Sistem Pengenal Tutur Bahasa Indonesia Berbasis Suku Kata Menggunakan MFCC, Wavelet Dan HMM

Syahroni Hidayat[1], Risanuri Hidayat[2], Teguh Bharata Adji[3]
Jurusan Teknik Elektro dan Teknologi Informasi
Fakultas Teknik, Universitas Gadjah Mada
Jl. Grafika no.2 Yogyakarta-55281, Indonesia
syahronihidayat788@gmail.com[1], risanuri@ugm.ac.id[2], adji@ugm.ac.id[3]

Abstract— **This paper presented the development of an automatic speech recognition (ASR) system based on Indonesian syllable using HMM classifier. The recognition rate for an ASR based on Indonesian syllable is still low. This problem might be occurs because the extraction process was applicated directly to the whole syllables. Therefore, in this research the feature extraction process is implemented to each of its constituent phonemes. MFCC and WPT were used as the feature extraction method. Feature of MFCC is obtained by applying frame of 512 sample for each phoneme. In Mel Frequency Warping process using 40 units of triangular filter banks. As for WPT feature extraction process, wavelet daubechies db3 and db7 were used with 5th level decomposition. Feature that extracted then randomly selected and established as a syllable's feature. The recognition accuracy using training data showed 100% accuracy for WPT feature, and 75% for MFCC feature. While using the external testing data the result showed the best accuracy are 100% for WPT db7, 83.33% for WPT db3 and 50% for MFCC. Whole best recognition results were obtained at the point of intersection for the consonants are of 1024 samples.**

Keywords-ASR; MFCC; Wavelet Packet Transform; HMM; Indonesian Syllable

Abstrak— **Pada paper ini disajikan tentang pengembangan sebuah sistem pengenal suara otomatis bahasa Indonesia berbasis suku kata menggunakan HMM. Pengenalan suara berbasis suku kata bahasa Indonesia masih memberikan akurasi yang sangat rendah. Salah satu penyebabnya adalah metode ekstraksi ciri secara langsung terhadap seluruh suara suku kata. Oleh karena itu, pada penelitian ini proses ekstaksi ciri suara suku kata diterapkan terhadap masing-masing fonem penyusunnya. Metode ekstraksi menggunakan metode Mel Frequency Cepstral Coefficient (MFCC) dan Wavelet Packet Transform (WPT). Ciri MFCC diperoleh dengan menerapkan frame sebesar 512 sampel terhadap masing-masing fonem. Pada proses Mel Frequency Warping digunakan 40 buah tapis segitiga. Adapun pada proses WPT digunakan wavelet daubechies db3 dan db7 dengan level dekomposisi 5. Ciri hasil ekstraksi kemudian dipilih secara random dan dibentuk sebagai ciri suku kata. Hasil pengenalan dengan data pelatihan menunjukkan akurasi terbaik sebesar 100% untuk metode WPT dan 75% untuk metode MFCC. Sedangkan dengan menggunakan data pengujian hasil akurasi terbaiknya adalah 100% untuk WPT db7, 83.33% untuk WPT db3 dan 50% untuk MFCC. Seluruh hasil pengenalan terbaik ini diperoleh pada titik potong panjang sampel konsonan sebesar 1024 sampel.**

Kata kunci- ASR, MFCC, Wavelet Packet Transform, HMM, Suku Kata Bahasa Indonesia

I. PENDAHULUAN

Penelitian tentang *Automatic Speech Recognition* (ASR) telah mencapai tingkat keberhasilan yang sangat tinggi. Namun pencapaian tersebut tidak berlaku untuk seluruh bahasa di dunia, salah satu contoh adalah bahasa Indonesia. Selain penelitian masih sangat terbatas, tingkat keberhasilannya pun masih belum sangat memuaskan. Pada [1]–[3] penelitian masih berkisar pada pembuatan Large Vocabulary Continuous Speech Recognition (LVCSR) bahasa Indonesia dan masih terkendala oleh masalah *Out Of Vocabulary* (OOV). Adapun [4], [5] melakukan penelitian tentang ASR terisolasi bahasa Indonesia yang terbatas untuk vokal dan beberapa kosa kata tertentu.

