EKSPERIMEN PENGENALAN UCAPAN AKSARA LAMPUNG DENGAN CMU SPHINX 4

Sri Ningsih [1], Saniati [2]

[1), 2)] *Informatika, Universitas Teknokrat Indonesia*
Jl. H. ZA. Pagaralam, No 9-11, Labuhanratu, Bandarlampung
Email : sriningsih00889@gmail.com[1)], saniati@teknokrat.ac.id[2)]

Abstrak

Aksara Lampung sebagai salah satu identitas masyarakat Lampung dalam perkembangannya terakhir ini minim apresiasi dari masyarakat. Upaya pelestarian dilakukan dengan mengenalkan budaya Lampung salah satunya aksara. Pengenalan aksara Lampung dapat melalui beberapa aspek yaitu mendengar, berbicara, membaca, dan menulis. Untuk mengaktifkan kemampuan dari aspek-aspek tersebut dapat memanfaatkan teknologi, salah satunya adalah speech recognition dalam membantu mempelajari pengucapan bahasa dan aksara. Eksperimen pengenalan ucapan ini memanfaatkan CMU Sphinx 4 sebagai library untuk membangun model pengenalan ucapan dengan berbasis Hidden Markov Model. Tahapan dalam eksperimen pengenalan ucapan ini adalah training dan testing. Masukan dalam pengenalan ucapan ini berupa suku kata aksara Lampung, kata, dan frasa. Dan keluaran dari masukan tersebut berupa teks suku kata, kata, dan frasa yang diucapkan. Hasil dari penelitian ini menunjukkan akurasi pengenalan ucapan suku kata sebesar 52,47%, pengenalan kata 74,03%, dan pengenalan frasa 57,83% dari 3 skenario eksperimen. Pengenalan ucapan terbaik adalah untuk kata, sedangkan untuk suku kata mendapatkan akurasi terendah, dikarenakan antar suku kata aksara Lampung hampir sama pelafalannya.

Kata Kunci: *Aksara Lampung, Pengenalan Ucapan, Suku Kata, Kata, Frasa, CMU Sphinx4, Hidden Markov Model.*

1. Pendahuluan

Pada setiap daerah di Indonesia memiliki bahasa yang menjadi simbol identitas budaya masing-masing daerah. Provinsi Lampung sendiri memiliki bahasa daerah yaitu Bahasa Lampung, Bahasa Lampung sebagai salah satu identitas masyarakat Lampung dalam perkembangannya terakhir ini dikhawatirkan perkembangannya karena minimnya apresiasi masyarakat terhadap bahasa Lampung termasuk sastra lisan maupun aksara Lampung. Upaya pelestarian dilakukan dengan mengenalkan budaya Lampung salah satunya aksara.

Pengenalan aksara Lampung dapat melalui beberapa aspek yaitu mendengar, berbicara, membaca, dan menulis. Untuk mengaktifkan kemampuan dari aspek-aspek tersebut pemanfaatan teknologi dapat digunakan salah satunya adalah *speech recognition* dalam membantu mempelajari bahasa dan aksara[1]. *Speech recognition* atau pengenalan ucapan adalah salah satu aplikasi yang memungkinkan sistem komputer dapat memahami dan mengenali kata-kata yang diucapkan dengan digitalisasi kata dan pencocokan sinyal digital tersebut dengan suatu pola yang tersimpan pada perangkat sistem komputer tersebut. *Speech recognition* juga memiliki cakupan aplikasi yang luas, seperti *Command recognition*, *Dictation*, *Interactive Voice Response* dan dapat juga digunakan untuk pembelajaran bahasa asing[2].

