

Rekognisi Pengucap Forensik

Webinar INAPR 25.03.2022
Miranti Indar Mandasari, Ph.D.

Outline

- Rekognisi pengucap sebagai bagian dari sistem biometrics
- Voiceprint/ sidik suara dan model source-filter
- Teknologi sistem rekognisi pengucap
- Aplikasi forensik dari rekognisi pengucap

Biometrics dan Rekognisi Pengucap

Biometrics



Biometrics



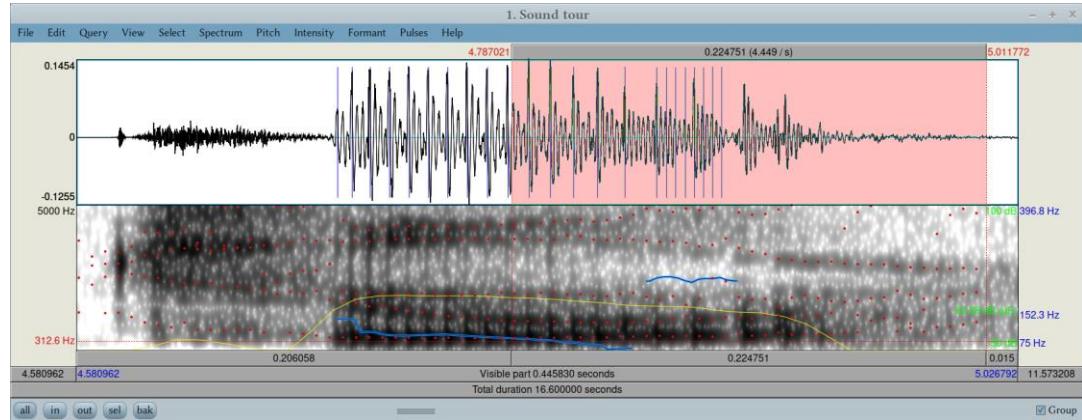
Biometrics are body measurements and calculations related to human characteristics. Biometric authentication is used in computer science as a form of identification and access control. It is also used to identify individuals in groups that are under surveillance. [Wikipedia](#)

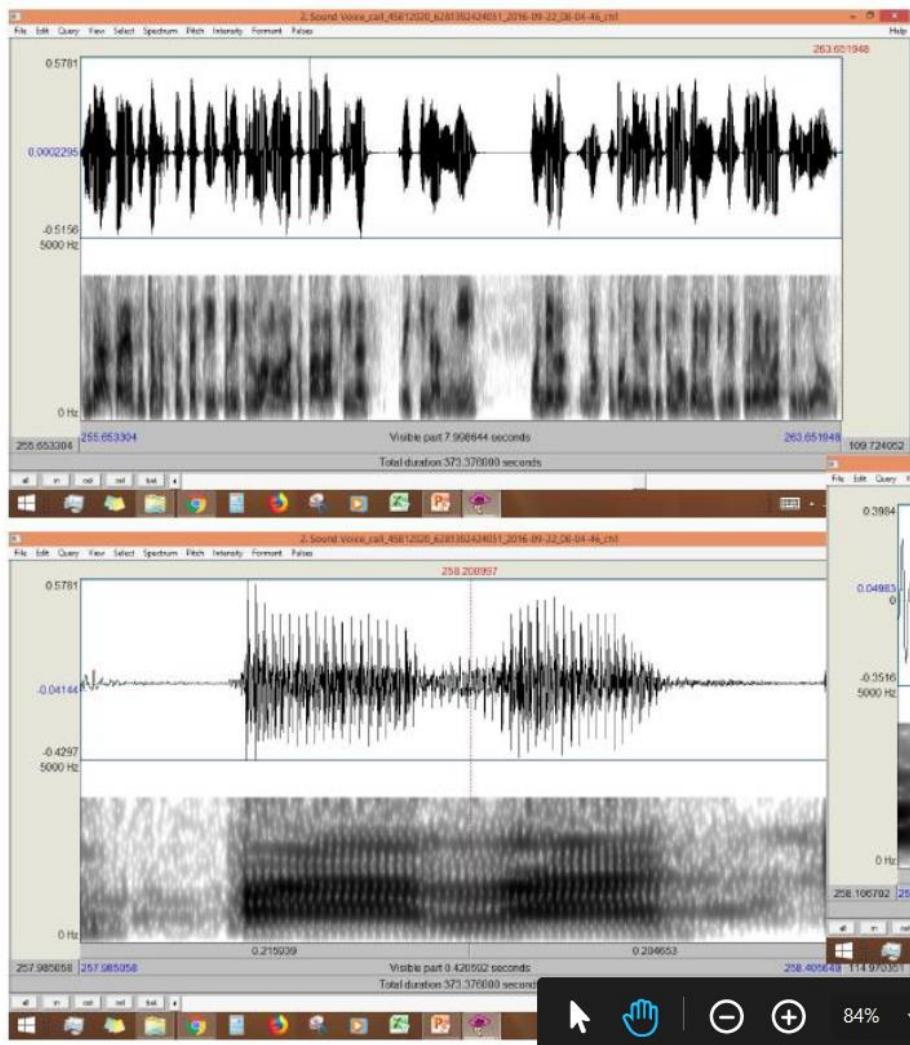
Voiceprint atau Sidik Suara

Fingerprint



Voiceprint





Time-Frequency Analysis Spectrogram

Berbagai Informasi pada Suara Ucap

Dari daerah mana dia berasal?

Accent Recognition

Bahasa apa yang dia gunakan?

Language Recognition

Apakah kata yang dia ucapkan?

Speech Recognition

Apakah dia sedih/
senang?

Emotion Recognition

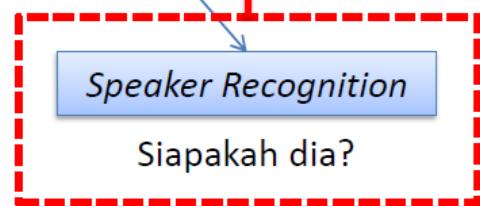
Apakah dia pria/
wanita?

