Tutorial on Understanding and Mitigating Bias in Emotion Recognition Systems

Dr. Woan-Shiuan Chien (Winnie)
National Tsing Hua University
wschien@gapp.nthu.edu.tw

Prof. Chi-Chun Lee (Jeremy)
National Tsing Hua University
cclee@ee.nthu.edu.tw

biic 人 本 訊 號 運 算 研 究 室
Behavioral Informatics and Interaction Computation Lab

國立清華大學
NATIONAL TSING HUA UNIVERSITY

Woan-Shiuan Chien (Winnie)

Postdoctoral Researcher

NATIONAL TSING HUA UNIVERSITY

Department of Electrical Engineering, NTHU, Taiwan

Professional Interests

Multimodal Signal Processing · Speech · Physiology
Affective Computing · Trustworthy AI

Education

National Tsing Hua University, Taiwan

PH.D IN DEPARTMENT OF ELECTRICAL ENGINEERING, 2025/03
Advisor: Chi-Chun Lee (Jeremy)
Dissertation: From Data Resource Impacts to Fairness Realization in Speech Emotion Recognition

Working Experiences

AIST, AIRC, Japan

AI STUDENT INTERN @
INTELLIGENT MEDIA PROCESSING RESEARCH TEAM
2024/01-2024/03

Honors and Awards

AWARD

NTHU Outstanding Postdoctoral Research Fellow (2025)
The Rising Stars Women in Engineering Workshop – Shortlisted Participants (2025)
Merry Electronics Co., Ltd.: Electroacoustics Thesis Award Finalist, Taiwan (2024)

SCHOLARSHIP

NSTC Outstanding Doctoral Students Fellowship, NSTC, Taiwan (2022-2023)
Elite-Well Doctoral Scholarship, Elite-Well Education Foundation, Taiwan (2025)
NTHU International Visiting Scholarship, National Tsing Hua University, Taiwan (2024, 2023)
Google Conference Scholarships (APAC), Google (2024, 2023)

TRAVEL GRANT

IEEE BSN Travel Awards, IEEE Engineering in Medicine and Biology Society (EMBS) (2024)
ACII 2023 Travel Bursary, AAAC (2023)
ICASSP 2023 Conference Travel Grant, IEEE Signal Processing Society (SPS) (2023)
PROGRESS Student Travel Awards, IEEE PROmotinG DiveRsity in Signal ProcESSing (2023)
ACLCLP Outstanding Students Conference Travel Grant (2024, 2023)

CHI-CHUN LEE (Jeremy)

Ph.D.
Electrical Engineering

University of Southern California (USA)

Professor / Associate Chair

Department of Electrical Engineering, NTHU, Taiwan

Joint Appointment

NTHU

Institute of Communications Engineering

College of Semiconductor Research

Biomed AI Ph.D. Program

International Intercollegiate Ph.D. Program

Precision Medicine Ph.D. Program

ACADEMIA SINICA

Center for Information Technology Innovation (Research Fellow)

Associate Editor

- IEEE Transactions on Audio, Speech and Language Processing (2025–)
- Journal of Computer Speech and Language (2021–)

Awards & Honors

- Novatek Distinguished Talent Chair -NTHU (2025)
- Outstanding Research Award -NSTC (2024)
- Young Innovator Award -FAOS (2020)
- Outstanding Young Electrical Engineer Award -CIEE (2020)
- K.T. Li Cornerstone Award -ICM (2024)
- K.T. Li Young Researcher Award -ICM (2021)
- Tsing Hua Talent Development Fund Outstanding Research Award -NTHU (2024)
- Outstanding Industry University Research Award -NTHU (2023)
- Industry Collaboration Excellence Award -NTHU (2023) (2021)

Tutorial Outline

- **9:15-9:30 - Setting the Stage: Why Fairness Matters in Affective Computing**
  - A human-centered perspective on fairness, bias, and ethical challenges in emotion AI systems.
- **9:30-9:45 - Sources of Bias & Case Study: Speech Emotion Recognition**
  - Where does bias come from? Annotation subjectivity, demographic gaps
  - Why is SER particularly sensitive to fairness issues? Speaker- and rater-side analysis, dataset evidence
- **9:45-10:30 - Break**
- **10:30-11:00 - From Data to Evaluation: Strategies for Fair Affective Systems**
  - Fairness-aware Data Practices: Inclusive annotation, dataset auditing, labeling diversity
  - Bias Mitigation Methods: Pre-, in-, and post-processing strategies
  - Evaluation Frameworks: Group vs. individual fairness, metrics and trade-offs
- **11:00-11:40 - Societal Implications, Open Problems and Bias Analysis in BIIC-Podcast**
  - Cross-cultural affect, affective feedback, trust in emotion AI
  - BIIC-Podcast: An intelligent infrastructure toward large scale naturalistic affective speech corpora collection

Outline

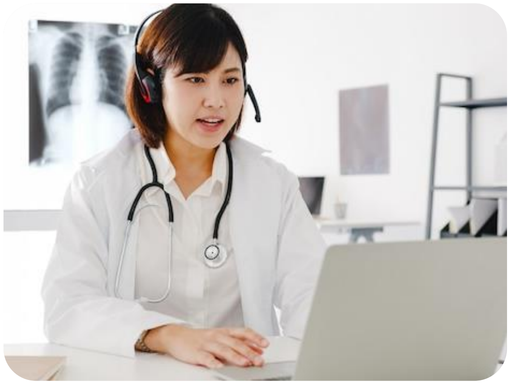
- Introduction
  - Why Fairness Matters in Affective Computing
  - Motivation of Bias in Emotion Recognition Systems
  - Relationship with AI Ethics
- Sources of Bias & Case Study: Speech Emotion Recognition
  - Biases and Fairness in Machine Learning
  - Where does bias come from? Annotation subjectivity, demographic gaps
  - Why is SER particularly sensitive to fairness issues? Speaker- and rater-side analysis, dataset evidence
- From Data to Evaluation: Strategies for Fair Affective Systems
  - Fairness-aware Data Practices: Inclusive annotation, dataset auditing, labeling diversity
  - Bias Mitigation Methods: Pre-, in-, and post-processing strategies
  - Evaluation Frameworks: Group vs. individual fairness, metrics and trade-offs
- Societal Implications, Open Problems and Bias Analysis in BIIC-Podcast

Learning Objective

- Recognize the sources and impacts of bias in emotion recognition systems
- Understand fairness concepts and their adaptation to affective computing
- Examine case studies of Speech Emotion Recognition to ground fairness issues
- Learn taxonomies of bias (speaker-side, rater-side, group vs. individual)
- Explore datasets, metrics, and protocols to evaluate and mitigate bias

Affective Computing are Everywhere

Healthcare Systems



Emotion-Aware Mental Health Monitoring

Automotive Systems

In-Car Emotion Recognition

Education Systems

Affective Tutoring and Feedback

Customer Service

Emotion-Aware Call Centers

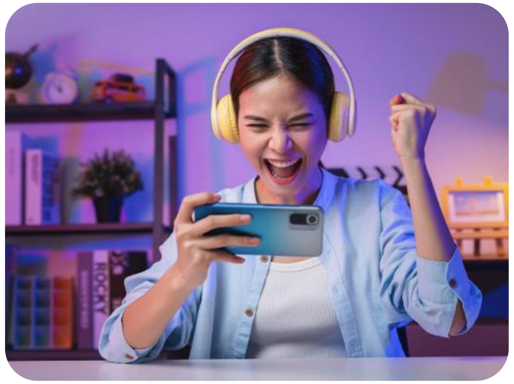
Social Media & Communication

Emotion Analytics for Online Interaction

Human–Robot Interaction

Emotionally Adaptive Robots

Entertainment & Gaming



Emotion-Responsive Games and Media

Virtual Assistants

Emotionally Intelligent Voice Agents

Social Impacts of Affective Systems

- Affective Systems are far more than just emotion recognition tools
  - They shape how emotions are interpreted, responded to, and valued in society
    - Emotional responses influence decisions, behaviors, and well-being
    - Affective AI mediates social relationships between humans and machines
- The Human–AI–Human Paradigm:
  - Users – Emotions Systems – Society
    Students – Emotions – Tutors
    Patients – Emotions – Clinicians
    Drivers – Emotions – Vehicles
    Customers – Emotions – Service Agents
    Citizens – Emotions – Social Media

Affective systems not only sense emotions — they also influence emotional norms, trust, and social fairness, creating feedback loops that reshape human–AI–human interaction.

Why Fairness Matters in Affective Computing

- Most affective systems are trained on some training data
  - Training data may encode social bias
  - Annotation labels may reflect subjective judgments or cultural bias
  - Model may echo or even reinforce the bias in training emotion-labeled human data

Fairness in affective computing is not just a technical concern — it determines whose emotions are correctly understood and whose are misinterpreted.

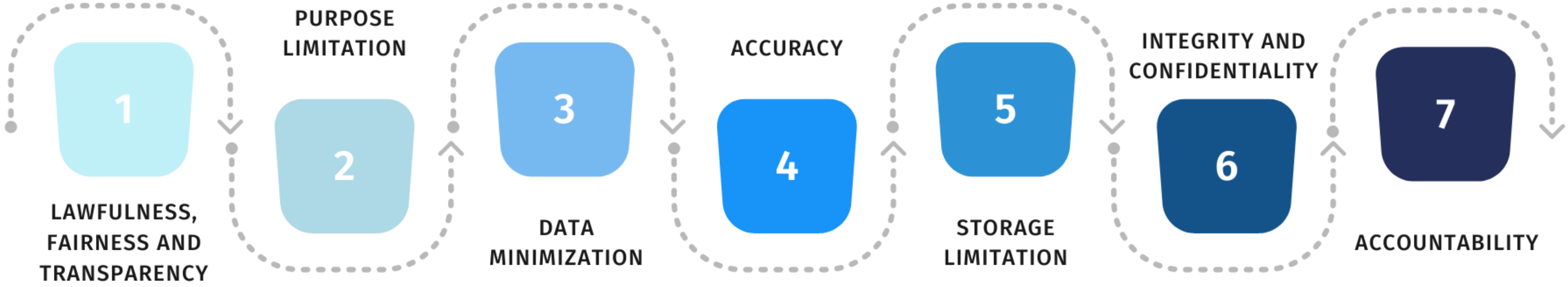
Potential Consequences of Unfairness in Affective Systems

Gender Bias	Exacerbation of Social Injustice	Risks in Mental Health Monitoring	Declining Trust in Technology
			
Emotion recognition systems may associate certain emotions with specific genders (e.g., women perceived as "sad" or "emotional," men as "angry" or "neutral"). Such bias perpetuates gender stereotypes and unequal treatment.	*When emotion AI is used in hiring, education, or law enforcement, biased affect interpretations can unfairly penalize marginalized groups and amplify existing inequities.*	*Emotion recognition errors can lead to misdiagnosis or overgeneralization, especially in stress or depression detection. This raises ethical and privacy concerns for individuals being continuously monitored.*	*Unfair or inconsistent emotion judgments can reduce user trust, making people feel misunderstood, surveilled, or discriminated against by AI systems.*

Unfair affective systems not only misinterpret emotions — they reshape how people are perceived, evaluated, and treated in society.