Bahasa Indonesia yang merupakan bahasa ibu Negara Kesatuan Republik Indonesia memiliki pola bahasa yang berbeda dengan bahasa Negara lain. Ia tidak terikat oleh penggunaan gender dan gaya suara. Sehingga bahasa Indonesia termasuk anggota bahasa aglutinasi, yang berarti memiliki sufiks dan prefiks yang kompleks yang melekat pada kata dasar. Kata dasar dalam bahasa Indonesia memiliki pola suku kata yang sederhana, umumnya berpola V, VK, KV dan KVK [6], [7].

Dalam penuturan sebuah kata bahasa Indonesia setiap suku katanya akan diiringi oleh sebuah hembusan nafas. Hal ini memungkinkan adanya jeda antar suku kata dan menghasilkan representasi sinyal suara kata bahasa Indonesia yang berbeda untuk tiap suku katanya. Sehingga suara tutur bahasa Indonesia dapat dikenali pada tingkat suku kata. Abriyono [7] telah melakukan penelitian ASR untuk mengubah suara suku kata bahasa Indonesia menjadi tulisan, namun akurasinya sangat rendah. Sedangkan Suyanto [8] mengajukan sebuah desain baru untuk membuat LVCSR bahasa Indonesia, yaitu berdasarkan pada penggabungan fonem dan suku kata. Meninjau pada kenyataan bahwa 92% kosa kata bahasa Indonesia tersusun dari 4 pola suku kata tersebut diatas, maka diharapkan database LVCSR yang begitu besar dapat digantikan dengan database suku kata yang lebih kecil namun dapat mencakup seluruh kata.

Tingkat pengenalan dipengaruhi oleh metode ekstraksi dan klasifikasi yang digunakan. Sejauh ini metode ekstraksi MFCC lebih banyak digunakan, seperti yang telah dilakukan oleh [1]–[4], [7] dan [9]. Adapun metode wavelet telah diaplikasikan oleh [5], [10] dan [11]. Kedua metode tersebut masing-masing memiliki kekurangan dan kelebihan seperti yang telah dijelaskan [12]. Aplikasi kedua metode tersebut juga masih berkisar pada fonem,

kosa kata dan kalimat. Adapun aplikasi MFCC pada domain suku kata telah dilakukan oleh [7] untuk bahasa Indonesia. Namun hasil pengenalannya masih sangat rendah. Sehingga perlu diterapkan metode ekstraksi lain yang dapat meningkatkan akurasi pengenalan suku kata, yaitu dengan metode ekstraksi wavelet.

Sedangkan untuk klasifikasi, *Hidden Markov Models* (HMM) yang bekerja berdasarkan proses stokastik memberikan hasil yang paling baik karena kemampuannya mengenali variabilitas data yang sangat kecil, seperti yang telah dilakukan oleh [1]–[3], [9]. Namun penggunaan *Gaussian Mixture Models* (GMM) sebagai model distribusi probabilitas observasi masih meninggalkan masalah pada masalah komputasi. Sehingga Budi [9] dan Buono [11] menggunakan probabilitas jarak Euclid sebagai pengganti GMM.

Pada penelitian ini dikembangkan sistem ASR bahasa Indonesia berbasis suku kata dengan metode ektraksi ciri MFCC dan wavelet. Klasifikasi menggunakan HMM dengan menerapkan metode probabilitas jarak Euclid.

II. METODOLOGI

A. *Voice Activity Detection* (VAD)

Sebelum melakukan proses ekstraksi ciri, sangatlah penting untuk menghilangkan derau yang menyertai sinyal suara. Salah satu cara yang dapat dilakukan adalah mendeteksi aktivitas suara dengan cara menghitung energi dari sinyal suara tersebut untuk membedakan antara sinyal derau dengan sinyal suara. Energi sinyal dihitung dengan persamaan berikut[7].

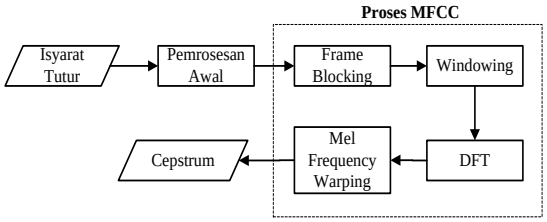
$$E = \frac{1}{N}\sum_{n-1}^{N}|x(n)| \quad (1)$$

Algoritma untuk deteksi aktivitas suara dijelaskan sebagai berikut.