Gender Recognition

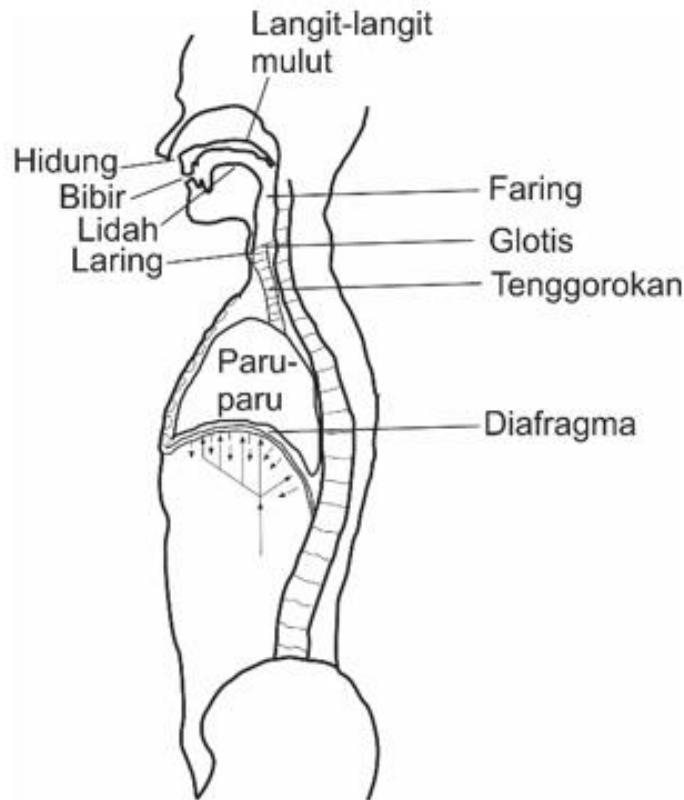
Sistem Rekognisi Pengucap

Speaker Recognition

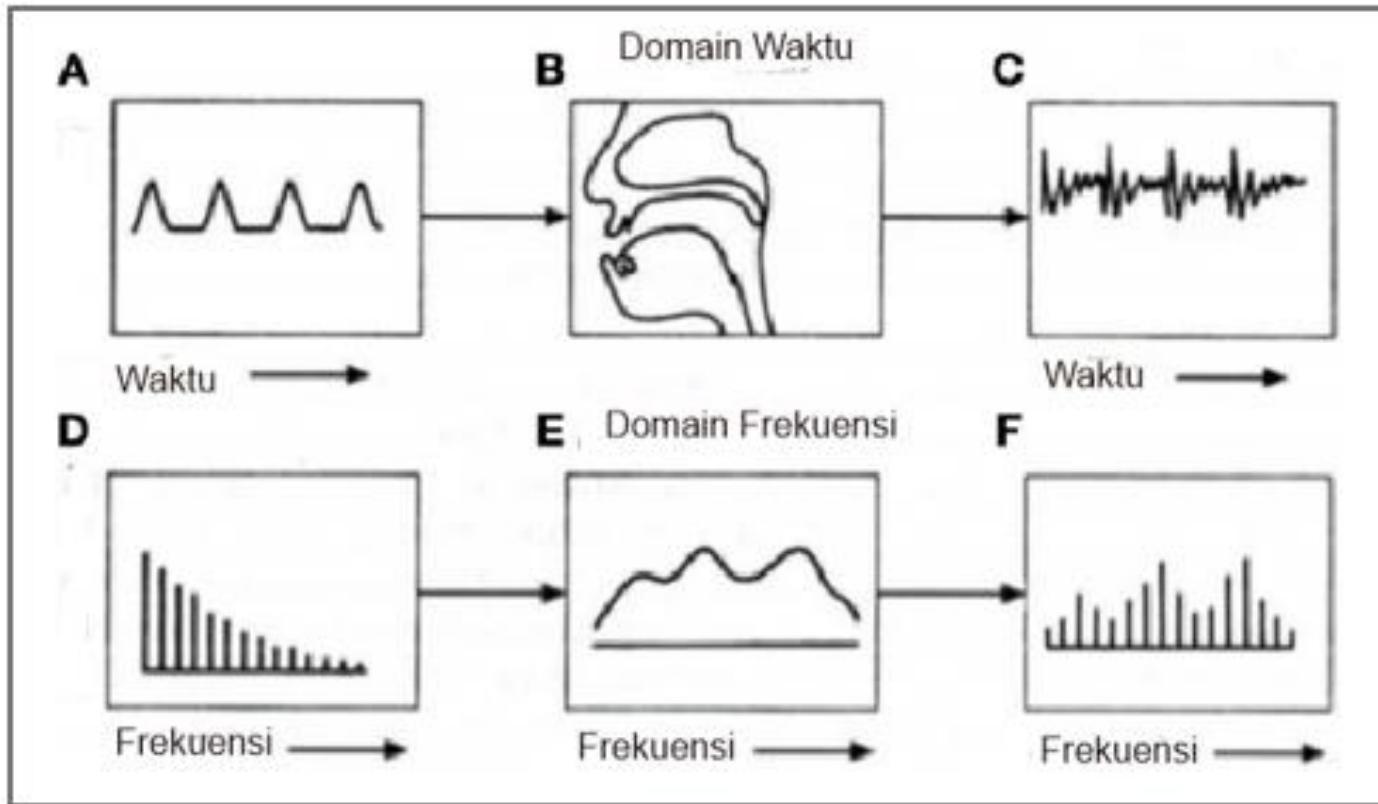
Siapakah dia?



Model Source-Filter



*Gambar II.2 Diagram dari organ yang terlibat dalam pembentukan suara ucapan
(Pickett, 1999)*



Gambar II.1 Model *source filter* (Solomon dkk, 2015)

Teknologi Sistem Rekognisi Pengucap

Sistem Rekognisi Pengucap

- Dalam Bahasa Inggris dikenal dengan
 - *Speaker recognition*, atau
 - *Voiceprint recognition*.
- Definisi:
 - *Speaker recognition is a process of recognizing the identity of a speaker from a given speech segment.*
 - Rekognisi Pengucap adalah sebuah proses dalam merekognisi (mengenali) identitas seorang pengucap dari suatu segmen suara ucap.

