Hearing versus Seeing Identical Twins

Li Zhang, Shenggao Zhu, Terence Sim,
Wee Kheng Leow, Hossein Nejati and Dong Guo

School of Computing, NUS

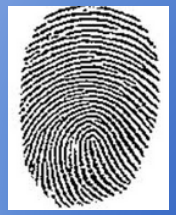
Social Party Scenario

Who’s who?

Can a classifier help?

Biometrics to Identify Twins

Fingerprint



Iris

Palmprint

Equal-error Rate (EER) ≈ 0.49
Require subject's cooperation

Ear

May be obscured by hair or hat

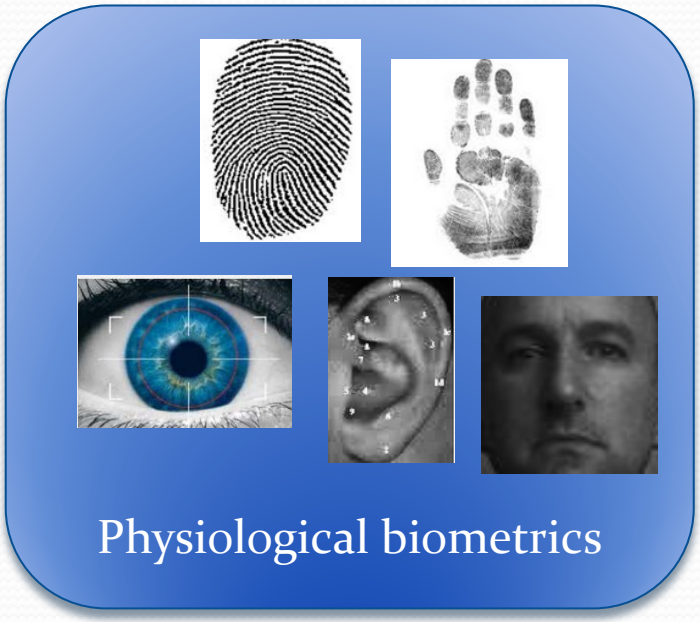
Face

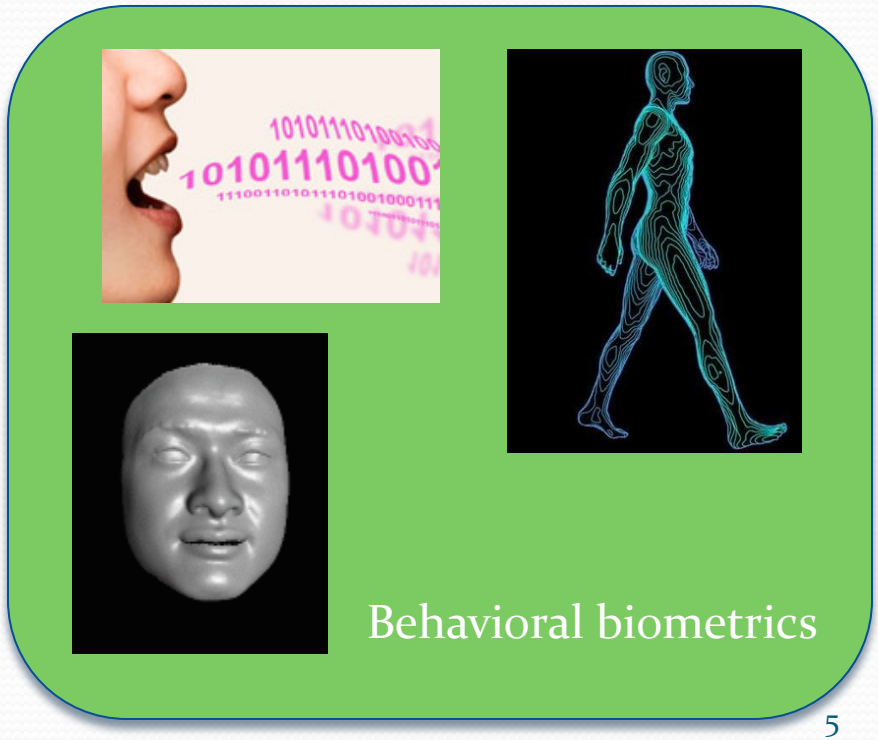
EER = 0.33

So how to recognize twins?

