# **IF4072 Pemrosesan Bahasa Alami (*Speech*)**

**Aplikasi PDF *Reader* dengan *Input* Suara**

### 

Disusun oleh :

1. 13515013 - Robby Syaifullah
2. 13515029 - Finiko Kasula Novenda
3. 13515040 - Patrick Nugroho H.
4. 13515091 - Adrian Hartanto P.

**Program Studi Teknik Informatika  
Sekolah Teknik Elektro dan Informatika  
Institut Teknologi Bandung  
2018**

## **Deskripsi Data**

Data berasal dari rekaman suara seluruh anggota ditambahkan dengan beberapa orang non-anggota. Total ada 7 orang, terdiri atas 4 anggota (semua pria) dan 3 orang non-anggota (2 pria, 1 wanita). Suara yang direkam berupa kata-kata perintah (*command*) untuk *file* pdf. Perintah-perintah yang digunakan adalah: “Tutup”, “Halaman Awal”, “Halaman Akhir”, “Halaman Sebelumnya”, dan “Halaman Selanjutnya”. Setiap perintah ini kemudian direkam dua kali untuk masing-masing orang. Total rekaman yang diambil adalah 70 rekaman. Suara direkam dengan menggunakan perlengkapan yang berbeda-beda untuk setiap anggota tim (berbeda alat rekam, ada yang menggunakan *earphone*, ada yang menggunakan *headset*), tetapi sama-sama menggunakan kakas Audacity dengan format yang sama.

## **Proses *Training* dan Parameter *Training***

Pada proses *training* yang dilakukan, ada 3 tahapan yang dilakukan yaitu preparasi data, membuat *monophone HMMs* dan membuat *tied-state triphones*. Tahapan preparasi data dilakukan untuk mempersiapkan data hasil rekaman (*format wav*) menjadi bentuk MFCC (*feature extraction*).

Pada preparasi data, dilakukan 5 langkah. Pertama-tama dibuat *grammar* dari data rekaman yang ada lalu dilakukan kompilasi dari *grammar* yang dibuat. Setelah itu dilakukan pembuatan kamus *pronounciation* dari *grammar*. Kamus ini berisi kata-kata terurut secara leksikal dalam *grammar* yang dimiliki. Langkah selanjutnya yaitu melakukan pengambilan rekaman data sebagai *dataset* untuk *training*. Jumlah rekaman yang digunakan untuk *training* adalah 63 dari 70 rekaman, dimana sisanya digunakan untuk *testing*.

*Grammar* (tugas.grammar) dari sistem adalah sebagai berikut.

|  |
| --- |
| S : NS\_B COMMAND NS\_E  COMMAND: CLOSE  COMMAND: PAGE DEST |

*Vocabulary* (tugas.voca) dari sistem adalah sebagai berikut.

|  |
| --- |
| % NS\_B  <s> sil  % NS\_E  </s> sil  % CLOSE  TUTUP t u t u p  % PAGE  HALAMAN h a l a m a n  % DEST  AWAL a w a l  AKHIR a h i r  AKHIR a kh i r  AKHIR a k h i r  SEBELUMNYA s e b e l u m ny a  SELANJUTNYA s e l a n j u t ny a |

*Dictionary* (*lexicon*) dari sistem adalah sebagai berikut.

|  |
| --- |
| AKHIR [AKHIR] a h i r  AKHIR [AKHIR] a kh i r  AKHIR [AKHIR] a k h i r  AWAL [AWAL] a w a l  HALAMAN [HALAMAN] h a l a m a n  SEBELUMNYA [SEBELUMNYA] s e b e l u m ny a  SELANJUTNYA [SELANJUTNYA] s e l a n j u t ny a  SENT-END [] sil  SENT-START [] sil  TUTUP [TUTUP] t u t u p |

Setelah itu dibuat *file* transkripsi untuk setiap data rekaman yang dimiliki, agar *file* teks yang berisikan perintah dalam bahasa alami menjadi format yang dikenali oleh kakas HTK. Langkah akhir pada tahapan ini yaitu melakukan parameterisasi pada rekaman mentah menjadi urutan vektor fitur. Langkah ini akan menghasilkan bentuk MFCC dari data rekaman yang dimiliki.

