# Pengenalan Pembicara dengan Ekstraksi Ciri MFCC Menggunakan Kuantisasi Vektor (VQ)

**Yoyo Somantri & Erik Haritman**
**dosen tek elektro fptk UPI.**

# Tujuan

Tujuan dari penelitian ini adalah untuk mengidentifikasi pembicara berdasarkan ekstraksi ciri ucapan *mel frequency cepstrum coeffisient* menggunakan *vector quantization.*

# Batasan Masalah

Masalah yang diteliti pada penelitian ini dibatasi pada :

- Ekstraksi ciri menggunakan *Mel Frequency Cepstral Coefficients* (MFCC)
- Menggunakan *text dependent speaker*.
- Algoritma pemecahan masalah menggunakan Kuantisasi Vektor.

# Pemodelan

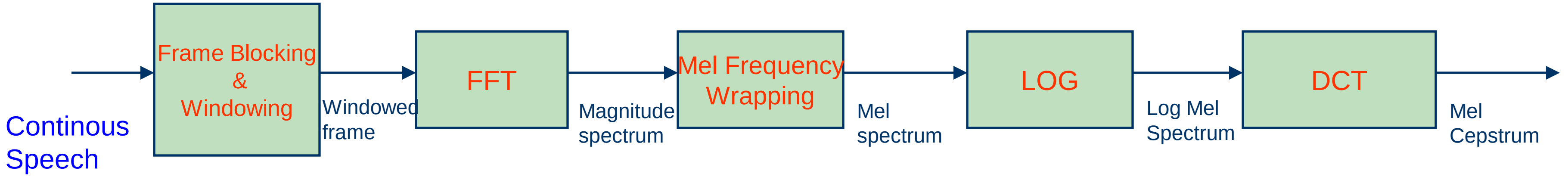

Nama Pembicara

# Dasar Teori

*Mel Frequency Cepstrum Coefficients* (MFCC)

Model ini didasarkan pada bukti-bukti bahwa komponen frekuensi rendah dari sinyal ucapan lebih penting dibandingkan dengan komponen frekuensi tingginya.

# Diagram Blok Proses MFCC




Continous Speech

$$w(n) = 0{,}54 - 0{,}46\cos\left(\frac{2\pi n}{N-1}\right),$$
$$0 \le n \le N-1$$

Hamming Window

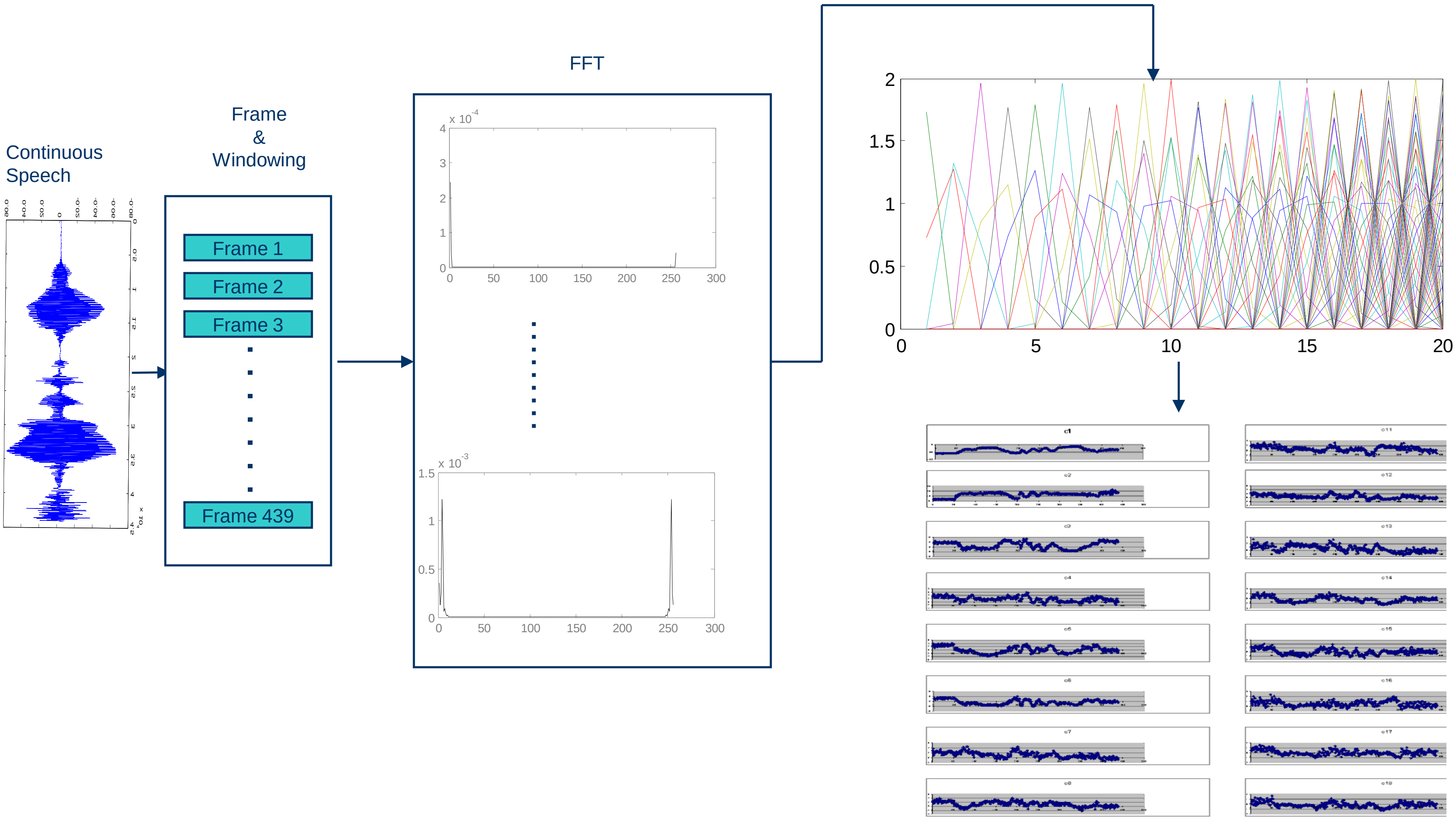
$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N},$$
$$n = 0,1,2,\ldots, N-1$$

