PENGUJIAN SISTEM

Pada bab ini akan dijelaskan tentang pengujian terhadap sistem dan aplikasi yang telah dibuat. Pengujian dilakukan dengan menggunakan dataset yang sudah pernah dilakukan terhadap masing – masing percobaan.

5.1. Pengujian Sistem

Pengujian sistem mengevaluasi performa daripada sistem *training* dan *testing* yang telah dibuat. Parameter pada sistem *training* adalah CTC-loss sedangkan pada sistem *testing* adalah CER *(Character Error Rate)*.

5.1.1. Pengujian Awal

5.1.1.1. Percobaan Pertama

Percobaan pertama dengan jumlah neuron 78, 104, 130, 156, 182, 208,234, 260, 286, 312. Pada percobaan pertama dilakukan dengan dataset suara dengan durasi kurang lebih 20 menit disuarakan dengan 4 orang dimana masing-masing berasal dari Jawa 2 orang, Makasar 1 orang, dan Maluku 1 orang. Fitur yang digunakan hanya MFCC dengan 1 konteks yaitu sebanyak 26. Hasil CTC Loss yang dihasilkan tidak menunjukkan ada indikasi error yang lebih kecil. Pada tiap hidden layer tidak dilakukan batch normalization maupun dropout. Pada Bi-RNN tidak dilakukan layer LSTM (Long-Short-Term-Memory). Nilai *Learning rate* adalah 0.001.

Pada tabel 5.1 dapat dilihat hingga *epoch* ke 50 menunjukkan perubahan CTC loss yang terjebak pada *lokal optimum*. Masalah lain adalah peningkatan neuron mengakibatkan waktu *training* yang meningkat dikarenakan pada generasi pertama hanya menggunakan komputasi CPU.

Tabel 5.1. Tabel *sample* setiap 10 *epoch* dalam 50 *epoch* hasil percobaan pertama

Epoch	**CTC-loss**	**Decode text**
0	2423.96	whwhwjwhwjwjwjjjwjjwjjwjjwjwjwjwjwjwjwjjwjjw jjwjjjwjjwjwwwwwwjwwjwj
10	284.578	ianapananasanasanas nas sasasasasa sa sasa sasasasas s s
20	281.987	piaiasasasasasasasasasasasas
30	248.184	iranaranasasasasasas sah sah asah sad sd us s
40	240.975	iranaerkanakanankanananianaaniksais s sas siaha aiah aiah aiah aiad miad uai
50	202.882	eira aerka aka a a ana iksa sas ah ah ah a ah miai a ni a i as

5.1.1.2. Percobaan Kedua

Percobaan kedua dengan jumlah neuron yang sama yaitu 128 namun fitur yang diambil adalah MFCC dengan konteks 3, 5 , dan 9. Dengan dataset suara yang sama yaitu durasi kurang lebih 20 menit disuarakan dengan 4 orang dimana masing-masing berasal dari Jawa 2 orang, Makasar 1 orang, dan Maluku 1 orang. Pada tiap hidden layer tidak dilakukan batch normalization maupun dropout. Pada Bi-RNN tidak dilakukan layer LSTM (Long-Short-Term-Memory). Nilai *Learning rate* adalah 0.001.

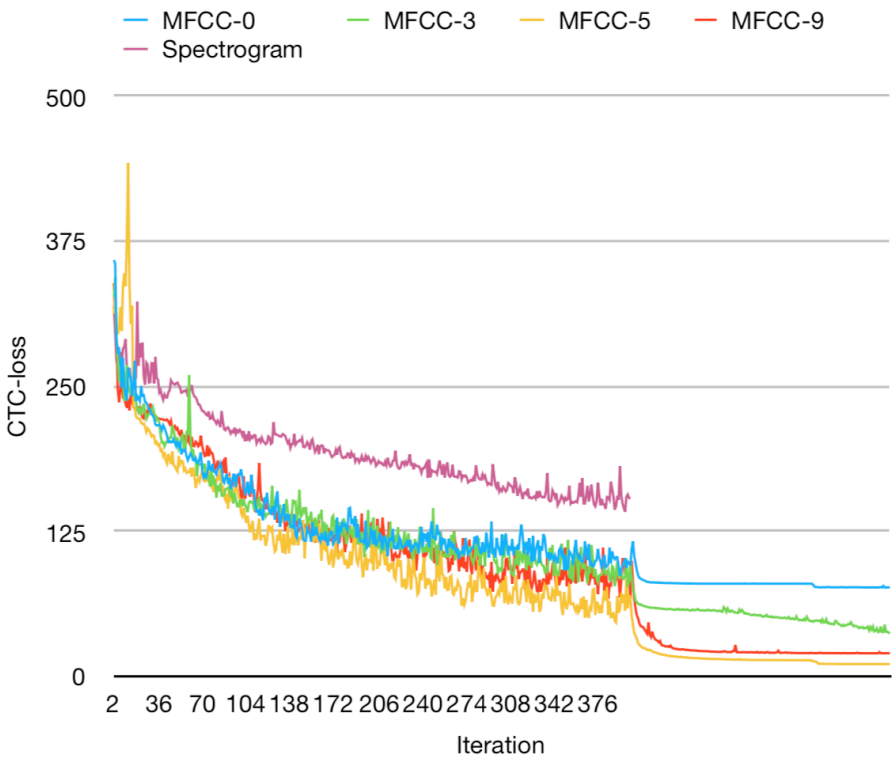
Tabel 5.2. Tabel sample setiap 10 *epoch* dalam 50 *epoch* hasil percobaan kedua

Epoch	**CTC-loss**	**Decode text**
0	960.439	cucudcuducuwcuwcucucuwuwuwuwuwuwuwfwutwfwutw utuwuwudtwuwtwudtvtudtuwtuwtuwtwuwuwuwuwuwuwuwu wuwuwuwuwududududududtueutueutueutuetutueuedudududcud tudtututewutwdtwuetwd
10	350.15	t ta a a a a a a a a a a a a a a a a a a
20	351.087	at a a a a a a a a a a a a a a a a
30	348.7	a a a
40	346.621	a a a
50	346.187	a a a

Pada tabel 5.2 dapat dilihat menunjukkan hasil yang lebih buruk dari percobaan pertama (bandingkan dengan tabel 5.1).

5.3 Percobaan Ketiga

Percobaan ketiga dengan jumlah neuron yang sama yaitu 128 dengan fitur yang sama dengan percobaan kedua yaitu MFCC dengan konteks 3,5, dan 9. Namun ditambah lagi dengan Spectrogram sebagai pembanding. Dengan dataset suara yang sama yaitu durasi kurang lebih 20 menit disuarakan dengan 4 orang dimana masing-masing berasal dari Jawa 2 orang, Makasar 1 orang, dan Maluku 1 orang. Pada tiap hidden layer tidak dilakukan batch normalization maupun dropout. Namun pada Bi-RNN dilakukan layer LSTM (Long-Short-Term-Memory). Nilai *Learning rate* adalah 0.001.Namun pada iterasi ke-400 keatas *learning rate* diturunkan menjadi 0.0001.



