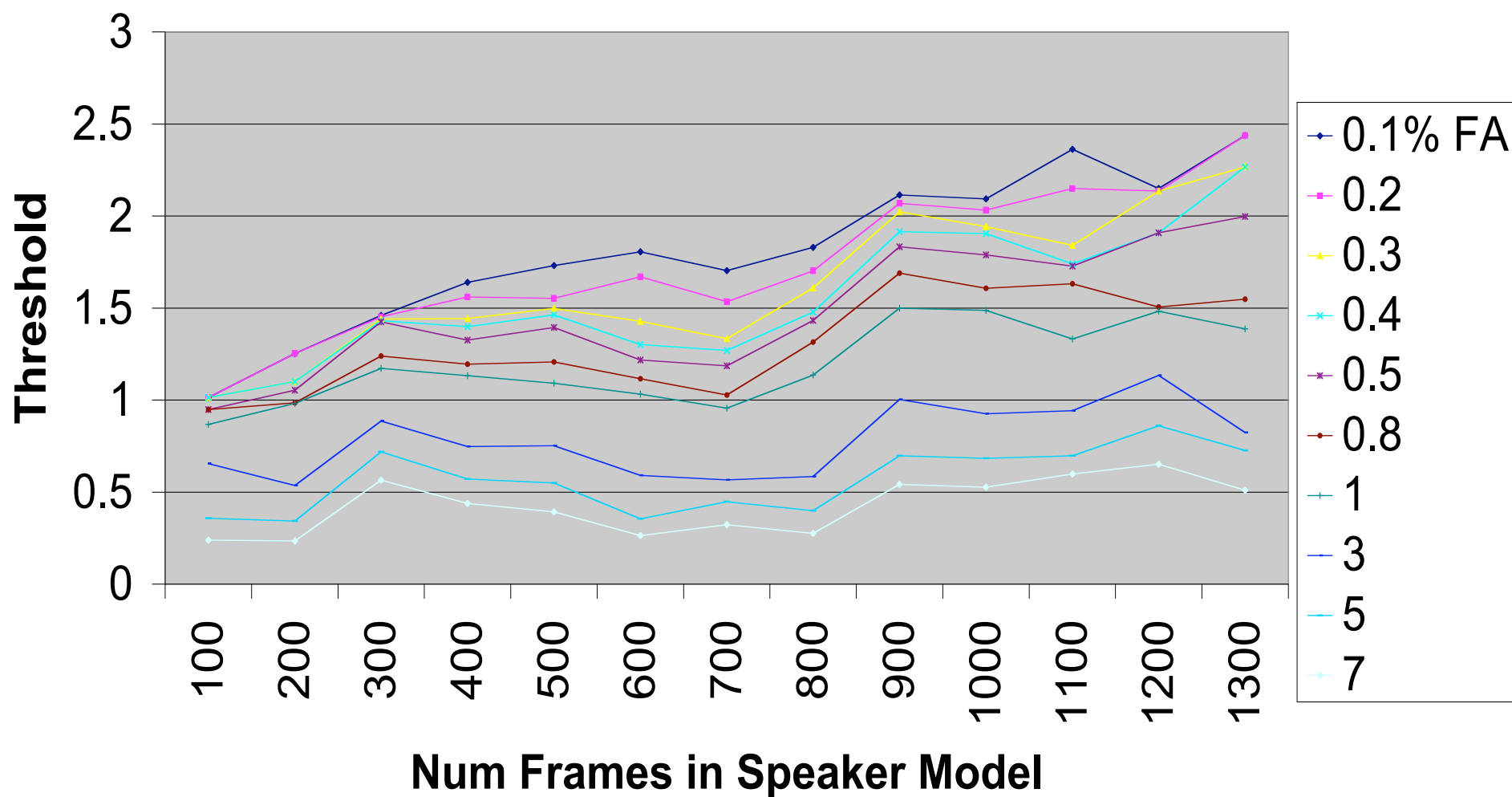
“It’s a Bug!” Or is it?

Canadian French Text



The trend is confirmed...