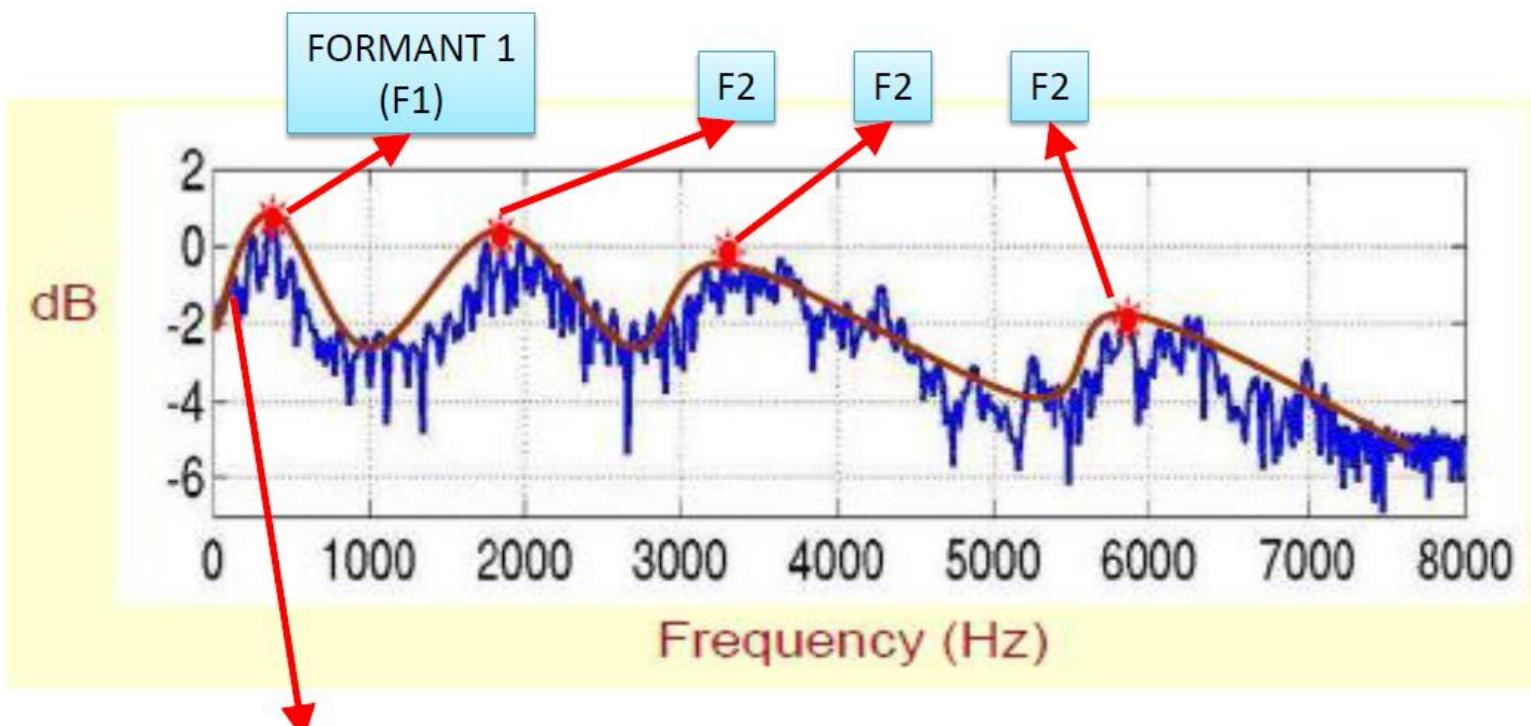
Tipe-tipe Sistem Rekognisi Pengucap

- Pendekatan fonetik-akustik
 - biasanya menggunakan metode manual
- Sistem rekognisi pengucap otomatis
- Sistem *hybrid*
 - Gabungan antara manual-otomatis

Rekognisi Pengucap Berbasis Fonetik-Akustik

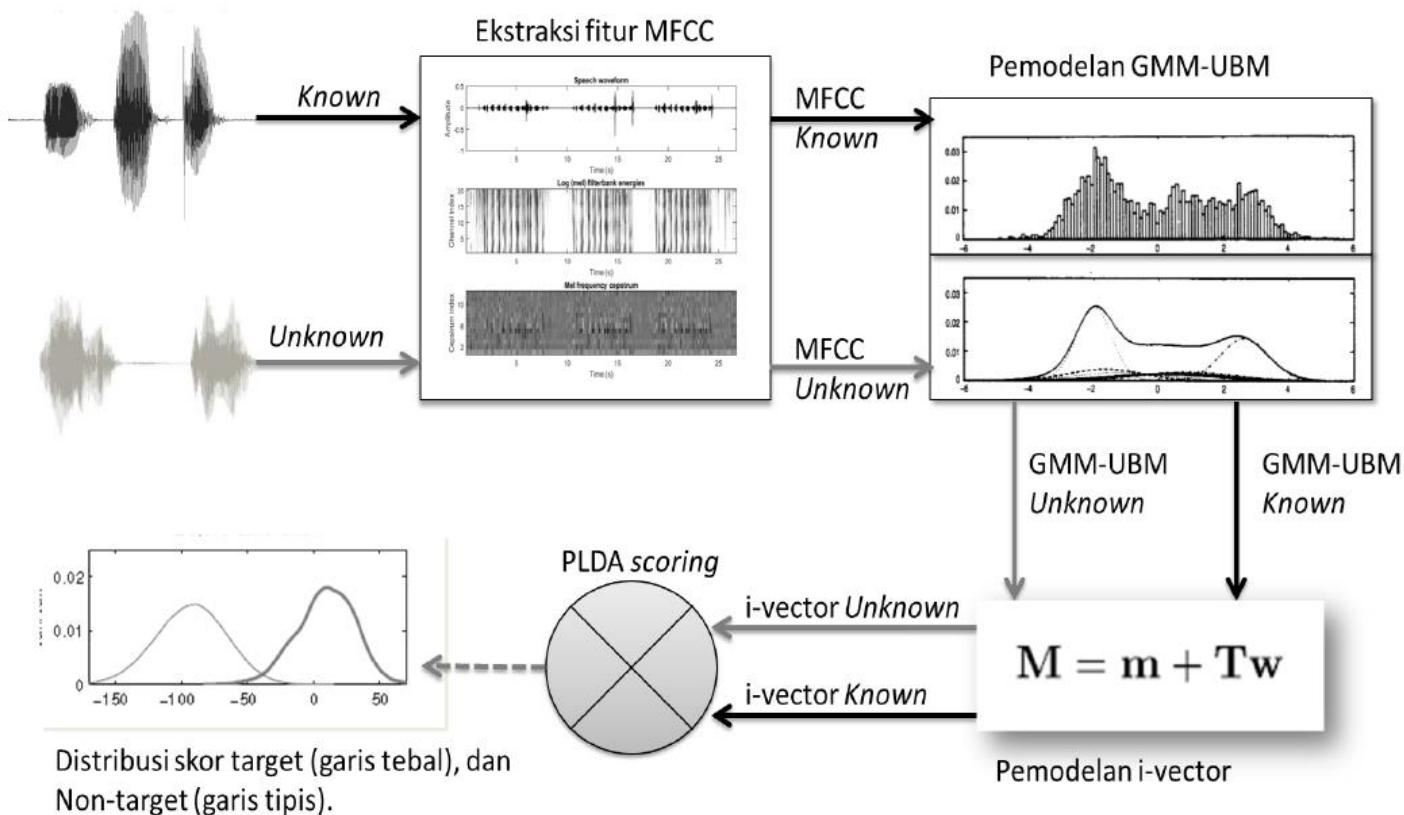
- Fitur berdasarkan properti akustik dari suara ucapan
 - *Pitch* (Frekuensi Fundamental)
 - *Formant & Bandwidth Formant*
 - Intensitas sinyal (Energi)
 - Durasi
- Analisis:
 - Menggunakan pendekatan statistik dasar
 - Level: kalimat, kata/frase, suku-kata, dan fonem
 - Biasanya: *text-dependent*