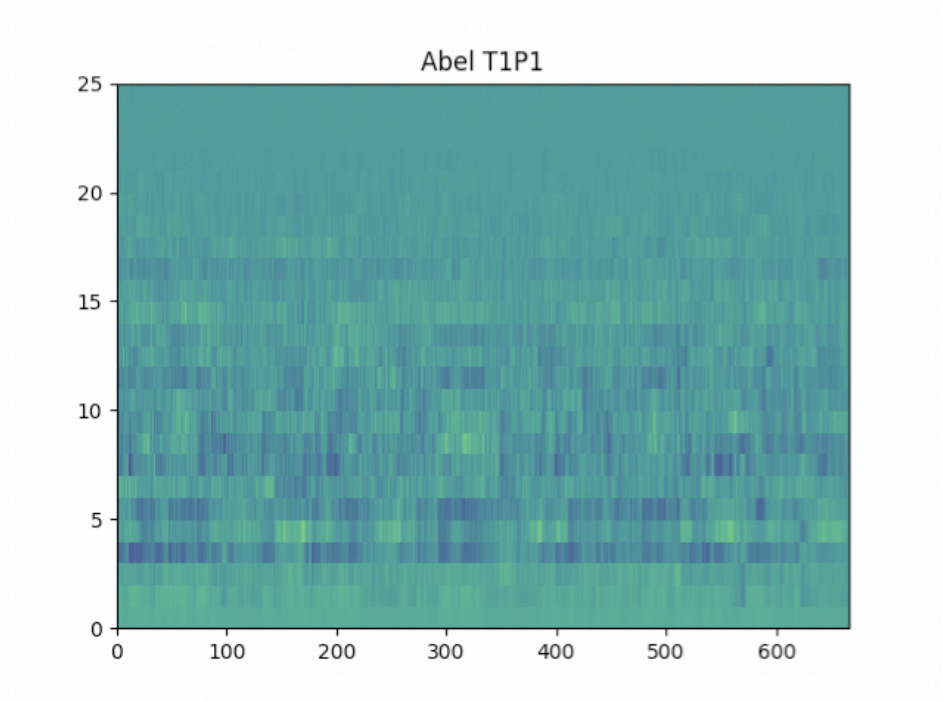
Gambar 5.1. Grafik percobaan ketiga dengan perbedaan fitur dan jumlah konteks

Grafik pada gambar 5.2 menunjukkan performa terbaik pada fitur MFCC dengan 5 konteks, Sedangkan pada MFCC dengan 9 konteks menunjukkan posisi kedua terbaik, diikuti dengan MFCC dengan 3 konteks. Namun pada Spectrogram tidak menunjukkan perubahan CTC-Loss yang signifikan bahkan dibandingkan dengan MFCC dengan 0 konteks terdapat perbedaan hingga 25 persen.

5.1.2. Percobaan Akhir

5.1.2.1. Percobaan Keempat

Pada percobaan keempat dijalankan dengan komputasi GPU agar dapat mengatasi masalah waktu *training* yang meningkat jika neuron ingin dinaikkan hingga 1024. Namun pada komputasi GPU juga tidak akan efisien jika menggunakan *training batch* sebanyak 1. Dalam percobaan ini *training batch* sebanyak 8, dengan fitur MFCC diubah menjadi 24 daripada percobaan pertama hingga keempat yaitu 26. Perubahan ini ditetapkan melihat 2 indeks terakhir yang tidak terlalu signifikan. Dapat dilihat pada gambar 5.2 bahwa suara manusia hanya memberikan frekuensi dari rentang skala mel 0-23.



Gambar 5.2. *Sample* MFCC dengan 0-konteks untuk *file* Abel T1P1

Pada percobaan keempat juga dilakukan *batch normalization* karena mengingat *training batch* lebih dari satu untuk meningkatkan akurasi pada *dataset*

testing. Dropout juga dilakukan untuk mempercepat learning untuk mencapai akurasi yang lebih tinggi. Pada percobaan ini terdapat banyak variasi dataset yang berbeda, mulai dari suara yang dibuat dari sistem Apple dan Bing (*speech synthesizer*) dinamakan *clean-synth*, suara direkam melalui rekaman studio dinamakan *clean-human*, dan suara direkam dengan *smartphone* dinamakan *noise-human*. Dataset yang digunakan pada percobaan ini dapat dilihat di Tabel 5.3. Beberapa parameter yang digunakan sebagai pembanding pada percobaan ini dapat dilihat di tabel 5.4.

Tabel 5.3. Tabel variasi dataset yang dilakukan pada percobaan ini

Dataset	**Total waktu**
clean synth	44 menit 3 detik
clean human	14 menit 6 detik
noise human	58 menit 48 detik

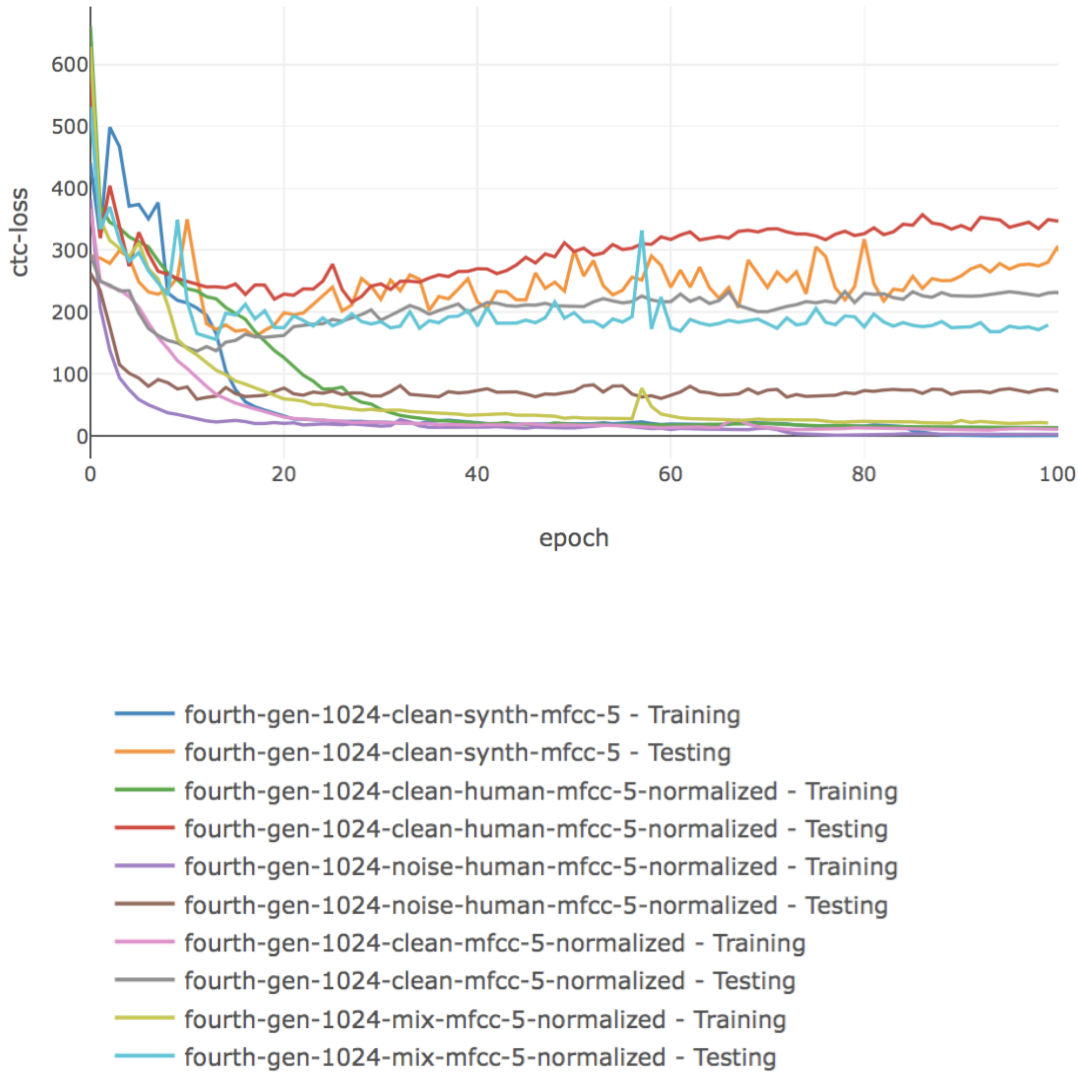
Tabel 5.4. Tabel parameter pembanding percobaan keempat