$$mel(f) = 2595 * \log 10(1 + f/700)$$

$$y(k) = w(k)\sum_{n=1}^{N} x(n)\cos\frac{\pi(2n-1)(k-1)}{2N},$$
$$k = 1,\ldots, N$$

*Dimana*

$$w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k=1 \\ \sqrt{\frac{2}{N}}, & 2 \le k \le N \end{cases}$$



Continuous Speech
Frame & Windowing
Frame 1
Frame 2
Frame 3
Frame 439
FFT
x 10^-4
x 10^-3
0
50
100
150
200
250
300
0
5
10
15
20
0.5
1
1.5
2

# Vector Quantization

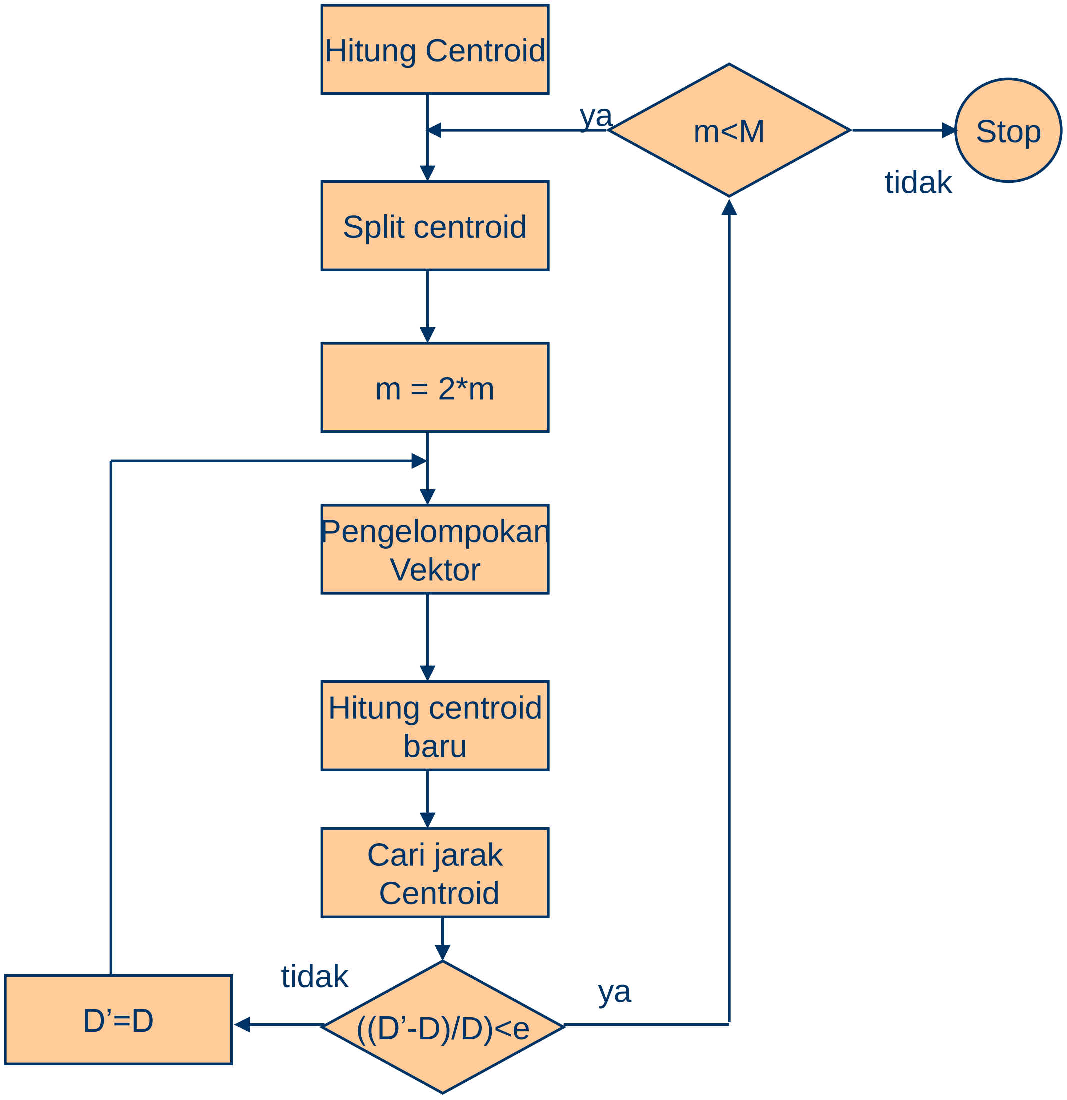
*Vector quantization* adalah proses pemetaan vektor dari ruang vektor ke bagian dari ruang vektor. Bagian dari ruang vektor ini disebut sebagai *cluster* dan direpresentasikan dengan pusat vektor atau *centroids*. Suatu himpunan *centroids* yang merepresentasikan seluruh ruang vektor disebut *codebook.*