Fairness in Affective Systems: an AI Ethics Perspective

- Affective Systems as responsible AI
  - Should ensure fair and respectful interpretation of human emotions
  - Provide equitable emotional decisions for all users, regardless of gender, culture, or accent



7 Principles of EU GDPR Regulation

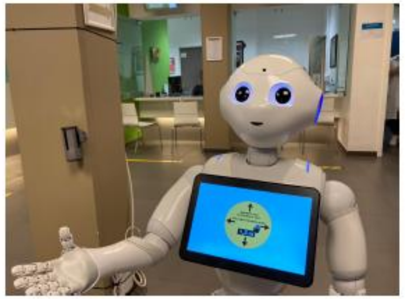
- Fairness often appears together with other responsible AI perspectives
  - e.g., transparency / explainability (honesty) of algorithmic decisions is the foundation of fairness

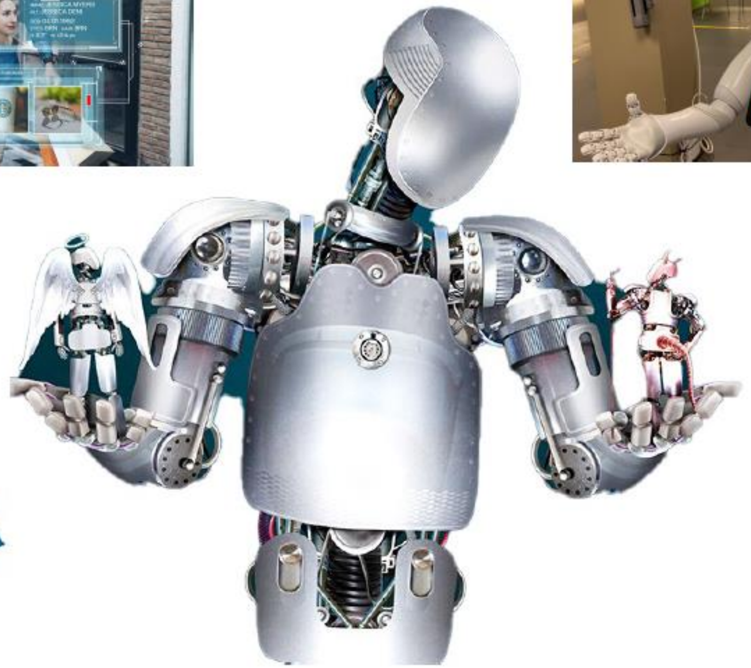
Outline

- Introduction
  - Why Fairness Matters in Affective Computing
  - Motivation of Bias in Emotion Recognition Systems
  - Relationship with AI Ethics
- Sources of Bias & Case Study: Speech Emotion Recognition
  - Biases and Fairness in Machine Learning
  - Where does bias come from? Annotation subjectivity, demographic gaps
  - Why is SER particularly sensitive to fairness issues? Speaker- and rater-side analysis, dataset evidence
- From Data to Evaluation: Strategies for Fair Affective Systems
  - Fairness-aware Data Practices: Inclusive annotation, dataset auditing, labeling diversity
  - Bias Mitigation Methods: Pre-, in-, and post-processing strategies
  - Evaluation Frameworks: Group vs. individual fairness, metrics and trade-offs
- Societal Implications, Open Problems and Bias Analysis in BIIC-Podcast

Biases and Fairness in Machine Learning – Motivations

- Fairness matters because it has impact on everyone's benefit.





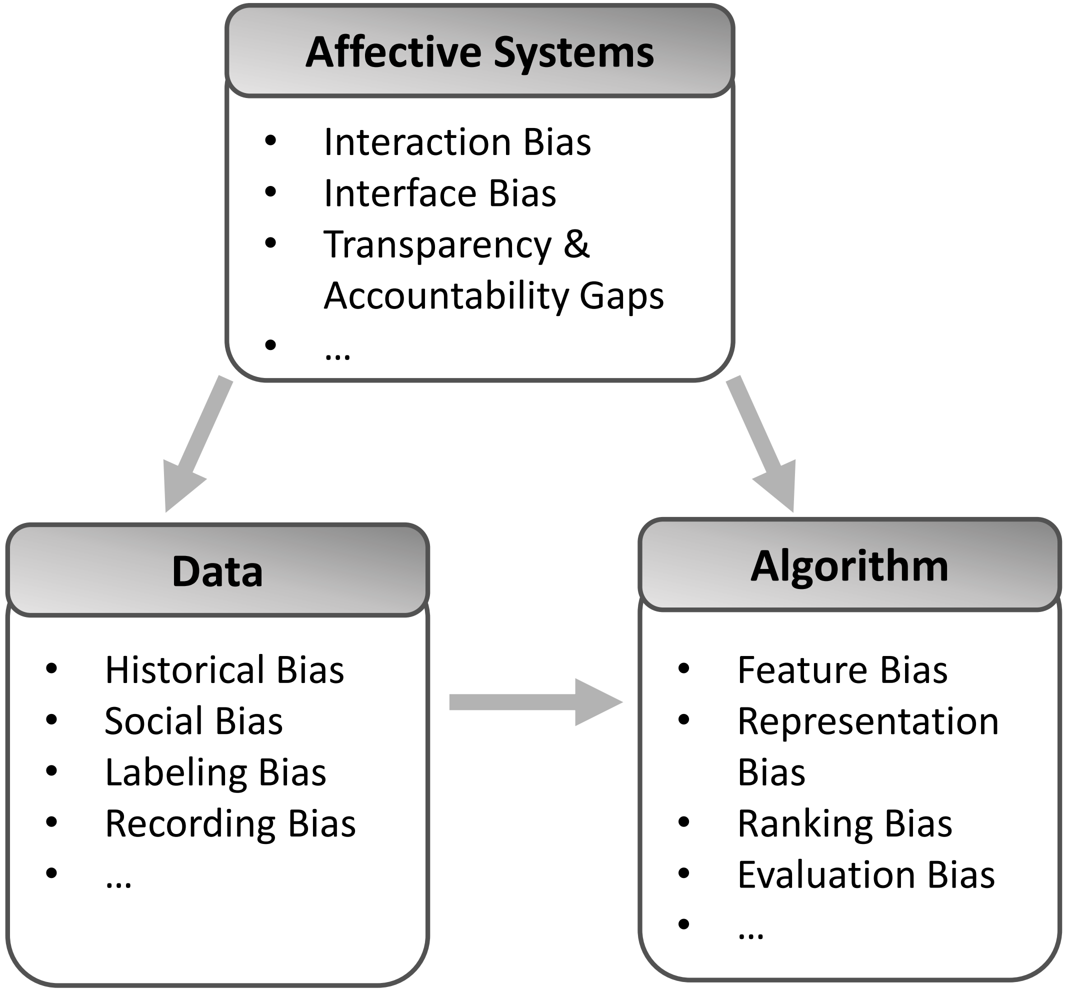
Biases and Fairness in Machine Learning – Causes

Data Bias

- Statistical Bias: non-random sample; record error
- Historical Bias: biased decision
- …

Algorithmic Bias

- Ranking Bias: exposure allocation
- Evaluation Bias: inappropriate benchmarks
- …



Biases and Fairness in Machine Learning – Definitions

Group Fairness	Statistical parity
Individual Fairness	Consistency, Counterfactual Fairness
Subgroup Fairness	Fairness holds over a large collection of subgroups defined by a class of functions

Biases and Fairness in Machine Learning – Methods

Pre-processing

Try to transform the data so that the underlying discrimination is removed.

Transform or rebalance data before training

- **Re-sampling / Re-weighting** – balance demographic groups in training data
- **Data Augmentation** – synthesize underrepresented samples (e.g., gender or language)
- **Label Correction / De-bias Annotation** – reduce subjective or noisy emotional labels
- **Representation Learning (Fair PCA, Domain Adaptation)** – learn latent features independent of sensitive attributes

In-processing

Try to modify the learning algorithms to remove discrimination during the model training process.

Modify learning algorithms to enforce fairness during training

- **Adversarial Debiasing** – train model to predict emotion while disentangling sensitive factors
- **Fairness Regularization / Constraint** – add fairness terms (e.g., demographic parity loss, equalized odds)
- **Sample Weighting** – penalize errors on minority or sensitive groups
- **Multi-task or Domain-Invariant Learning** – jointly learn emotion + fairness objectives

Post-processing

Perform after training by accessing a holdout set which was not involved during the training of the model.

