Combined RSA and DRA For Noise-Robust Continuous Speech Recognition System

Yiming SUN

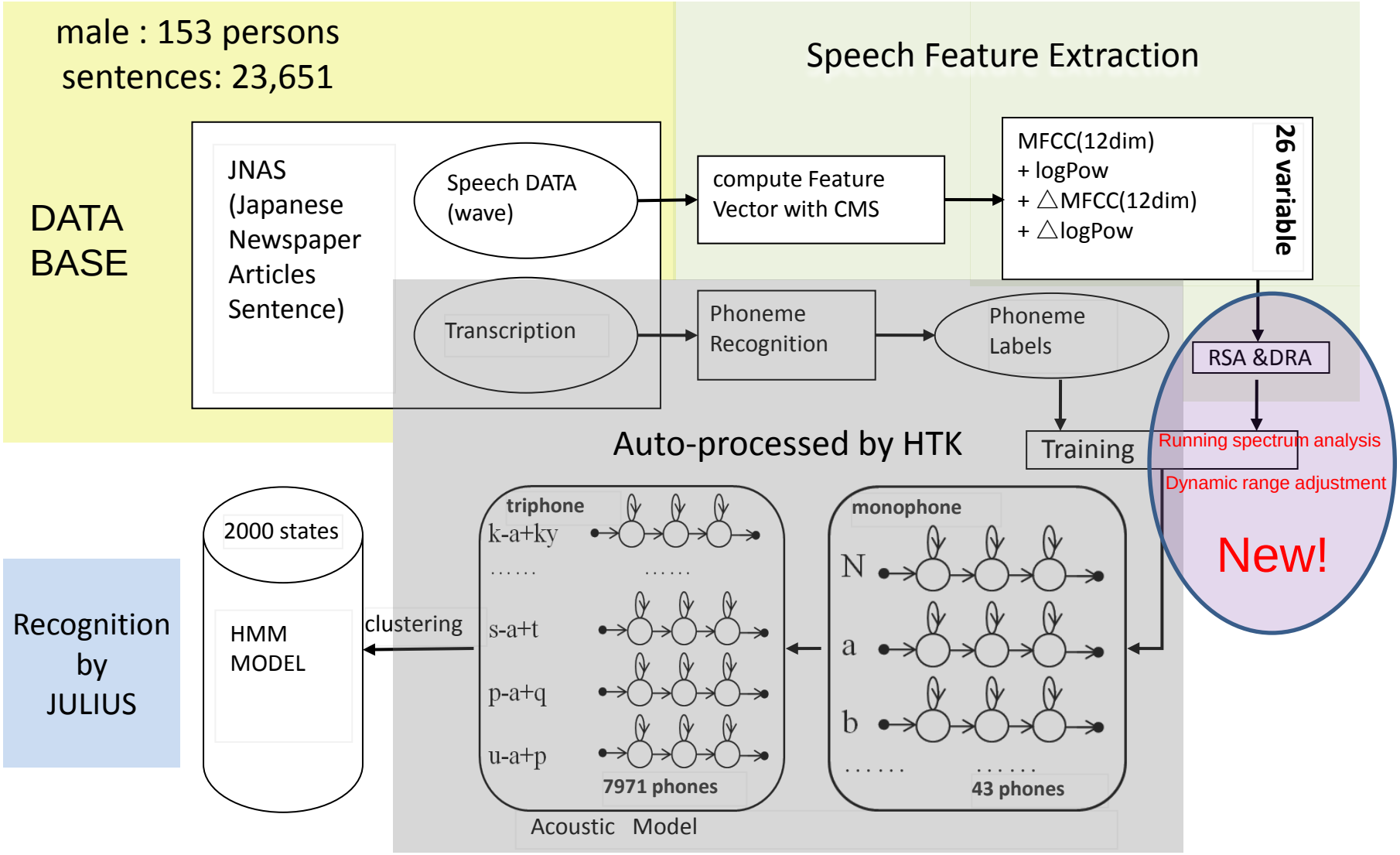
Information Communication Networks Laboratory
Graduate School of Information Science and Technology
Hokkaido University, Sapporo, Japan
sunny@icn.ist.hokudai.ac.jp