# Clustering vektor training

Vektor akustik diekstrak dari masukan sinyal ucapan pembicara dan membentuk suatu himpunan vektor training.

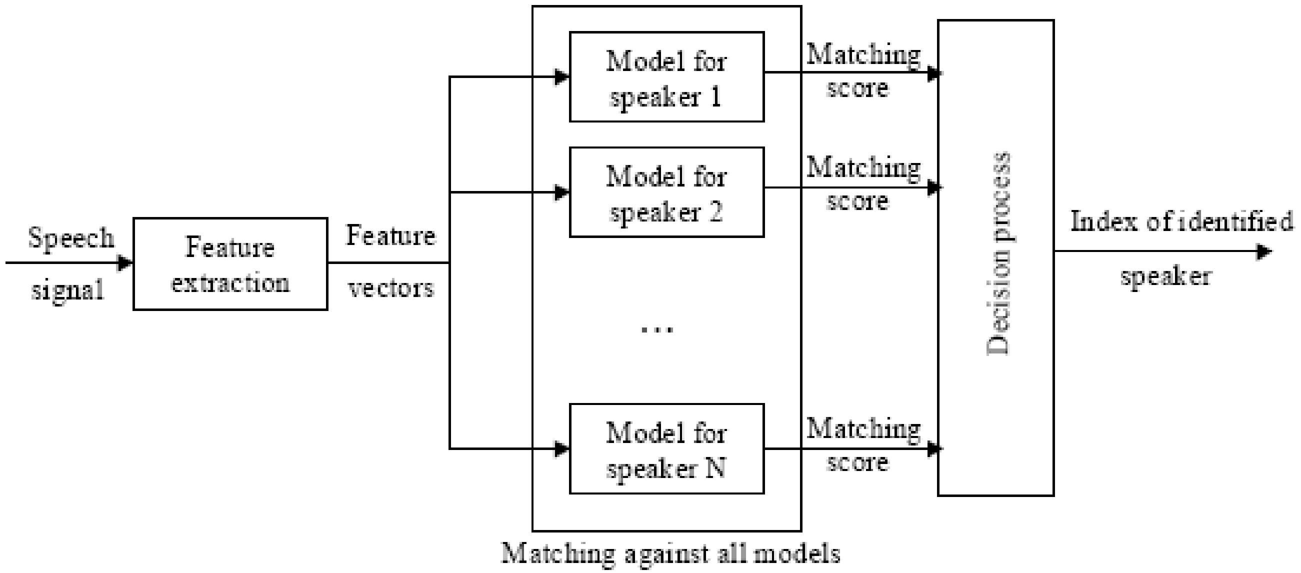
Algoritma LBG(Linde, Buzo dan Gray,1980)digunakan untuk clustering himpunan vektor training L kedalam himpunan vektor codebook M.