Adjust model outputs or decisions after training

- **Threshold Adjustment / Calibration** – tune decision boundaries per group to equalize outcomes
- **Re-ranking or Re-scoring** – reorder predictions for group balance
- **Confidence Reweighting** – lower confidence for uncertain or biased regions
- **Fairness Auditing & Explainability** – analyze disparities, interpret emotion model behaviors

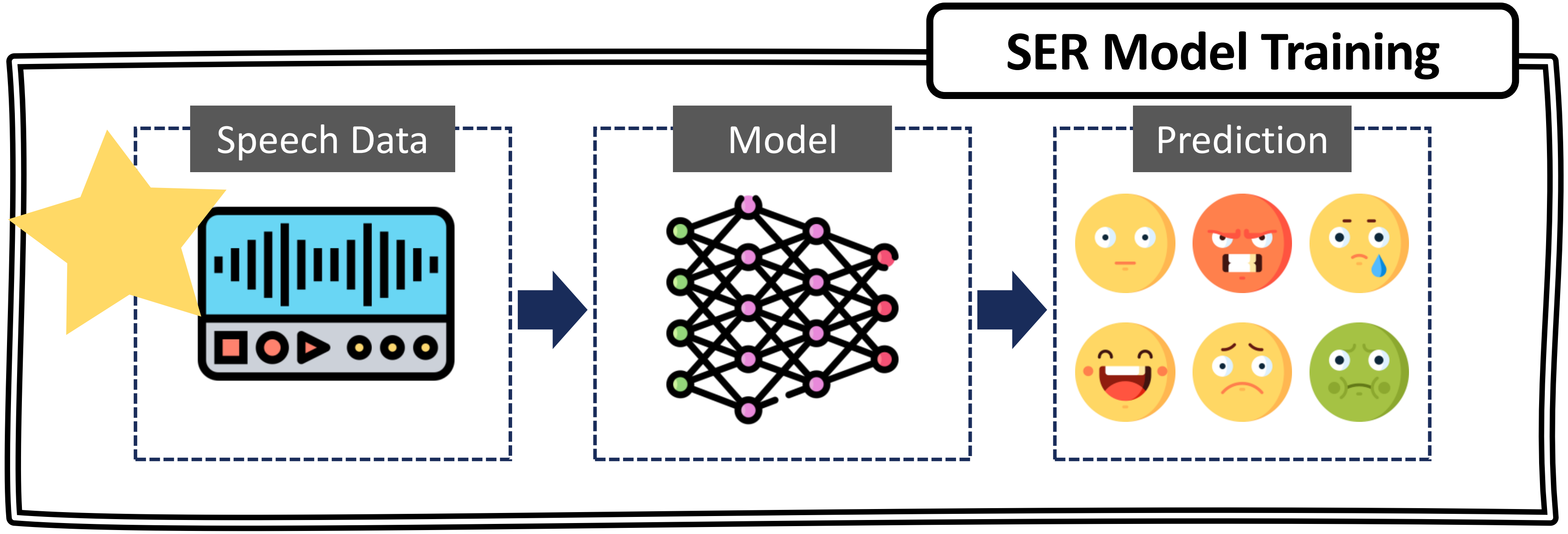
What Exactly Are the Sources of Bias in Emotion Recognition Systems?

- Case Study: ***Speech Emotion Recognition (SER)***

Causes	Definitions	Method
Labeling Bias Speaker Bias	Group Fairness Individual Fairness	In-Processing Debiasing

What Exactly Are the Sources of Bias in SER?

- How to train an SER system?

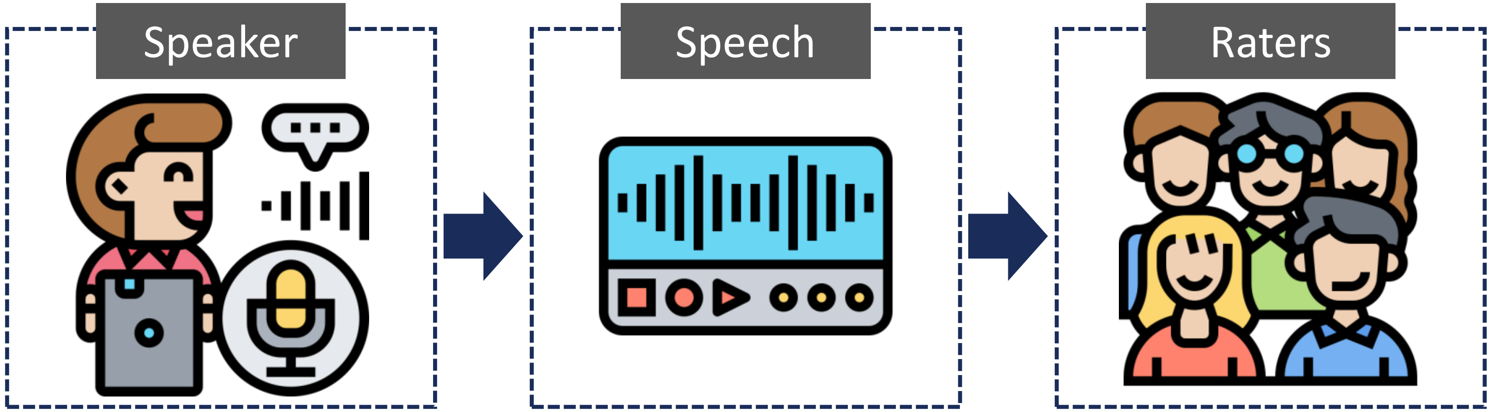


- Building the database is crucial, as many influential factors originate directly from the data.
- The algorithm learns from the data we provide, meaning its outcomes are shaped by the quality and characteristics of the dataset.

→ **How to construct an emotion database?**

What Exactly Are the Sources of Bias in SER?

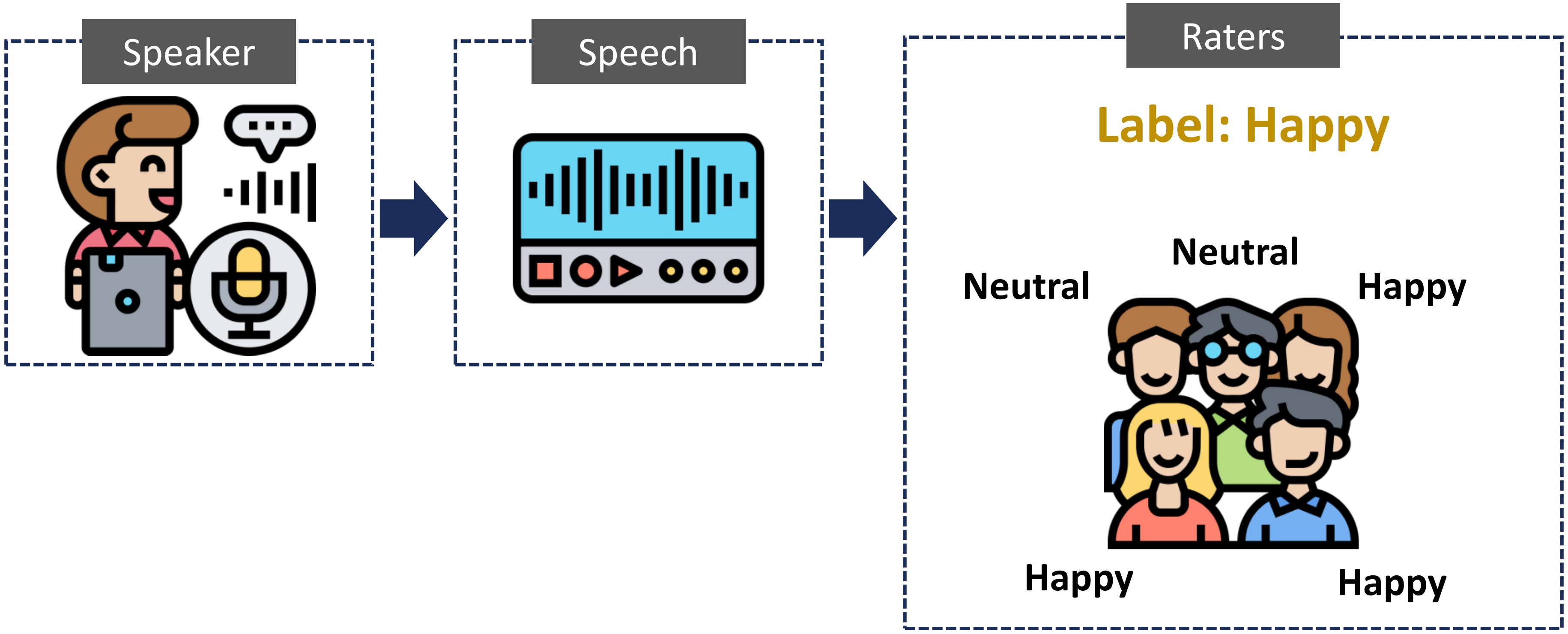
- How to construct an emotion database?



[9] S. G. Upadhyay, W.-S. Chien, and others, “An intelligent infrastructure toward large scale naturalistic affective speech corpora collection,” in 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1–8, IEEE, 2023.[7] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and classification models,” Digital signal processing, vol. 22, no. 6, pp. 1154–1160, 2012.