Ekstraksi Fitur Akustik: *Pitch dan Formant*



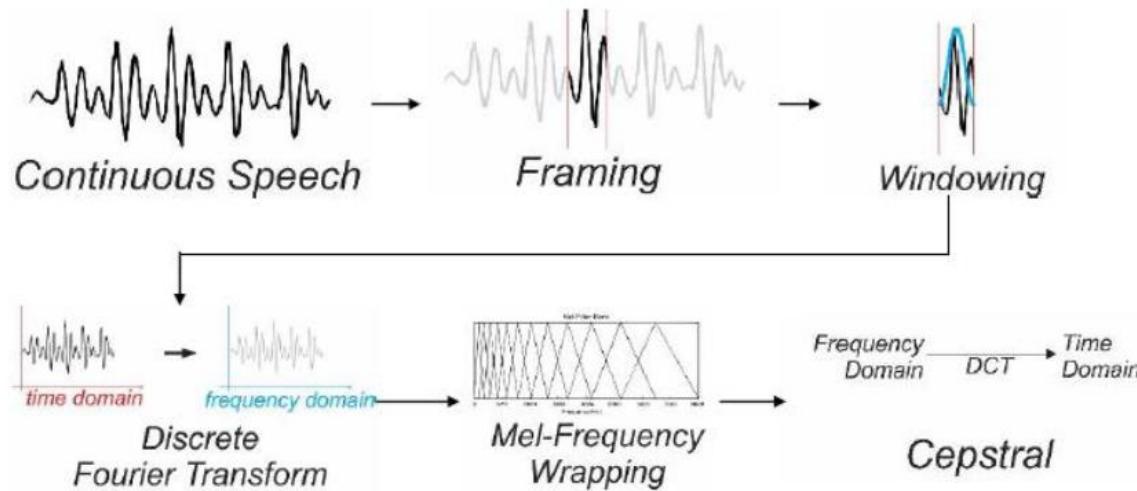
PITCH: fundamental frequency.

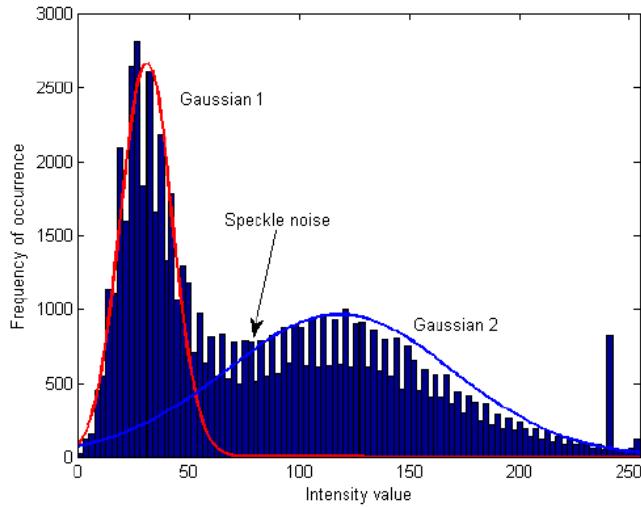
Sistem Rekognisi Pengucap Modern



MFCC Extractor

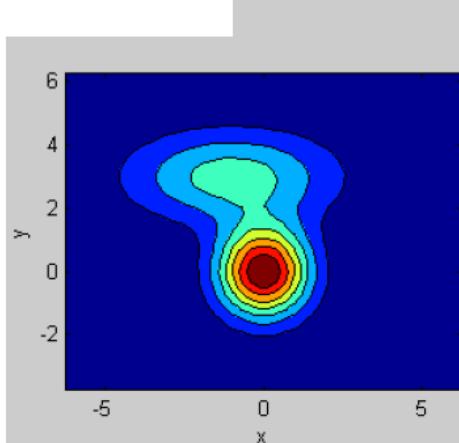
(*Mel-Frequency Cepstral Coefficient*)



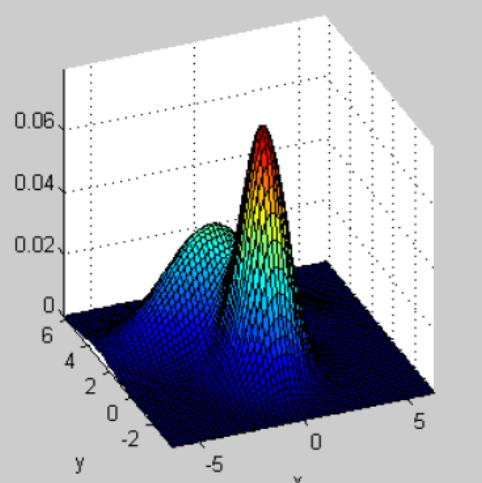


Gaussian mixture model (GMM)