# Algotitma LBG

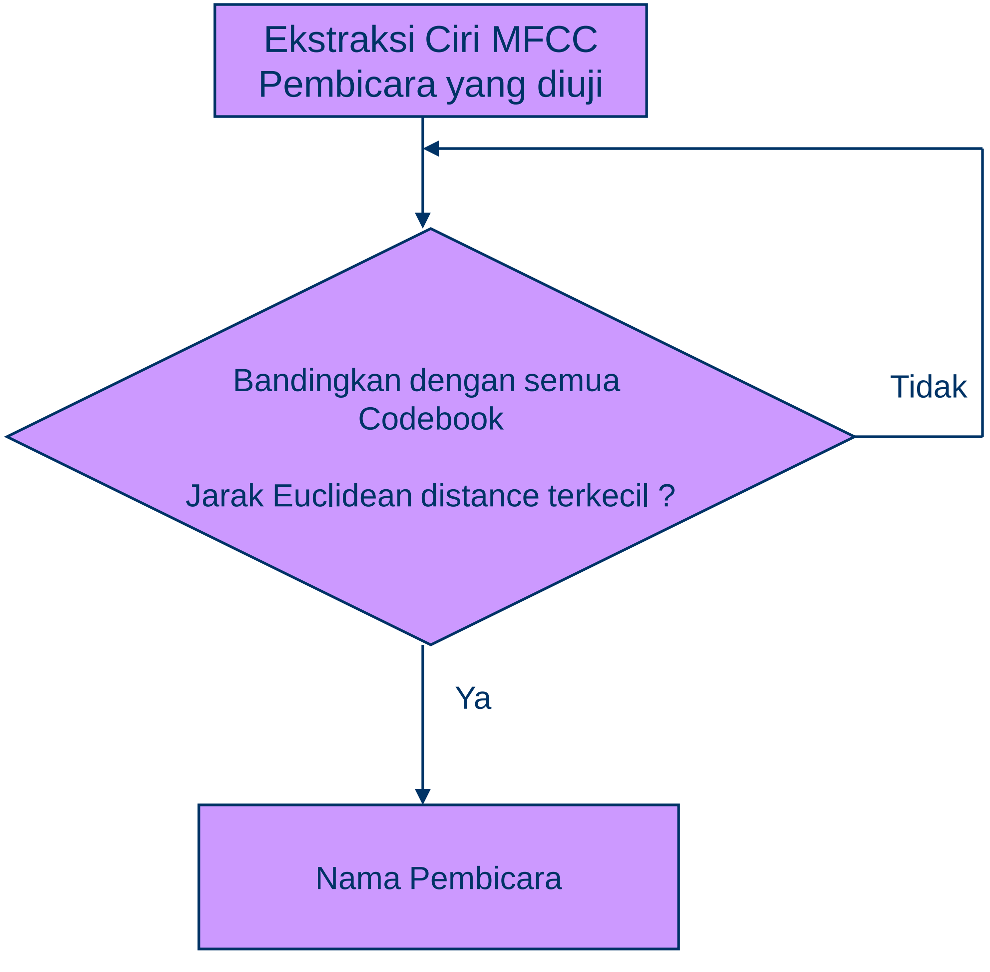


# Speaker Identifikasi menggunakan VQ

Pembicara yang dikenal berdasarkan harga distorsi kuantisasi yang paling minimum/ harga jarak Euclidean distance



# Diagram Alir Proses identifikasi



# Data Pembicara

| No | Identitas | Jenis Kelamin | Usia | Daerah asal | Sample 1 | Sample 2 | Sample 3 |
|---|---|---|---|---|---|---|---|
| 1 | Pembicara 1 | Pria | 25 | jabar | | | |
| 2 | Pembicara 2 | Pria | 30 | jakarta | | | |
| 3 | Pembicara 3 | Pria | 25 | jabar | | | |
| 4 | Pembicara 4 | wanita | 30 | Jabar | | | |
| 5 | Pembicara 5 | wanita | 18 | jabar | | | |
| 6 | Pembicara 6 | pria | 33 | jabar | | | |
| 7 | Pembicara 7 | pria | 40 | jabar | | | |
| 8 | Pembicara 8 | pria | 18 | jabar | | | |
| 9 | Pembicara 9 | pria | 18 | jabar | | | |
| 10 | Pembicara 10 | pria | 18 | jabar | | | |
| 11 | Pembicara 11 | pria | 21 | jabar | | | |
| 12 | Pembicara 12 | pria | 19 | jabar | | | |
| 13 | Pembicara 13 | pria | 19 | jabar | | | |
| 14 | Pembicara 14 | pria | 19 | jabar | | | |
| 15 | Pembicara 15 | pria | 19 | padang | | | |
| 16 | Pembicara 16 | pria | 19 | jabar | | | |
| 17 | Pembicara 17 | pria | 19 | jabar | | | |
| 18 | Pembicara 18 | wanita | 18 | jabar | | | |
| 19 | Pembicara 19 | wanita | 18 | jabar | | | |
| 20 | Pembicara 20 | wanita | 18 | jabar | | | |
| 21 | Pembicara 21 | wanita | 18 | jabar | | | |
| 22 | Pembicara 22 | wanita | 18 | jabar | | | |
| 23 | Pembicara 23 | wanita | 18 | jabar | | | |
| 24 | Pembicara 24 | pria | 18 | kalimantan | | | |
| 25 | Pembicara 25 | pria | 18 | sulawesi | | | |

