

Gender Effects in Speech Perception: An ETSI TR 103 950-Based Analysis of Listening Quality and Perceptual Dimensions

Presented by: Wafaa Wardah





Do telecom technologies
affect male and female
speech differently?




Pitch

Speaking style

Speed

...



Do telecom technologies
affect male and female
speech differently?



Overall Quality?

Noisiness?

Coloration?

Discontinuity?

Loudness?

Do telecom technologies
affect male and female
speech differently?

MALE Speakers

$$\text{MOS}_{\text{Reference}} - \text{MOS}_{\text{Condition}} = \text{Difference}_{\text{Condition}}$$

**Significantly different
differences across all
conditions**



FEMALE Speakers

$$\text{MOS}_{\text{Reference}} - \text{MOS}_{\text{Condition}} = \text{Difference}_{\text{Condition}}$$

Significantly different differences across all conditions, across dimensions?

MALE Speakers

Overall MOS

?

FEMALE Speakers

Overall MOS

Significantly different differences across all conditions, across dimensions?

MALE Speakers

Overall MOS

Noisiness

Coloration

Discontinuity

Loudness

?

?

?

?

?

FEMALE Speakers

Overall MOS

Noisiness

Coloration

Discontinuity

Loudness

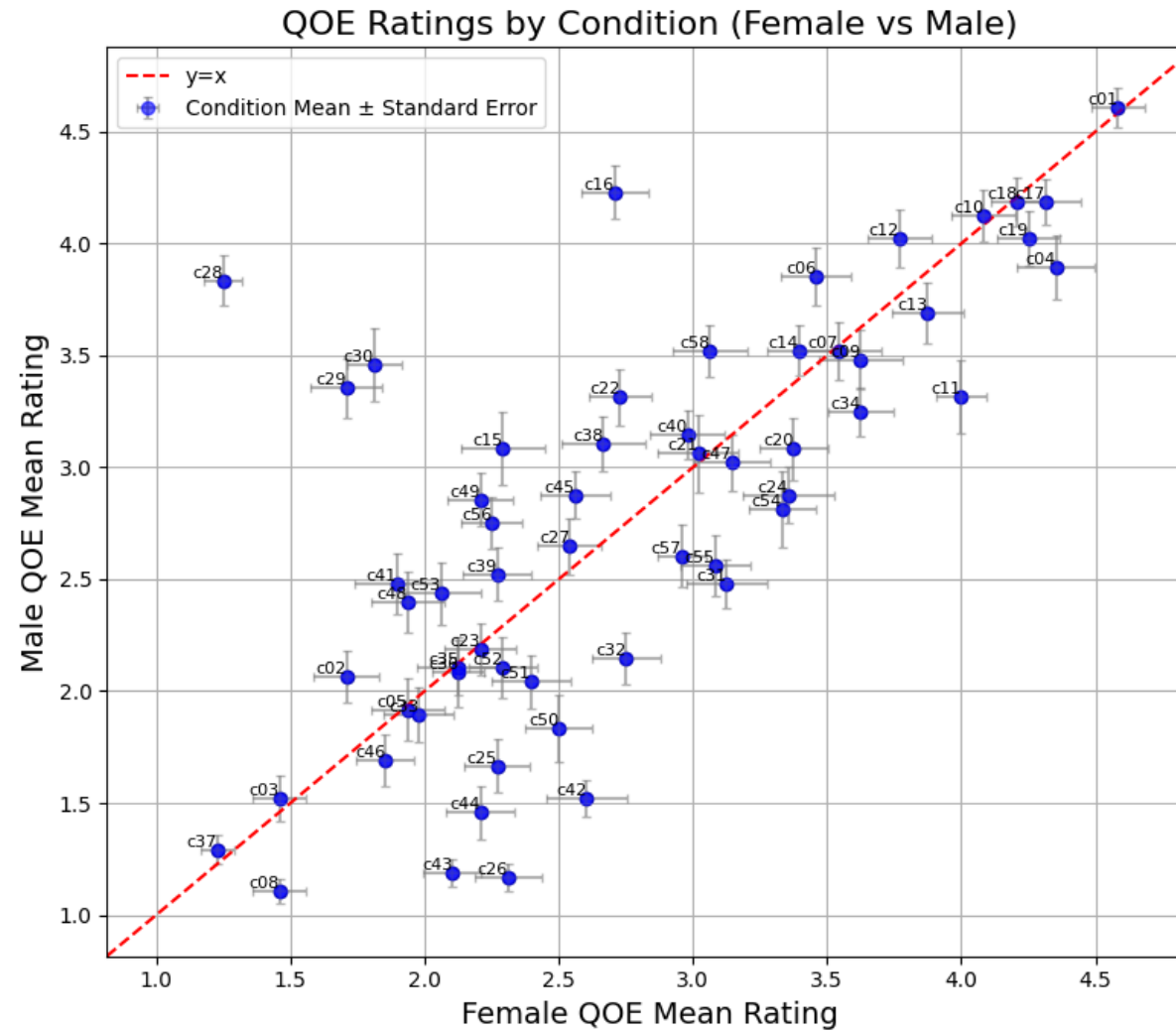
Dataset

- 232 speech files – 6-12 seconds long
- 57 conditions (plus reference)
- 4 files per condition - 2 male, 2 female speakers
- 24 votes per file - 96 votes per condition
- Language: German