#	**Pembanding**
1	Variasi dataset (synth, clean, dan noise)
2	Variasi label (grafem dan fonem)
3	Variasi n-konteks (5 dan 9)

Semua dijalankan dengan insialisasi *weight* dan *bias* yang sama. Dengan perbandingan 9:1 antara *training* dan *testing dataset.* Parameter uji yang digunakan adalah CTC-loss pada training dan testing, Character Error Rate pada testing, dan waktu training dan testing.

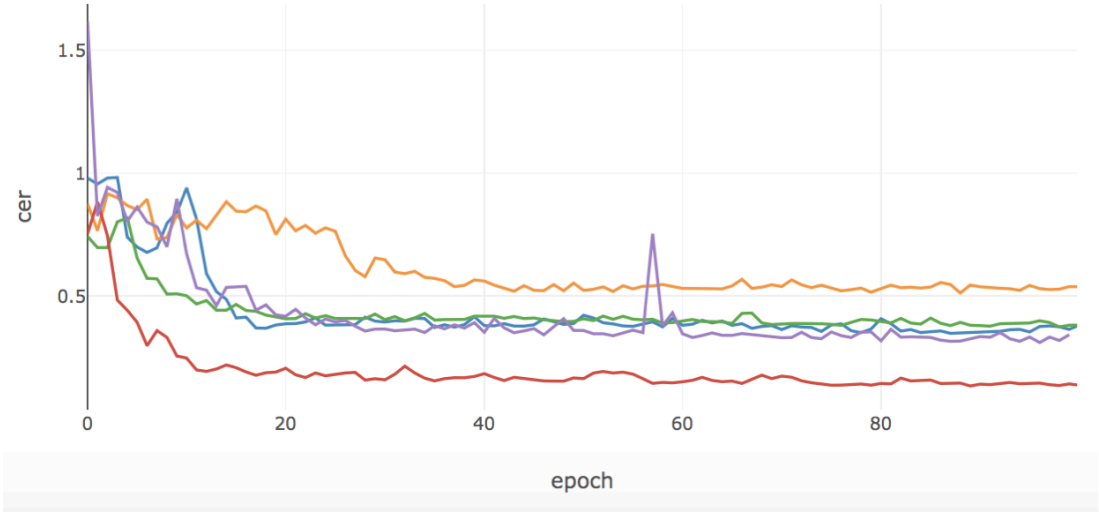
5.1.2.2. Variasi Dataset

Pada percobaan ini akan dibandingkan hasil dari dataset suara *clean synth*, *clean human*, *noise human, clean* dan *mix*. Grafik pembanding untuk *CTC-loss* pada *training* dan *testing* dataset dapat dilihat di gambar 5.3.



Gambar 5.3. Grafik CTC-loss pada training dan testing

Berikut grafik pembanding untuk CER *(Character Error Rate)* pada Testing dataset :



Gambar 5.4. Grafik CER pada testing dataset

Grafik pada Gambar 5.4 menunjukkan performa terbaik pada dataset *clean*, bahkan menambahkan variasi dataset dengan menggabung *clean synth* dan *clean human* mampu meningkatkan akurasi berlaku juga pada dataset *mix* yaitu penggabungan dari keseluruhan dataset. Performa terburuk pada dataset *clean human*. Tabel 5.5 menunjukkan bahwa dataset *clean* memiliki CTC-loss dan CER yang paling rendah.

Tabel 5.5. Tabel pembanding CTC-loss dan CER dengan variasi dataset

Dataset	**CTC-loss**		**CER**
	Training	**Testing**	
clean-synth	74.164	237.842	0.356
clean-human	42.082	335.573	0.494
clean	**11.751**	**76.774**	**0.120**
noise-human	19.114	226.643	0.345
mix	65.194	199.726	0.311

Tabel 5.6 menunjukkan waktu yang dibutuhkan dalam rata-rata setiap batch dalam satu epoch.

Tabel 5.6. Tabel kompleksitas waktu yang dibutuhkan setiap iterasi

Dataset	**Training (detik)**	**Testing (detik)**
clean-synth	3.026	0.385
clean-human	2.811	0.382
clean	2.881	0.371
noise-human	**2.621**	0.368
mix	2.911	**0.364**