# Prosedur Percobaan

- Sampel yang sama diambil pada waktu yang berbeda Sample suara ‘hallo apa kabar’ yang di pecah menjadi 2 ucapan, yaitu ‘hallo’ dan ‘apakabar’.
- Masing-masing ucapan diuji pada Kuantisasi vektor

# Hasil

Perbandingan Hasil Uji Coba Vector Quantization

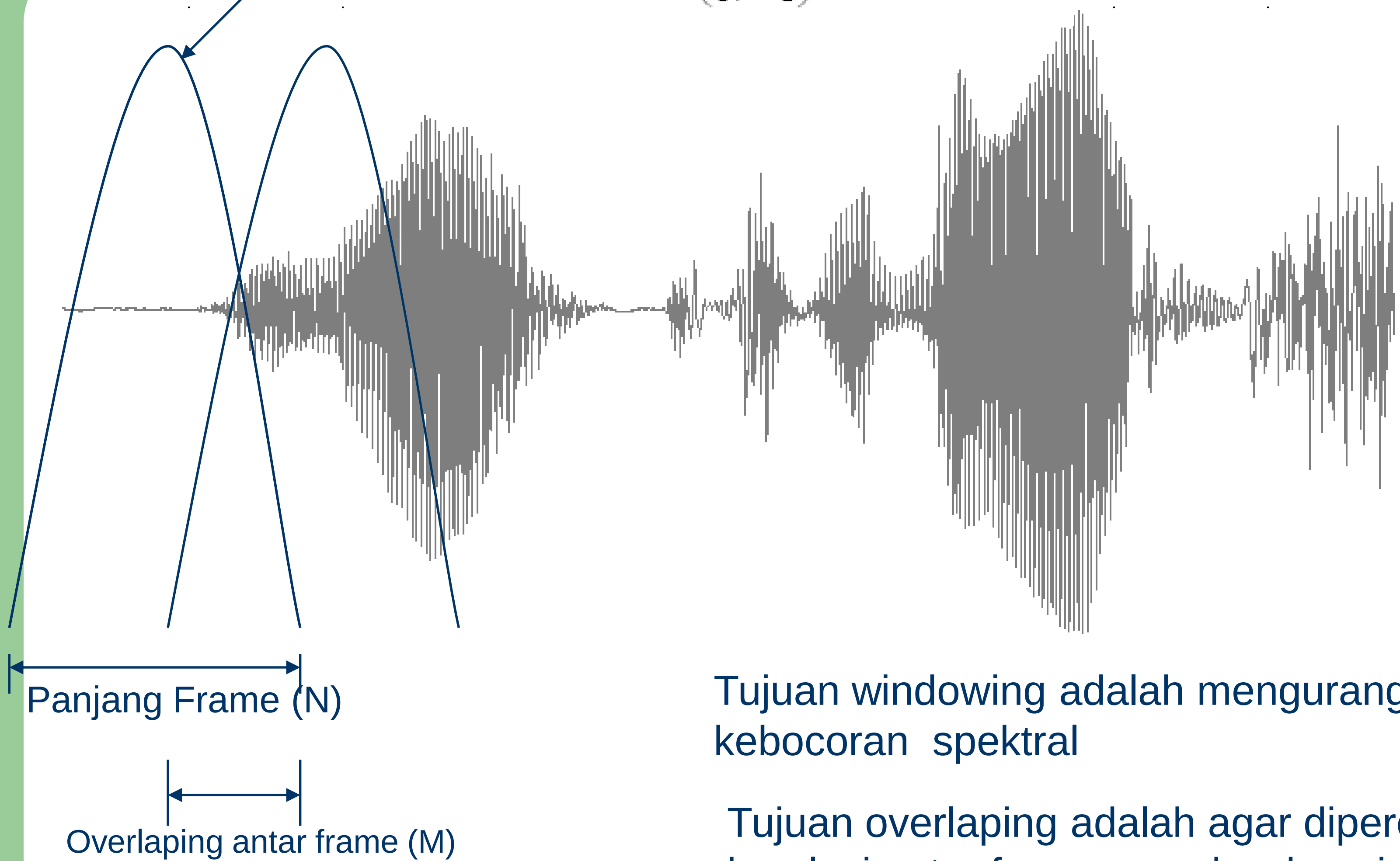
| Voice Print | 4 centroid | | | 8 centroid | | | 16 centroid | | | 32 centroid | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | t1 | t2 | t3 | t1 | t2 | t3 | t1 | t2 | t3 | t1 | t2 | t3 |
| Hallo Apa Kabar | 100% | 80% | 64% | 100% | 92% | 72% | 100% | 92% | 72% | **100%** | **96%** | **80%** |
| Hallo | 100% | 56% | 60% | 100% | 60% | 64% | 60% | 48% | 44% | **100%** | **64%** | **72%** |
| Apa Kabar | 100% | 80% | 72% | 100% | 88% | 76% | **100%** | **92%** | **80%** | 100% | 84% | 84% |

# Kesimpulan

- Ekstraksi ciri MFCC dapat digunakan masukan untuk mengenali identitas pembicara.
- Kuantisasi vektor dapat mengenali pembicara berdasarkan eksstraksi ciri MFCC dengan hasil yang cukup baik.