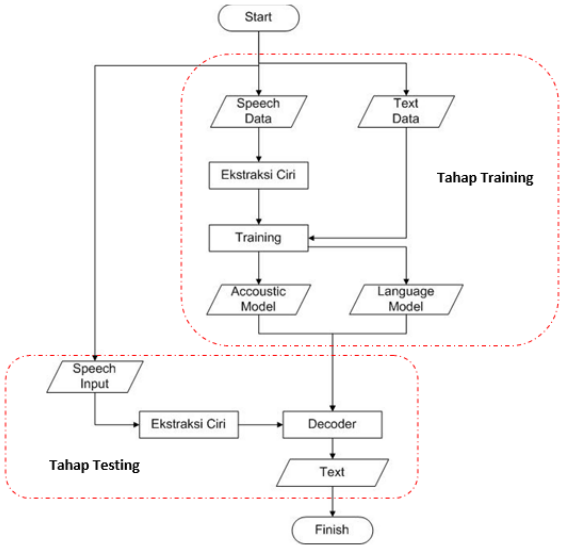
Kemajuan sistem *speech recognition* banyak memunculkan program-program *open source* salah satunya yaitu CMU Sphinx. CMU Sphinx yang dikembangkan oleh salah satu peneliti dari *Carnegie Mellon University* merupakan software pengenalan suara yang bisa digunakan untuk mengidentifikasi suara ucapan untuk kemudian melakukan berbagai tugas yang diperintahkan. Di dalam penelitian ini penulis tertarik untuk menggunakan CMU Sphinx-4 seperti pada penelitian Walker [3]. CMU Sphinx-4 merupakan pengembangan perangkat terbaru dalam CMU Sphinx dimana keseluruhan sistem dibangun dengan bahasa pemrograman Java. Penelitian ini menggunakan CMU Sphinx-4 karena memiliki beberapa kelebihan yang mendukung, yaitu adanya tutorial yang disediakan oleh Sphinx, *open source*, serta tersedianya forum diskusi di *sourceforge*. Selain itu beberapa penelitian terkait berhasil menerapkan pada beberapa bahasa berbeda seperti Bangla[4], Hindi[5] serta Indonesia[6].

Diketahui bahwa pelafalan aksara Lampung memiliki struktur pelafalan yang berbeda dengan bahasa Inggris sehingga pada penelitian ini penulis melakukan eksperimen dengan CMU Sphinx-4 untuk mengenali lafal aksara Lampung.

2. Rancangan Arsitektur

Arsitektur dari eksperimen pengenalan ucapan untuk aksara Lampung ini, secara umum dibagi dalam dua tahap yaitu *training* dan *testing* seperti pada gambar 1. Dua tahap ini juga digunakan pada riset Fitriliana [7].
Tahap *training* membutuhkan dua data yaitu *speech* data dan *text* data. Speech data diperlukan untuk menghasilkan *acoustic model* dari proses *training*.

Sebelum digunakan untuk proses *training*, *speech data* melalui proses ekstraksi ciri. Pada proses ekstraksi ciri penggunaan *Mel Frequency Cepstral Coefficient* (MFCC) menghasilkan vector ciri yang digunakan untuk proses *training*. Dimana proses ini akan melewati tahapan-tahapan yaitu *pre-emphasis*, *frame blocking*, *windowing*, *power spectrum*, dan *mel spectrum*. Sedangkan untuk *text* data digunakan pada proses *training* dan menghasilkan *Language Model*.



Gambar 1. Arsitektur eksperimen pengenalan ucapan untuk aksara Lampung.

Pada tahap *testing*, *acoustic model* dan *language model* yang dihasilkan oleh proses *training* digunakan pada proses *decoder*. Proses *decoder* ini, dilakukan pengenalan suara dari hasil ekstraksi ciri *speech* input dan menghasilkan *output* berupa text.

Penerapan arsitektur pada fase *training* dilakukan dengan menyiapkan sejumlah data seperti Data Teks, *Speech Corpus*, *Transcription File*, *Language Dictionary*, *Language Model*, *Control File*, *Filler Dictionary* dan *Phone File*.

Data teks yang digunakan adalah 20 suku kata aksara Lampung (Ka, ga, nga, pa, ba, ma, ta, da, na, ca, ja, nya, ya, a, la, ra, sa, wa, ha, gha) dan 40 kata umum dalam Bahasa Lampung (Adik apak, beli, bighu, cerai, gelagh, ghibu, ghik, ghuwa, handak, indui, jama, kabagh, kamagh, kapan, kawai, kemaman, kota, lambam, lampung, lapah, lulus, mengan, mi, mid, mobil, motogh, nginum, niku, pakai, pedom, pekon, pulau, sai, sampai, sikam, sore, suluh, taman, umar), dimana 40 kata-kata ini dapat disusun menjadi beberapa frasa.