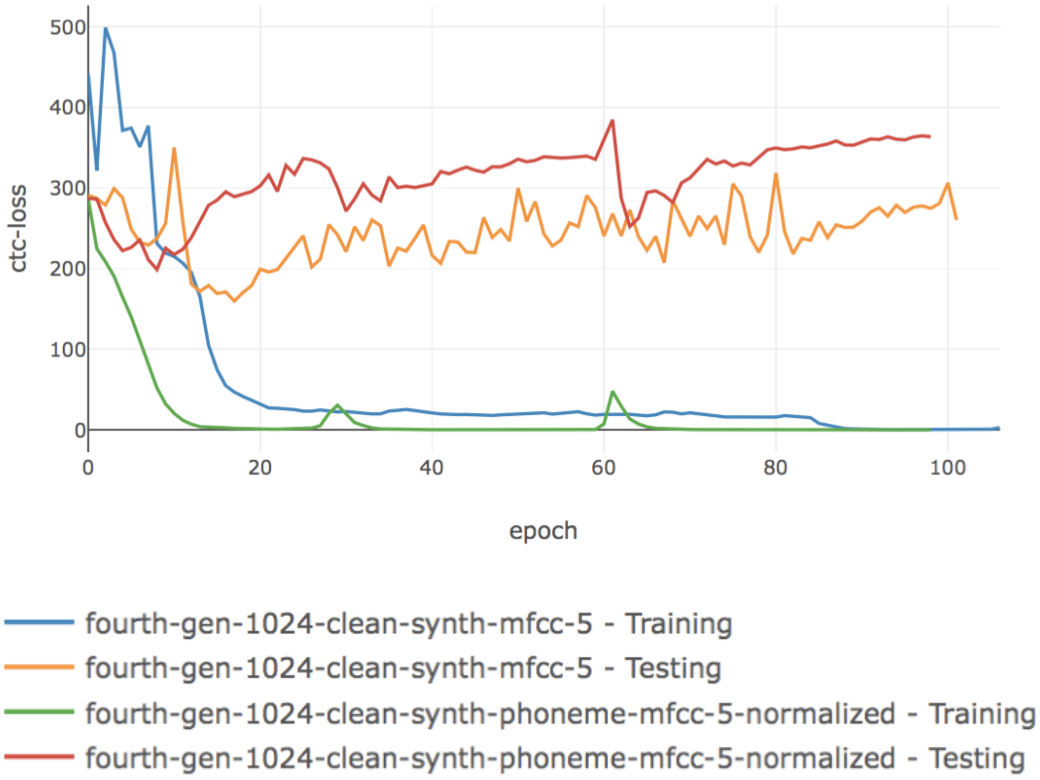
5.1.2.3. Variasi Label

Pada percobaan ini akan dilakukan per bagian dataset *clean synth*, *clean human*, *clean*, dan *noise human*. Fonem yang digunakan pada perubahan dari grafem dapat dilihat pada Tabel 5.7.

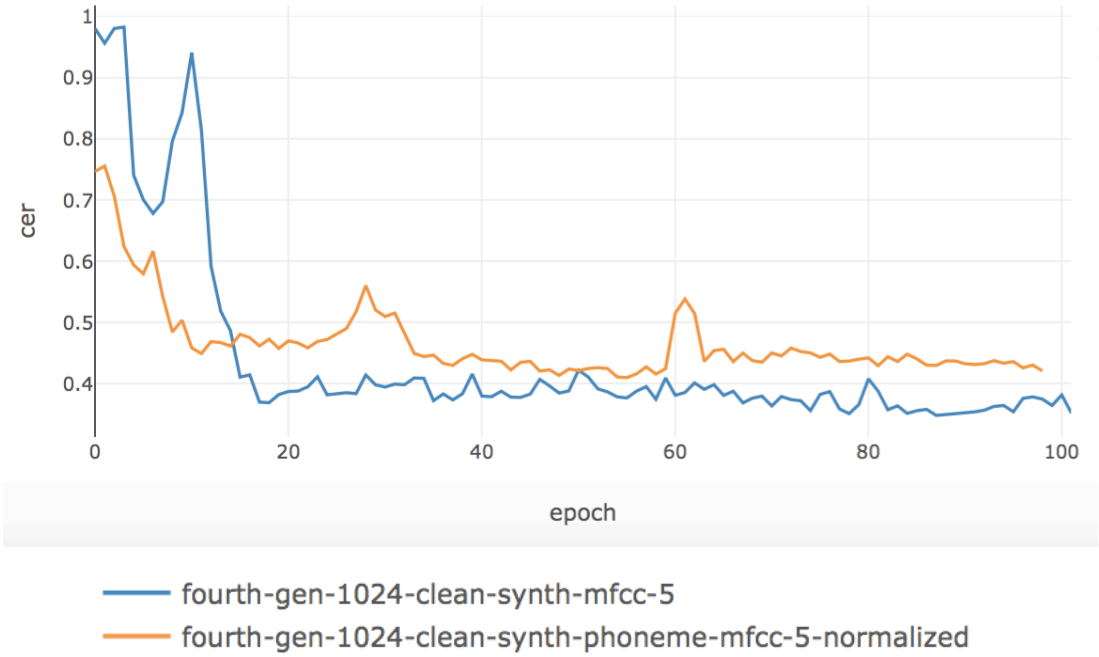
Tabel 5.7. Tabel perbandingan grafem dan fonem

Teks	**Grafem**	**Fonem**
e pada menang	e	ə
ng	ng	ŋ
ny	ny	ñ
sy	sy	S
kh	kh	x

Dari setiap dataset akan digunakan perbedaan label pada set karakter pada label grafem dan fonem. Grafik pembanding CTC-loss untuk dataset *clean synth* dapat dilihat pada gambar 5.5.



Gambar 5.5. Grafik CTC-loss pada *training* dan *testing*

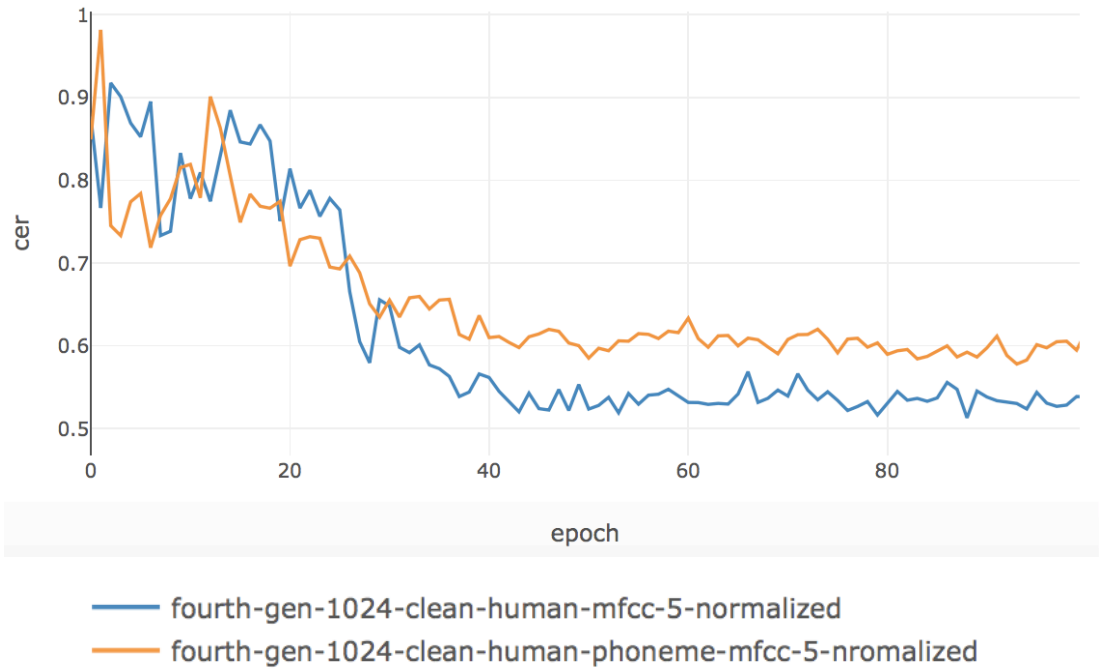


Gambar 5.6. grafik CER pada testing dataset

Percobaan menggunakan dataset *clean human,* Grafik pembanding CTC-loss untuk dataset *clean human* dapat dilihat pada gambar 5.7.