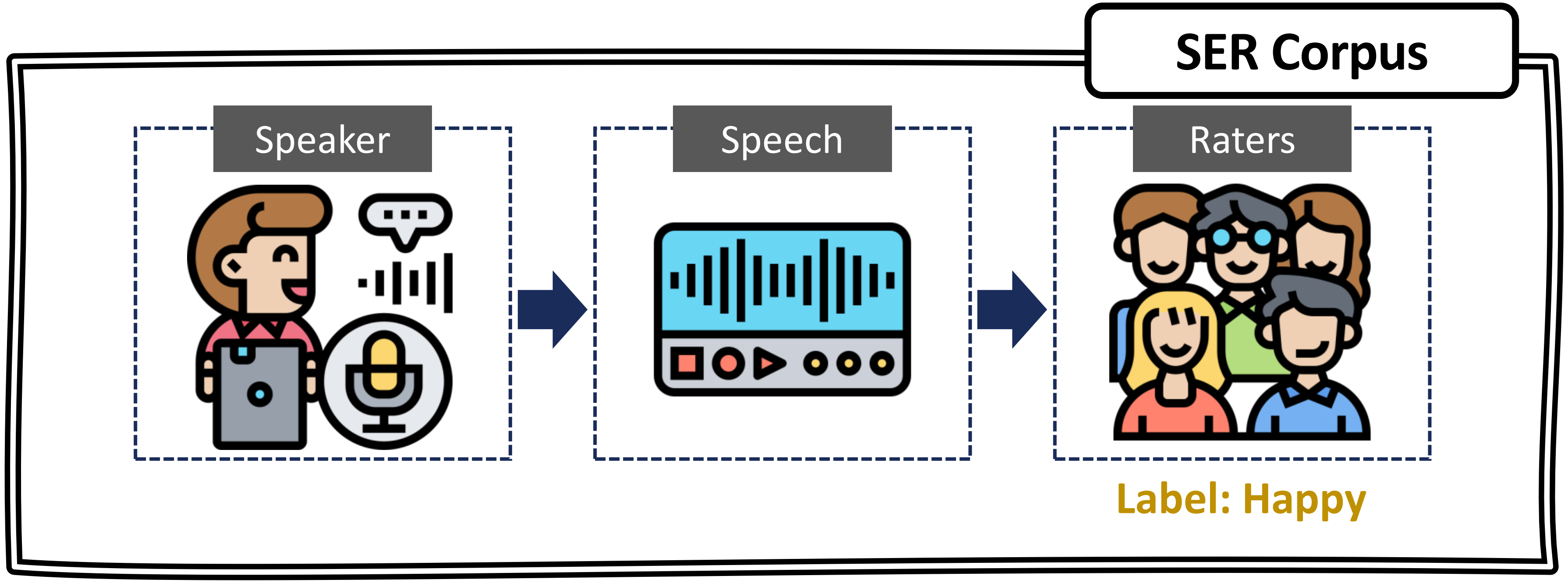
What Exactly Are the Sources of Bias in SER?

- Emotion Label is followed by the plurality voting.



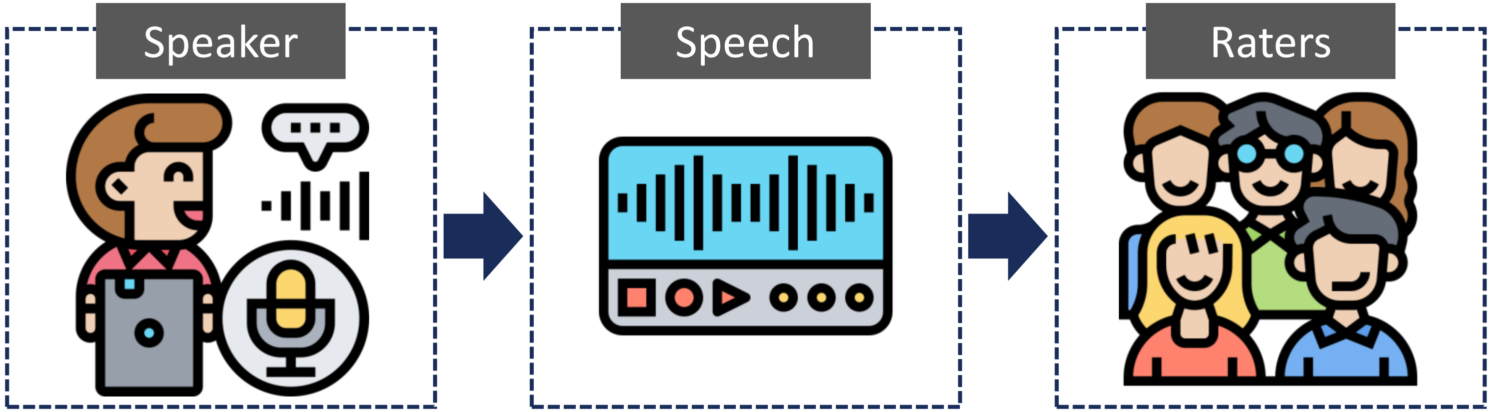
What Exactly Are the Sources of Bias in SER?

- Emotion Label is followed by the plurality voting.



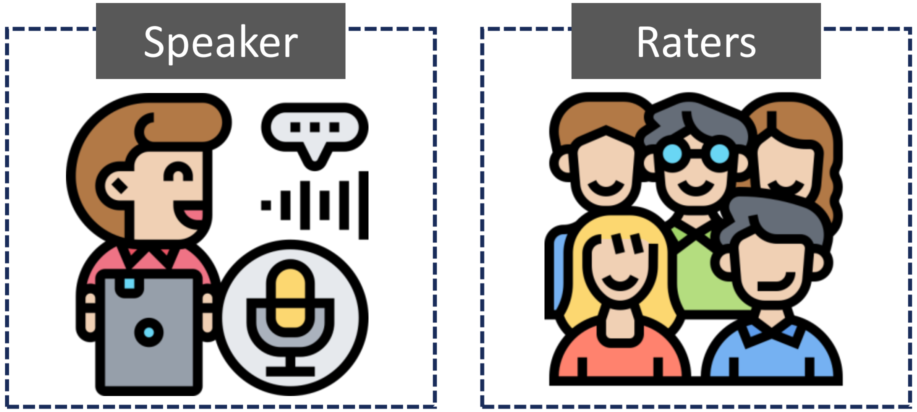
What Exactly Are the Sources of Bias in SER?

- Human *speakers* engaging in spoken dialogs
  with human *raters* providing ground truth labels



What Exactly Are the Sources of Bias in SER?

Acknowledgment of Human Diversity



→ **Induce Bias and Fairness Issue**

→ **Especially from Gender-wise Bias**

Demographic Factors / Individual Differences / Subjectivity

What Exactly Are the Sources of Bias in SER?

Gender Factors

Speaker

Raters

- Speakers differ in acoustic cues by gender [10]
  - Female voices exhibit higher f0 values and less intensity compared to male voices

BIIC-PODCAST_0043_0100.wav

BIIC-PODCAST_0036_0090.wav

- Raters' emotional perception varies by gender [11]
  - Females sometimes report higher sensitivity to emotional cues and may judge certain emotions (e.g., sadness or fear) more intensely than males

worker_0006	worker_0008	worker_0010	worker_0033	worker_0034
Neutral	Happy	Neutral	Happy	Happy

➔Consensus Label: Happy

[10] A. Groyecka-Bernard and others, "Do voice-based judgments of socially relevant speaker traits differ across speech types?," Journal of Speech, Language, and Hearing Research, vol. 65, no. 10, pp. 3674–3694, 2022.
[11] M. Swerts and E. Krahmer, "Gender-related differences in the production and perception of emotion," in Ninth Annual Conference of the International Speech Communication Association, 2008.

What Exactly Are the Sources of Bias in SER?

Gender Factors

- Rater-gender biases affect the consensus labels

Raters

Label: Happy

Neutral
Neutral
Happy

Happy
Happy

Male Raters

Male Label: Neutral

Neutral
Neutral
Happy

Happy
Happy

Female Raters

Female Label: Happy

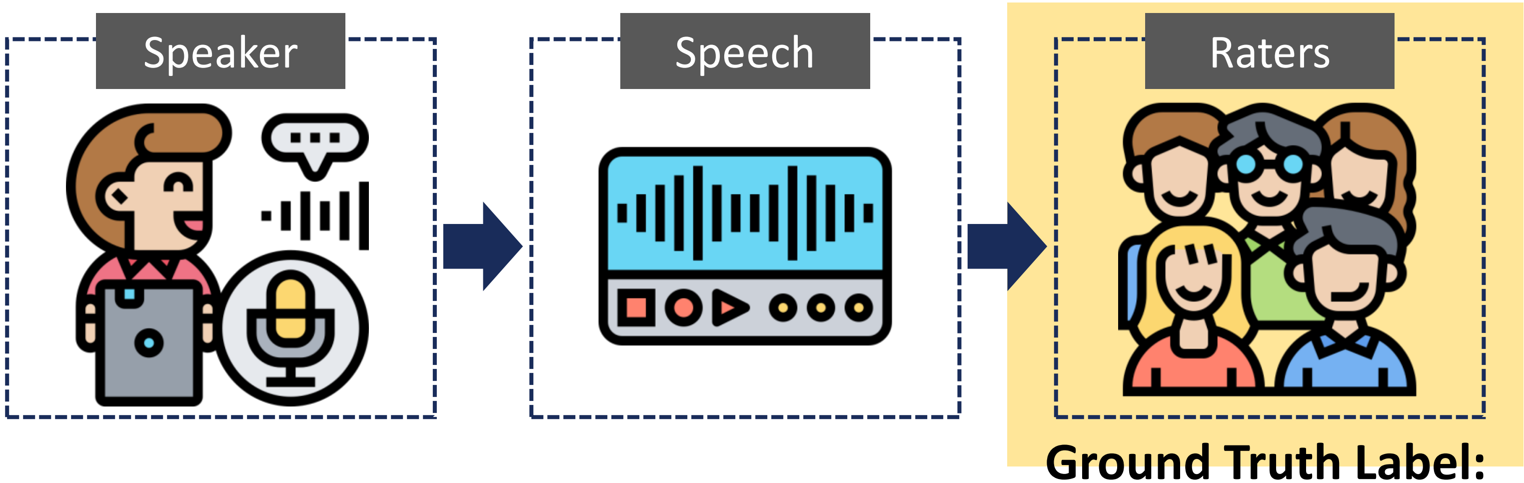
Neutral
Neutral
Happy

Happy
Happy

What Exactly Are the Sources of Bias in SER?