UK English Text



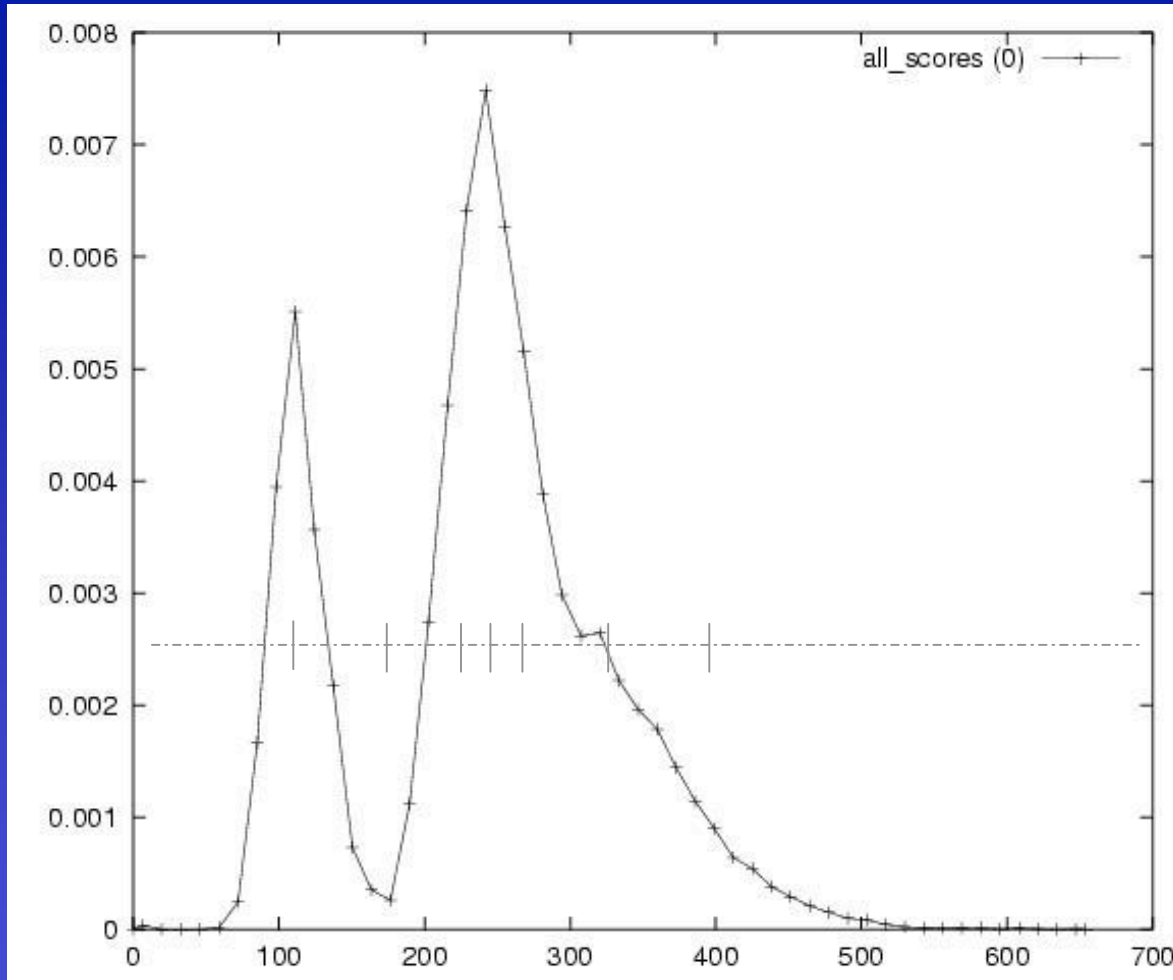
What's the Missing Parameter?

- Password length!
- Japanese digits DB uniform, whereas the two others are not
- Contain both short and long passwords
- In general: verification performance lower for shorter passwords
- Hence, need higher threshold
- Can we parameterize?
 - Threshold = G(Target FA, Num. Frames, Password Length)

Calibration

- Japanese text AND digit
- Both short and long passwords
- Trained and tested 10K speaker models
- Divided password length region between 0 and 700 into 8 bins