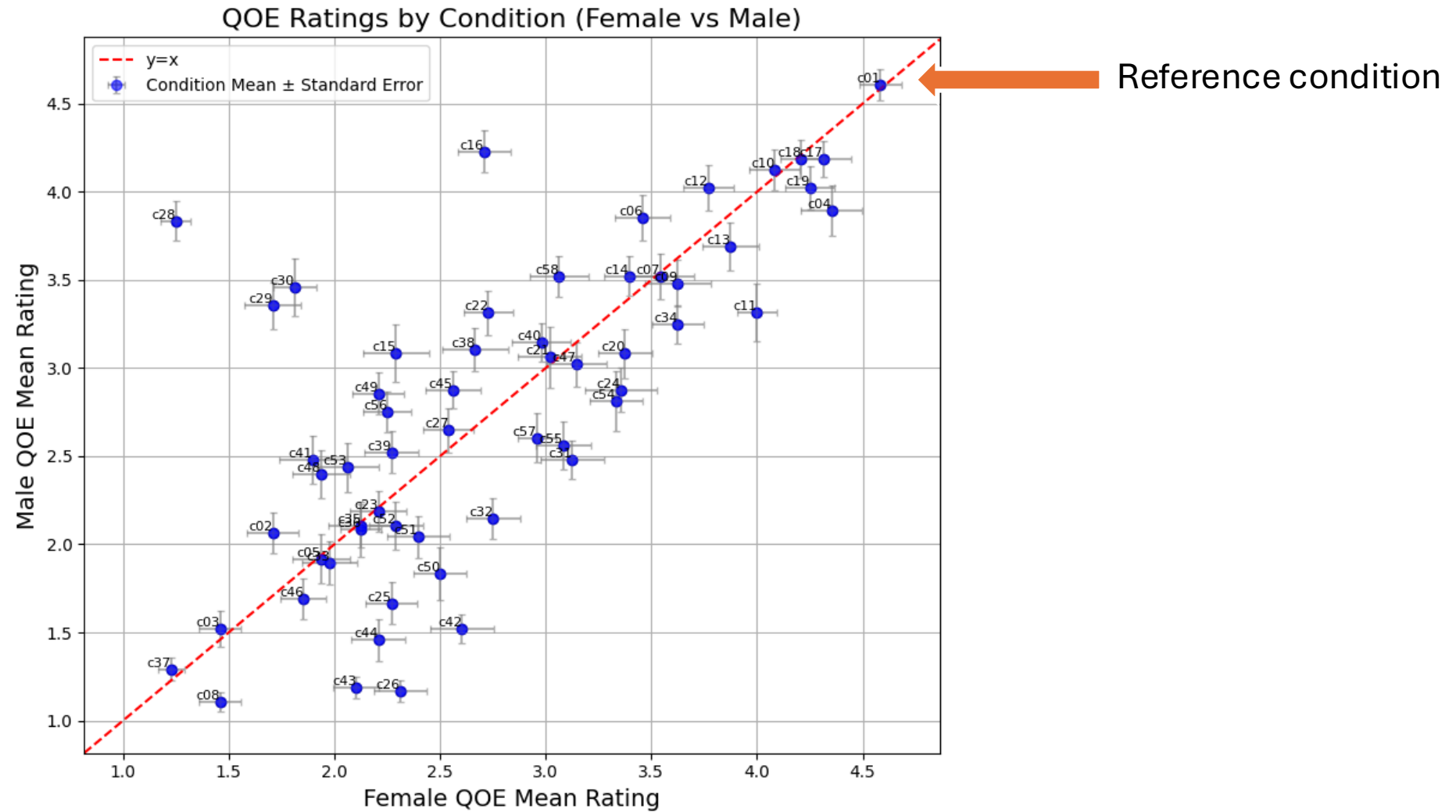


The NISQA_TEST_LIVETALK Dataset from: G. Mittag, B. Naderi, A. Chehadi, and S. Möller “NISQA -- A Deep CNN-Self-Attention Model for Multidimensional Speech Quality Prediction with Crowdsourced Datasets,” 2021.