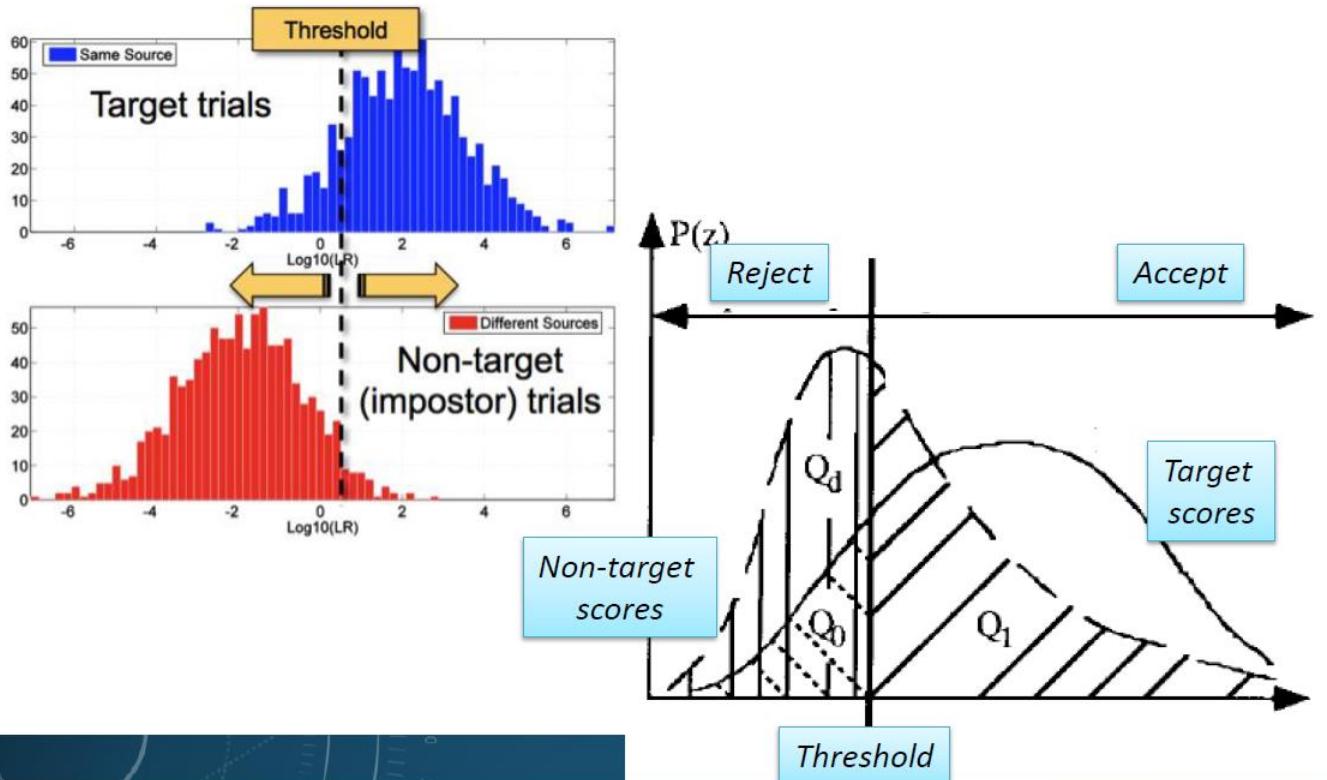
<https://i.stack.imgur.com/onXZr.png>



<https://www.quora.com/What-is-an-intuitive-explanation-of-Gaussian-mixture-models>

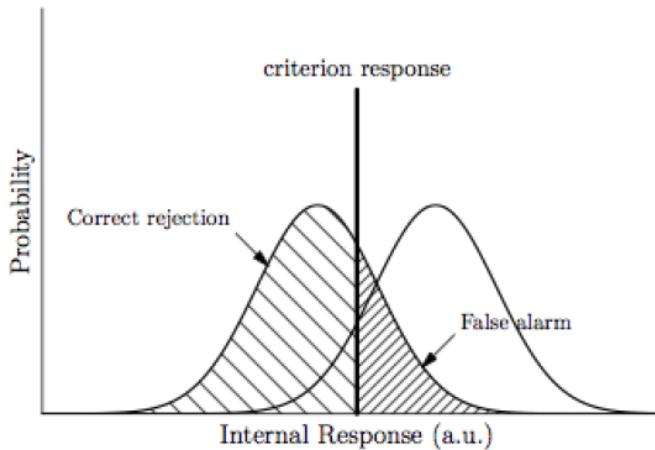
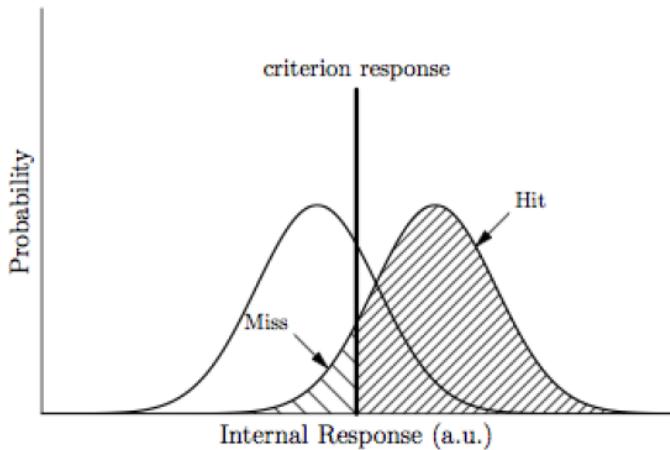
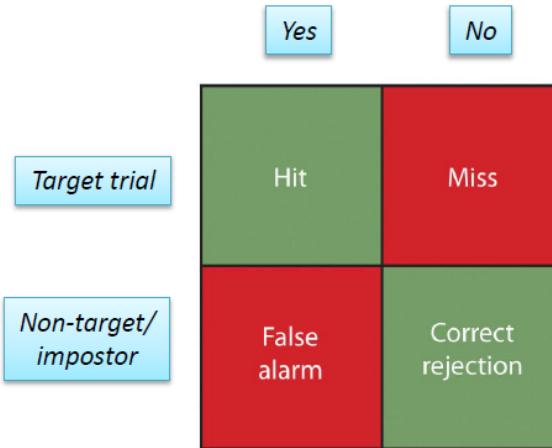


Membuat Keputusan *Binary*



Tipe Error:

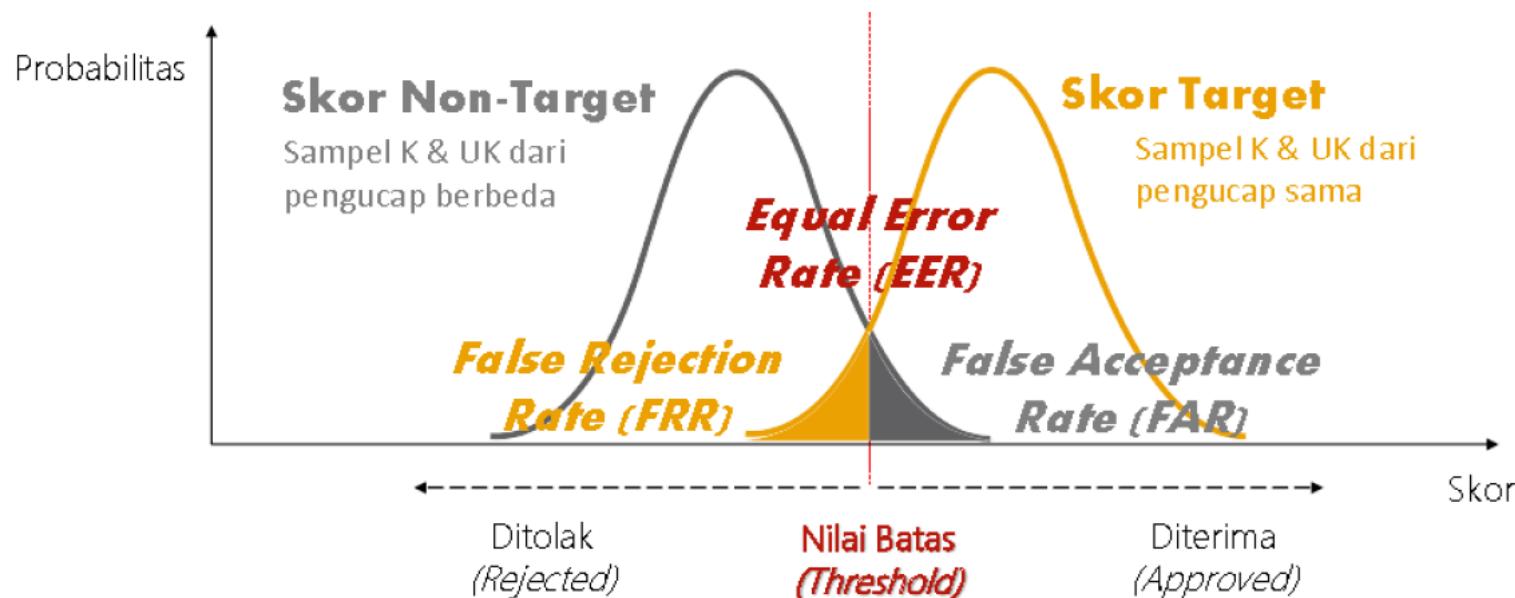
- *False alarm*
- *Miss detection*



Metrics

- EER
- Cllr

Evaluasi Sistem



Aplikasi Forensik

Komersial

- Akses pada
 - Perangkat elektronik
 - Akun Bank
 - Ruangan rahasia

- Diarisasi pengucap untuk:
 - Notulensi otomatis saat rapat

Forensik

- Verifikasi pengucap
 - Verifikasi suara tersangka

- Identifikasi pengucap
 - Mencari identitas tersangka

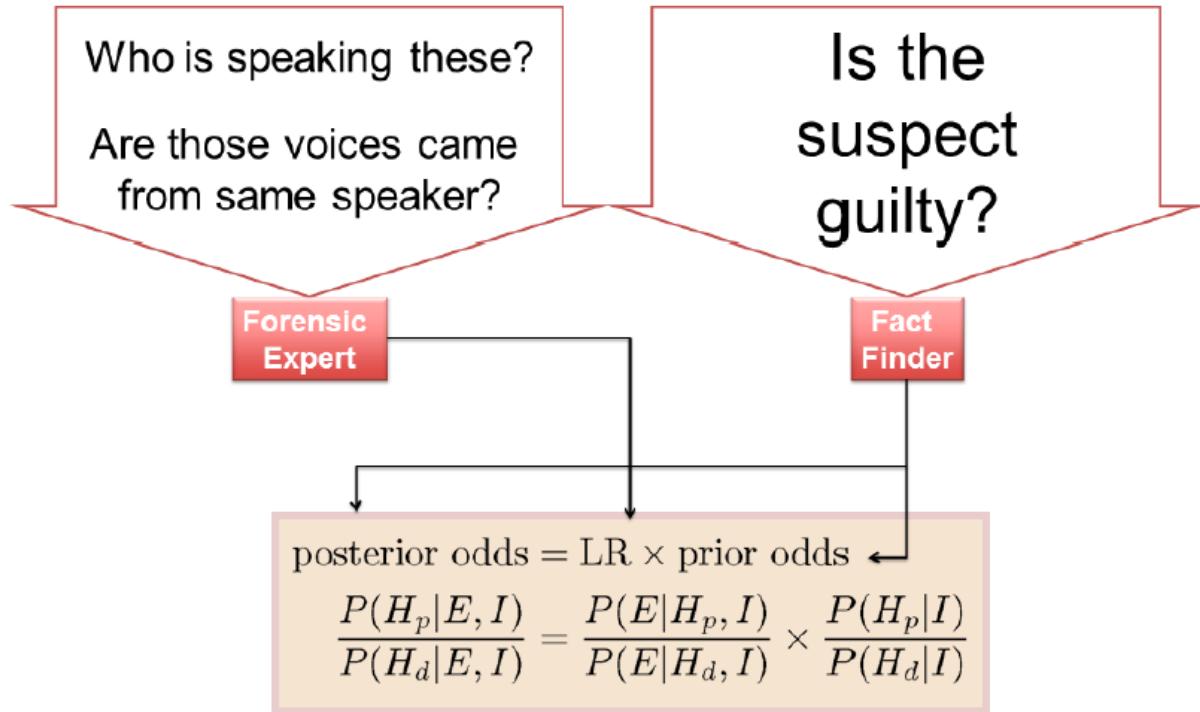
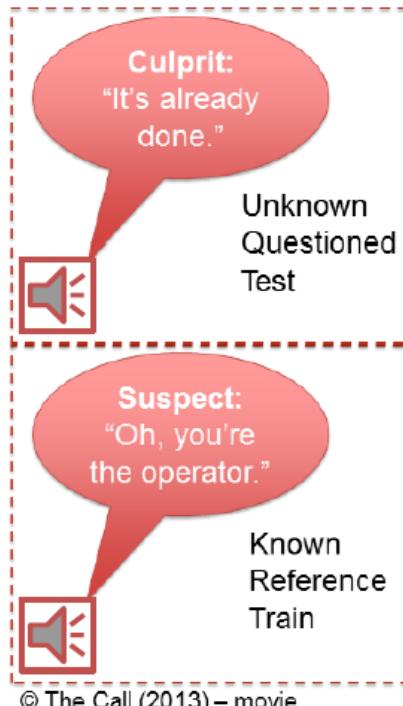
- Diarisasi pengucap
 - Pembuatan transkrip otomatis

Aplikasi dari
Sistem Rekognisi Pengucap

Forensik Suara Ucap

- Aplikasi sistem rekognisi pengucap untuk aplikasi forensik:
 - *Speaker profiling:*
 - Sistem rekognisi aksen
 - Sistem rekognisi bahasa
 - Sistem rekognisi gender
 - Mengetahui konten percakapan
 - Sistem rekognisi suara ucap (*speech recognition*)
 - Identitas pengucap – Siapa yang berbicara?
 - Sistem rekognisi pengucap (*speaker recognition*)