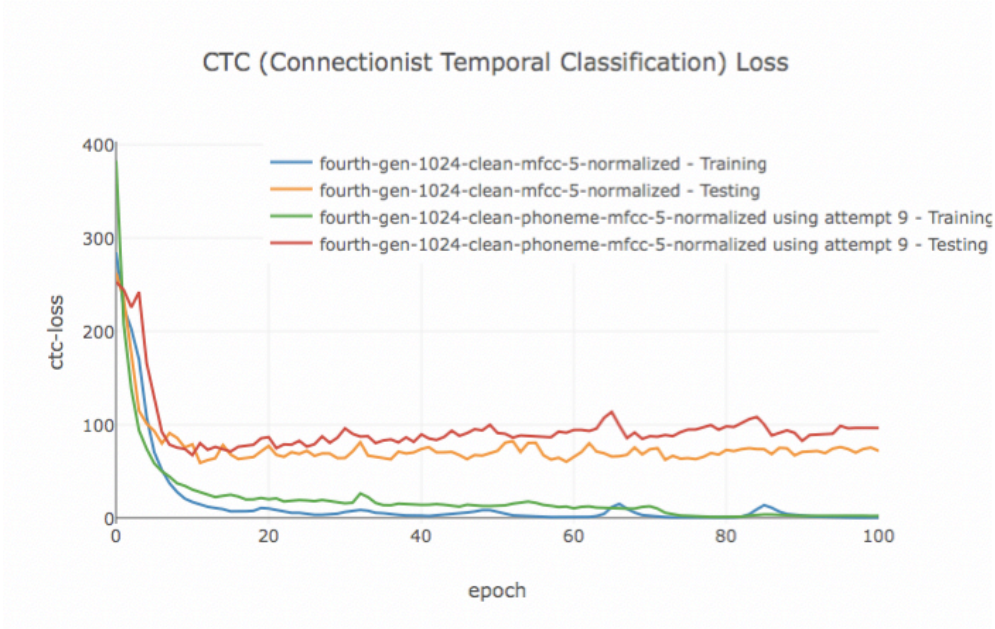
Gambar 5.7. Grafik pembanding CTC-loss pada label grafem dan fonem untuk dataset *clean human*.

Grafik pembanding untuk CER *(Character Error Rate)* pada Testing dataset dapat dilihat pada gambar 5.8



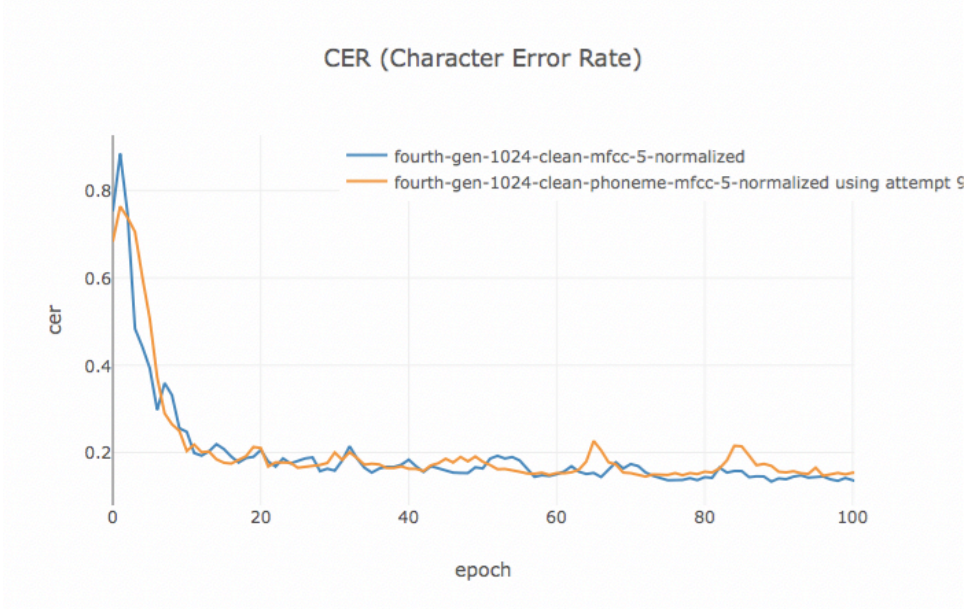
Gambar 5.8. Grafik pembanding CER pada label grafem dan fonem untuk dataset *clean human*.

Percobaan menggunakan dataset *clean human,* Grafik pembanding CTC-loss untuk dataset *clean* dapat dilihat pada gambar 5.9.



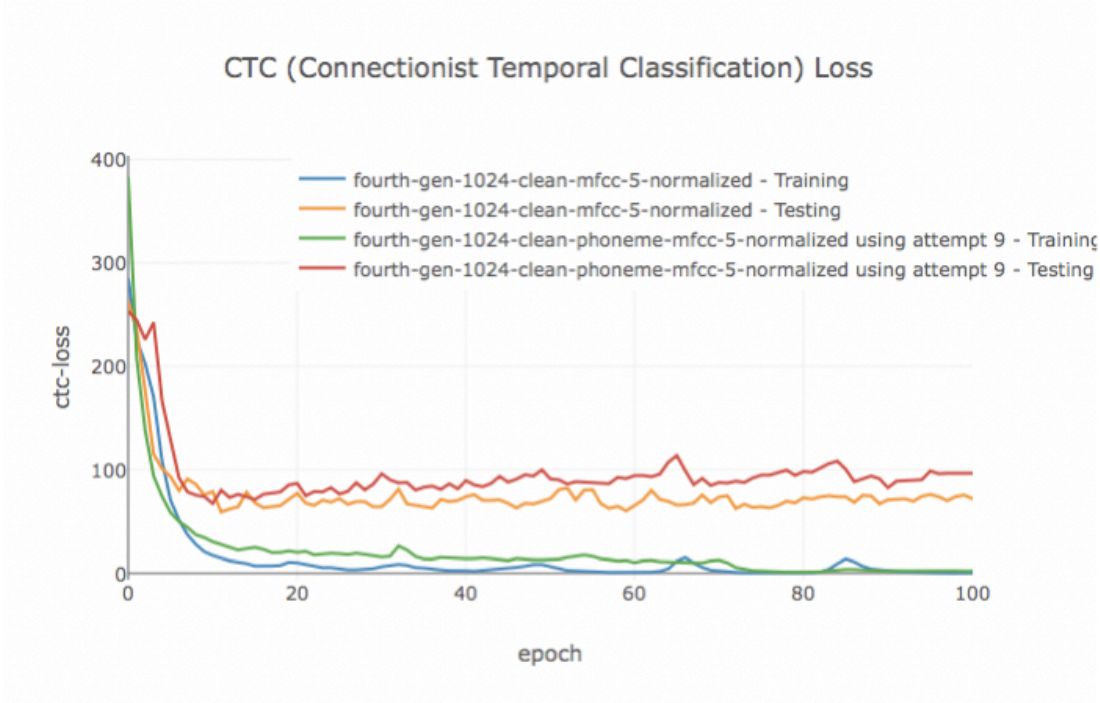
Gambar 5.9. Grafik pembanding CTC-loss pada label grafem dan fonem untuk dataset *clean*.