Rater-Gender Biases

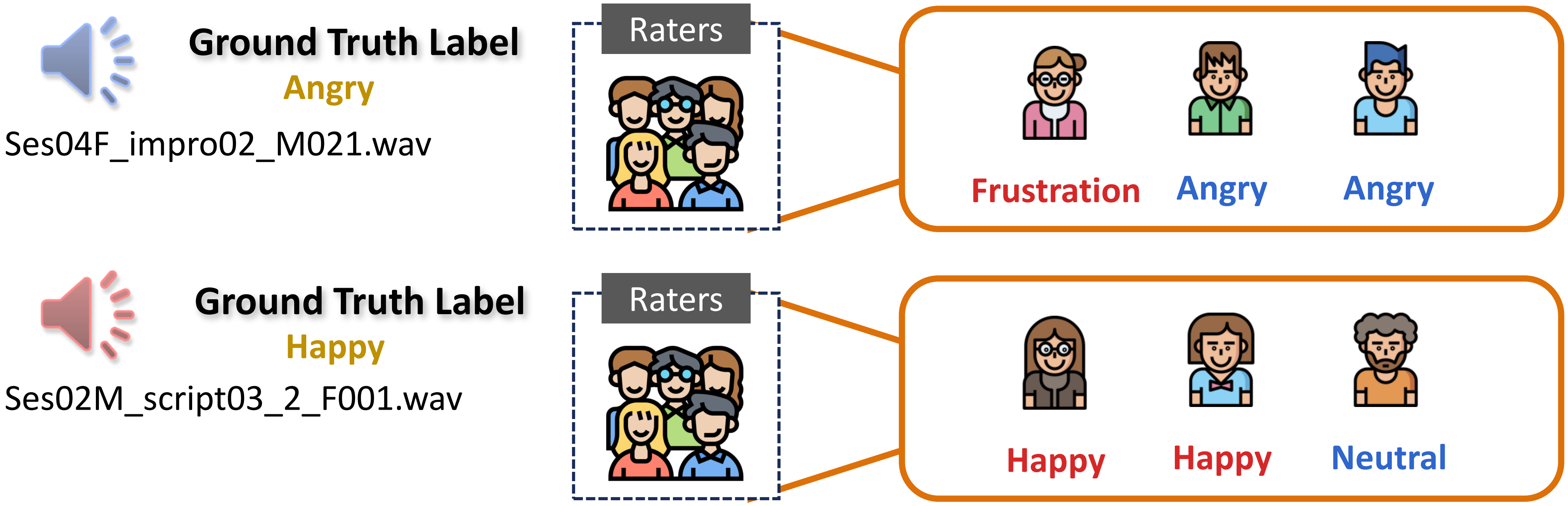
- One of the unique fairness issues in SER is caused by the inherently biased emotion perception given by the raters as ground truth labels. → Mitigating rater-gender biases



Ground Truth Label: Happy

Biases and Fairness in SER – Motivation

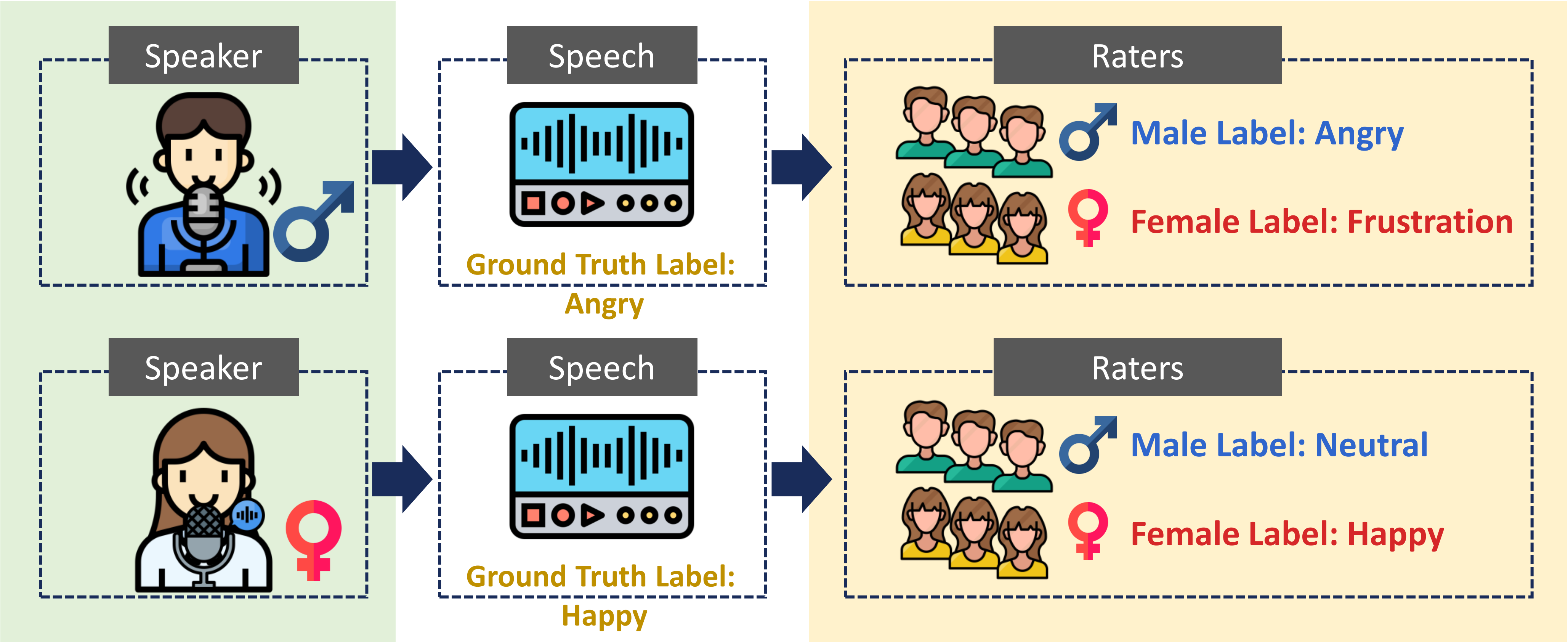
- Examples from IEMOCAP database



Biases and Fairness in SER – Background

Speaker-side and Rater-side

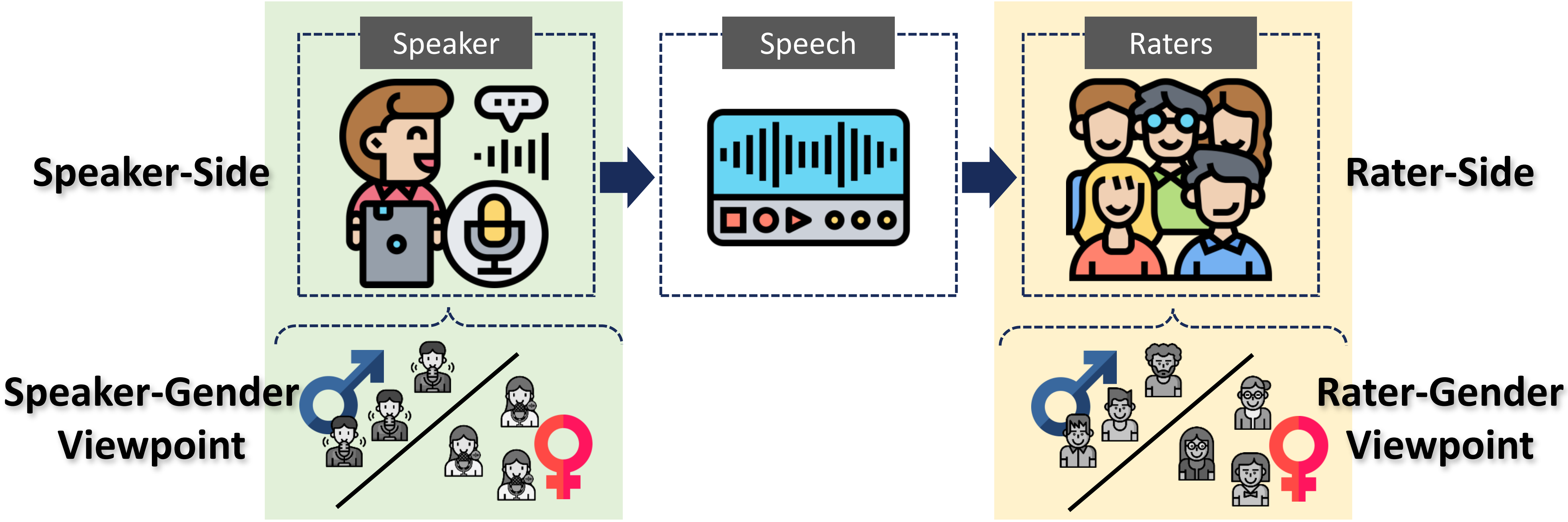
- A typical SER model is constructed by learning on datasets comprised of human *speakers* engaging in spoken dialogs with human *raters* providing ground truth labels. → Compound biases



Biases and Fairness in SER – Background

Speaker-side and Rater-side

- Ensure gender viewpoint fairness
- Learn gender-debiasing representation for either speaker-side or rater-side



Tutorial Outline

9:15-9:30 - Setting the Stage: Why Fairness Matters in Affective Computing

A human-centered perspective on fairness, bias, and ethical challenges in emotion AI systems.

9:30-9:45 - Sources of Bias & Case Study: Speech Emotion Recognition

Where does bias come from? Annotation subjectivity, demographic gaps

Why is SER particularly sensitive to fairness issues? Speaker- and rater-side analysis, dataset evidence

9:45-10:30 - Break

10:30-11:00 - From Data to Evaluation: Strategies for Fair Affective Systems

Fairness-aware Data Practices: Inclusive annotation, dataset auditing, labeling diversity

Bias Mitigation Methods: Pre-, in-, and post-processing strategies

Evaluation Frameworks: Group vs. individual fairness, metrics and trade-offs

11:00-11:40 - Societal Implications, Open Problems and Bias Analysis in BIIC-Podcast

Cross-cultural affect, affective feedback, trust in emotion AI

BIIC-Podcast: An intelligent infrastructure toward large scale naturalistic affective speech corpora collection