Key Idea

Twins may look alike, but will behave differently





Voice for Twins Identification

Voice biometrics a.k.a. speaker recognition has proven to be effective for the General Population (non-twins)

But does it work for Twins?

Does Voice biometrics work for Twins?

Yes!

Oui!

Is Voice better than Face?

Yes, and EER is lowered!

Can Voice and Face be combined?

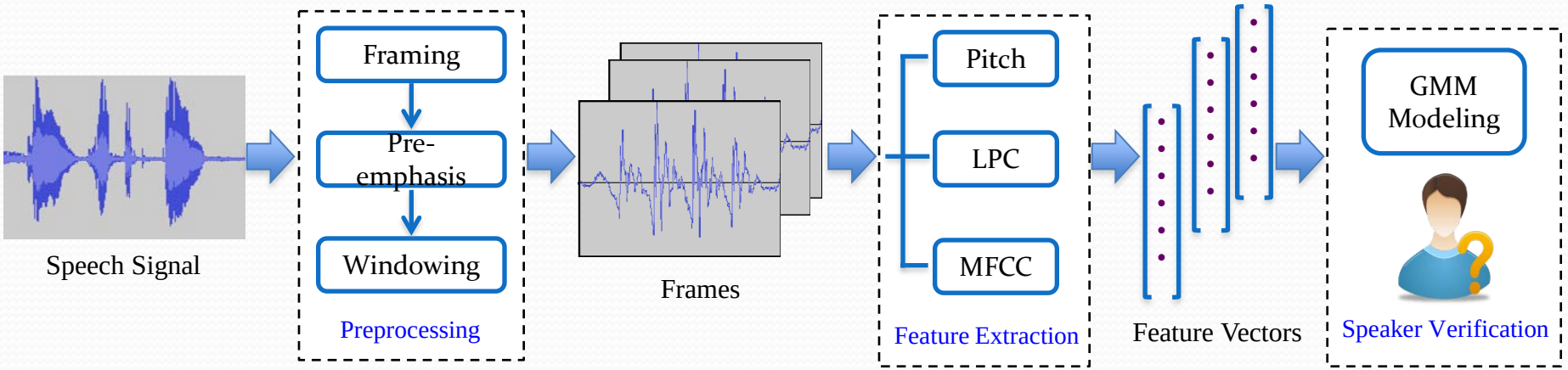
Which Voice feature to use?

MFCC works best

NUS Twins Database

- 39 Pairs of Identical Twins
- Multiple races: Russian, Chinese and Canadian
- Frontal images and videos
- Voice recording

Building an acoustic model



- Feature Extraction
- GMM Modeling
- Classification using likelihood ratio

Class conditional pdf

$$p(x|s) = \sum_{i=1}^{M} w_i b_i(x|s)$$

x : input feature vector (Pitch, LPC, MFCC in each frame)
b_i : Gaussian pdf
w_i : mixture weight
s : subject

Verification

Given a probe, ψ, and a claimed identity S,

H0: ψ is from S.

H1: ψ is the twin sibling of S.