Setelah menentukan data teks untuk pengenalan ucapan, maka dibutuhkan data suara untuk penelitian ini. Peneliti melakukan proses perekaman dari tiap kata yang telah ditentukan pada data teks. Dalam penelitian ini, perekaman suara dilakukan oleh satu orang, yaitu penulis. Proses rekaman dilakukan dengan perangkat lunak Audacity.

File transkrip diperlukan untuk merepresentasikan apa yang pembicara ucapkan dalam file audio. Dialog pembicara dicatat sehingga kata yang diucapkan/direkam sama persis dengan kata yang dituliskan, dengan diapit silencetag (tag dimulai degan <s> dan diakhiri dengan </s>), kemudian diikuti oleh file ID yang merepresentasikan suara yang direkam. *Transcription file* ada dua macam yaitu satu *transcription file* untuk proses *training* dan satu *transcription file* untuk proses *testing*.

Language dictionary merupakan file yang berisi setiap pasangan kata beserta pelafalan (fonem) nya. Fonem merupakan konsep penting dalam representasi ucapan [8] , dan setiap kata yang melalui proses *training* dan *testing* akan dicantumkan kedalam *language dictionary*.

Language model (model bahasa) berfungsi untuk mendeskripsikan peluang kata yang akan dipanggil saat kata atau frasa diucapkan. Penggunaan *language model* ini dapat meningkatkan akurasi pengenalan suara karena model ini memberikan sistem, pengetahuan tentang konteks kata yang akan dikenali.

Control file merupakan sebuah *file* teks yang mengandung semua nama dari *file* audio dengan tanpa mencantumkan ekstensi dari file tersebut (.wav).

Filler dictionary merupakan *Language Dictionary* yang berisi *non-speech sounds* yang dipetakan sesuai dengan *non-speech sound unit*.

Phone file merupakan sebuah teks sederhana yang memberi penjelasan pada trainer bahwa fonem yang dicantumkan merupakan bagian dari file training. File terdiri dari satu fonem pada tiap baris, harus sesuai dengan fonem dalam *Language Dictionary* (aksara.dic) dan tidak diperbolehkan adanya duplikasi.

Proses training dilakukan menggunakan data training sebanyak 300 file suara ucapan penulis yang merepresentasikan 60 kata dalam *dictionary*.

Setelah *acoustic model* dan *language model* telah didapatkan dari tahap *training*, kemudian dilakukan tahap *testing*. Eksperimen tersebut dilakukan pada tiga jenis skenario yang bervariasi dari jenis kata (suku kata, kata, dan frasa), sumber suara penguji, serta *microphone*. Sedangkan lingkungan lainnya juga dibatasi seperti lingkungan, jumlah percobaan dan jarak *microphone* terhadap sumber suara. Berikut tiga jenis skenario tersebut.

Skenario Eksperimen 1:

a. Sumber suara penguji : *trained speaker* (suara peneliti)
b. Lingkungan : *closed room*
c. *Microphone* : Genius MIC01A
d. Kata yang diuji coba : 20 suku kata aksara Lampung, 40 kata, 20 frasa
e. Jumlah percobaan : 30 kali pengucapan setiap suku kata, kata, dan frasa
f. Jarak *microphone* dan sumber suara 5 cm

Skenario Eksperimen 2:
a. Sumber suara penguji : *trained speaker* (suara peneliti)
b. Lingkungan : *closed room*
c. *Microphone* : *Realtek Audio 6.0.1.6128*
d. Kata yang diuji coba : 20 suku kata aksara Lampung, 40 kata, 20 frasa
e. Jumlah percobaan : 30 kali pengucapan setiap suku kata, kata, dan frasa
f. Jarak *microphone* dan sumber suara 5 cm

Skenario Eksperimen 3:
a. Sumber suara penguji : un*trained speaker* (*female speaker*)
b. Lingkungan : *closed room*
c. *Microphone* : Genius MIC01A
d. Kata yang diuji coba : 10 suku kata aksara Lampung, 40 kata, 20 frasa
e. Jumlah percobaan : 30 kali pengucapan setiap suku kata, kata, dan frasa
f. Jarak *microphone* dan sumber suara 5 cm

3. Hasil Pengujian

Dari hasil uji coba menurut skenario eksperimen yang telah ditetapkan, kemudian dibandingkan berdasarkan jenis kata yang digunakan yaitu suku kata, kata dan frasa. Tabel 1 menunjukkan hasil eksperimen dengan 3 skenario terhadap pengenalan suku kata pada aksara Lampung.