- Sampel isyarat suara diblok dengan panjang *frame* (lfr) 4 ms dengan pergeseran (sfr) 2 ms.
- Kalikan masing-masing *frame* dengan jendela kotak w_{rec}.
- Hitung energi per-*frame* (EF).
- Tenentukan nilai ambang (*threshold*) TH dengan mengalikan nilai maksimum EF dengan bobot tertentu. Besar bobot bergantung pada nilai EF maksimum. Jika $EF_{max} > 1$ maka bobot = $10^{-2}/4$, namun jika $EF_{max} < 1$ maka bobot = $10^{-1}/4$. Nilai ambang terbobot ini membuat nilai ambang dinamis, sehingga memungkinkan pemotongan terjadi pada titik yang tepat.
- Tenentukan *frame-frame* yang berisi isyarat suara dengan membandingkan besar EF dengan TH. Jika EF > TH maka EF = 1, namun jika EF < TH maka EF = 0.
- Menentukan jumlah suku kata (SK), indeks awal (I_{awal}) dan Indeks akhir (I_{akhir}) dengan cara berikut:
  a. Tentukan SK = 0.
  b. Mulai dari *frame* ke-2 sampai akhir, cari nilai *frame* yang memenuhi kriteria EF(i-1) = 1 dan EF(i) = 0.
  c. Jika memenuhi, maka SK = SK + 1 dan I_{akhir} = i.
  d. Untuk indeks awal (I_{awal}) , tentukan SK = 0.
  e. Mulai dari *frame* ke-2 sampai akhir, cari nilai *frame* yang memenuhi kriteria EF(i-1) = 0 dan EF(i) = 1.
  f. Jika memenuhi, maka SK = SK + 1 dan I_{akhir} = i. Untuk menghindari kehilangan data karena posisi awal isyarat suara yang terlalu di awal dan di akhir maka dapat ditentukan nilai I_{akhir} = jumlah *frame* maksimum dan I_{awal} = 1.
- Tentukan isyarat suku kata baru (SKB) dengan cara
  a. Tentukan jumlah SK.
  b. Jika SK = 1, maka SKB = I_{awal}*sfr : I_{akhir}*(sfr-1)
  c. Jika SK > 1, maka SKB = I_{awal}(i)*sfr : I_{akhir}(i)*(sfr-1)

B. *Mel Frequency Cepstral Coefficient* (MFCC)

Proses ekstraksi ciri MFCC merupakan proses pengambilan ciri yang berdasar pada transformasi fourier diskrit. Secara ringkas tahapan-tahapannya ditunjukkan pda Gambar 1.

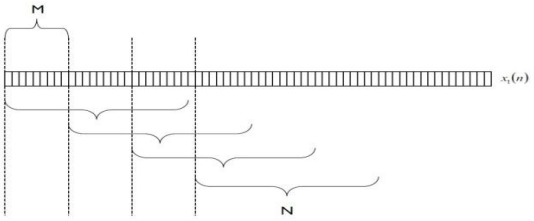


Gambar 1. Proses MFCC

1) *Frame Blocking dan Windowing*

Isyarat suara merupakan isyarat non-stasioner, artinya sifat-sifat satistiknya selalu berubah terhadap waktu. Sehingga tidaklah memungkinkan mengekstraksi ciri spektral dari suara tutur sekaligus. Oleh karena itu, ciri spektral isyarat suara diekstrak melalui sebuah *window* isyarat suara yang mencirikan bagian suara tertentu sehingga dengannya dapat dibuat suatu asumsi bahwa isyarat suara tersebut stasioner[14].

Proses *windowing* dicirikan oleh tiga parameter, yaitu lebar jendela, offset antar jendela dan bentuk jendela. Hasil *windowing* disebut sebagai *frame*, dengan panjang (*frame size*) dan besar pergeseran (*frame shift*) tertentu dalam satuan milidetik. Rentang pergeseran *frame* berkisar antara 1/3 – 1/2 kali panjang *frame*. Gambar 2 menunjukkan proses framing isyarat suara.