# FRAME BLOCKING dan WINDOWING

Window = $w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1$




Tujuan windowing adalah mengurangi kebocoran spektral

Tujuan overlaping adalah agar diperolah korelasi antar frame yang berdampingan

Pada penelitian ini dipilih N = 256 (30ms)
M = 100

Sehingga jumlah Frame = ((I-N)/M)+1
= ((44091 – 256)/100) + 1 = 439

# Fast Fourier Transform (FFT)

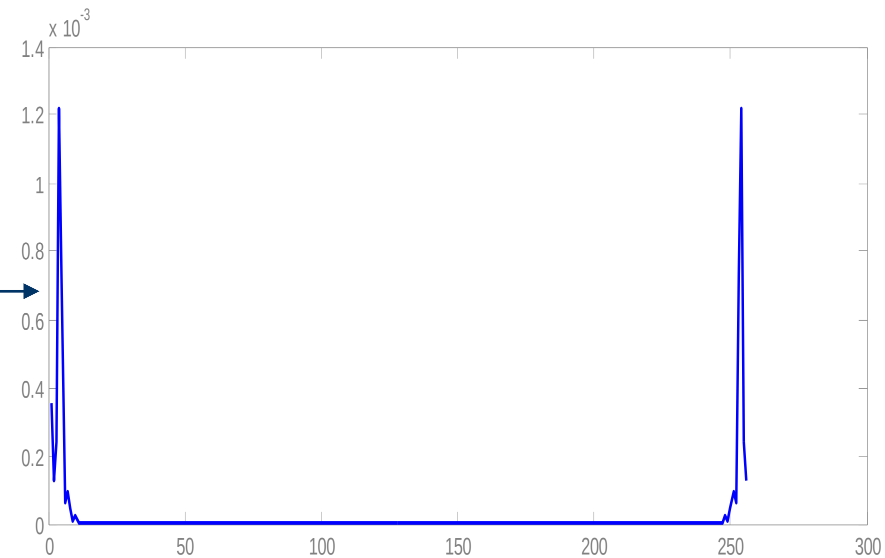
FFT akan mengkonversi masing-masing frame dari domain waktu ke domain frekuensi. FFT adalah sebuah algoritma yang cepat untuk mengimplementasikan Discrete Fourier Transform (DFT).

FRAME 439 →

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N},$$

$$n = 0,1,2,\ldots, N-1$$

→



# Mel Frequency Wrapping

Skala mel frequency adalah frekuensi linier pada daerah di bawah 1 KHz dan logaritmik untuk daerah di atas 1 KHz. Sebagai pendekatan diberikan formula sebagai berikut :

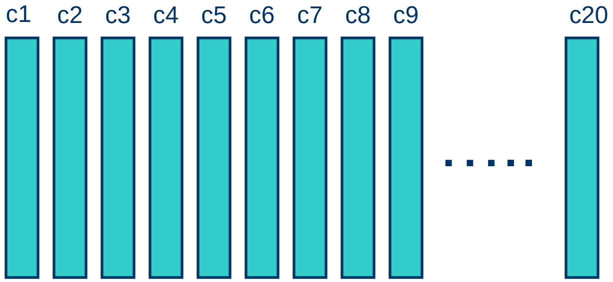
$$mel(f) = 2595 * \log_{10}(1 + f / 700)$$

# cepstrum

Spektrum Log mel dikonversi kembali ke dalam waktu. Hasilnya disebut sebagai mel frequency cepstrum coefficients (MFCC). Oleh karena koefisien mel spekrum merupakan bilangan real, kita dapat mengkonversinya dalam domain waktu  menggunakan discrete cosine transform (DCT) , kita dapat menghitung MFCC sebagai :

$$\tilde{c}_n = \sum_{k=1}^{K} (\log \tilde{S}_k) \cos\left[ n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right], \qquad n = 1,2,\ldots,K$$