Tabel 1. Tabel hasil pengenalan suku kata

Eksperimen	Jumlah Percobaan	Jumlah Akurat	Persentase Akurasi
Eksperimen 1	600	382	63,67%
Eksperimen 2	600	281	46,83%
Eksperimen 3	300	124	41,33%

Dari tabel di atas didapatkan selisih perbandingan yang tidak begitu jauh. Pada eksperimen pertama didapatkan hasil akurasi 63,67% dimana akurasi pengenalan suku kata yang paling tinggi dibanding eksperimen 2 dan eksperimen 3 yang mendapatkan akurasi 46,83% dan 41,33%. Suku kata yang memiliki akurasi terbaik dari hasil uji coba banyak terdapat pada eksperimen 1 dimana ada 6 suku kata (ka, ba, da, ja, nya, dan sa) yang memiliki tingkat akurasi mencapai 100%. Untuk kata terendah akurasinya terdapat 3 suku kata (ga, na, dan ca) yang memiliki tingkat akurasi 0% dan 2 suku kata yang memiliki akurasi hanya 3,33% (a) dan 6,67% (wa). Sedangkan untuk hasil pengenalan suku kata pada eksperimen 2 dimana menggunakan *microphone* yang berbeda dari eksperimen 1, tidak terdapat akurasi pengenalan yang berhasil mencapai 100% dan suku kata yang memiliki akurasi tertinggi yaitu "sa" 96,67%.

Untuk eksperimen terakhir yaitu eksperimen 3 yang menggunakan suara untrained speaker sebagai suara penguji dengan *microphone* yang sama seperti eksperimen 1, didapatkan hasil yang tidak begitu baik juga untuk tingkat akurasi pengenalannya. Dari 10 suku kata pilihan yaitu 5 suku kata berakurasi baik dan 5 suku kata berakurasi buruk, terdapat 2 suku kata yang memiliki tingkat akurasi baik yaitu "ja" dan "sa" 93,33%. Dan suku kata dengan akurasi buruk adalah "ga", "ca", dan "wa" yang memiliki akurasi 0%.

Berdasarkan persentase akurasi dari hasil uji coba pengenalan suku kata dari masing-masing eksperimen yang telah dilakukan. Skenario eksperimen pengenalan suku kata lebih baik akurasinya pada eksperimen 1. Hal ini disebabkan *microphone* yang digunakan lebih baik daripada *microphone* pada eksperimen 2. Walaupun pada eksperimen 3 menggunakan *microphone* seperti eksperimen 1, suara penguji lebih baik akurasinya dengan menggunakan suara *trained speaker* dibanding suara *untrained speaker*.

Tabel 2. Tabel hasil pengenalan kata

Eksperimen	Jumlah Percobaan	Jumlah Akurat	Persentase Akurasi
Eksperimen 1	1200	1009	84,08%
Eksperimen 2	1200	814	67,83%
Eksperimen 3	1200	842	70,17%

Pada tabel 2 dapat dilihat persentase akurasi pengenalan kata pada beberapa eksperimen memiliki hasil yang jauh berbeda antara eksperimen 1 dibanding eksperimen 2 dan eksperimen 3. Akurasi pengenalan kata pada eksperimen 1 memiliki akurasi yang tertinggi dibanding eksperimen lainnya. Kata yang memiliki akurasi hingga 100 % cukup banyak yaitu ada 10 kata "ghuwa", "indui", "sai ", "suluh", "mengan", "mi", "nginum", "kamagh", "kemaman", dan "lapah". Dan hasil yang paling buruk tingkat akurasinya yaitu kata "apak" (0%), "kawai" (16,67%) dan "mid" (16,67%).