Gambar 2. Proses *frame blocking*

Fungsi *window* yang digunakan adalah yang bernilai maksimum 1 untuk daerah di dalam jendela dan nol untuk daerah yang lain. Jendela bergerak sepanjang isyarat suara dan mengekstraksi bentuk isyarat yang berada di dalamnya. Pada penelitian ini digunakan fungsi *window* hamming yang dinyatakan oleh persamaan berikut

$$w[n] = \begin{cases} \alpha - \beta \cos\left(\frac{2\pi n}{L}\right), 0 \le n \le L-1 \\ 0 \qquad\qquad yang\ lain \end{cases} \quad (2)$$

Dan proses perkalian nilai isyarat *frame* N pada waktu n, *s*[n], dengan nilai fungsi jendela pada waktu n, *w*[n]:

$$y[n] = w[n]s[n] \quad (3)$$

2) Transformasi Fourier Diskrit

Transformasi Fourier Diskrit (DFT) digunakan untuk mengekstraksi informasi spektral dari isyarat terjendela yaitu untuk mengetahui besar energi yang terkandung pada pita frekuensi berbeda [9]. Karena menggunakan *frame* terjendela, maka penerapannya akan memungkinkan akan terjadinya kehilangan informasi. Transformasi yang digunakan dinyatakan pada persamaan (4) berikut.

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j2\pi nk/N} \quad (4)$$

Karena hasil transformasi fourier simetris, maka hanya setengahnya saja yang digunakan untuk proses selanjutnya.

3) Mel Frequency Warping

Proses Mel Frequency Warping adalah proses pembungkusan spektrum isyarat dengan menggunakan filterbank segitiga. Karena isyarat suara berbeda dengan persepsi pendengaran manusia, dimana isyarat suara tidaklah memiliki frekuensi dengan skala yang linier. Oleh karena itu, dibutuhkan penyesuaian dengan persepsi pendengaran manusia yang bersifat linier dalam proses ekstraksi ciri agar dapat meningkatkan performa pengenalan. Sebagai acuan , penskalaan antara frekuensi dalam Hz dan skala mel bersifat linier pada frekuensi di bawah 1000 Hz dan bersifat logaritmik pada frekuensi diatasnya. Untuk merubah frekuensi suara menjadi frekuensi mel digunakan persamaan berikut[15]:

$$mel(f) = 1125\, ln\left(1 + \frac{f_{Hz}}{700}\right) \quad (5)$$

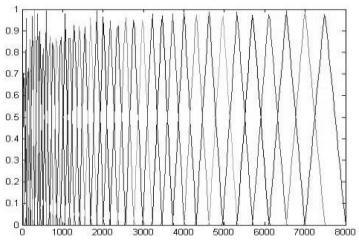
Dan untuk inversnya menggunakan persamaan

$$f_{Hz}(mel) = 700\left(exp\left(\frac{mel}{1125}\right) - 1\right) \quad (6)$$

Setelah diperoleh titik tengan frekuensinya kemudian dibentuk M buah filter segitiga dengan persamaan berikut:

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \frac{(k - f[m-1])}{(f[m] - f[m-1])}, & f[m-1] \le k \le f[m] \\ \frac{(f[m+1] - k)}{(f[m+1] - f[m])}, & f[m] \le k \le f[m+1] \\ 0 & k > f[m+1] \end{cases} \quad (7)$$

Bentuk filterbank segitiga berdasarkan persamaan diatas yang digunakan pada penelitian ini ditunjukkan pada gambar 3:



Gambar 3. Filterbank segitiga

4) Transformasi Cosinus Diskrit

Pada tahapan MFCC proses *cepstrum* adalah proses merubah frekuensi mel kembali ke kawasan waktu. Hasil dari proses ini berupa koefisein mel *cepstrum* (MFCC). Karena koefisien spektrum mel adalah bilangan nyata, maka dapat diubah ke kawasan waktu menggunakan transformasi Cosinus Diskrit berikut ini[15]:

$$c[n] = \sum_{m=0}^{M-1} S[m]\cos\left(\pi n(m+1/2)/M\right) \qquad 0 \le n < M \quad (8)$$

Dimana *S*[*m*] adalah hasil dari perkalian spektrum dengan konjugetnya.