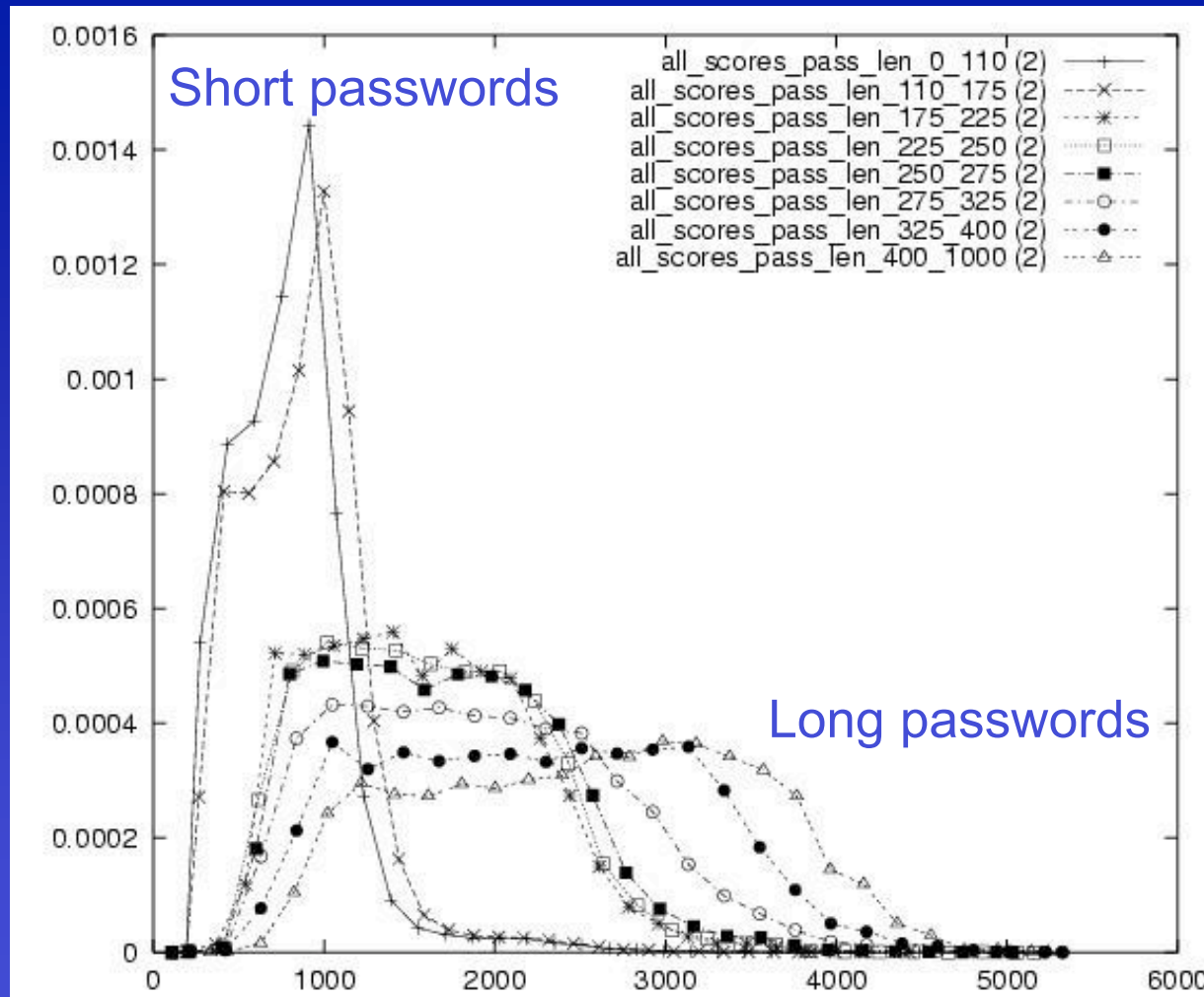
Password Length distribution



Password Length

Bins: between 120K
to 196K attempts each

Num. Training Frames in Speaker Model For each password length

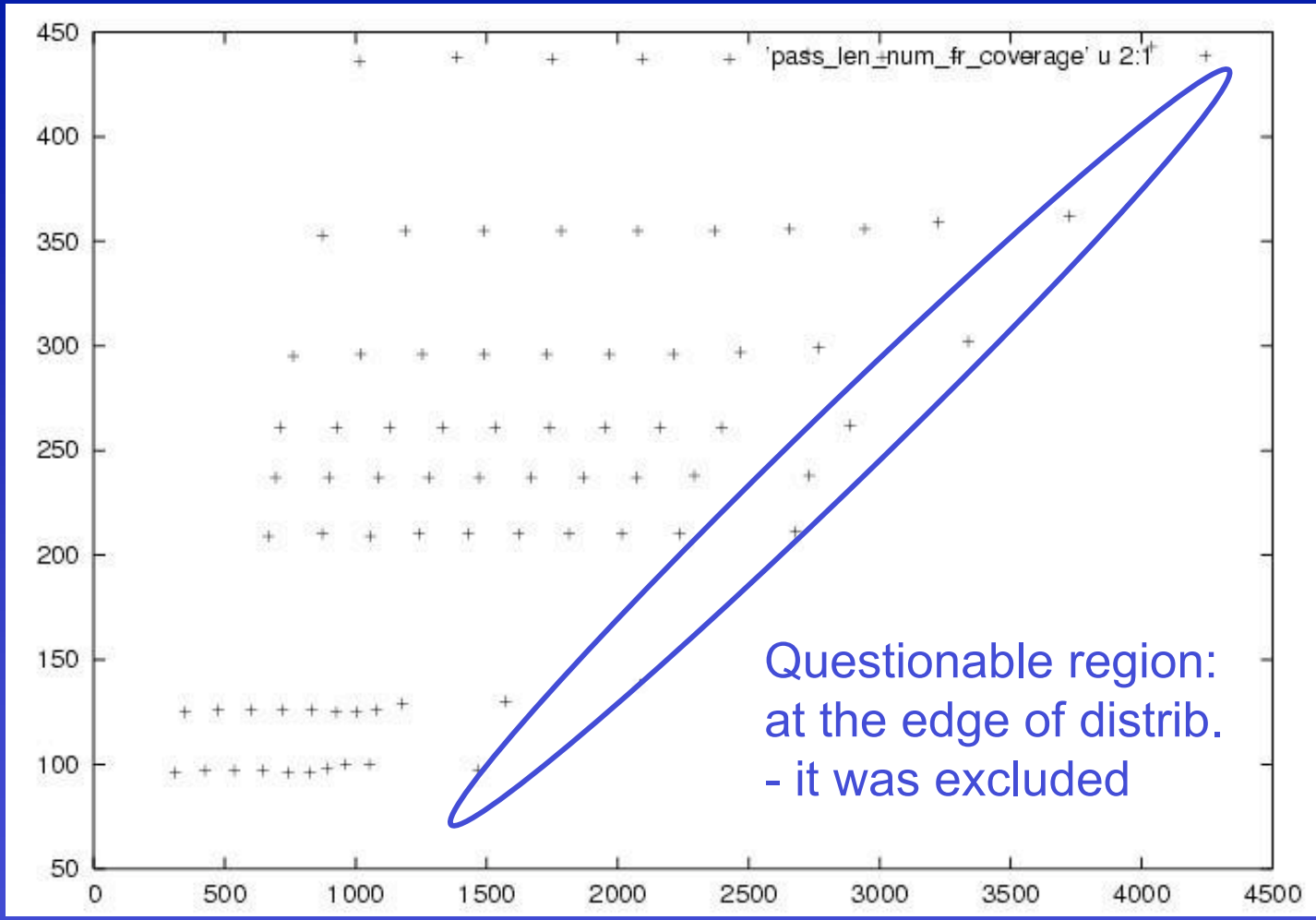


Num. Training Frames in Speaker Model

- Divide data each bin into 10 sub-bins according to the number of frames in speaker model
- Calculated Average Password Length (APL) and Average Num Frames (ANF)

Coverage of 2D space Num. Training Frames vs. Password Len.