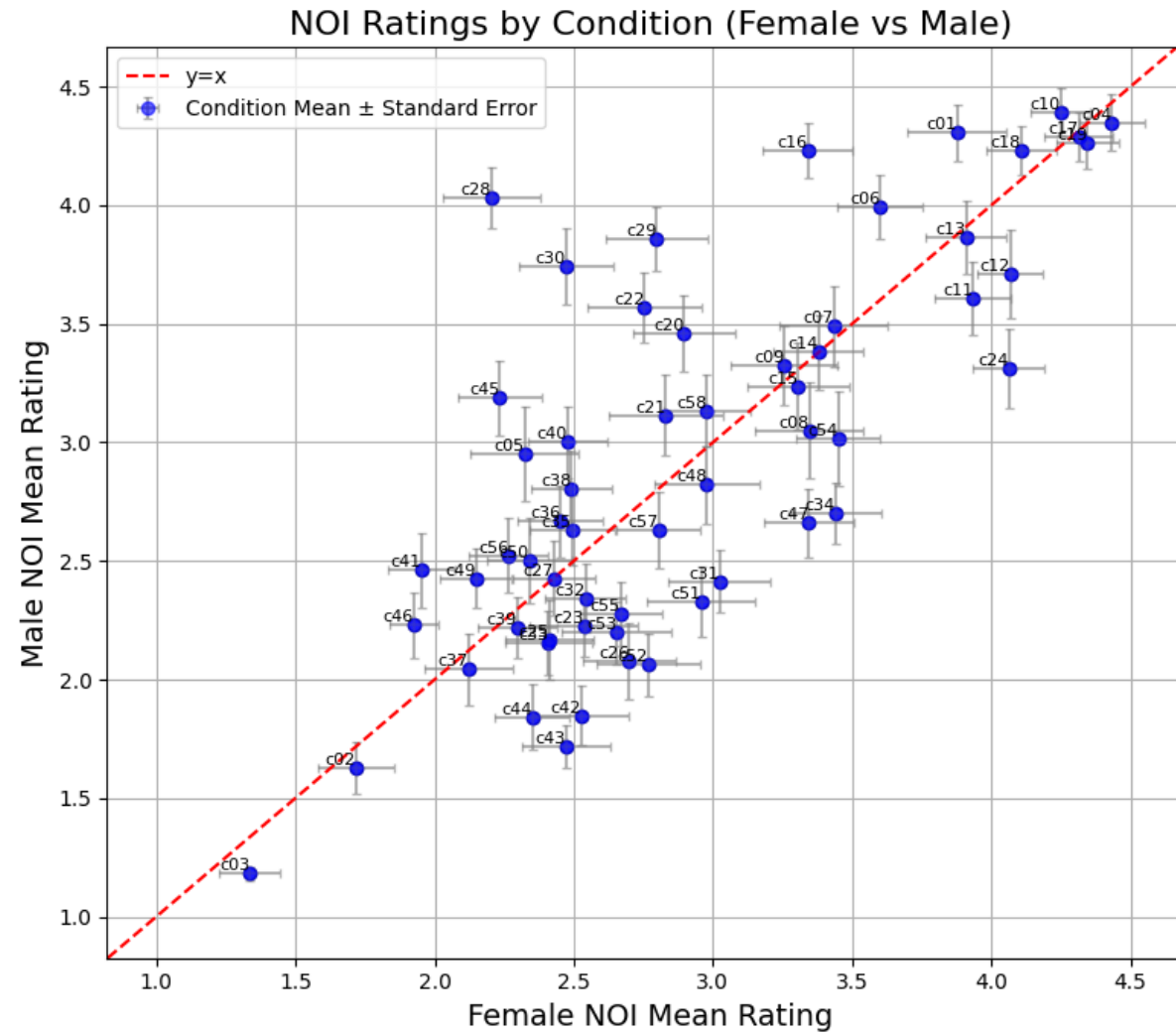
Overall Quality MOS



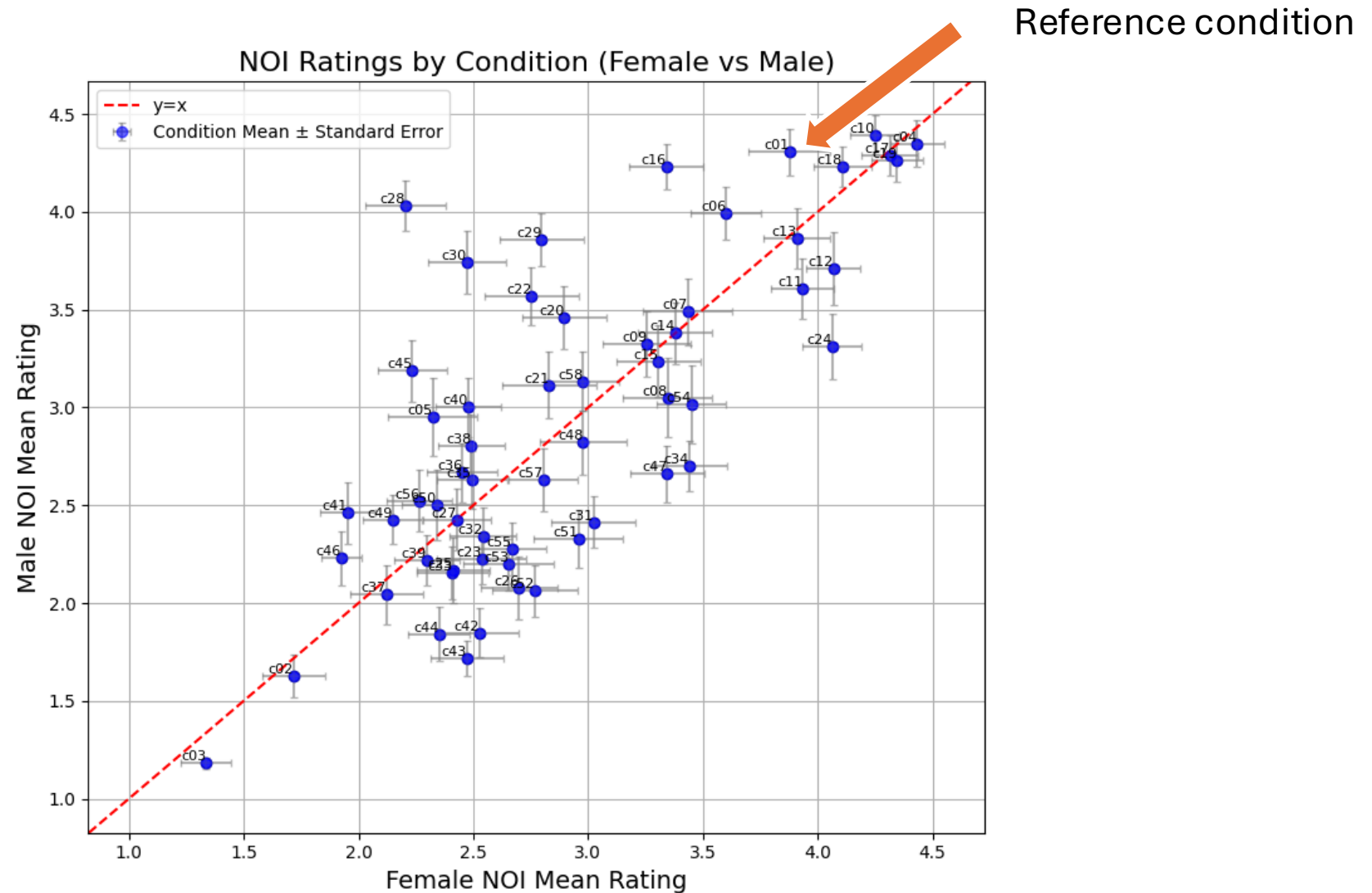
Overall Quality MOS



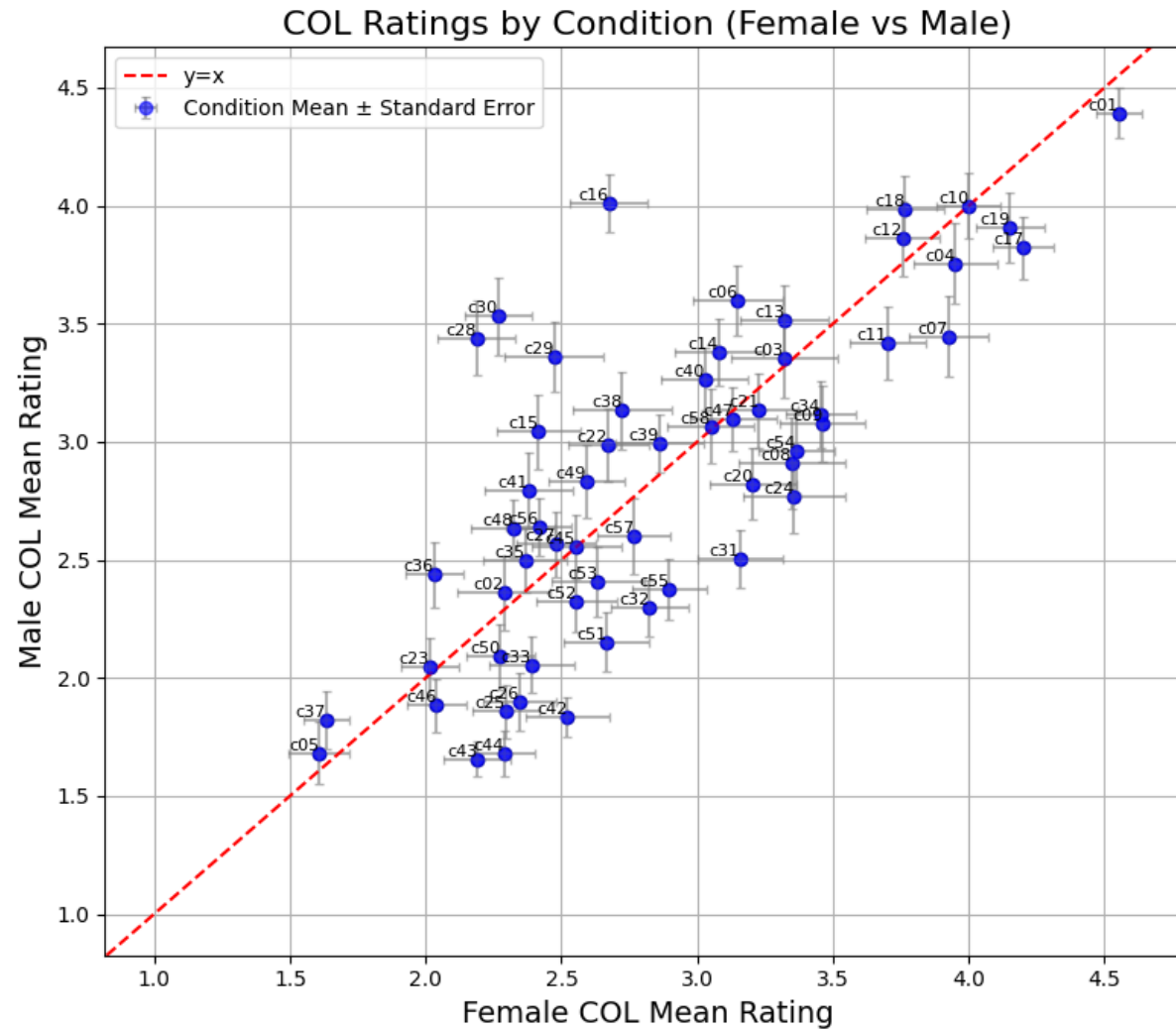
Noisiness



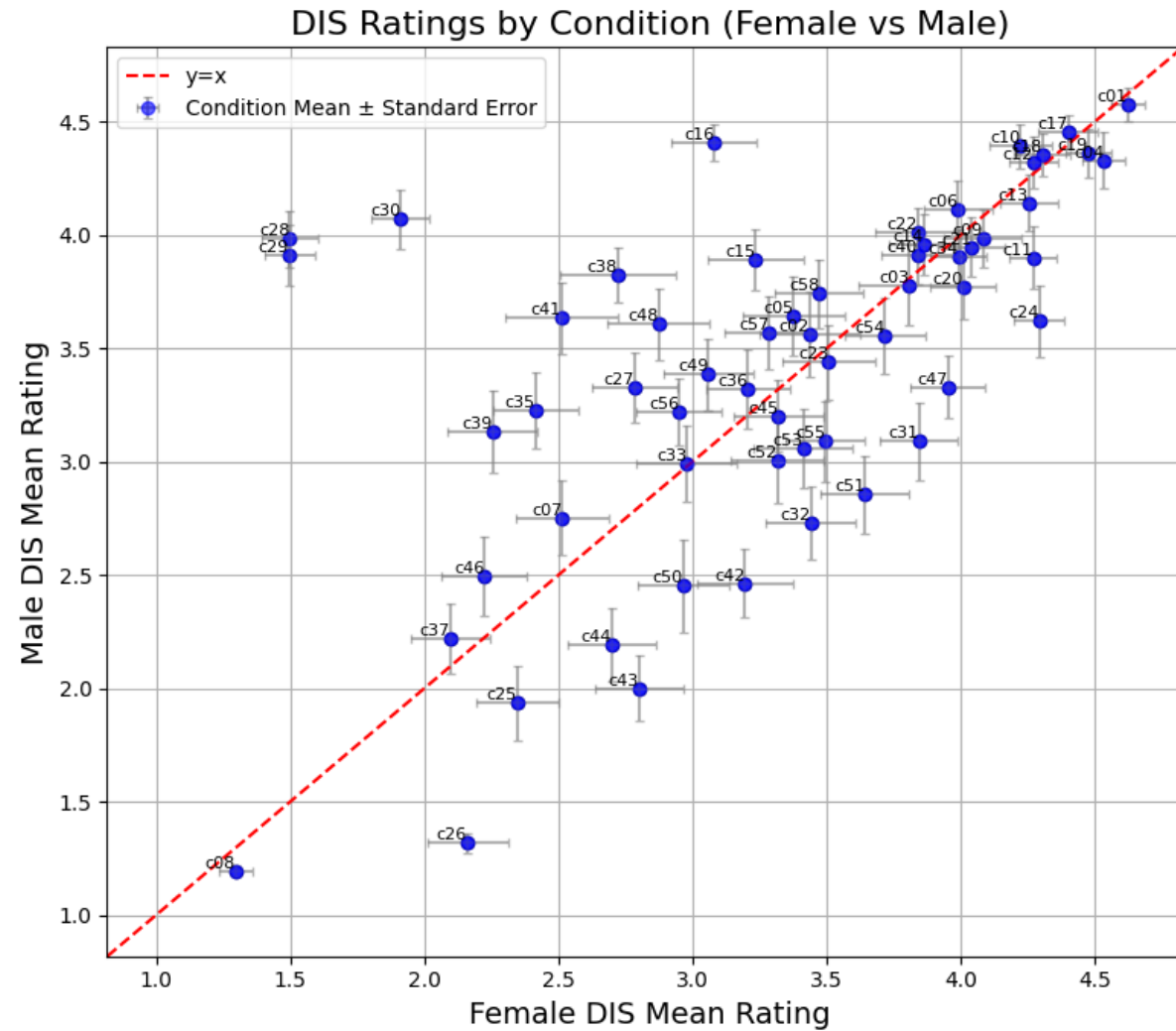
Noisiness



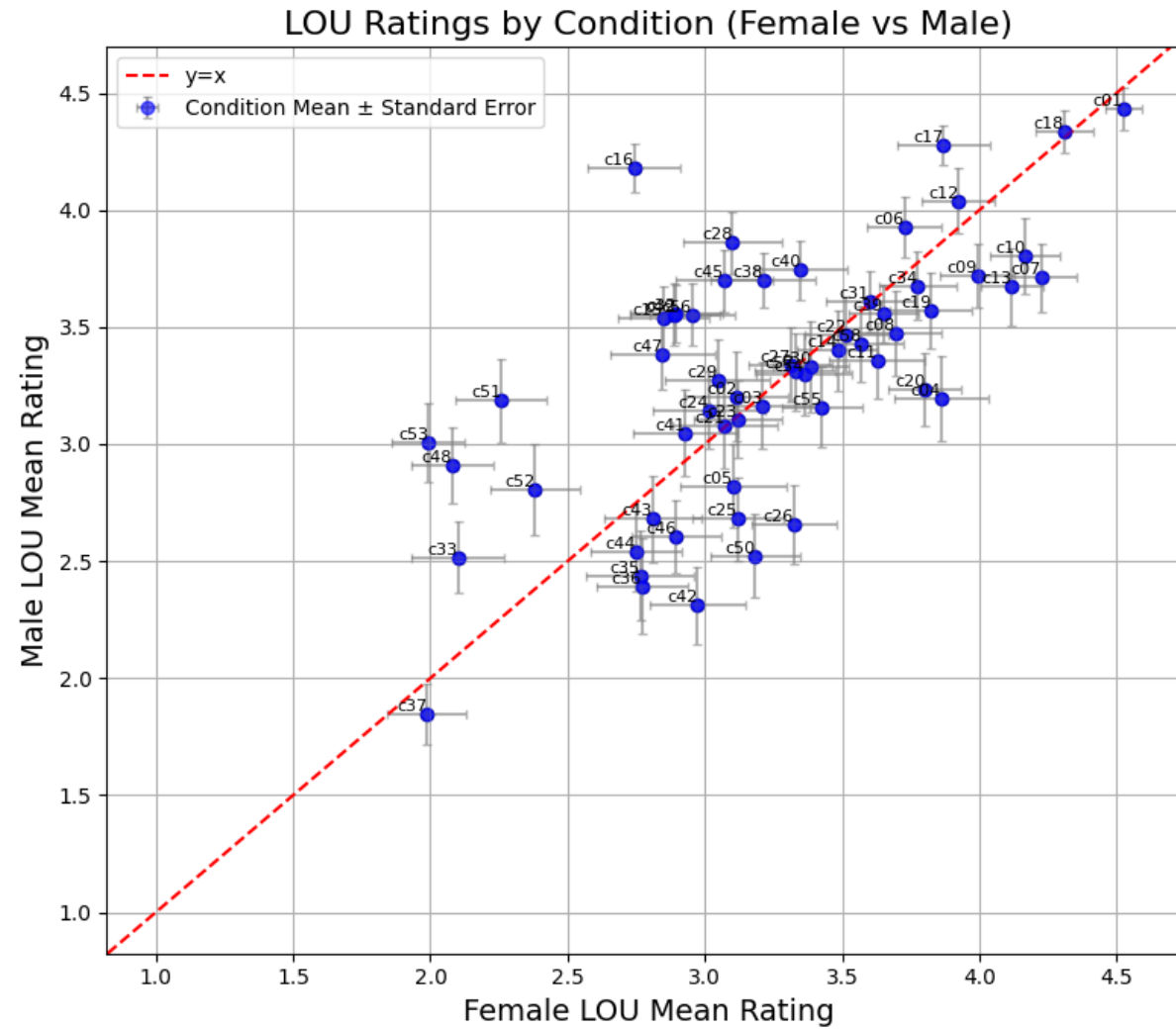
Coloration



Discontinuity



Loudness



Results: Reference Condition Gender Neutrality