Overview

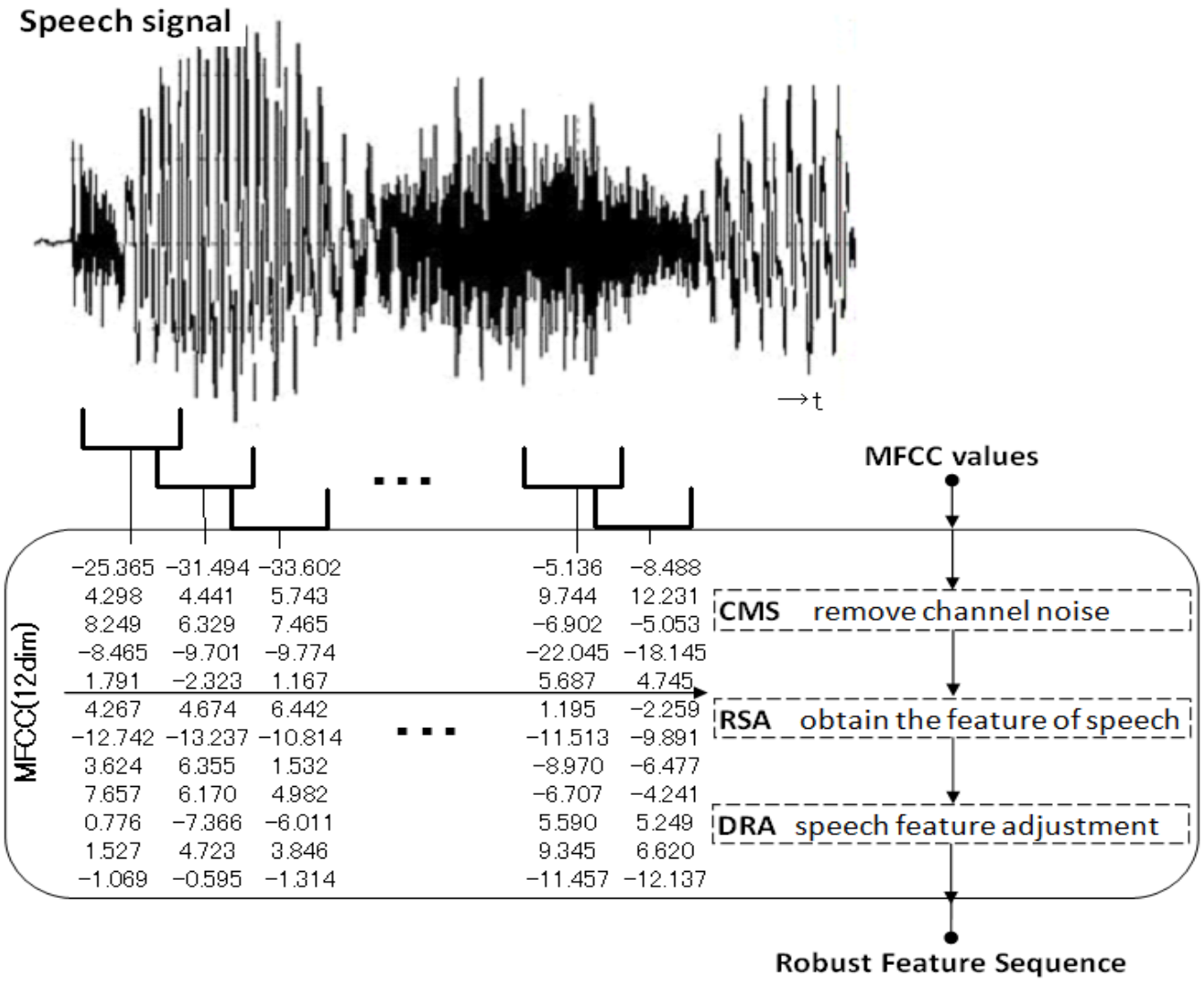
北海道大学
HOKKAIDO UNIVERSITY

- Robust continuous speech recognition (CSR) System
- Speech Feature Extraction
  - CMS: cepstrum means subtraction
  - RSA: running spectrum analysis
  - DRA: dynamic range adjustment
- Results