Rekognisi Pengucap Forensik



Membuat Keputusan Forensik

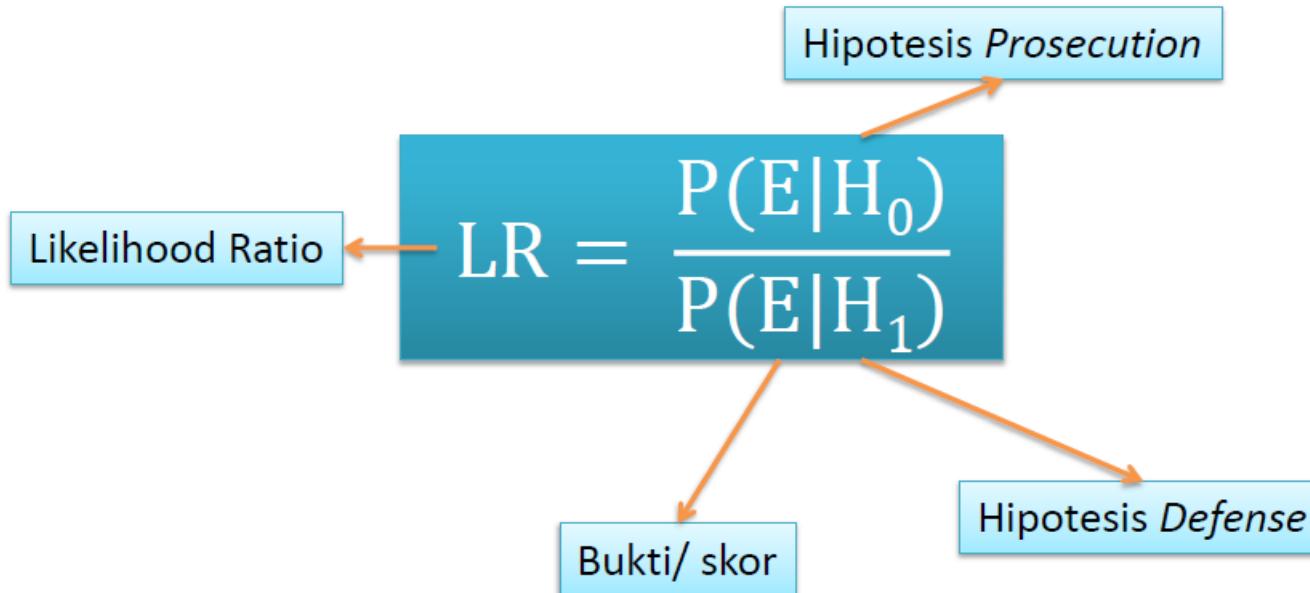
Menggunakan kerangka *likelihood ratio*

<i>Posterior knowledge</i>	<i>Evidence</i>	<i>Prior knowledge</i>
$\frac{P(H_0 E)}{P(H_1 E)}$	$= \frac{P(E H_0)}{P(E H_1)} \times$	$\frac{P(H_0)}{P(H_1)}$
<i>Posterior odds</i>	<i>Likelihood ratio</i>	<i>Prior odds</i>
Wewenang Pengadilan	Wewenang Expert	Wewenang Pengadilan

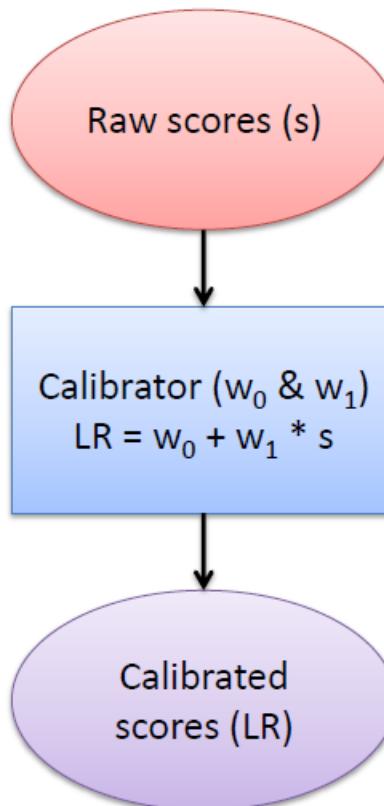
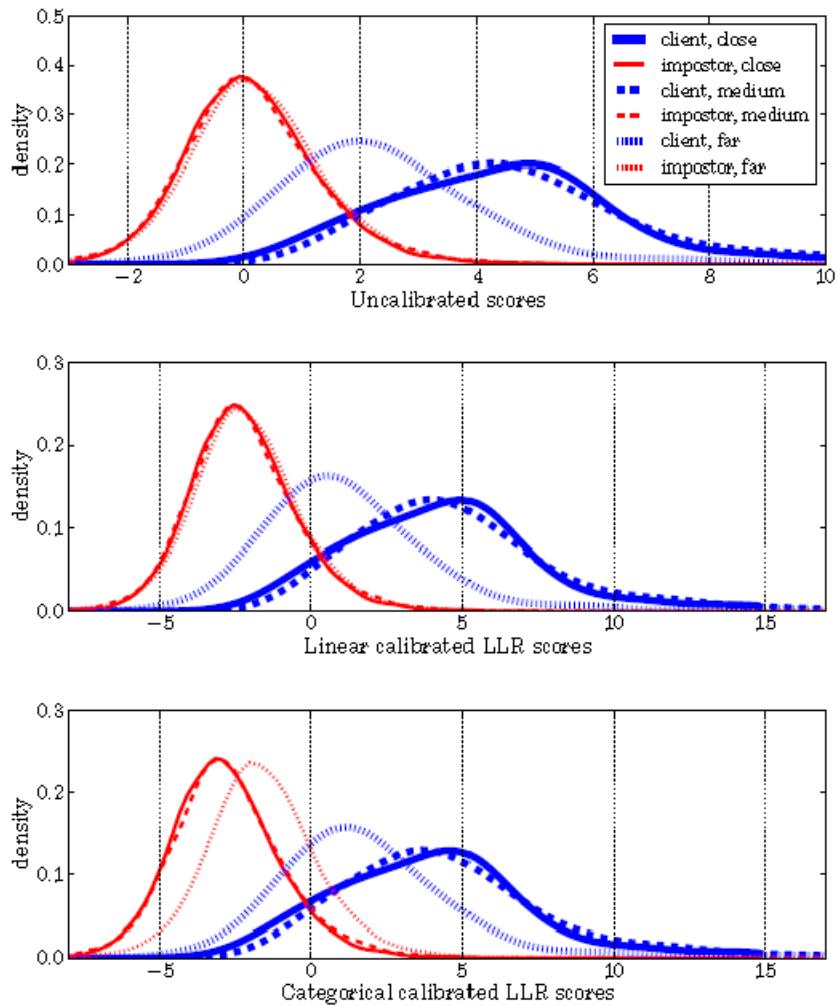
Likelihood Ratio

Likelihood ratio (LR) adalah probabilitas relatif dari suatu bukti/skor E terhadap kedua hipotesis:

- H_0 : suara unknown dan known berasal dari pengucap yang sama, dan
- H_1 : suara unknown dan known berasal dari pengucap yang berbeda



Kalibrasi



SPEAKER RECOGNITION SYSTEMS
IN FORENSIC CONDITIONS:

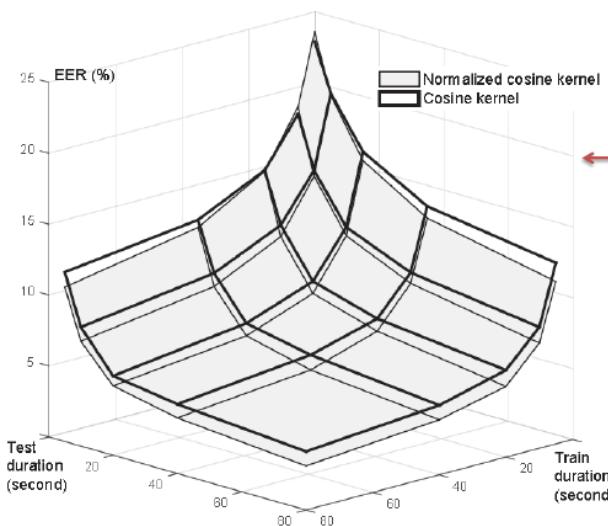
THE CALIBRATION AND EVALUATION
OF THE LIKELIHOOD RATIO



MIRANTI INDAR MANDASARI

Study #1: Modern automatic speaker recognition systems evaluation in various forensically relevant conditions

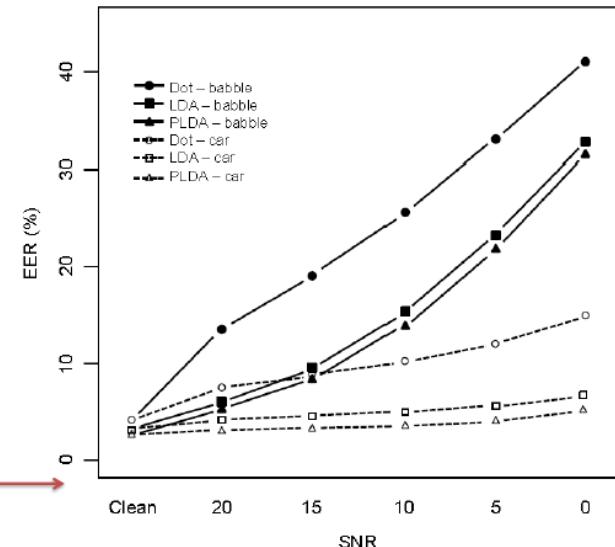
Short-duration



Symmetrical behavior in the performance of i-vector based system.