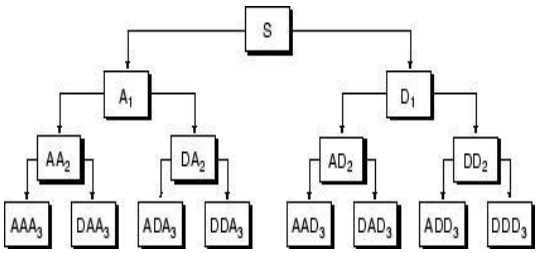
C. Wavelet Packet Transform (WPT)

Wavelet adalah gelombang singkat dengan durasi terbatas yang memiliki nilai rata-rata nol. Wavelet ini mengkonsentrasikan energinya dalam ruang dan waktu sehingga cocok untuk menganalisis isyarat yang sifatnya sementara saja. Ada dua jenis wavelet yaitu Transformasi wavelet kontinyu dan Transformasi wavelet diskrit[5]. Wavelet Packet Transform merupakan salah satu pengembangan dari transformasi wavelet diskrit yang mendekomposisi baik sisi detail maupun aproksimasi. Proses dekomposisi WPT ditunjukkan oleh gambar 4, dan persamaan dekomposisinya adalah:

$$y_h[k] = \sum_n x[n]\, h[2k - n] \quad (9)$$

$$y_l[k] = \sum_n x[n] g[2k - n] \quad (10)$$

Dimana y_h[k] adalah detail dari informasi isyarat atau *D*, dan y_l[k] merupakan aproksimasi kasar dari fungsi penskalaan atau *A*. *x*[*n*] adalah isyarat asli dengan h[n] dan g[n] masing-masing koefisien HPF dan koefisien LPF.



Gambar 4. Proses dekomposisi dengan WPT

Proses dekomposisi dengan WPT akan menghasilkan sub-band sebanyak 2^j dimana j adalah level dekomposisi. Untuk membuat vektor ciri, energi pada tiap sub-band level ke-j dihitung menggunakan persamaan berikut:

$$E_i = \sqrt{\sum_{k=1}^{N} |X_i(k)|^2} \quad (11)$$

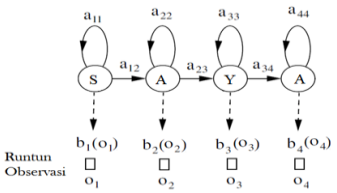
Normalisasi energi diperlukan agar nilai vektor ciri yang diperoleh memiliki standar yang sama. Normalisasi energi dilakukan dengan cara membagi setiap total energi sub-band dengan total energi sub-band yang digunakan, dirumuskan oleh kedua persamaan berikut.

$$E_{tot} = \sqrt{\sum_{i=1}^{I} E_i^2} \quad (12)$$

$$V_{energi} = \frac{E_i}{E_{tot}} \quad (13)$$

D. *Hidden Markov Models* (HMM)

Hidden Markov Models menurut Rabiner 1989 didasarkan pada proses stokastik ganda, pertama adalah proses stokastik yang menghasilkan *state* yang tidak dapat diamati, dan kedua adalah proses yang menghasilkan runtun observasi yang dapat diamati[16]. Ada dua tipe HMM yaitu HMM ergodic dan HMM kiri-kanan. Untuk memodelkan isyarat suara digunakan HMM kiri-kanan karena suara tidak dapat berulang ke *state* sebelumnya.



Gambar 4. HMM kiri kanan

Sebuah HMM dimodelkan oleh elemen-elemen berikut:

$$\lambda=(A,B,\pi) \quad (14)$$

Dan untuk tipe HMM kiri kanan, model tersebut dijabarkan sebagai berikut:

- $A=\{a_{ij}\}$, probabilitas transisi *state*

  $a_{ij}=0$ untuk $j<i$ dan $j>i+1$, $a_{NN}=1$,

  $a_{Nj}=0,\ j<N$. (15)

- $\pi=\{\pi_i\}$, distribusi probabilitas *state* awal

  $\pi_i=1$ untuk $i=1$, dan $\pi_i=0$ untuk $i\neq0$ (16)

- $B= b_j(k)$, distribusi probabilitas observasi.