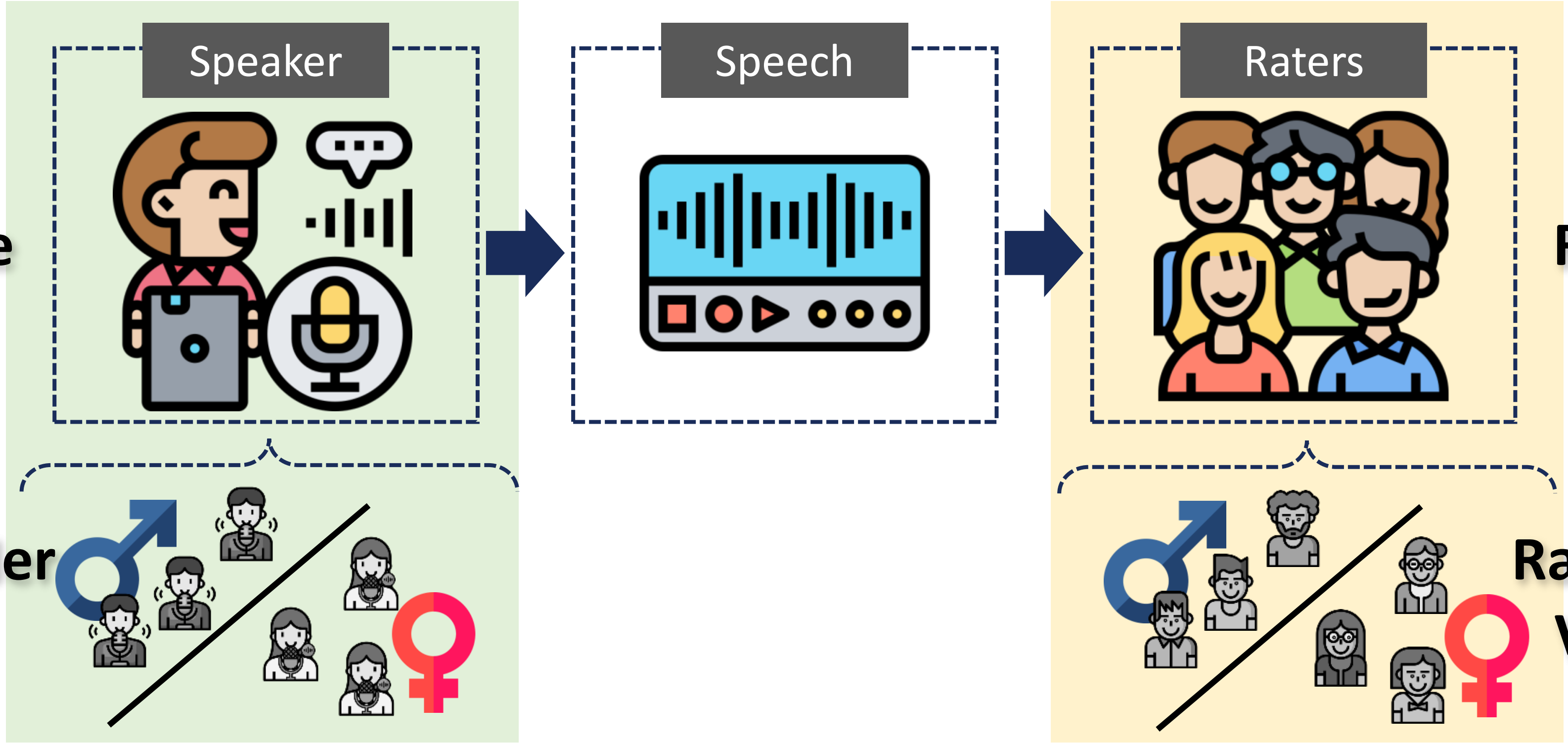
Outline

- Introduction
  - Why Fairness Matters in Affective Computing
  - Motivation of Bias in Emotion Recognition Systems
  - Relationship with AI Ethics
- Sources of Bias & Case Study: Speech Emotion Recognition
  - Biases and Fairness in Machine Learning
  - Where does bias come from? Annotation subjectivity, demographic gaps
  - Why is SER particularly sensitive to fairness issues? Speaker- and rater-side analysis, dataset evidence
- From Data to Evaluation: Strategies for Fair Affective Systems
  - Fairness-aware Data Practices: Inclusive annotation, dataset auditing, labeling diversity
  - Bias Mitigation Methods: Pre-, in-, and post-processing strategies
  - Evaluation Frameworks: Group vs. individual fairness, metrics and trade-offs
- Societal Implications, Open Problems and Bias Analysis in BIIC-Podcast

Fairness-aware Data Practices

Speaker-Rater Data

Speaker-Side



Rater-Side

Speaker-Gender Viewpoint

Rater-Gender Viewpoint

Guiding Question!!

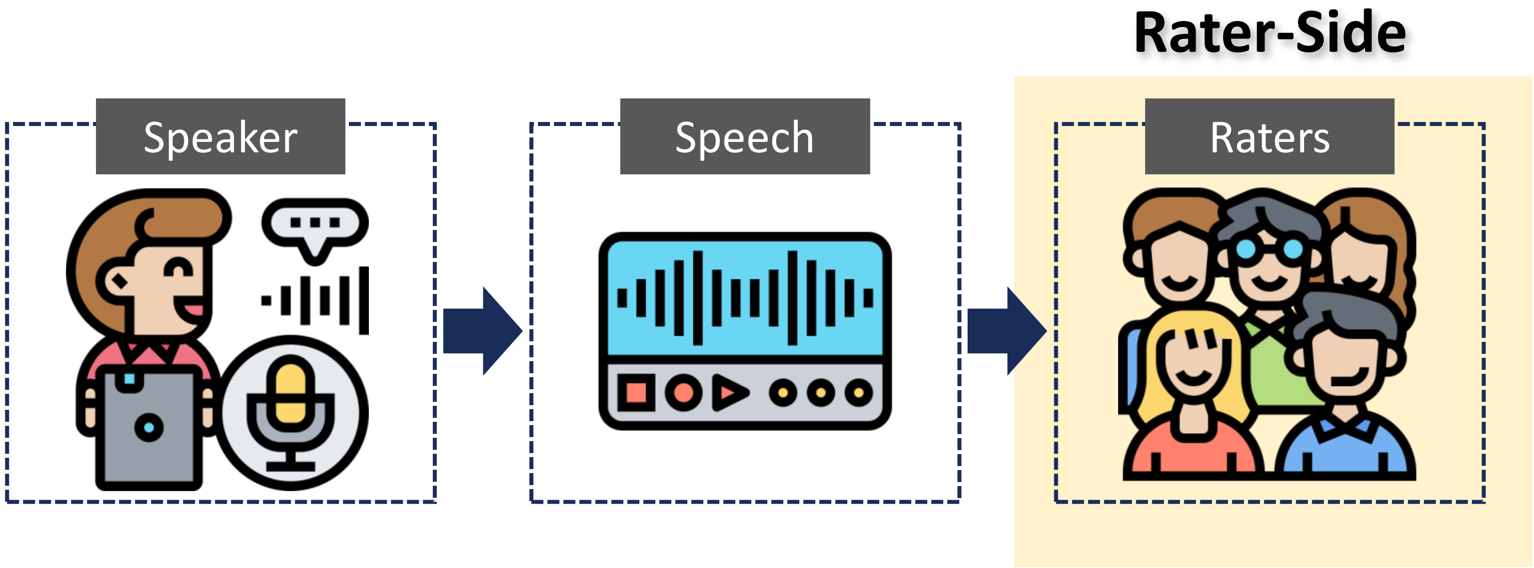
How would the **Rating Biases** arising from *group* or *individual* perspectives manifest in emotional corpora?

Fairness-aware Data Practices

Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024.*

Rater Labeling Biases

- A unique fairness issue in SER stems from the biased emotion perception of human raters as ground truth labels.

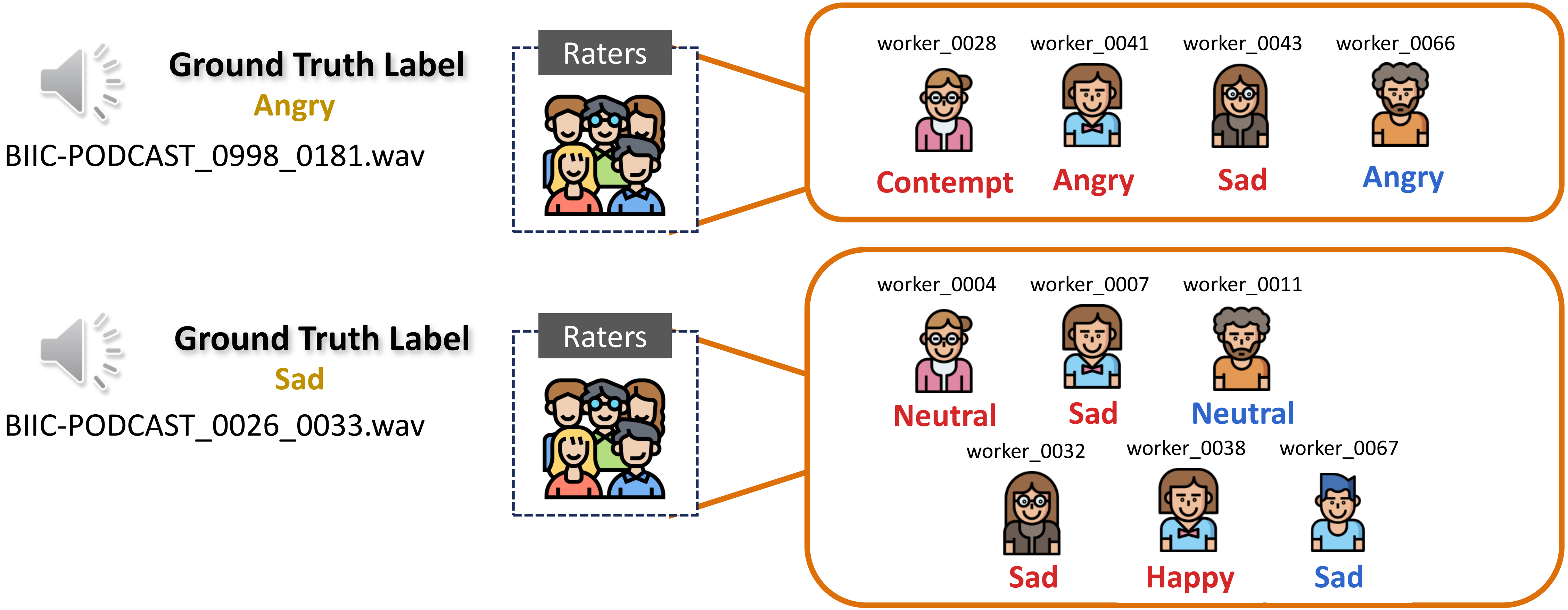


Fairness-aware Data Practices

Woan-Shiuan Chien and Chi-Chun Lee, “**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**” in *Proceeding of Conference of the International Speech Communication Association (Interspeech ‘24), 2024*.