Pada tahapan pembuatan *monophone HMMs*, ada 3 langkah yang dilakukan. Dilakukan pendefinisian model prototipe yang akan digunakan dengan metode *Hidden Markov Model* (HMM). Dilakukan pula konfigurasi untuk *file model* prototipe HMM seperti ukuran *window*, penggunaan *hamming*, dan sebagainya. Selanjutnya dari *file monophone grammar* yang telah ada, dibuat bentuk *flat start monophone* dengan menambahkan model prototipe yang telah dibangkitkan sebelumnya. *Flat start monophone* diestimasi ulang sebanyak 9 kali untuk mendapatkan mode yang lebih baik.

Setelah itu, dibuat HMM *models* yang memiliki *short pause* (“sp”) dalam modelnya. Jadi dilakukan perbaikan pada model HMM agar mengandung “sp”. Langkah terakhir adalah melakukan *realigning* terhadap data *training*. Dilakukan pengecekan untuk setiap pengucapan kata yang ada agar tidak menyebabkan *error* pada tahapan selanjutnya.

Tahapan terakhir adalah membuat *tied-state* *triphone* yang terdiri dari 2 langkah, yakni mengonversi model *monophones* menjadi model *triphone* dan membuat *tied-state* dari *triphone* tersebut. Bentuk *tied-state* penting untuk memastikan semua *state* yang terdistribusi dapat diestimasi dengan baik.

## **Proses *Decoding***

Pada HTK, proses *decoding* menggunakan HCopy dan HVite. Pertama-tama, rekaman suara berformat \*.wav dikonversi menjadi bentuk \*.mfc menggunakan HCopy, konfigurasi konversi sama dengan yang digunakan untuk *training*. Selanjutnya dilakukan perapihan (*viterbi alignment*) menggunakan HVite berdasarkan model HMM sebelumnya. Untuk meningkatkan probabilitas terdeteksinya rekaman yang akan dicocokkan dengan grammar yang ada, dilakukan *grammar scale factor*.

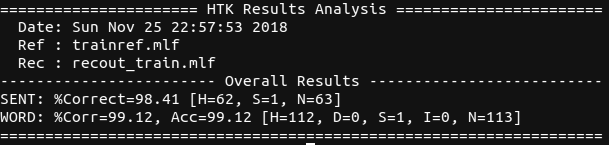
## **Hasil Evaluasi**

Gambar 4.1 berikut adalah hasil evaluasi HTK pada *closed experimental condition* (diuji dengan data latih). Sistem mampu mengenali seluruh data latih dengan tepat, kecuali satu data yaitu data ke 46 (suara pria). Suara pria tersebut “Halaman selanjutnya” tetapi sistem memprediksi “Halaman sebelumnya”. Hal ini karena *lexicon* yang kurang sempurna dalam menangani semua pelafalan.

Kesalahan dapat diketahui karena melihat prediksi pada *file* recout\_train.mlf dan membandingkannya dengan label sebenarnya pada *file* trainref.mlf.

Kalimat yang dikenali dengan benar : 98,41%

Jumlah kata yang dikenali dengan benar : 99,12%

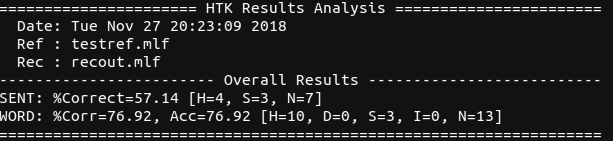


Gambar 4.1 Hasil Evaluasi *Closed Experimental*

Gambar 4.2 berikut adalah hasil evaluasi HTK pada *open experimental condition* (diuji dengan data di luar data latih). Sistem mampu mengenali 4 dari 7 data uji. Sistem gagal mengenali data uji suara perempuan.

Kalimat yang dikenali dengan benar : 57,14%

Jumlah kata yang dikenali dengan benar : 76,92%



Gambar 4.2 Hasil Evaluasi *Open Experimental*

Berikut adalah data uji yang digunakan pada *Open Experimental.*

|  |
| --- |
| \*/test1 TUTUP  \*/test2 HALAMAN SELANJUTNYA  \*/test3 HALAMAN SELANJUTNYA  \*/test4 HALAMAN SEBELUMNYA  \*/test5 HALAMAN SEBELUMNYA  \*/test6 HALAMAN AWAL  \*/test7 HALAMAN AKHIR |

Berikut adalah tabel perbandingan hasil prediksi pada *Open Experimental* (membandingkan isi *file* recout.mlf dengan testref.mlf).