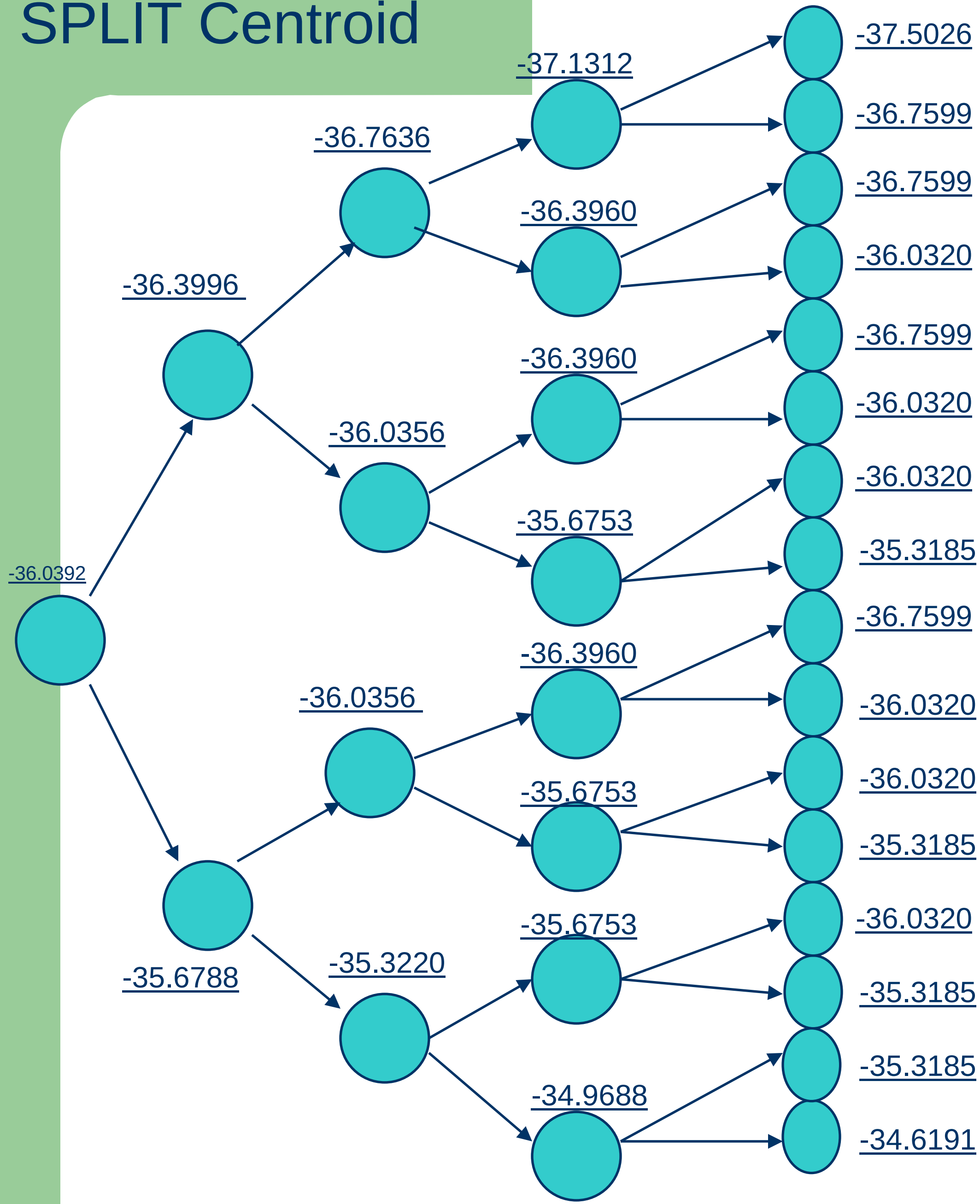
# Menghitung Centroid




Rata-rata dari masing-masing MFCC =

-36.0392
13.0268
1.6690
1.7284
0.0889
-0.7103
1.1984
1.0933
-0.7756
0.5638
0.8033
0.5879
0.6014
-0.2787
0.0177
0.5424
0.0506
-0.0939
0.2578
-0.0676

# SPLIT Centroid



# Pengelompokan vektor

- Pengelompokan berdasarkan jarak terdekat antara vektor ciri(MFCC) dengan centroid yang telah displit.
- Menggunakan Euclidean Distance :

$$d(i,j) = \sqrt{(|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 + ... + |x_{i_p} - x_{j_p}|^2)}$$

Misal vektor ciri MFCC

|  | | | | | | |
|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 | = c1 |
| V = | 6 | 7 | 8 | 9 | 10 | = c2 |
|  | 11 | 12 | 13 | 14 | 15 | = c3 |

Centroid masing-masing coeffisient yaitu :

$$r = \begin{matrix} 3 \\ 8 \\ 13 \end{matrix}$$

Centroid kita split 2x (4 code), maka centroid menjadi :

$r^{+} = r(1 + \varepsilon)$

$r^{-} = r(1 - \varepsilon)$

dipilih $\varepsilon = 0{,}01$

| | | | | |
|---|---|---|---|---|
|  | 3.0603 | 2.9997 | 2.9997 | 2.9403 |
| r = | 8.1608 | 7.9992 | 7.9992 | 7.8408 |
|  | 13.2613 | 12.9987 | 12.9987 | 12.7413 |