Hasil eksperimen 2 yang menggunakan *microphone* Realtek Audio 6.0.1.6128 memiliki akurasi pengenalan kata 67,83%, kata dengan akurasi terbaik adalah "sai" (100%) dan kata dengan akurasi terendah adalah "apak" (0%). Eksperimen 3 dimana sumber suara penguji adalah untrained speaker serta lingkungan dan *microphone* sama dengan eksperimen 1, tetap memiliki akurasi pengenalan yang kurang baik dari 40 kata yang diuji coba. Kata yang memiliki akurasi terbaik adalah "sai" (100%) dan "ghuwa" (96,67%). Sedangkan untuk akurasi pengenalan kata yang memiliki akurasi buruk yaitu kata "apak" dan "mid" (0%), "ghibu" (6,67%), dan "kawai" (13,33%).

Tabel 3. Tabel hasil pengenalan frasa

Eksperimen	Jumlah Percobaan	Jumlah Akurat	Persentase Akurasi
Eksperimen 1	600	392	66%
Eksperimen 2	600	321	53,50%
Eksperimen 3	600	328	54,67%

Hasil pengenalan frasa dari eksperimen yang telah dilakukan, akurasi tertinggi adalah eksperimen 1 yaitu 66%, dan eksperimen 2 mendapatkan akurasi terendah yaitu 53,50%. Pada eksperimen 1 frasa dengan akurasi tertinggi dengan akurasi mencapai 100% ada 3 frasa (lihat tabel 4.12) dan akurasi pengenalan frasa terendah adalah frasa "mobil apak" dan "apak mid kota" dengan persentase 0%. Eksperimen 2 memiliki persentase akurasi 53,50%, dimana akurasi frasa tertinggi pada eksperimen 2 terdapat 1 frasa yang mencapai akurasi pengenalan 83,33% "beli mobil bighu". Sedangkan akurasi terendah terdapat 2 frasa dengan akurasi 0% yaitu "mobil apak" dan "apak mid kota". Akurasi tertinggi pada eksperimen 3 hanya terdapat 1 frasa yang mencapai 96,67% "beli mobil bighu". Dan akurasi terendah yaitu 0% terdapat 2 frasa yaitu "mobil apak" dan "apak mid kota".

Berdasarkan hasil ini semakin memperjelas bahwa *microphone* dan sumber suara penguji berpengaruh dalam akurasi pengenalan suku kata, kata, dan frasa. Dimana eksperimen terbaik untuk pengenalan suku kata, kata, dan frasa adalah eksperimen 1 dengan *microphone* Genius MIC01A serta suara penguji adalah *trained speaker*. Sedangkan eksperimen 3 yang menggunakan *microphone* yang sama dengan eksperimen 1 dan *untrained speaker* sebagai suara penguji memiliki hasil akurasi yang rendah. Tetapi walaupun memiliki akurasi rendah hasil implementasi testing pengenalan ucapan ini dapat mengenali ucapan yang tidak hanya bersumber pada ucapan *trained speaker*.

Hasil keseluruhan dari rata-rata akurasi pengenalan ucapan pada suku kata, kata, dan frasa dapat diamati pada tabel 4.

Tabel 4. Tabel hasil pengenalan suku kata, kata, dan frasa

Jenis Ucapan	Jumlah Percobaan	Jumlah Akurat	Persentase Akurasi
Suku Kata	1500	787	52,47%
Kata	3600	2665	74,03%
Frasa	1800	1041	57,83%

Pada pengenalan suku kata, akurasi pengenalan ucapan didapatkan 52.47% dari total jumlah percobaan pengucapan 1500 kali pengucapan dengan 787 pengucapan yang tepat dikenali. Ada beberapa suku kata yang mendapatkan akurasi pengenalan suku kata 0 % diantaranya yaitu suku kata "ga", "na", dan "ca". Suku kata "ga" sendiri sering dikenali sebagai "da" dan "ja". Sedangkan suku kata "na" sering dikenali sebagai "ma". Dan untuk suku kata ca sering dikenali sebagai "sa" dan "ja". Pengenalan ucapan suku kata juga tingkat akurasinya lebih rendah dibanding akurasi pengenalan kata dan frasa, hal ini dipengaruhi dari cara pengucapan yang hampir sama antara suku kata satu dan lainnya. Dari hasil pengenalan suku kata ini juga peneliti harus lebih baik lagi dalam melakukan perekaman suara dan perancangan *acoustic model* untuk mengatasi masalah pengenalan ucapan yang memiliki kemiripan dalam pengucapannya.

