Gender variation effects on ASR system performance

Presenter: Afsheen

Problem Statement

To design an automatic speech recognition system that gives best recognition results for both male and female speakers.

Parameters effecting ASR performance

1. Gender
2. Accent
3. Age
4. Emotion
5. Health
6. Noise

Suppress gender variation effects on Automatic Speech Recognition Performance

Possible Solutions

Solution	Training data	Testing data
1.Train a single gender specific system	Male/Female	Male + Female
2.Train two acoustic models, one model for each gender	Male	Male
	Female	Female
3.Train a single model on both male and female data	Male + Female	Male+ Female

Gender Variation Effects on ASR System Performance

Literature Survey

Paper	Training Corpus	Testing Corpus	Experiments		Results	Support
			Training data	Testing data		
[1]	100 sentences uttered by 30 male and 30 female speakers.	100 sentences uttered by 10 male and 10 female speakers.	Male	M+F	57	PS.3
			Female	M+F	69	
			M+F	M+F	87	
[2]	2496 talks uttered by 1508 male and 988 female speakers.	20 talks uttered by 15 male and 5 female speakers.	GD	GD	71.58	PS.3
			M+F	M+F	71.78	
[3]	7037 sentences uttered by 7 male +	1002 sentences uttered by either one	GD	GD	93.88	PS.3
			M+F	M+F	96.29	

Gender Variation Effects on ASR System Performance

Paper	Training Corpus	Testing Corpus	Experiments		Results	Support
			Training data	Testing data		
[4]	390 sentences uttered by 56 male and 36 female speakers.	390 sentences uttered by 26 male and 19 female speakers.	Male	Male	38.01	PS. 2
			Female	Female	34.74	
			M+F	M+F	36.29	
[5]	10 sentences uttered by 630 male and female speakers.		Male	Male	36.35	PS.3
			Female	Female	36.12	
			M+F	M+F	40.10	

[4] Statistical Evaluation of the Effect of Gender on Prosodic Parameters and their Influence on Gender Dependent Speech Recognition
[5] A study on pitch variation on the use of DWT with SVM for Speaker dependent phoneme.

Location based ASR system

Vocabulary	Training Corpus	Testing Corpus	Experiments		Results
			Training data	Testing data	
48 place names	14 place names uttered by 48 male and 48 female speakers.	2 place names uttered by 48 male and 48 female speakers.	Male	Male	47.50
				Female	26.00
			Female	Male	24.00
				Female	57.50
			M+F	Male	88.50
				Female	97.00
				M+F	92.75

Conclusion

To get good accuracy results for both male and female data it is suggested to train a single ASR on mixed (male + female) data rather than separately training the two for each gender.