Grafik pembanding untuk CER *(Character Error Rate)* pada Testing dataset dapat dilihat pada gambar 5.10.



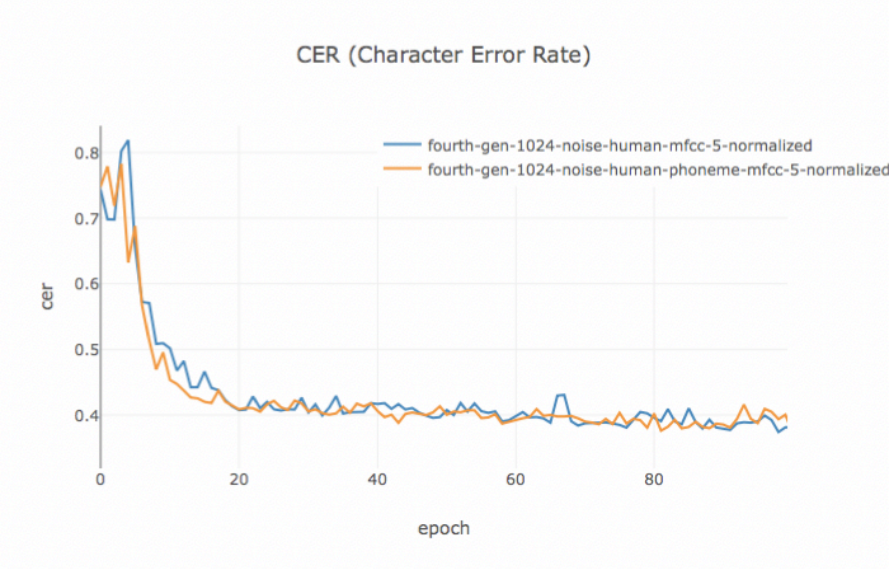
Gambar 5.10. Grafik pembanding CER pada label grafem dan fonem untuk dataset *clean*.

Grafik pembanding CTC-loss untuk dataset *noise human* dapat dilihat pada gambar 5.11.



Gambar 5.11. Grafik pembanding CTC-loss pada label grafem dan fonem untuk dataset *noise human*.

Grafik pembanding untuk CER *(Character Error Rate)* pada Testing dataset dapat dilihat pada gambar 5.12.



Gambar 5.12. Grafik pembanding CER pada label grafem dan fonem untuk dataset *noise human*.

Grafik pada gambar 5.8 menunjukkan performa terbaik adalah dataset yang menggunakan label grafem daripada menggunakan label fonem. Namun grafik pada gambar 5.10 dan 5.12 tidak mengalami perbedaan yang signifikan antara label grafem ataupun fonem. Label fonem dalam kasus tertentu menyebabkan hasil akurasi yang lebih kecil. Hasil perbandingan rata-rata CTC-loss dan minimum CER dapat dilihat pada tabel 5.8.

Tabel 5.8. Tabel pembanding rata-rata CTC-loss dan minimum CER yang dicapai dengan variasi label grafem dan fonem

Dataset	CTC-loss		CER
	Training	Testing	
clean-synth-grapheme	74.164	237.842	0.356
clean-synth-phoneme	17.654	312.295	0.413
clean-human-grapheme	42.082	335.573	0.494
clean-human-phoneme	38.640	390.884	0.581
clean-grapheme	11.751	**76.774**	**0.120**
clean-phoneme	**9.652**	96.837	0.143
noise-human-grapheme	19.114	226.643	0.345
noise-human-phoneme	14.932	245.620	0.353

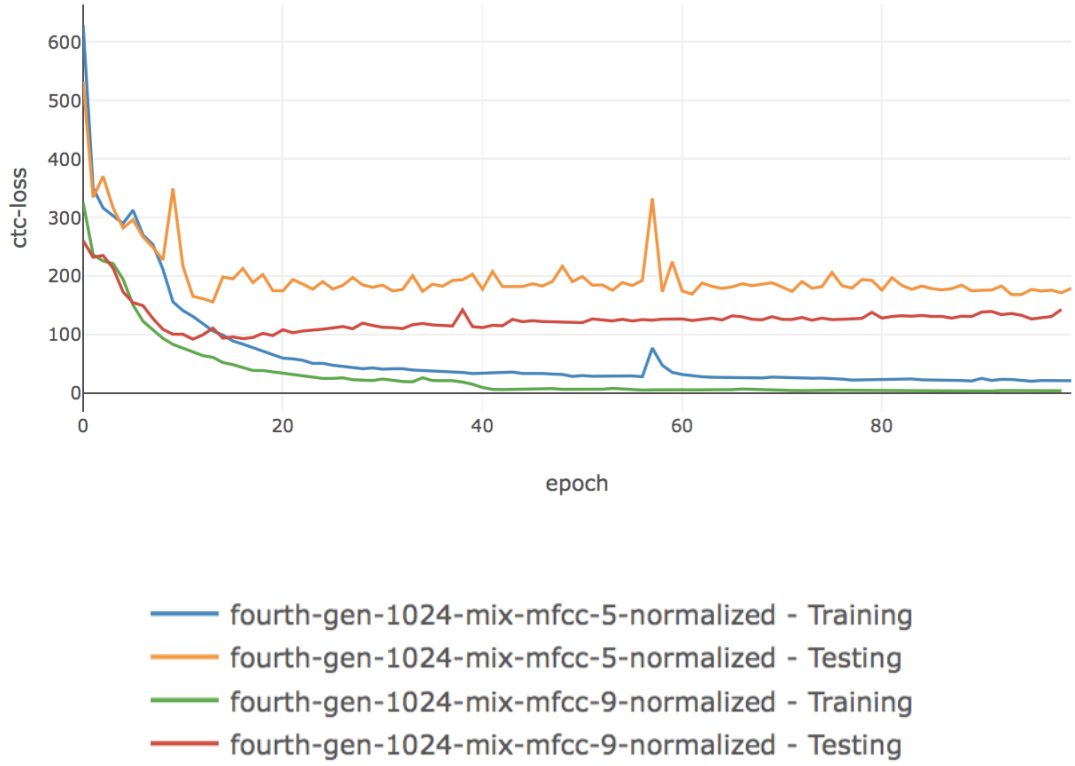
Tabel 5.9 menunjukkan waktu yang dibutuhkan dalam rata-rata setiap batch dalam satu epoch.