Rater Labeling Biases

- Examples of rater labeling differences from BIIC-Podcast database



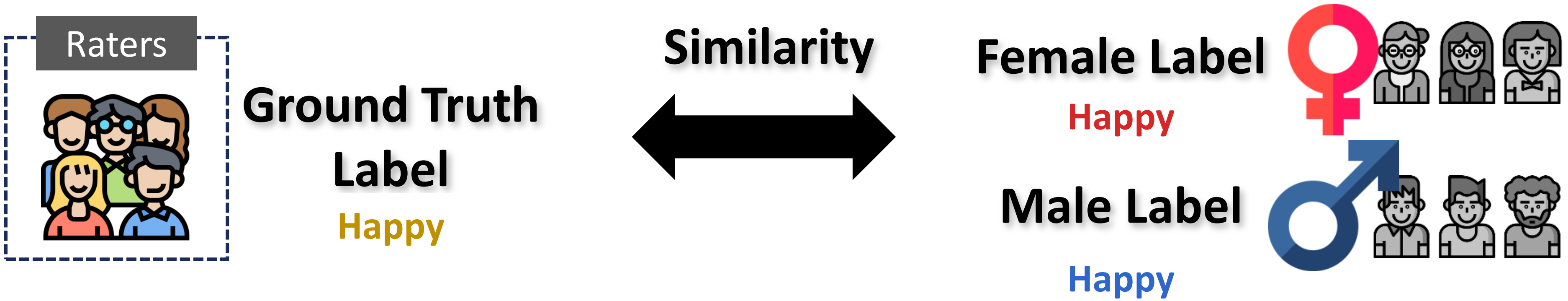
Example

Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024*.

Speech Emotion Corpora

- **IEMOCAP**: 6 unique raters (2 males and 4 females) who provide emotion ratings
- **BIIC-Podcast**: 89 unique raters (30 males and 59 females) who provide emotion ratings
- Emotion: consensus labels are obtained with the plurality rule for primary emotions
- Study sets:
  - S_C: the **rater-gender** unbiased set
    - both ♂ and ♀ have identical emotion perceptions to the ground truth labels
  - S_{NC}: the **rater-gender** biased set
    - the ground truth labels align with the emotion annotation given by either ♂ or ♀ rater only

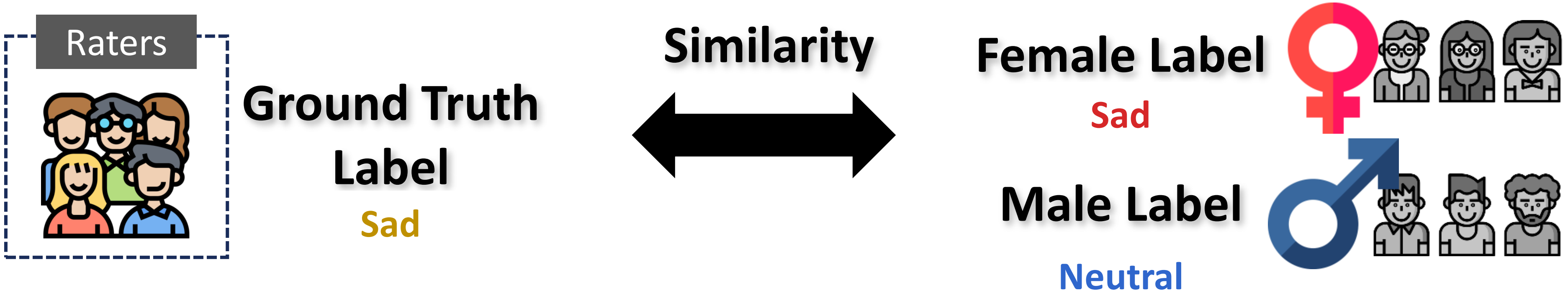
	IEMOCAP					BIIC-Podcast				
	Overall	Neu.	Hap.	Ang.	Sad.	Overall	Neu.	Hap.	Ang.	Sad.
Data Distribution (Numbers)										
S_C	2593	383	1187	471	552	30733	11828	13122	2293	3490
S_{NC}	3025	1323	446	628	628	30736	12726	10888	4035	3087



Fairness-aware Data Practices

Differences in Rater Labeling

- Gender-based Rating Differences: **Label Similarity**
  - Measure the consistency between the consensus ratings by male and female raters against the established ground truth labels.



		IEMOCAP					BIIC-Podcast				
		Overall	Neu.	Hap.	Ang.	Sad.	Overall	Neu.	Hap.	Ang.	Sad.
Label Similarity (%)											
Group (Male)	All Data	80.66	90.04	91.73	90.81	85.83	63.56	77.22	73.65	86.55	72.02
	S_{NC}	67.72	87.30	69.73	85.03	77.55	56.22	51.65	42.17	60.22	42.60
Group (Female)	All Data	59.85	34.82	80.96	50.77	53.80	70.03	68.58	88.29	73.21	80.77
	S_{NC}	32.28	12.70	30.27	14.97	22.45	43.78	48.35	57.83	39.78	57.40

Fairness-aware Data Practices

Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024*.

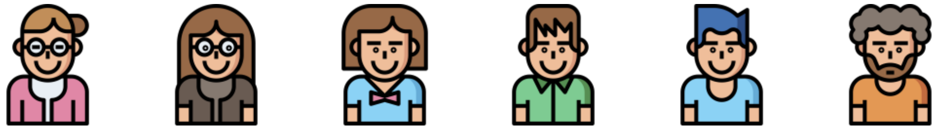
Differences in Rater Labeling

- Individual Rating Differences: **Inter-Annotator Agreement**
  - Employ Fleiss' Kappa (κ) statistics to evaluate the consistency among raters' ratings.
    - Both datasets exhibit fair agreement (κ values ranging from 0.2 to 0.4) for each emotional category.

Raters

Ground Truth Label

Happy



Happy Neutral Happy Happy Angry Sad

		IEMOCAP					BIIC-Podcast				
		Overall	Neu.	Hap.	Ang.	Sad.	Overall	Neu.	Hap.	Ang.	Sad.
Inter-Annotator Agreement (κ)											
Individual	All Data	0.446	0.328	0.306	0.294	0.312	0.421	0.226	0.247	0.218	0.224
Group-level (Male)	All Data	0.467	0.348	0.360	0.402	0.316	0.372	0.212	0.218	0.194	0.226
Group-level (Female)	All Data	0.434	0.305	0.342	0.318	0.288	0.413	0.231	0.210	0.220	0.216

Guiding Question!!

If bias is inevitable, can we *learn* to make the model ignore it?

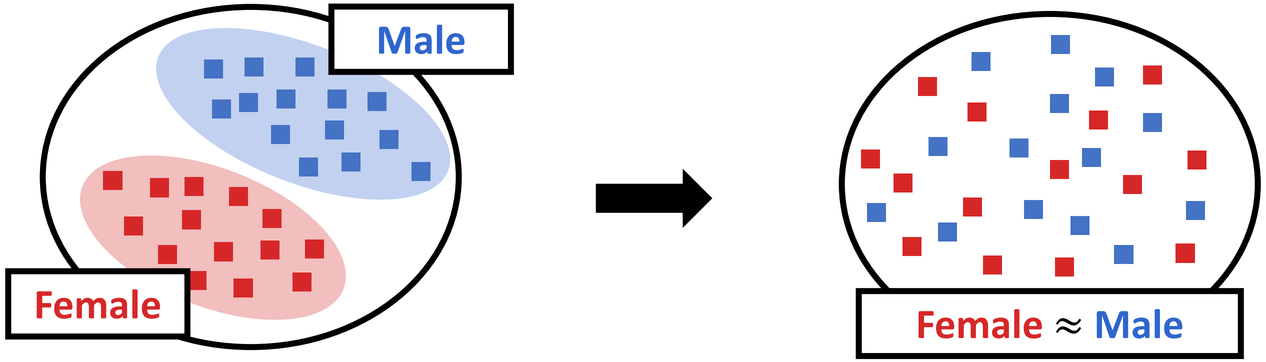
How can we mitigate ***Gender-Based*** bias?

Bias Mitigation Methods

Woan-Shiuan Chien and Chi-Chun Lee, "**Achieving Fair Speech Emotion Recognition via Perceptual Fairness.**" in *Proceeding of the 48th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '23), 2023*.

Rater-sided Fair Representation Learning (Fair_{rat})

- Satisfy Group Fairness: Achieve equitable outcomes across groups (predefined attributes)
  - Related work: Adversarial strategy and Fairness constraint



Y. Ganin, E. Ustinova, H. Ajakan, and others, "Domain-adversarial training of neural networks," The journal of machine learning research, vol. 17, no. 1, pp. 2096–2030, 2016.
C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in Proceedings of the 3rd innovations in theoretical computer science conference, pp. 214–226, 2012.

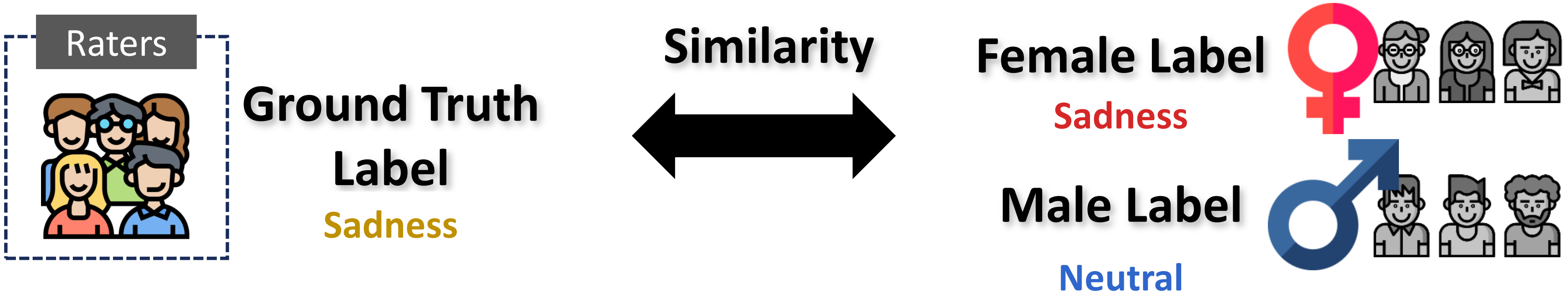
Example

Woan-Shiuan Chien and Chi-Chun Lee, "**Achieving Fair Speech Emotion Recognition via Perceptual Fairness.**" in *Proceeding of the 48th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '23), 2023*.