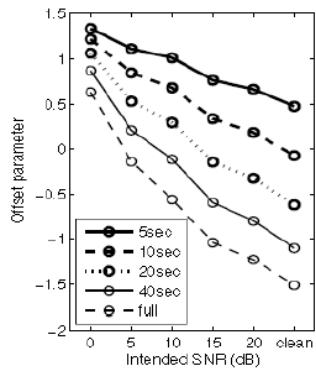
Noise robustness progresses along with improved algorithms.

Noisy speech

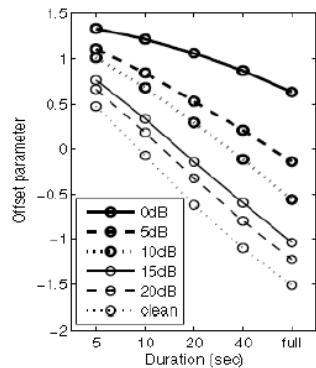


Study #2: Enhancing the system performance and reliability in the aforementioned forensic conditions

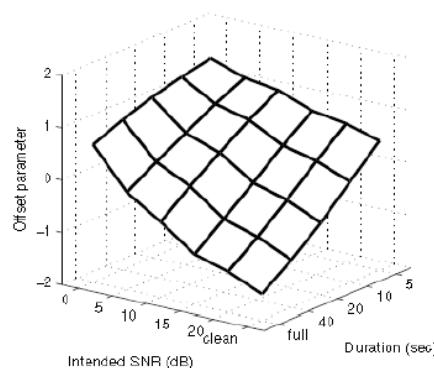
Quality Measure Function (QMF)



(a) Per duration



(b) Per SNR level



(c) Offset parameters in 2D

$$\begin{array}{l}
 \text{Linear calibration} \\
 \text{QMF factor} \\
 \downarrow \quad \downarrow \\
 x = w_0 + w_1 s + Q(d_m, d_t, w_2, \dots) \\
 \downarrow \quad \downarrow \\
 \text{Offset & Scaling parameters}
 \end{array}$$

$$Q_d = w_0 + w_d \log_2(d)$$

$$Q_n = w_0 + w_n \gamma$$

$$Q_1 = Q_d + Q_n$$

$$= w_0 + w_d \log_2(d) + w_n \gamma$$

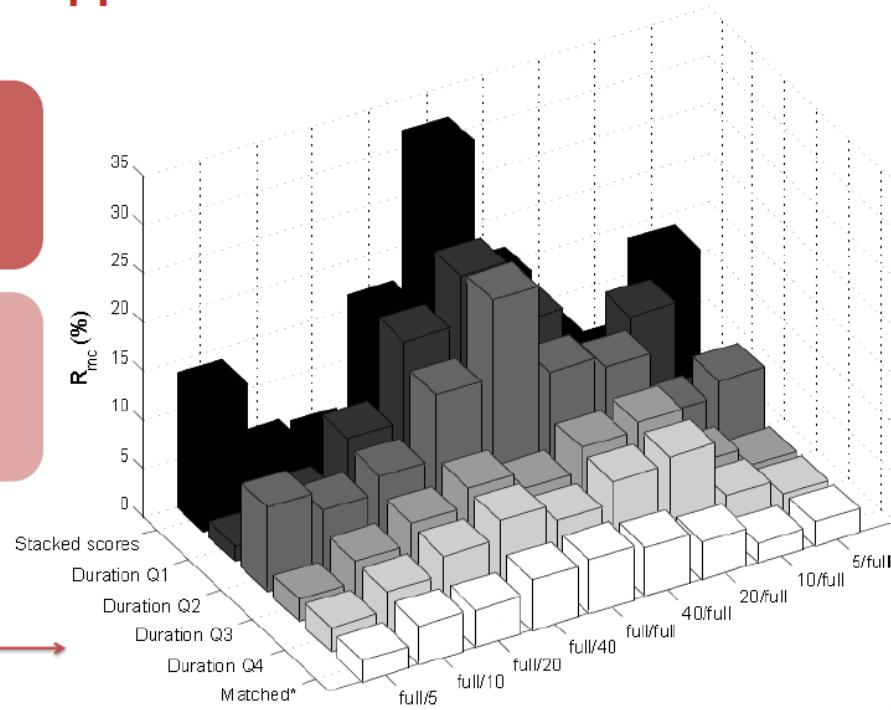
$$Q_2 = Q_d + Q_n + \alpha Q_d Q_n$$

$$= w_0 + w_d \log_2(d) + w_n \gamma + w_{dn} \gamma \log_2(d)$$

Study #3: System performance evaluation after the application of the proposed QMF calibration approach

QMF handles the duration and noise variability problems well

QMF works in extrapolation or unseen data condition



in conclusion

Kondisi Saat ini

- Kebanyakan: *text-* dan *channel-independent*
- Beberapa: *gender-* dan *language-independent*
- Pada kondisi terkontrol:
 - Performa baik, *equal error rate* $\leq 1.00\%$
- Cara meningkatkan performa:
 - *Fusion* pada saat ekstraksi fitur, pemodelan, atau pada level skor
 - Menggunakan *multi-modal biometrics* (*face*, *fingerprints*, etc.)
- Aplikasi pada dunia forensik:
 - Masih terbatas
 - Memerlukan proses **kalibrasi** *likelihood ratio*

Tantangan

- Tantangan utama: kondisi *mismatched* antara segmen suara ucapan *known* dan *unknown*:
 - *Channel* atau media perekaman,
 - *Level noise*,
 - Durasi,
 - Kata dan/atau bahasa yang terucap,
 - Kondisi emosi dan kesehatan pengucap,
 - Gaya berbicara, dan
 - etc.
- Tantangan lainnya: *noise robustness*, suara ucapan yang tersamarkan (*disguised speech*: *whisper*, *vocal effort*, etc.) & *voice aging*

e-mail me at:

miranti.indar.mandasari@gmail.com