Robust CSR System

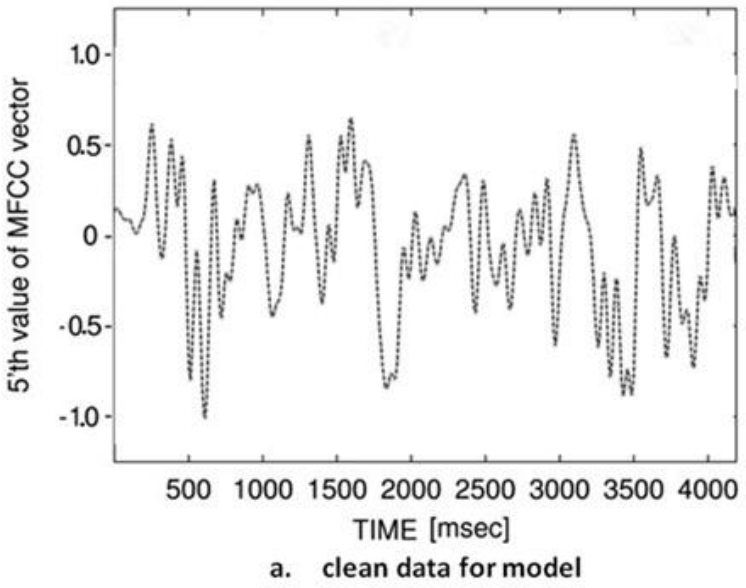


Speech Feature Extraction Flow Sheet

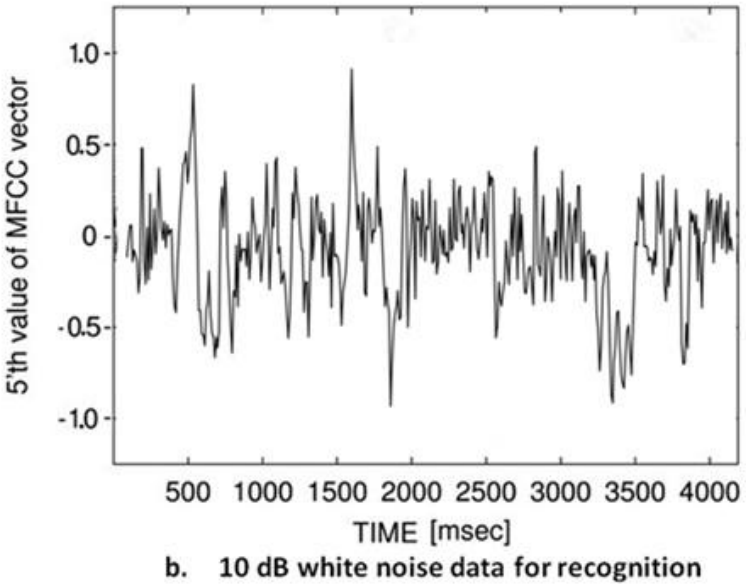


RSA and DRA

- RSA can smooth the influences of the speech feature values
- DRA significantly reduces noise corruption and preserving important speech characteristics.
- The range of feature values approaches those of the clean feature values after RSA and DRA.



a. clean data for model



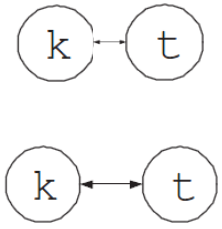
b. 10 dB white noise data for recognition

RSA and DRA

- Use Mahalanobis distance to measure the separation of two groups of objects, phoneme i and phoneme j, the equation is given by the following:

$$D_{ij} = \sqrt{\frac{K\sum_{k=1}^{K}(\mu_{ik} - \mu_{jk})^2}{\sum_{k=1}^{K}\sigma_{ik}^2 + \sum_{k=1}^{K}\sigma_{jk}^2}}$$

- In high SNR condition, combined RSA and DRA method ? the distance between phonemes.



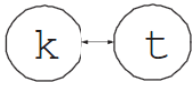
?