Umumnya pada HMM sebaran distribusi probabilitas observasi dimodelkan dengan *Gaussian Mixture Model* (GMM). Namun pemodelan ini meninggalkan masalah komputasi, sehingga dapat digunakan persamaan distribus jarak Euclid sebagai alternatif yang dinyatakan oleh:

$$b_j(o_t) = \frac{1}{1+d_j(o_t)} \quad (17)$$

$$d_j(o_t) = \sqrt{\sum_{k=1}^{M} (o_{tk} - \mu_{jk})^2} \quad (18)$$

III. JALAN PENELITIAN

Pada penelitian ini data yang diolah berupa data rekaman suara suku kata bahasa Indonesia yang direkam dari 8 orang pembicara. Data suara dari 6 orang pembicara digunakan untuk pelatihan sedangkan untuk pengujian berupa rekaman data suara dari 2 orang pembicara lainnya beserta salah seorang dari 6 orang pembicara yang rekaman suaranya telah diikutsertakan sebagai data pelatihan. Perekaman dilakukan menggunakan Matlab R2009a ditempat terbuka dengan durasi 1 detik dan pengulangan 5 kali. Frekuensi sampling yang digunakan adalah 16000 Hz, PCM-16 bit, mono dan disimpan dalam format *wav*. Suku kata yang digunakan berpola KV yang merupakan kombinasi dari fonem konsonan *g,k,l,r* dan vokal *a, i, u, e, e2, o*. Sehingga total terdapat 24 buah suku kata yang digunakan pada penelitian ini. Sehingga terdapat total 720 data suara rekaman suku kata yang digunakan untuk pelatihan dengan masing-masing suku kata diwakili oleh 30 data suara rekaman. Sedangkan untuk pengujian digunakan total sebanyak 72 suara rekaman suku kata.

Untuk memisahkan suara dari derau digunakan algoritma VAD. Suara yang telah dipisahkan selanjutnya diekstraksi dengan menggunakan metode MFCC dan WPT. Pada proses ekstraksi ciri MFCC, suara disampel dengan panjang *frame* 512 dengan overlap 50%. Pada proses *warping* digunakan 40 buah *filterbank* segitiga, untuk memperoleh *cepstrum*. *Cepstrum* kemudian diinvers dengan DCT untuk mendapatkan koefisien mel *cepstrum* (MFCC). Pada penelitian ini hanya 12 koefisien saja yang digunakan sebagai koefisien ciri, yaitu mulai dari koefisien ciri ke-3 sampai dengan koefisien ciri ke-14.

Adapun pada proses ekstraksi ciri WPT panjang *frame* bergantung pada jumlah *state* yang digunakan. *Frame* suara ditransformasi menggunakan wavelet daubechies orde 3 (db3) dan orde 7 (db7) dengan level dekomposisi j = 5. Kemudian dibentuk sebuah vektor ciri dengan menghitung energi dari tiap subband pada hasil dekomposisi level ke-5. Karena proses WPT mendekomposisi baik detail maupun aproksimasi, maka

setiap *frame* akan diwakili oleh 2^j buah ciri. Hasil ekstraksi kemudian diklasifikasi dengan menggunakan HMM dengan probabilitas jarak Euclid sebagai model distribusi probabilitas observasi. Jumlah *state* yang digunakan sebanyak 2 buah, menyesuaikan dengan jumlah fonem pada suku kata yang digunakan.

Pengenalan suku kata dengan mengekstraksi ciri langsung dari seluruh suku kata masih memberikan akurasi yang sangat rendah, seperti yang telah dilakukan oleh Abriyono [7]. Oleh karena itu, pada penelitian ini ekstraksi ciri suara suku kata dilakukan terhadap masing-masing fonem penyusunnya. Sebagai rujukan dalam membagi data suara suku kata menjadi fonemnya masing-masing, digunakan penelitian yang telah dilakukan oleh Farooq [10], dimana di dalamnya disebutkan bahwa durasi pengucapan sebuah fonem konsonan dalam bahasa inggris adalah 32 ms untuk frekuensi sampling 16000 Hz, atau sekitar 512 sampel. Maka pada kasus ini panjang 512 sampel juga digunakan sebagai rujukan panjang sampel minimum sebuah fonem konsonan sedangkan sisanya diasumsikan sebagai panjang sampel vokal.