Classification by likelihood ratio

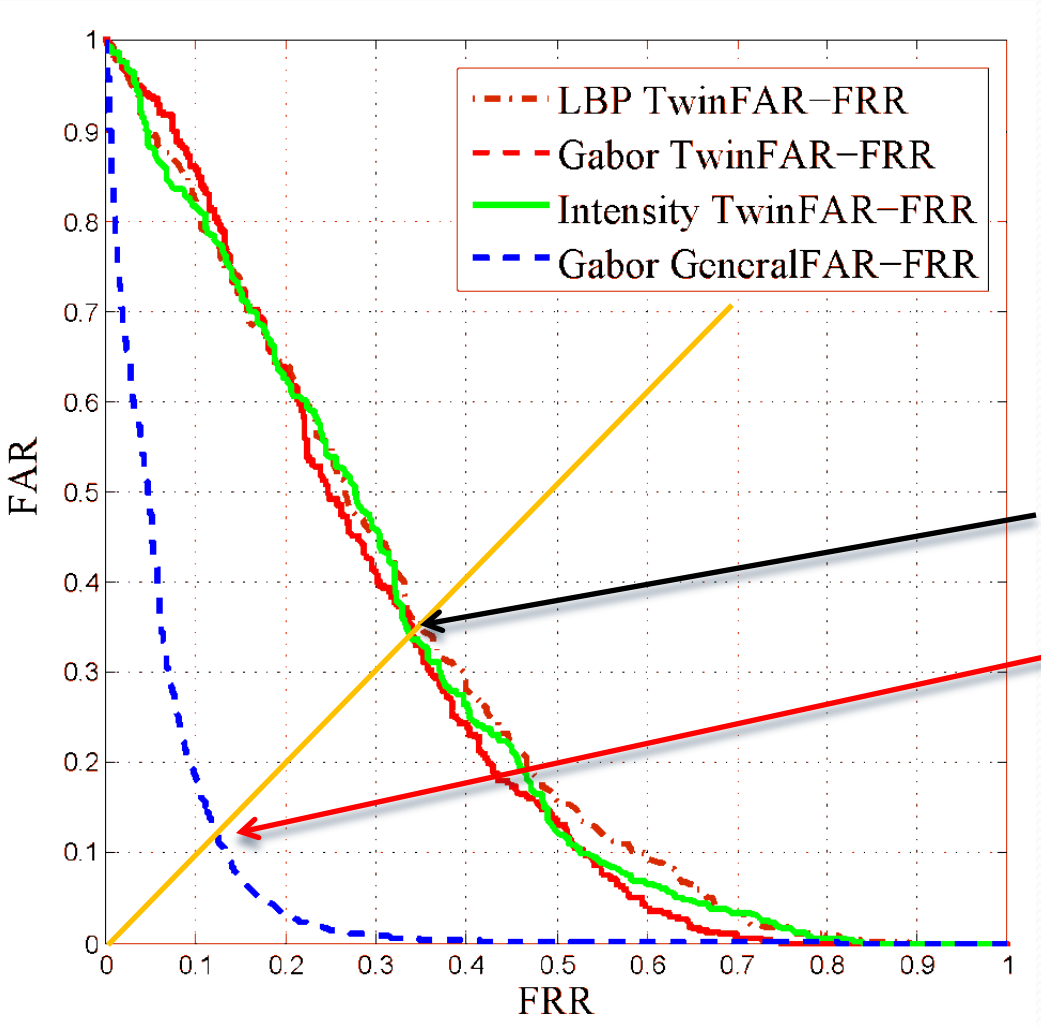
$$LR = \frac{p(\psi|H0)}{p(\psi|H1)}$$

If LR > ξ, we accept H0;
otherwise, we reject H0.

Evaluation

- Twin-FAR: Percentage of mistaking one twin imposter for his sibling
- Twin-FRR: Percentage of rejecting genuine twin
- Twin Equal Error Rate (Twin-EER): where Twin-FAR = Twin-FRR
- General-EER: for non-twins