RSA and DRA

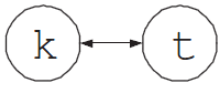
- Use Mahalanobis distance to measure the separation of two groups of objects, phoneme i and phoneme j, the equation is given by the following:

$$D_{ij} = \sqrt{\frac{K\sum_{k=1}^{K}(\mu_{ik} - \mu_{jk})^2}{\sum_{k=1}^{K}\sigma_{ik}^2 + \sum_{k=1}^{K}\sigma_{jk}^2}}$$

- In high SNR condition, combined RSA and DRA method increased the distance between phonemes.



Use RSA only

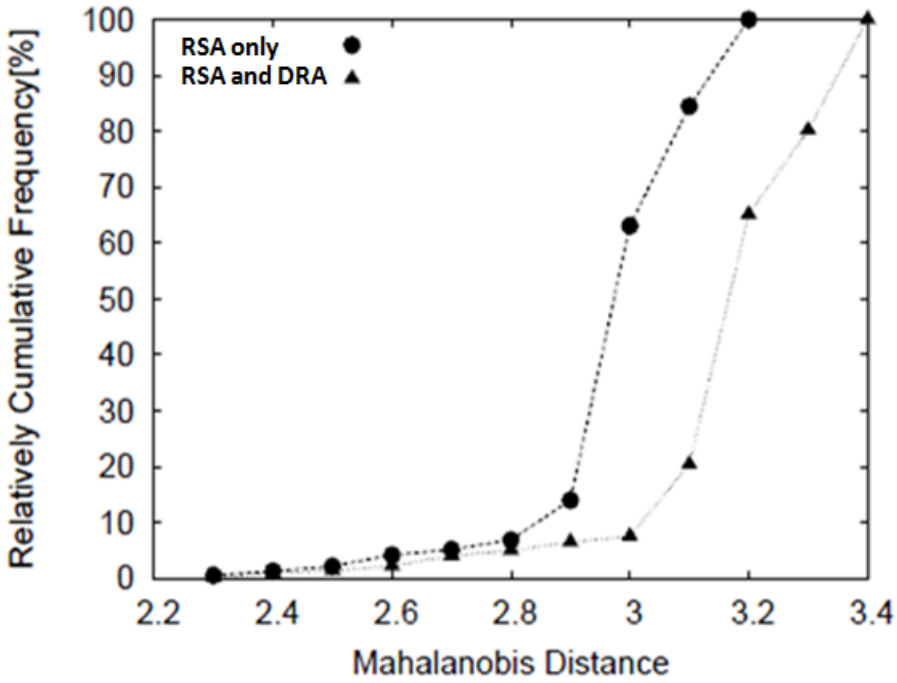


Combined RSA and DRA

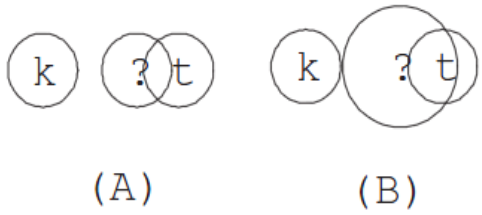
RSA and DRA

北海道大学
HOKKAIDO UNIVERSITY

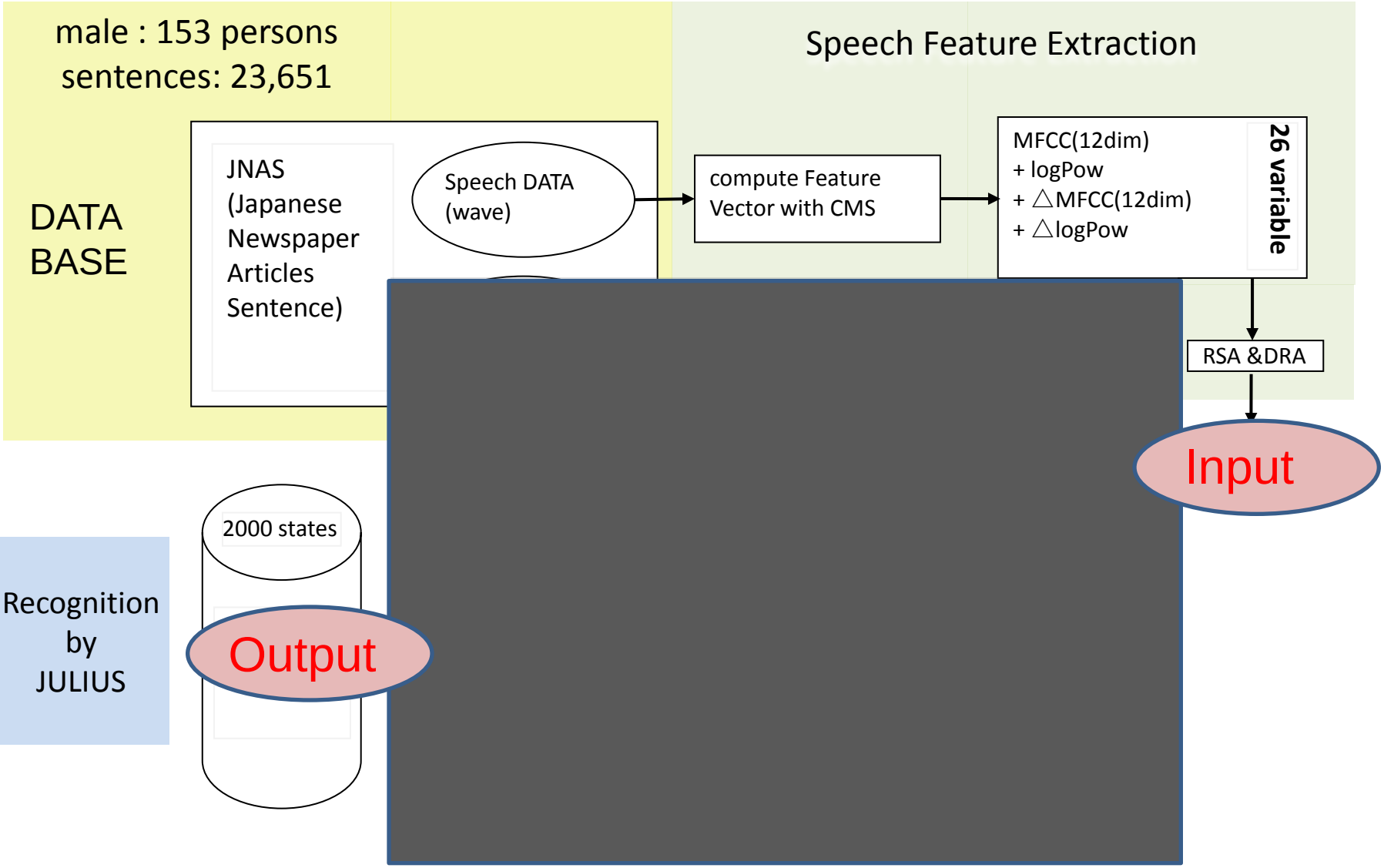
- Mahalanobis Distance compare using RSA only with combined RSA and DRA.



- A is the conditions of clean data.
- B is the high SNR (10 dB) conditions.



Robust CSR System



Result

$$\text{Percent Accuracy} = \frac{N - D - S - I}{N} \times 100\%$$

- Shows accurately the total performance

$$\text{Percent Correct} = \frac{N - D - S}{N} \times 100\%$$

- Shows the correct word recognition rate

N: Total number of words
D:Deletion errors
S:Substitution errors
I: Insertion errors

Conditions: 20 dB Known data 50 sentences

	RSA+DRA		RSA	
	Corr	Acc	Corr	Acc
babble	77.53	72.47	76.46	73.01
buccaneer1	74.87	73.01	73.40	71.68
buccaneer2	76.60	75.00	73.94	72.74
destroyerenginer	78.72	76.46	77.39	75.93
destroyerops	82.98	81.12	82.85	81.78
f16	77.79	75.93	74.60	72.74
factory1	78.72	77.53	77.79	76.60
factory2	78.72	77.53	77.79	76.60
hfchannel	66.49	63.70	57.31	55.05
leopard	85.24	80.98	86.57	83.11
m109	88.30	87.23	89.89	88.96
machinegun	83.64	77.53	80.98	74.47
pink	79.26	77.66	76.99	76.06
volvo	89.76	88.16	91.09	89.89
white	68.35	65.82	69.95	68.62
average	79.13	76.68	77.8	75.82