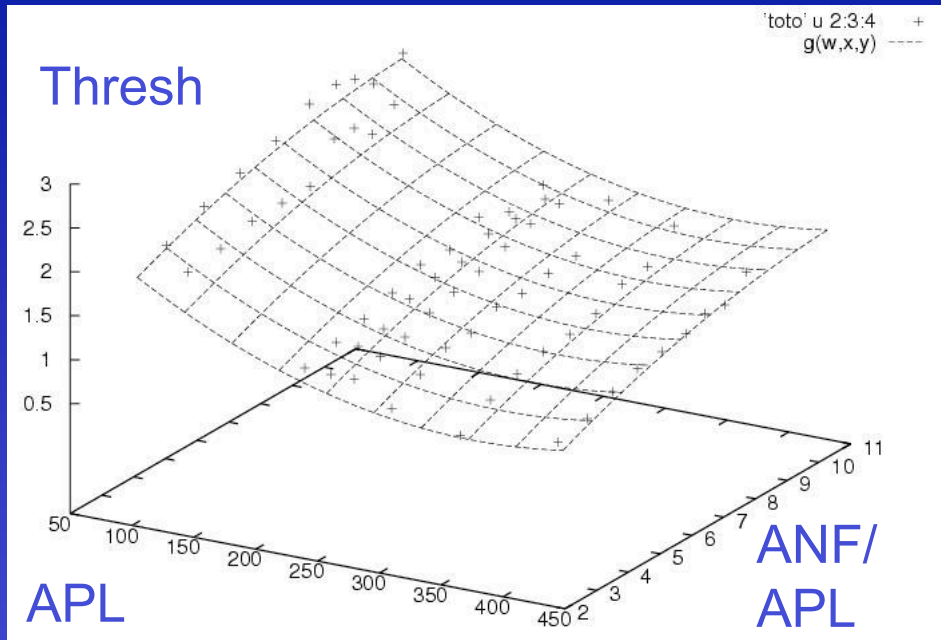
Ave. Password Length



Ave. Num. Training Frames in Speaker Model

Result of fitting TargetFA = 0.1%

TargetFA=0.1%

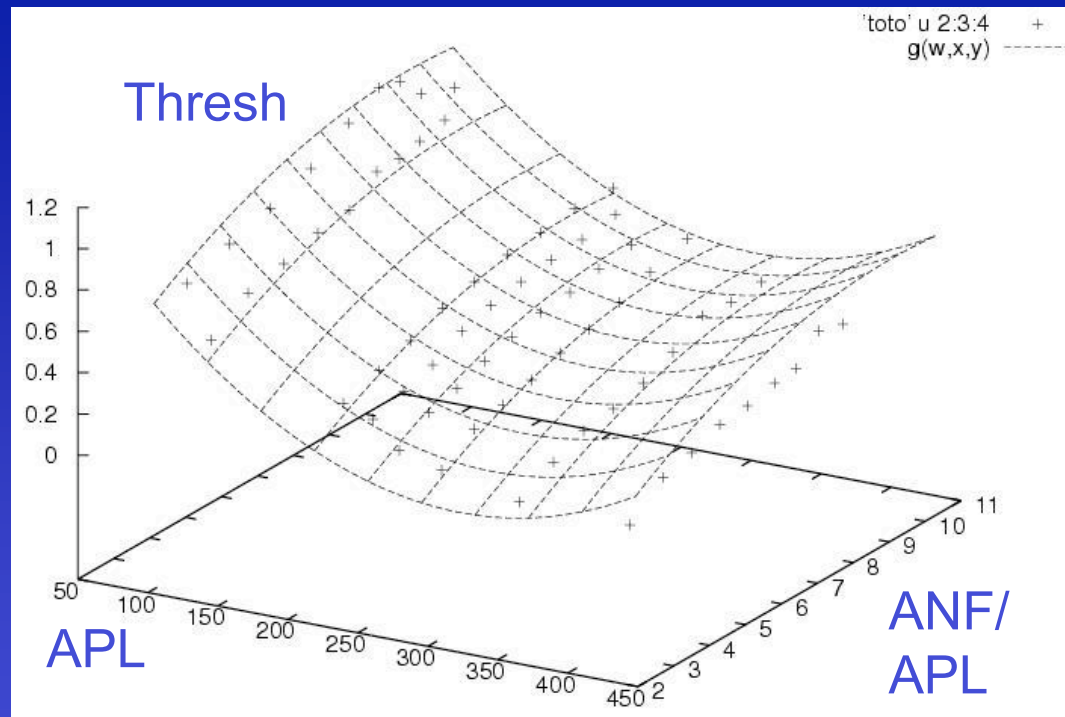


- Fit a second degree polynomial with three parameters (10 free coefficients)

Fit looks good!

