DT2112 Lab2 Report
Practical exercise in automatic speech recognition

Group member names

- Sp1:
- Sp2:
- Sp3:

Grammar explanation and graph (four digits):

Copy the content of your `four_digits.grm` definition and include the graph you obtained from it. What kind of utterances does this grammar allow? How was this grammar obtained by the rules you defined in `four_digits.grm`?

Grammar explanation and graph (digit loop):

Copy the content of your `digit_loop.grm` definition and include the graph you obtained from it. What kind of utterances does this grammar allow? How was this grammar obtained by the rules you defined in `digit_loop.grm`?

Words and their phonetic transcirption

Report the words in your dictionary and the phonetic transcriptions you have defined.

Feature extraction parameters

These are defined in the configuration file `config/features.cfg`, where times are given in 100 ns units. Convert these values into the units specified in the third column of the table. Print the contents using the command: `cat config/features.cfg`.

The sampling frequency of the recordings is not specified in `config/features.cfg`. To extract it, look at one of the files you have recorded. You can use the command `file` to get information about the file, or use the file browser. In the second case right click on the file and choose the "Properties" option and then the "Audio" tab. The recordings are for each group member in the `Sp$n/train_data` and `Sp$n/test_data` directories

Parameter	Hint	Value
Sampling frequency (kHz):	check recordings	(kHz)
Analysis window (ms):	WINDOWSIZE (100 ns units)	(ms)
Frame interval (ms):	TARGETRATE (100 ns units)	(ms)
Pre-emphasis coeff.	PREEMPHCOEFF	
Filterbank # channels	NUMCHANS	
Energy normalization	ENORMALISE	
# cepstrum coefficients	NUMCEPS	
Hamming	USEHAMMING	

Answer the following questions:

How many speech samples are contained in one analysis window?

How much do consequent analysis windows overlap?

In a typical four digit utterance that you have recorded, how many analysis windows are used?

Acoustic model parameters

Check the prototype model definition in the file `proto.mmf` and answer the following questions with the help of the HTK Book:
What kind of features are used? (hint: defined in the `~o` macro)

What is the size of the feature vector?

How many states are used per phoneme?

Draw the topology (states and transitions) of the prototype model (Hint: `TransP` is the transition

probability matrix).

Recognition evaluation

Speaker dependent results: Matching the test speaker against his/her own trained models. Cross-speaker results: Matching the test speaker against another speaker's models.

4-digits	Training speaker(s)	Test speaker(s)	Accuracy %	#Ins	#Del
Speaker dependent	Sp1	Sp1			
Speaker dependent	Sp2	Sp2			
Speaker dependent	Sp3	Sp3			
Cross-speaker	Sp1	Sp2			
Cross-speaker	Sp1	Sp3			
Cross-speaker	Sp2	Sp1			
Cross-speaker	Sp2	Sp3			
Cross-speaker	Sp3	Sp1			
Cross-speaker	Sp3	Sp2			

digit-loop	Training speaker(s)	Test speaker(s)	Accuracy %	#Ins	#Del
Speaker dependent	Sp1	Sp1			
Speaker dependent	Sp2	Sp2			
Speaker dependent	Sp3	Sp3			
Cross-speaker	Sp1	Sp2			
Cross-speaker	Sp1	Sp3			
Cross-speaker	Sp2	Sp1			
Cross-speaker	Sp2	Sp3			
Cross-speaker	Sp3	Sp1			
Cross-speaker	Sp3	Sp2			

Discussion of the results

Common digit confusions:

These might have been caused by:

Compare and discuss the difference between the cross-speaker and the speaker-dependent results.

How do the performance and types of errors differ between recognition of fixed and non-restricted number of digits (digit loop)?

Experience with the live recogniser

how does the recogniser behave when you use it live?

what is the effect of varying the speaking rate?

what is the effect of varying the insertion penalty?

how does the recogniser cope with long pauses between words? Can you change the behaviour using the insertion penalty?

what happens when you include an optional silence "word" between each digits in the grammar?

what words did you add to the dictionary and grammar? How did the recogniser perform with them?