Tabel 5.9. Tabel pembanding CTC-loss dan CER dengan variasi label

Dataset	Training (detik)	Testing (detik)
clean-synth-grapheme	3.026	0.385
clean-synth-phoneme	2.977	**0.358**
clean-human-grapheme	2.811	0.382
clean-human-phoneme	2.981	0.365
clean-grapheme	2.881	0.355
clean-phoneme	2.974	0.385
noise-human-grapheme	**2.621**	0.368
noise-human-phoneme	2.665	0.361

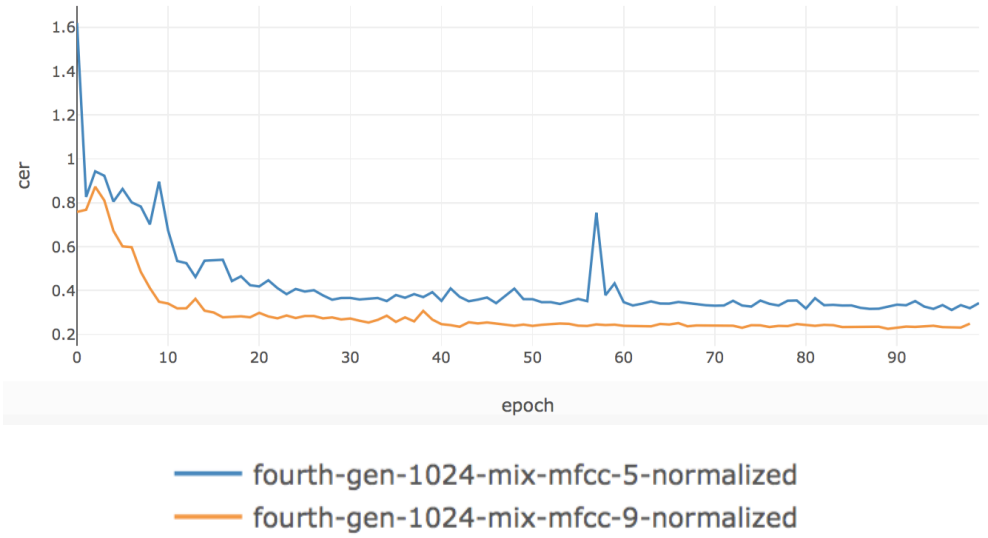
5.1.2.4. Variasi N-Konteks

Pada percobaan ini akan dilakukan pada dataset *mix*. Fitur sebagai pembanding adalah MFCC dengan 5 konteks dan 9 konteks. Grafik CTC-loss dari masing-masing fitur dapat dilihat pada gambar 5.13.



Gambar 5.13. Grafik pembanding CTC-loss pada variasi fitur MFCC 5 konteks dan 9 konteks.

Grafik pembanding untuk CER *(Character Error Rate)* pada Testing dataset dapat dilihat pada gambar 5.14.



Gambar 5.14. Grafik pembanding CER pada variasi fitur MFCC 5 konteks dan 9 konteks.

Grafik pada gambar 5.14 menunjukkan performa terbaik adalah pada MFCC dengan 9 konteks. Menaikkan konteks dapat meningkatkan akurasi. Hasil rata-rata CTC-loss dan minimum CER dapat dilihat pada tabel 5.10.

Tabel 5.10. Tabel pembanding rata-rata CTC-loss dan minimum CER yang dicapai dengan variasi fitur MFCC-5 dan MFCC-9

Dataset	CTC-loss		CER
	Training	Testing	
mix-mfcc-5	65.194	199.726	0.311
mix-mfcc-9	**30.627**	**126.273**	**0.255**

Tabel 5.11 menunjukkan waktu yang dibutuhkan dalam rata-rata setiap batch dalam satu epoch.

Tabel 5.11. Tabel kompleksitas waktu yang dibutuhkan setiap iterasi

Dataset	Training	Testing
mix-mfcc-5	2.911	**0.364**
mix-mfcc-9	**2.757**	0.503

5.1.3. Pengujian Aplikasi

Pada implementasi ini akan diambil *tensorflow_model* dengan CER terkecil untuk di jadikan sebagai model pada sistem yang telah dijelaskan pada subbab 4.5 dan 4.6. *Tensorflow model* yang digunakan adalah *clean*, *noise human*, *mix* MFCC dengan 5 konteks dan 9 konteks. Percobaan dilakukan dengan merekam dari 5 sample kalimat yang diambil berdasarkan urutan CER terendah pada testing dataset. Kalimat disuarakan ulang oleh 1 orang melalui smartphone Android yang terkoneksi dengan jaringan dan dikondisikan tidak ada suara penggangu *(noise background)*. Hasil *output* adalah kata yang dihasilkan *neural network (decode text)* dan hasil yang telah di koreksi oleh *language model (LM text).* Sebagai parameter uji dilakukan penghitungan WER *(Word Error Rate)* dari setiap kalimat setelah dihasilkan oleh *language model.* Hasil uji dapat dilihat pada tabel 5.12.

Tabel 5.12. Tabel hasil uji sistem *test_server.py* pada *tensorflow model* dataset *clean*.

Target	doakan hadiri bertumbuh bersama
Decode text	dekoarkaha kingti esnmumu kersas
LM text	dengarkan tinggi tersenyum keras
WER	1.0
Target	sebenarnya pantai suluban hanyalah nama lain untuk pantai uluwatu
Decode text	sebenahda asu tukn a sei ntu kngtau duatu
LM text	sebenarnya masuk tukang pei itu kerbau suatu
WER	0.889
Target	doa adalah nafas kehidupan orang percaya
Decode text	duah hahahda eauaskn u u lon bejada
LM text	buah bahagia kawasan lot berada
WER	1.0
Target	tiga september akan membahas hidupmu adalah injil yang terbuka
Decode text	di ke setkeam bur ke mem besidrunu adala inji angtunkng
LM text	di ke semacam buru ke jemi bersembunyi adalah injil angkuhanmu
WER	1.0
Target	sebagian besar wisatawan yang datang akan berselancar atau berjemur
Decode text	seku ya busa buisoatkwan ya bebom hotusa ja tatu tujulusu
LM text	sekuat ya bisa wisatawan ya belum solusinya jam satu eksklusif
WER	1.0

Dapat dilihat pada tabel 5.12 bahwa dataset suara *clean* dengan MFCC 5 konteks dimana terdiri dari rekaman studio dan hasil *speech synthesiser* dari Apple dan Bing tidak dapat mengenali kata-kata dengan kondisi data suara yang direkam langsung melalui smartphone.