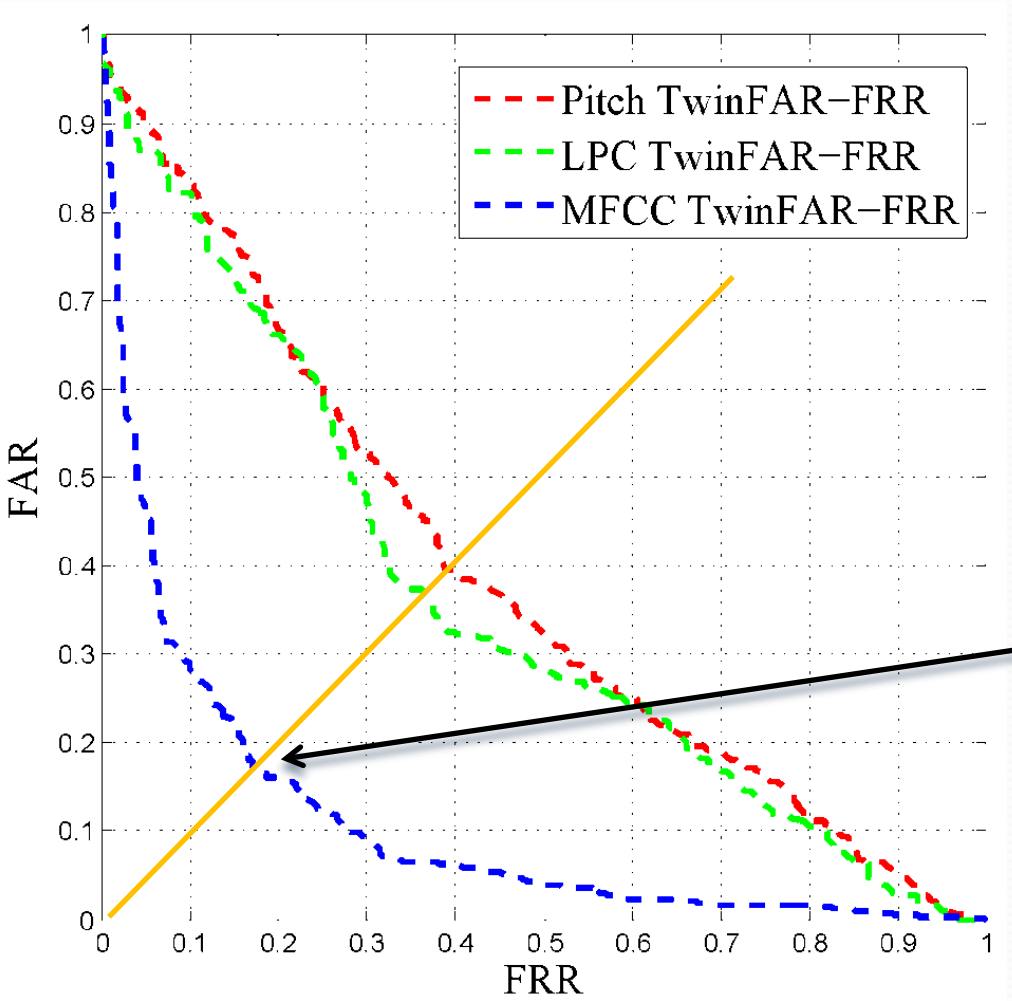
Experiments: face



Best Twin-EER: 0.338(Gabor)