Agar dapat lebih tepat dalam mengestimasi panjang minimum sampel fonem konsonan bahasa Indonesia, maka digunakan variasi panjang konsonan-vokal berikut ini:

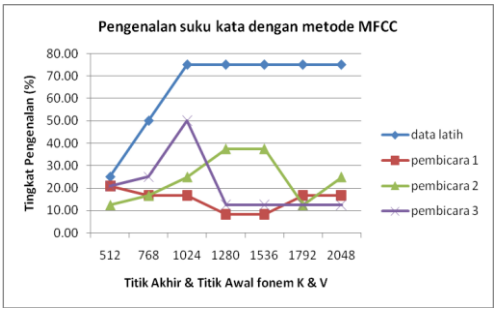
Tabel 1. Kombinasi panjang K dan V

Panjang Konsonan	Panjang Vokal
1 – 512	513 – akhir
1 – 768	769 – akhir
1 – 1024	1025 – akhir
1 – 1280	1281 – akhir
1 – 1536	1537 – akhir
1 – 1792	1793 – akhir
1 – 2048	2049 – akhir

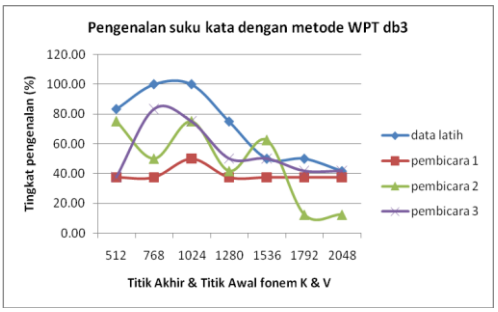
Setiap sampel fonem konsonan dan vokal baik untuk data pelatihan maupun pengujian diekstraksi ciri MFCC dan WPT-nya. Kemudian hasil ekstraksi ciri fonem konsonan dan vokal data pelatihan dibentuk menjadi sebuah model fonem konsonan dan vokal menggunakan HMM dengan jumlah *state* 2. Adapun untuk pengujian pertama-tama dibuat model suku kata gabungan data latih dan model suku kata gabungan data uji yang dibentuk dengan memilih secara random ciri MFCC dan WPT untuk masing-masing fonem konsonan dan vokal. Untuk evaluasi, model suku kata gabungan akan bernilai benar (1) jika model fonem konsonan dan model fonem vokal sama-sama bernilai benar, dan akan bernilai salah (0) jika salah satunya bernilai salah. Kemudian persentase kebenaran hasil pengenalan dihitung berdasarkan pada jumlah total suku kata gabungan yang bernilai benar dibagi dengan jumlah suku kata asli yang digunakan dikalikan 100. Pengaruh panjang sampel konsonan-vokal dan metode ekstraksi ciri dianalisa terhadap hasil pengenalan.

IV. HASIL DAN PEMBAHASAN

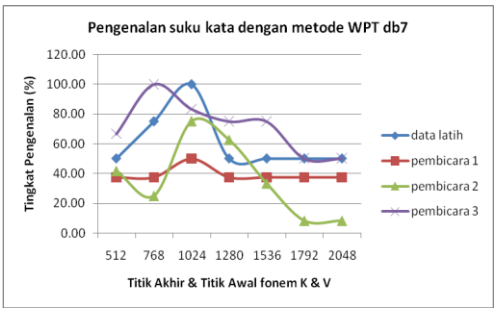
Gambar 5 sampai gambar 7 menunjukkan hasil pengenalan suku kata untuk ekstraksi ciri MFCC, WPT db3 dan WPT db7 terhadap pengujian menggunakan data latih dan data uji. Secara terurut tingkat pengenalan tertinggi untuk pengujian dengan data latih adalah 75% untuk metode MFCC, dan 100% untuk metode WPT db3 dan db7. Sedangkan untuk pengujian dengan data uji, pengenalan suku kata dari masing-masing pembicara memberikan hasil yang berbeda-beda. Pada metode MFCC, secara terurut pengenalan tertinggi suku kata untuk masing-masing pembicara adalah 20.83%, 37.50% dan 50%. Adapun pada metode WPT db3 tingkat pengenalan tertingginya secara terurut adalah 50%, 75% dan 83.33%. Dan pada metode WPT db7 tingkat pengenalannya adalah 50%, 75% dan 100%. Dua pembicara pertama merupakan pembicara yang data suaranya tidak diikutsertakan dalam pelatihan. Adapun data suara pembicara ke 3 diikutsertakan pada pelatihan. Namun dapat dilihat dari ketiga grafik bahwa pembicara 2 dapat memberikan tingkat pengenalan tertinggi yang relatif seimbang dengan pembicara 3, terutama pada metode WPT.