Cari jarak antar matrik kolom menggunakan rumus Euclidean Distance :

$$d(i,j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \ldots + |x_{ip} - x_{jp}|^2)}$$

| | | | | | |
|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 |
| V = | 6 | 7 | 8 | 9 | 10 |
|  | 11 | 12 | 13 | 14 | 15 |

dengan

| | | | | | |
|---|---|---|---|---|---|
|  | 3.0603 | 3.0603 | 3.0603 | 3.0603 | 3.0603 |
| r(1) = | 8.1608 | 8.1608 | 8.1608 | 8.1608 | 8.1608 |
|  | 13.2613 | 13.2613 | 13.2613 | 13.2613 | 13.2613 |

$$d(1,1) = \sqrt{\left(|1-3{,}0603|^2 + |6-8{,}1608|^2 + |11-13{,}2613|^2\right)}$$

$$= \sqrt{14{,}0274} = 3{,}7453$$

$$d(2,1) = \sqrt{\left(|2-3{,}0603|^2 + |7-8{,}1608|^2 + |12-13{,}2613|^2\right)}$$

$$= \sqrt{4{,}0626} = 2{,}0156$$

$$d(3,1) = \sqrt{\left(|3-3{,}0603|^2 + |8-8{,}1608|^2 + |13-13{,}2613|^2\right)}$$

$$= \sqrt{0{,}0978} = 0{,}3127$$

$$d(4,1) = \sqrt{\left(|4-3{,}0603|^2 + |9-8{,}1608|^2 + |14-13{,}2613|^2\right)}$$

$$= \sqrt{2{,}1330} = 1{,}4605$$

$$d(5,1) = \sqrt{\left(|5-3{,}0603|^2 + |10-8{,}1608|^2 + |15-13{,}2613|^2\right)}$$

$$= \sqrt{10{,}1682} = 3{,}1888$$

$$d(\,:,1)' = \begin{matrix} 3.7453 \\ 2.0156 \\ 0.3127 \\ 1.4605 \\ 3.1888 \end{matrix}$$

Dengan cara yang sama maka diperoleh :

| | | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| V = | 6 | 7 | 8 | 9 | 10 |
| | 11 | 12 | 13 | 14 | 15 |

dengan

| | | | | | |
|---|---|---|---|---|---|
| | 2.9997 | 2.9997 | 2.9997 | 2.9997 | 2.9997 |
| R(2) = | 7.9992 | 7.9992 | 7.9992 | 7.9992 | 7.9992 |
| | 12.9987 | 12.9987 | 12.9987 | 12.9987 | 12.9987 |

| | |
|---|---|
| | 3.4627 |
| | 1.7307 |
| d ( :,2)'= | 0.0016 |
| | 1.7334 |
| | 3.4655 |

Dengan cara yang sama akan diperoleh :

| | | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| V = | 6 | 7 | 8 | 9 | 10 |
| | 11 | 12 | 13 | 14 | 15 |

dengan

| | | | | | |
|---|---|---|---|---|---|
| | 2.9997 | 2.9997 | 2.9997 | 2.9997 | 2.9997 |
| r(3) = | 7.9992 | 7.9992 | 7.9992 | 7.9992 | 7.9992 |
| | 12.9987 | 12.9987 | 12.9987 | 12.9987 | 12.9987 |

| | |
|---|---|
| | 3.4627 |
| | 1.7307 |
| d ( :,3)'= | 0.0016 |
| | 1.7334 |
| | 3.4655 |

Dengan cara yang sama akan diperoleh :

| | | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| V = | 6 | 7 | 8 | 9 | 10 |
| | 11 | 12 | 13 | 14 | 15 |

dengan

| | | | | | |
|---|---|---|---|---|---|
| | 2.9403 | 2.9403 | 2.9403 | 2.9403 | 2.9403 |
| R(4) = | 7.8408 | 7.8408 | 7.8408 | 7.8408 | 7.8408 |
| | 12.7413 | 12.7413 | 12.7413 | 12.7413 | 12.7413 |

| | |
|---|---|
| | 3.1915 |
| | 1.4631 |
| d ( :,4)'= | 0.3096 |
| | 2.0127 |
| | 3.7425 |

| | | | | |
|---|---|---|---|---|
| | 14.0274 | 11.9904 | 11.9904 | 10.1854 |
| | 4.0626 | 2.9952 | 2.9952 | 2.1406 |
| d 2 = | 0.0978 | 0.0000 | 0.0000 | 0.0958 |
| | 2.1330 | 3.0048 | 3.0048 | 4.0510 |
| | 10.1682 | 12.0096 | 12.0096 | 14.0062 |

| | | | | |
|---|---|---|---|---|
| | 3.7453 | 3.4627 | 3.4627 | 3.1915 |
| | 2.0156 | 1.7307 | 1.7307 | 1.4631 |
| d = | 0.3127 | 0.0016 | 0.0016 | 0.3096 |
| | 1.4605 | 1.7334 | 1.7334 | 2.0127 |
| | 3.1888 | 3.4655 | 3.4655 | 3.7425 |