General-EER: 0.122

Experiments: voice



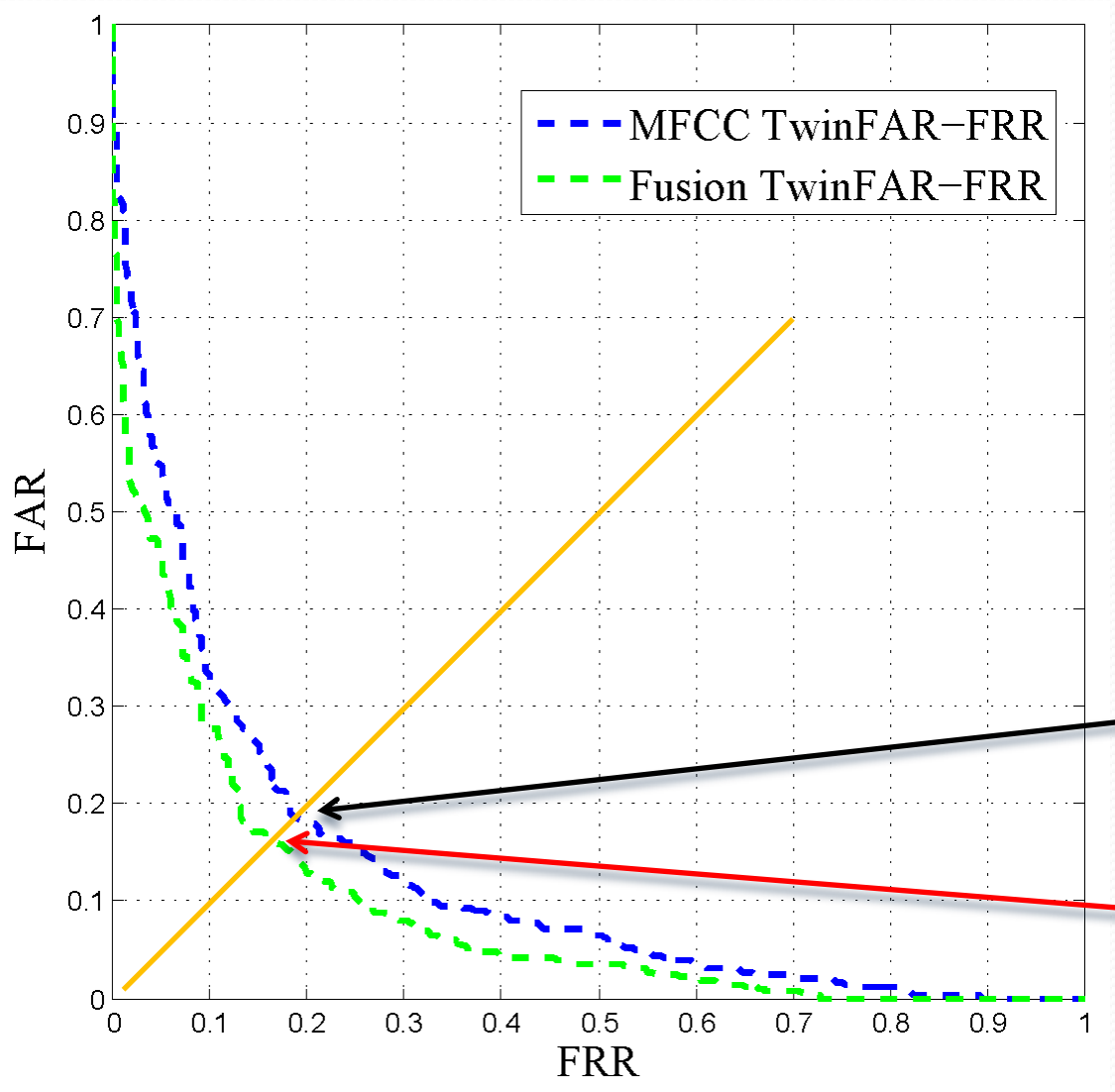
So Voice is better than Face!

MFCC is the best feature

Best Twin-EER: 0.171 (MFCC)

General-EER: 0.05, from [10]

Experiments: voice + face



Score-level fusion lowers EER

Voice Twin-EER: 0.171

Fusion Twin-EER: 0.160

Conclusion

Does Voice biometrics work for Twins?

Yes!

Oui!

Is Voice better than Face?

Yes, and EER is lowered!

Can Voice and Face be combined?

Which Voice feature to use?

MFCC works best

The next time you go to a party:

Thank you

Methodology

Feature Extraction

1. Framing
   To divide audio into successive overlapping frames
   frame size： 23mileseconds with 50% overlapping in our work

2. Hamming window
   To smooth out the discontinuity between the end and start

3. Thresholding
   To filter out the silent frames

4. Feature extraction from each frames
   LPC, Pitch, MFCC extracted separately

Introduction

Example of twins