Gambar 5. Grafik tingkat pengenalan dengan metode MFCC



Gambar 6. Grafik tingkat pengenalan dengan metode WPT db3



Gambar 7. Grafik tingkat pengenalan dengan metode WPT db7

Sedangkan untuk pengaruh variasi penentuan panjang sampel fonem konsonan dan fonem vokal secara keseluruhan memberikan hasil yang relatif sama, kecuali pada hasil pengenalan data latih MFCC. Dimana penambahan panjang sampel fonem konsonan menunjukkan adanya peningkatan tingkat akurasi untuk ketiga metode yang berpuncak pada titik 1024 sampel. Adapun setelahnya tren menunjukkan adanya penurunan tingkat pengenalan.

V. KESIMPULAN

Dari hasil dan pembahasan di atas dapat disimpulkan bahwa panjang sampel untuk sebuah fonem konsonan bahasa Indonesia adalah sekitar 1024 sampel. Dan metode ekstraksi ciri dengan pemisahan dapat diterapkan untuk pengenalan suku kata. Selanjutnya dari kedua metode tersebut, dari penelitian ini dapat disimpulkan bahwa metode WPT memberikan hasil yang lebih baik dibandingkan dengan metode MFCC.

REFERENCES

[1] S. Sakti, E. Kelana, H. Riza, and S. Sakai, "Development of Indonesian Large Vocabulary Continuous Speech Recognition System within A-STAR Project," in *TCAST*, 2008, pp. 19–24.

[2] D. P. Lestari, K. Iwano, and S. Furui, "A Large Vocabulary Continuous Speech Recognition System for Indonesian Language," in *15th Indonesian Scientific Conference in Japan Proceedings*, 2006, pp. 17–22.

[3] V. Ferdiansyah and A. Purwarianti, "Indonesian Automatic Speech Recognition System Using English-Based Acoustic Model," *Am. J. Signal Process.*, vol. 2, no. 4, pp. 60–63, Aug. 2012.

[4] U. Sutisna, "Pengenalan Tutur Kata Terisolasi Menggunakan MFCC dan ANFIS," Universitas Gadjah Mada, 2013.

[5] A. Asni B, "Ekstraksi Ciri Dan Pengenalan Tutur Vokal Bahasa Indonesia Menggunakan Metode Discrete Wavelet Transform (DWT) dan Dynamic Time Warping (DTW)," Universitas Gadjah Mada, 2014.

[6] H. Alwi, S. Dardjowidjojo, H. Lapoliwa, and A. M. Moeliono, *Tata Bahasa Baku Bahasa Indonesia : Edisi Ketiga*. Jakarta: Balai Pustaka, 2014.

[7] Abriyono and A. Harjoko, "Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST," *IJCCS*, vol. 6, no. 2, pp. 23–34, 2012.

[8] Suyanto and S. Hartati, "Design of Indonesian LVCSR Using Combined Phoneme and Syllable Models," in *The 7th International Conference on Information & Communication Technology and Systems (ICTS)*, 2013, pp. 191–196.

[9] B. Darmawan, "Ekstraksi CIri Suara Untuk Pengenalan Pembicara Menggunakan MFCC dan Hidden Markov Models," Universitas Gadjah Mada, 2011.

[10] O. Farooq and S. Datta, "Phoneme recognition using wavelet based features," in *Information Sciences*, 2003, vol. 150, no. 1–2, pp. 5–15.

[11] B. T. Tan, M. F. M. Fu, A. Spray, and P. Dermody, "The Use of Wavelet Transforms in Phoneme Recognition," *Proceeding Fourth Int. Conf. Spok. Lang. Process. ICSLP '96*, vol. 4, 1996.

[12] M. A. Anusuya and S. K. Katti, *Front end analysis of speech recognition: a review*, vol. 14, no. 2. 2011.

[13] A. Buono and B. Kusumoputro, "Pengembangan Model HMM Berbasis Maksimum Lokal Menggunakan Jarak Euclid Untuk Sistem Identifikasi Pembicara.pdf," in *National Conference on Computer Science & Information Technology*, 2007, pp. 49–54.

[14] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, 2008.

[15] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Prentice Hall, 2001.

[16] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2. pp. 257–286, 1989.