Percobaan kedua dilakukan menggunakan *tensorflow model* dengan CER terendah pada dataset *noise human* dapat dilihat pada tabel 5.13.

Tabel 5.13. Tabel hasil uji sistem *test_server.py* pada *tensorflow model* dataset *noise human*.

Target	perpustakaan terkenal sebagai tempat
Decode text	kperpustakaran terkenang sebgi tompa
LM text	perpustakaan terkadang sebagai tempat
WER	0.250
Target	terkenal sebagai perpustakaan tempat
Decode text	tekena sebeukei perpstakan taemupa
LM text	terkenal sebelumnya perpustakan tempat
WER	0.250
Target	tapi ternyata ada juga yang dalam satu keluarga terdapat
Decode text	teti junya ata ad juga jyang dala satu kelurga u trdakpa
LM text	tetapi punya kata ada juga yang dalam satu keluarga terdapat
WER	0.334
Target	di keluarganya ada tiga agama berbeda yang dianut saudara saudaranya
Decode text	i kluaraya pada tia ama erbida yagkian sadara sadaranyauh
LM text	keluarnya pada tiga sama berbeda bagian saudara saudaranya
WER	0.6
Target	empat kereta pustaka kereta pustaka ini diresmikan pada tahun dua ribu sebelas kereta pustaka ini letaknya berpindah pindah gak hanya satu tempat saja
Decode text	ampa kereka pstaka kerta pstaka isni jidesmikan are taun keuarikusebelas kata pustaka ini betanya ertindakindah ka hanya satu tepasacah
LM text	sampai mereka pustaka kereta pustaka disini diresmikan area tahun walaikumsalam kata pustaka ini bertanya pertunangan ka hanya satu permasalahan
WER	0.789

Dapat dilihat pada tabel 5.14 dataset *noise human* dengan MFCC 5 konteks dimana suara direkam melalui smartphone langsung dapat mengenali beberapa kata meskipun kata tersebut diacak urutannya (lihat contoh kalimat 1 dan 2).

Percobaan ketiga dilakukan menggunakan *tensorflow model* dengan CER terendah pada dataset *mix* dengan MFCC 5 konteks dapat dilihat pada tabel 5.14.

Tabel 5.14. Tabel hasil uji sistem *test_server.py* pada *tensorflow model* dataset *mix* dengan MFCC *5 konteks.*

Target	perpustakaan terkenal sebagai tempat
Decode text	parpustaka toena a suka i tema
LM text	perpustakan terkenal suka tema
WER	0.75
Target	terkenal sebagai perpustakaan tempat
Decode text	tkerkan mang sebat tempa e pustak ks an e
LM text	kerjaan emang hebat tempat pustaka aksi dan
WER	1.5
Target	komisi kesaksian dan pelayanan akan mengadakan persekutuan doa sektor
Decode text	komisi ke sasian dar belayae na a kan akahdaokan uskut ua oas e trhuauh
LM text	komisi ke saksikan dari pelayan nah kan diadakan takut dua pas terhadap
WER	1.2
Target	sebenarnya pantai suluban hanyalah nama lain untuk pantai uluwatu
Decode text	seben ayan panfa sulu ban nyalana mala a in uentuk pan ta unuatu bs
LM text	seberang akan pantai dulu ban nyawanya malam in untuk dan tak uluwatu
WER	1.0
Target	empat kereta pustaka kereta pustaka ini diresmikan pada tahun dua ribu sebelas
Decode text	tmpa e ta pustakara stakainidli s ikan pal ta u ua i pus bela
LM text	tempat tak pustaka mengklarifikasi ikan hal tak dua puas belas
WER	1.0

Dapat dilihat pada tabel 5.14 dataset *mix* dengan MFCC 5 konteks, pada kasus ini beberapa kata dapat dikenali dengan baik, namun beberapa imbuhan seperti ke-, -kan dan - an mendapati hasil yang ambigu antara diberi spasi atau tidak. Ini merupakan kesulitan yang menjadi penyebab besarnya juga error yang terjadi. Masalah ambiguitas pada penelitian ini tidak ditinjau lebih jauh karena tidak masuk dalam ruang lingkup.

Percobaan ketiga dilakukan menggunakan *tensorflow model* dengan CER terendah pada dataset *mix* dengan MFCC 9 konteks dapat dilihat pada tabel 5.15.

Tabel 5.15. Tabel hasil uji sistem *test_server.py* pada *tensorflow model* dataset *mix* dengan MFCC 9 konteks.

Target	perpustakaan terkenal sebagai tempat
Decode text	perpustakan cearto na se beor i sepot
LM text	perpustakan berarti nah ase belom sepotong
WER	1.5
Target	terkenal sebagai perpustakaan tempat
Decode text	terbenta seba kni terfat perpustakan
LM text	terbenam sebab kini terpaut perpustakan
WER	1.25
Target	komisi kesaksian dan pelayanan akan mengadakan persekutuan doa sektor
Decode text	komisi kesaksian dan pelayanan akan mengadakan prsekua klasektori
LM text	komisi kesaksian dan pelayanan akan mengadakan persekutuan kalajengking
WER	0.222
Target	sebenarnya pantai suluban hanyalah nama lain untuk pantai uluwatu
Decode text	sebendanya ba besu dan banyalaena ma wa ingutuk pang sain pulua tu
LM text	sebenarnya iba besar dan banyaknya mau wah mengulur yang lain puluh tuh
WER	1.222
Target	empat kereta pustaka kereta pustaka ini diresmikan pada tahun dua ribu sebelas
Decode text	empat dirata bustat eata pustaka ini didisndikan pauda taun dua ribusebueleas
LM text	empat dirasa buatan kata pustaka ini dirindukan pada tahun dua disebutkan
WER	0.5

Dapat dilihat tabel 5.15 dataset *mix* dengan MFCC 9 konteks, pada kasus ini dalam mengenali kata lebih laik daripada *mix* MFCC 5 konteks (bandingkan dengan tabel 5.15), namun dataset noise human masih lebih baik (bandingkan dengan tabel 5.16). Tabel perbandingan rata - rata WER dapat dilihat pada tabel 5.17.

Tabel 5.16. Tabel 5.17 hasil perbandingan rata-rata WER

Tensorflow model	**WER**
clean dengan MFCC 5 konteks	0.9778
noise human dengan MFCC 5 konteks	0.4446
mix dengan MFCC 5 konteks	1.09
mix dengan MFCC 9 konteks	0.9388