Dimension	MOS_F _{REF}	STD_F _{REF}	MOS_M _{REF}	STD_M _{REF}	p-value
Overall Quality	4.583	0.679	4.604	0.611	0.875
NOI	3.876	1.218	4.307	0.821	0.048
COL	4.552	0.576	4.392	0.725	0.237
DIS	4.621	0.430	4.573	0.501	0.617
LOU	4.524	0.449	4.432	0.608	0.400

Results: Balance Across All Conditions

Dimension	Mean (M-F)	STD	p-value	Conclusion
Overall Quality	-0.018	0.673	0.872	Gender-balanced
NOI	0.414	0.538	< 0.001	Gender-biased (female worse)
COL	-0.150	0.467	0.103	Gender-balanced
DIS	-0.172	0.722	0.132	Gender-balanced
LOU	-0.147	0.468	0.111	Gender-balanced

Significant Advantageous Conditions

Female speaker advantage:

inside building
metro-station
car
VoIP(Skype/Facebook)

Male speaker advantage:

shopping-centre
mobile-network
café
elevator

Note: There were no obvious and clear patterns identified.

Second study

	Overall Quality		Noisiness		Coloration		Discontinuity		Loudness	
	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2
Mean	-0.018	-0.062	0.414	0.093	-0.150	-0.116	-0.172	-0.108	-0.147	-0.065
STD	0.673	0.307	0.538	0.217	0.467	0.215	0.722	0.269	0.468	0.218
N	57	59	57	59	57	59	57	59	57	59
P-value	0.443	0.127	0.000	0.002	0.05	0.000	0.241	0.003	0.069	0.025
Conclusion	Balanced	Balanced	Misbalanced	Misbalanced	Balanced	Misbalanced	Balanced	Misbalanced	Balanced	Misbalanced

Conclusions

- Perceptual dimensions exhibit differing speaker gender related effects compared to overall quality MOS.
- A more systematic approach is required for identifying persistent patterns.
- Subjective rating of perceptual dimensions can vary even for reference condition.





Future Outlook Can Explore

Whether **listeners of different genders** perceive degradation differently – using this same methodology.

How **specific distortions interact with vocal characteristics** across genders.

Optimizing **reference material selection** to avoid baseline bias.



Thank you.