Pada pengenalan ucapan untuk kata didapatkan hasil akurasi pengenalan yang paling baik yaitu 74,03%. Ada beberapa kata yang memiliki tingkat akurasi pengenalan yang buruk yaitu "apak", "mid", dan "kawai". Untuk pengenalan kata "apak" sering diterjemahkan menjadi "a pa" dan "pa". Untuk pengenalan kata "mid" sering diterjemahkan sebagai "mi". Dan pengenalan kata "kawai" sering diterhjemahkan dengan "pakai", "sampai", dan "sai". Pengucapan kata yang memiliki cara pengucapan yang hampir sama/mirip akan menyebabkan penerjemahan pada pengenalan ucapan akan berbeda dengan kata yang diucapkan. Selain itu hal tersebut juga akan memengaruhi pada proses pengenalan frasa, dimana kata yang memiliki akurasi kurang baik akan menerjemahkan frasa yang tidak sesuai. Seperti kata "apak" yang berpengaruh pada pengenalan frasa "mobil apak".

Akurasi pengenalan untuk frasa dalam uji coba didapatkan hasil 57.83% dari total 1800 kali pengucapan pada frasa. Terdapat beberapa frasa yang tidak dapat dikenali, seperti frasa "mobil apak" dan "apak mid kota". Hal ini dipengaruhi oleh kata "apak" yang memiliki akurasi pengenalan yang buruk, sehingga frasa yang memiliki kata "apak" akan memiliki akurasi yang buruk juga.

4. Kesimpulan

Berdasarkan hasil eksperimen pengenalan ucapan untuk aksara Lampung, disimpulkan bahwa pengenalan ucapan terbaik terjadi pada jenis pengucapan kata, sedangkan akurasi pengenalan terendah terjadi pada jenis suku kata pelafalan ucapan suku kata aksara Lampung dikarenakan kemiripan pelafalan suku kata yang cukup pendek tersebut. Selain itu dari segi lain yaitu *microphone* dan sumber suara penguji mempengaruhi akurasi pengenalan ucapan. *Microphone* Genius MIC01A dan suara *trained speaker* mendapatkan akurasi yang lebih baik.

Daftar Pustaka

[1] S.M. Abdou, dkk., "Computer Aided Pronounciation Learning System Using Speech Recognition Techniques", Cairo University, 2006.
[2] Azmi, Mohamed M., dkk., "Syllable-Based Automatic Arabic Speech Recognition Techniques", Cairo University, 2008.
[3] Walker, Willie, dkk., "Sphinx-4: A Flexible Open Source Framework for Speech Recognition. Oakland", Sun Microsystem, 2004.
[4] Chowdhury, Shammur Absar., "Implementation of Speech Recognition System for Bangla", Dhaka: School of Engineering and Computer Science BRAC University, 2010.
[5] Gupta, Ripul., "Speech Recognition for Hindi", Mumbai, Indian Institute of Technology, 2008.
[6] Ferdiansyah, Veri dan Purwarianti, Ayu., "Indonesian Automatic Speech Recognition System Using English-Based Acoustic Model", Bandung, Institut Teknologi Bandung, 2011.
[7] Fitriliana, K. Rahmadi, A. Siska, "Pengenalan Ucapan Metode MFCC-HMM untuk Perintah Gerak Robot Mobil Penjejak Identifikasi Warna". Jurusan Teknik Elektro, Fakultas Teknik, Universitas Andalas, Sumatra Barat, 2013.
[8] Sipayung, Juan Rio. "Aplikasi Pemilih Menu Berpenggerak Suara Pada Sistem Operasi GNU/Linux", Universitas Sumatera Utara, Medan, 2010.