|  |  |
| --- | --- |
| Prediksi (recout.mlf) | Label (testref.mlf) |
| #!MLF!#  "\*/test1.rec"  400000 2400000 TUTUP -1652.525635  .  "\*/test2.rec"  1200000 3400000 HALAMAN -1585.088745  3400000 9500000 SELANJUTNYA -4714.974121  .  "\*/test3.rec"  600000 5700000 HALAMAN -3233.895264  5700000 12800000 AKHIR -4527.174316  .  "\*/test4.rec"  800000 6800000 HALAMAN -3812.187256  6800000 12400000 AWAL -3476.558105  .  "\*/test5.rec"  2100000 4400000 HALAMAN -1454.819458  4400000 9700000 SEBELUMNYA -3323.539551  .  "\*/test6.rec"  500000 5900000 HALAMAN -3230.759033  5900000 9600000 AKHIR -2171.000244  .  "\*/test7.rec"  900000 3700000 HALAMAN -1658.029419  3700000 8500000 AKHIR -2812.020996  . | #!MLF!#  "\*/test1.lab"  TUTUP  .  "\*/test2.lab"  HALAMAN  SELANJUTNYA  .  "\*/test3.lab"  HALAMAN  SELANJUTNYA  .  "\*/test4.lab"  HALAMAN  SEBELUMNYA  .  "\*/test5.lab"  HALAMAN  SEBELUMNYA  .  "\*/test6.lab"  HALAMAN  AWAL  .  "\*/test7.lab"  HALAMAN  AKHIR  . |

Tabel 4.1 Perbandingan Label dengan Prediksi pada *Open Experimental*

## **Analisis Permasalahan Sistem dan Batasan Yang Ditemukan pada Sistem**

Berikut adalah analisis dari hasil sistem.

1. Karena data rekaman yang digunakan saat *training* tidak seimbang (jumlah suara wanita hanya 1 orang, sedangkan pria 6 orang), maka pengenalan suara lebih condong untuk mengenali suara pria daripada wanita.
2. Suara wanita memiliki karakteristik yang berbeda dengan suara laki-laki.
3. Poin 1 menyebabkan suara wanita kurang dapat dikenali oleh sistem, terlihat dari hasil pengujian di mana ketiga kalimat yang salah dikenali adalah data suara wanita.

Ketiga hal ini ditunjukkan dari hasil *open experimental condition* di mana terdapat 3 kesalahan dari 7 data uji (ketiga kesalahan pada suara wanita yaitu pada data 3, 4, dan 6).

Berikut ini adalah batasan yang ditemukan pada sistem.

1. Kata yang disebutkan harus benar-benar jelas.
2. Sistem tidak menangani *input* dengan *noise* yang terlalu keras.
3. Fonem yang didefinisikan masih belum mencakup seluruh kemungkinan pengucapan pada domain.
4. Sistem tidak menangani pengucapan yang tidak baku.

## **Rekomendasi Untuk Pekerjaan Selanjutnya**

Terdapat beberapa rekomendasi yang dapat dilakukan untuk meningkatkan hasil dari eksperimen ini.

* Data-data yang digunakan agar mencakup seluruh fonem sesuai domain dengan cakupan yang lebih luas.
* Untuk data-data rekaman dapat dibuat lebih seimbang dalam hal jenis kelamin, suku dan lain-lainnya (tidak hanya didominasi oleh salah satu jenis).
* *Dictionary* pada sistem dapat ditingkatkan sehingga perintah yang diberikan dapat lebih beragam.
* Dapat juga dilakukan *training* dengan model lain seperti LSTM agar model yang dihasilkan dapat dibandingkan.

## **Pembagian Kerja**

* 1. 13515013 - Robby Syaifullah : Merekam dan mengolah data rekam latih dan uji, membuat kamus *lexicon*, membuat *vocabulary,* membuat model *monophone* HMM dan HMM dengan “sp”.
  2. 13515029 - Finiko Kasula Novenda : Merekam dan mengolah data rekam latih dan uji, membuat *grammar*, membuat *vocabulary*, membuat file transkripsi serta membuat model *monophone* HMM, HMM dengan “sp” dan *tied-state triphone*.
  3. 13515040 - Patrick Nugroho H. : Merekam dan mengolah data rekam latih dan uji, membuat *grammar*, membuat *vocabulary*, membuat kamus *lexicon*, membuat model *monophone* HMM, HMM dengan “sp” dan *tied-state triphone*, serta membuat fungsionalitas live test pada sistem.
  4. 13515091 - Adrian Hartanto P. : Merekam dan mengolah data rekam latih, serta membuat model *monophone* HMM dan HMM dengan “sp”.