Result

Conditions: 15 dB　Known data 50 sentences

Unknown data 180 sentences

	RSA+DRA		RSA	
	Corr	Acc	Corr	Acc
babble	63.30	55.45	62.63	59.31
buccaneer1	55.19	52.66	41.49	40.65
buccaneer2	59.44	56.65	44.02	41.89
destroyerenginer	63.30	59.84	56.78	54.79
destroyerops	74.20	71.94	66.36	64.63
f16	64.23	62.23	54.39	52.26
factory1	65.03	63.30	55.72	54.26
factory2	78.06	76.60	77.26	75.53
hfchannel	47.61	43.48	35.77	32.85
leopard	82.05	76.33	82.18	77.53
m109	84.84	83.78	81.38	80.05
machinegun	78.32	69.15	78.86	69.81
pink	65.56	64.49	49.07	47.07
volvo	89.89	88.83	91.36	90.16
white	46.41	41.89	39.23	36.04
average	67.83	64.44	61.10	58.40

	RSA+DRA		RSA	
	Corr	Acc	Corr	Acc
babble	65.23	60.33	65.27	63.08
buccaneer1	54.52	53.17	41.25	40.08
buccaneer2	52.53	50.04	40.05	38.54
destroyerenginer	58.27	55.80	53.09	51.36
destroyerops	68.44	65.67	63.16	61.49
f16	60.18	58.41	51.58	49.96
factory1	61.16	59.16	54.22	52.83
factory2	73.49	71.42	71.42	70.02
hfchannel	45.55	42.27	39.22	37.37
leopard	75.57	72.4	76.24	72.51
m109	77.11	75.11	78.17	76.89
machinegun	74.63	66.01	74.96	66.10
pink	57.81	55.73	45.29	43.78
volvo	79.71	77.45	81.98	79.90
white	39.56	37.14	36.73	35.33
average	62.92	60.01	58.18	55.95

Result

Conditions: 10 dB　Known data 50 sentences

Unknown data 180 sentences

	RSA+DRA		RSA	
	Corr	Acc	Corr	Acc
babble	36.97	26.20	36.04	32.05
buccaneer1	29.26	26.86	15.56	14.63
buccaneer2	34.31	32.31	18.08	16.87
destroyerenginer	40.16	36.70	24.07	23.14
destroyerops	51.20	48.80	34.18	33.24
f16	37.63	34.57	24.07	23.14
factory1	42.15	38.83	25.93	25.00
factory2	61.84	58.64	52.66	51.06
hfchannel	24.73	22.34	15.29	13.96
leopard	79.39	74.87	77.93	72.87
m109	75.13	73.27	61.30	59.97
machinegun	78.46	68.88	77.53	66.89
pink	36.57	34.71	19.28	18.09
volvo	89.89	88.70	88.96	87.50
white	26.60	23.67	17.55	16.22
average	49.62	45.96	39.23	36.98

	RSA+DRA		RSA	
	Corr	Acc	Corr	Acc
babble	44.49	35.9	38.8	36.69
buccaneer1	30.35	28.62	14.48	13.88
buccaneer2	30.51	28.88	18.93	18.29
destroyerenginer	38.84	37.07	21.08	20.51
destroyerops	49.41	46.57	34.92	34.28
f16	41.86	40.61	19.42	18.55
factory1	39.63	38.46	23.23	22.89
factory2	63.12	61.24	54.86	53.51
hfchannel	23.00	21.08	14.29	13.57
leopard	74.74	71.98	72.21	68.55
m109	69.04	66.78	59.31	57.88
machinegun	73.19	60.48	72.13	59.54
pink	34.62	33.14	17.87	17.16
volvo	80.09	78.02	78.21	76.90
white	23.30	21.49	16.25	15.61
average	47.75	44.69	37.06	35.19

Thank you!

Question?