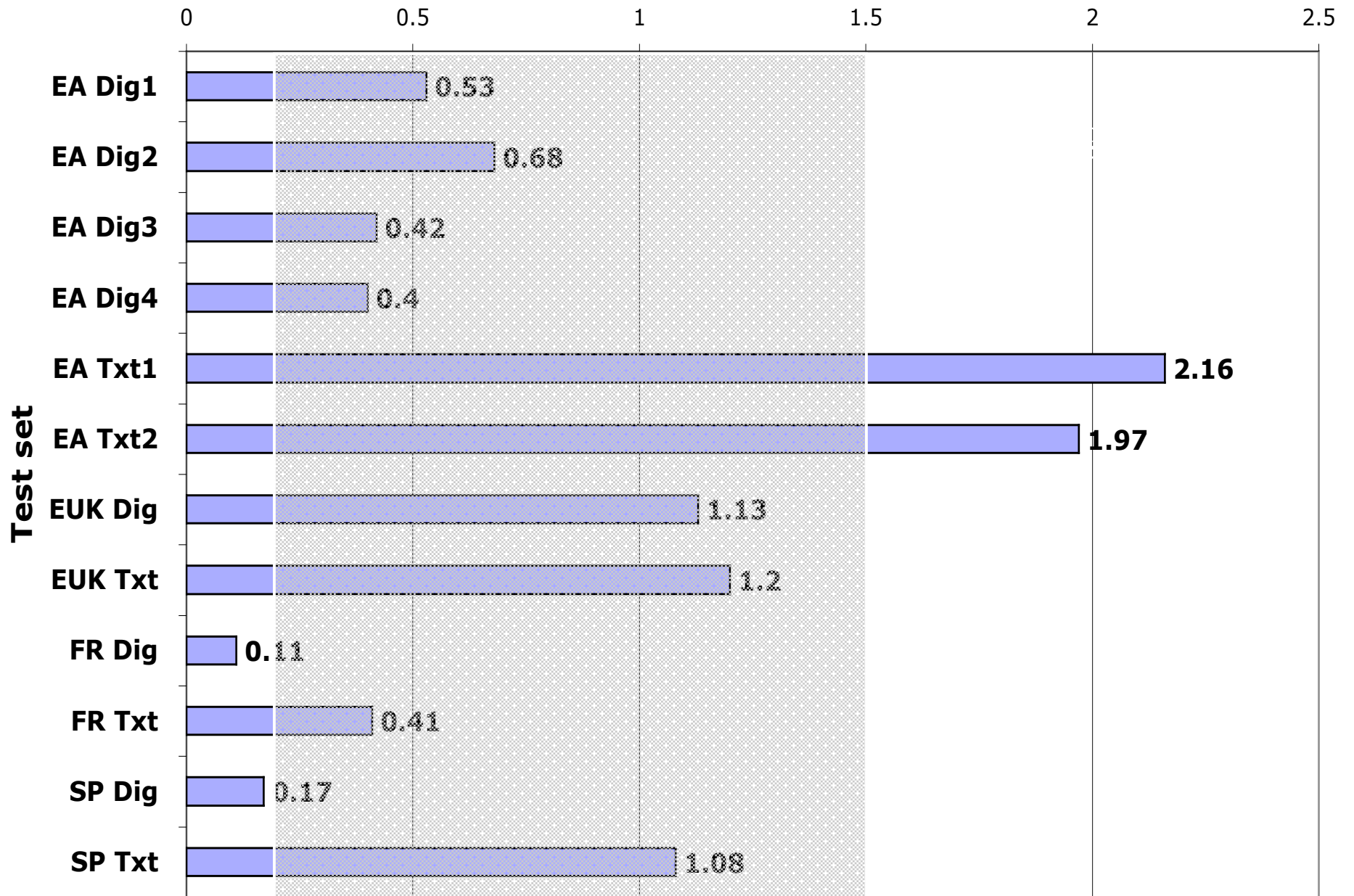
Result of fitting TargetFA = 5%

TargetFA=5%

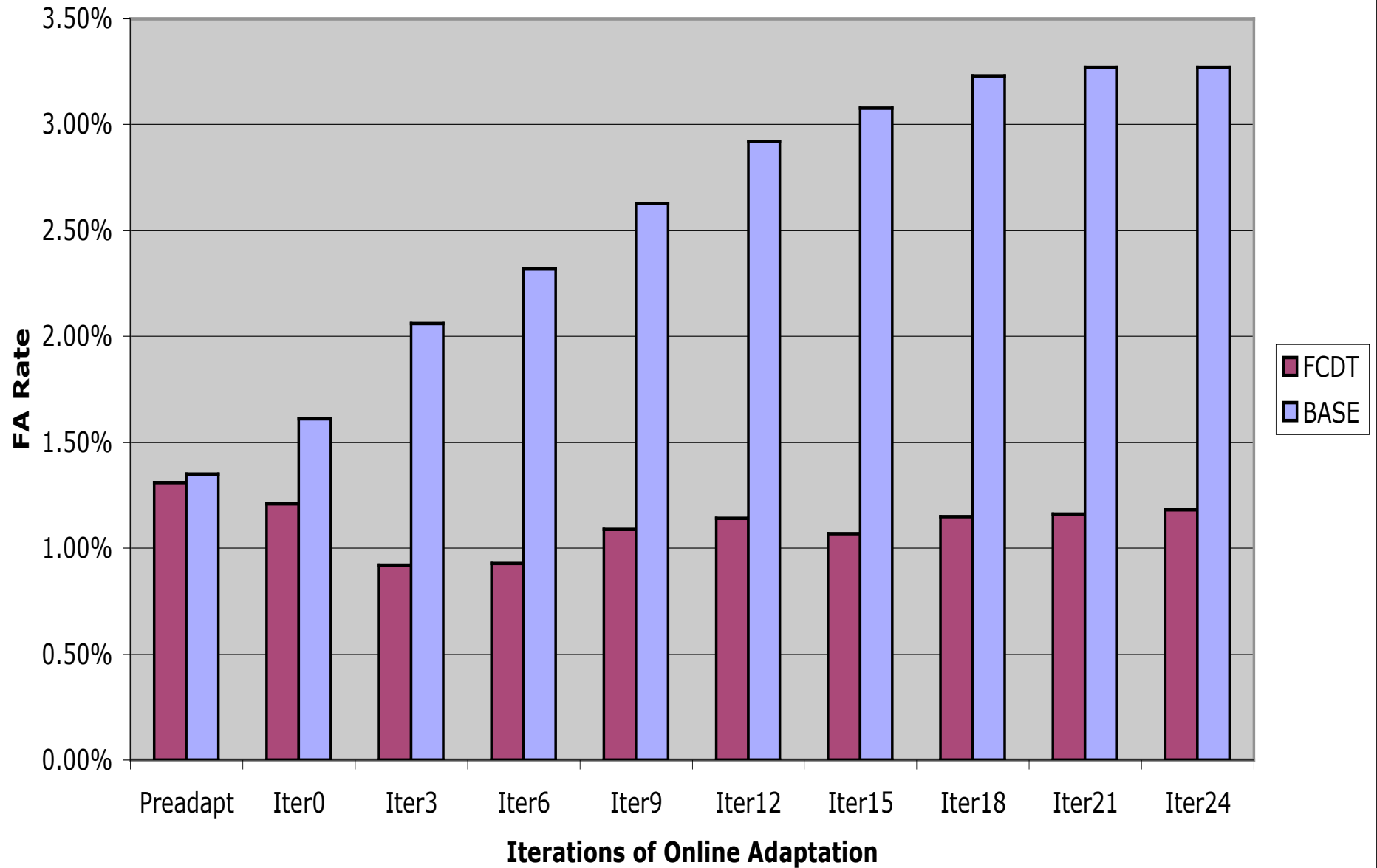


Fit looks good!

Actual FA rates for FA goal 0.2-1.5%



Effects of Online Adaptation on FA



In Conclusion

- Setting *a priori* thresholds is a challenging and important problem for field applications
- Threshold can be parameterized as function of
 - Target FA
 - Number of training frames in speaker model
 - Password length
- Threshold parameters can be calibrated on one database and applied to others successfully
 - Thresholds achieved within target FA range
 - FA rate kept constant during online adaptation
 - Overall performance (EER) does not degrade