IEMOCAP dataset

- Study sets:
  - S_{ALL}: the **speaker-gender** biased set (the whole dataset)
  - S_{C}: the **rater-gender** unbiased set
    - both ♂ and ♀ have identical emotion perceptions to the ground truth labels
  - S_{NC}: the **rater-gender** biased set
    - the ground truth labels align with the emotion annotation given by either ♂ or ♀ rater only

	Overall	Neutral	Happiness	Anger	Sadness
S_{ALL}	7362	1706	1633	1099	1080
S_{C}	3038	383	1187	471	552
S_{NC}	4324	1323	446	628	628



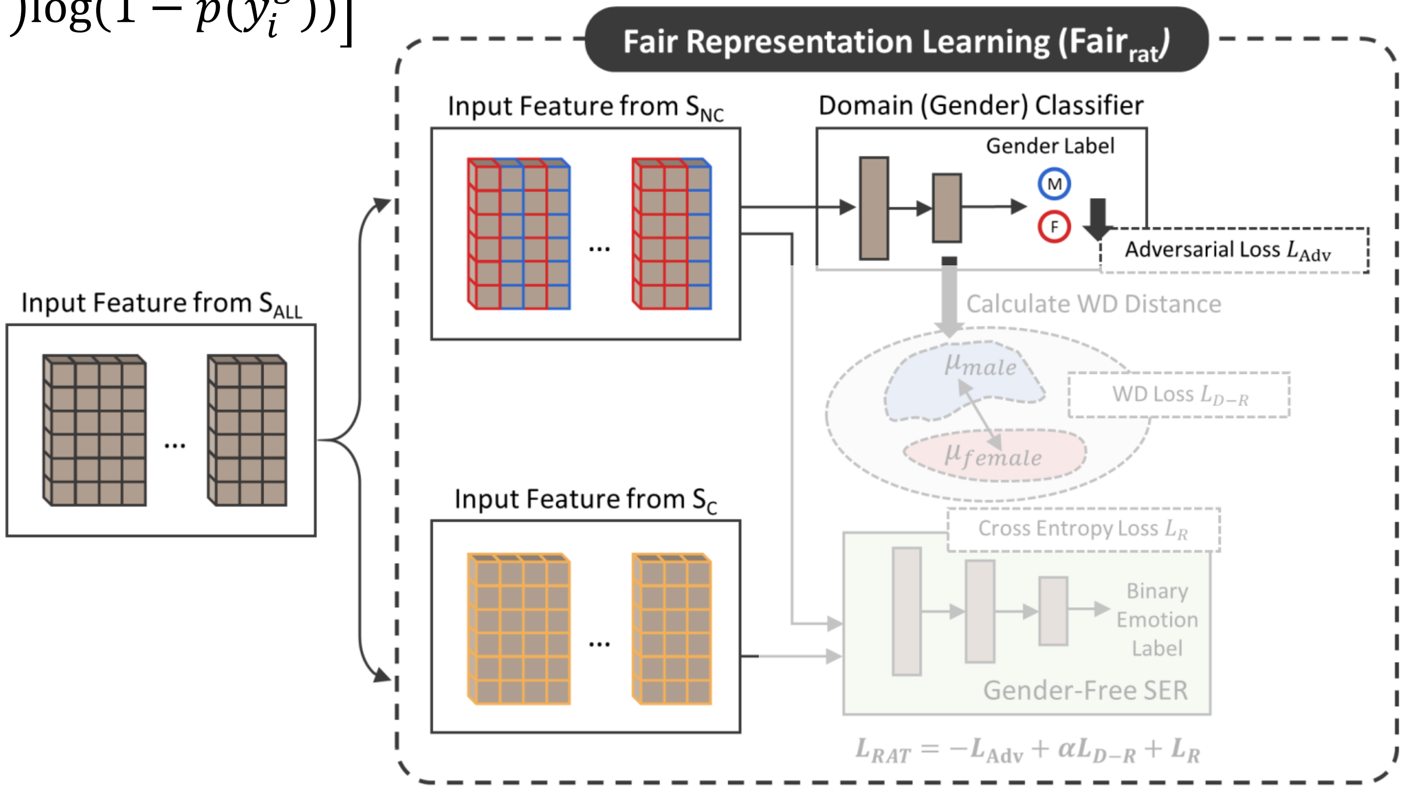
Bias Mitigation Methods

Woan-Shiuan Chien and Chi-Chun Lee, "**Achieving Fair Speech Emotion Recognition via Perceptual Fairness.**" in *Proceeding of the 48th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '23), 2023*.

Rater-sided Fair Representation Learning (Fair$_{rat}$)

- Direct eliminate gender information by learning unbiased representation latent embedding

$$L_{Adv} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i^g \log\left(p(y_i^g)\right) + (1 - y_i^g)\log(1 - p(y_i^g))\right]$$



Bias Mitigation Methods

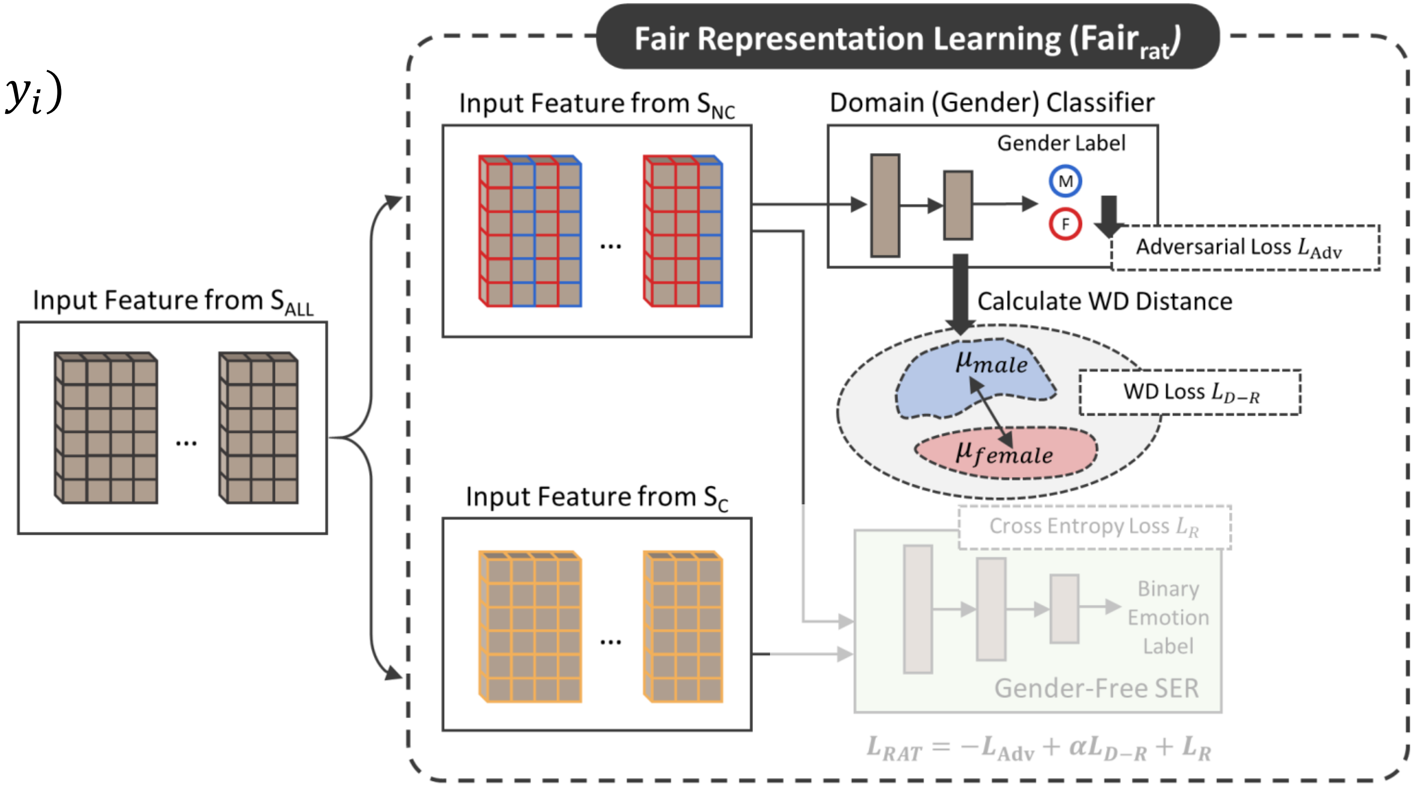
Woan-Shiuan Chien and Chi-Chun Lee, "**Achieving Fair Speech Emotion Recognition via Perceptual Fairness.**" in *Proceeding of the 48th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '23), 2023*.

Rater-sided Fair Representation Learning (Fair$_{rat}$)

- Impose fairness constraints on the distribution of instances in the feature space

$$L_{D-R} = W_1(\mu_{male}, \mu_{female})$$

$$= \min_{\pi \in \Pi(\mu_{male}, \mu_{female})} \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{i,j}\, d(x_i, y_i)$$



Bias Mitigation Methods

Woan-Shiuan Chien and Chi-Chun Lee, "**Achieving Fair Speech Emotion Recognition via Perceptual Fairness.**" in *Proceeding of the 48th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '23), 2023*.

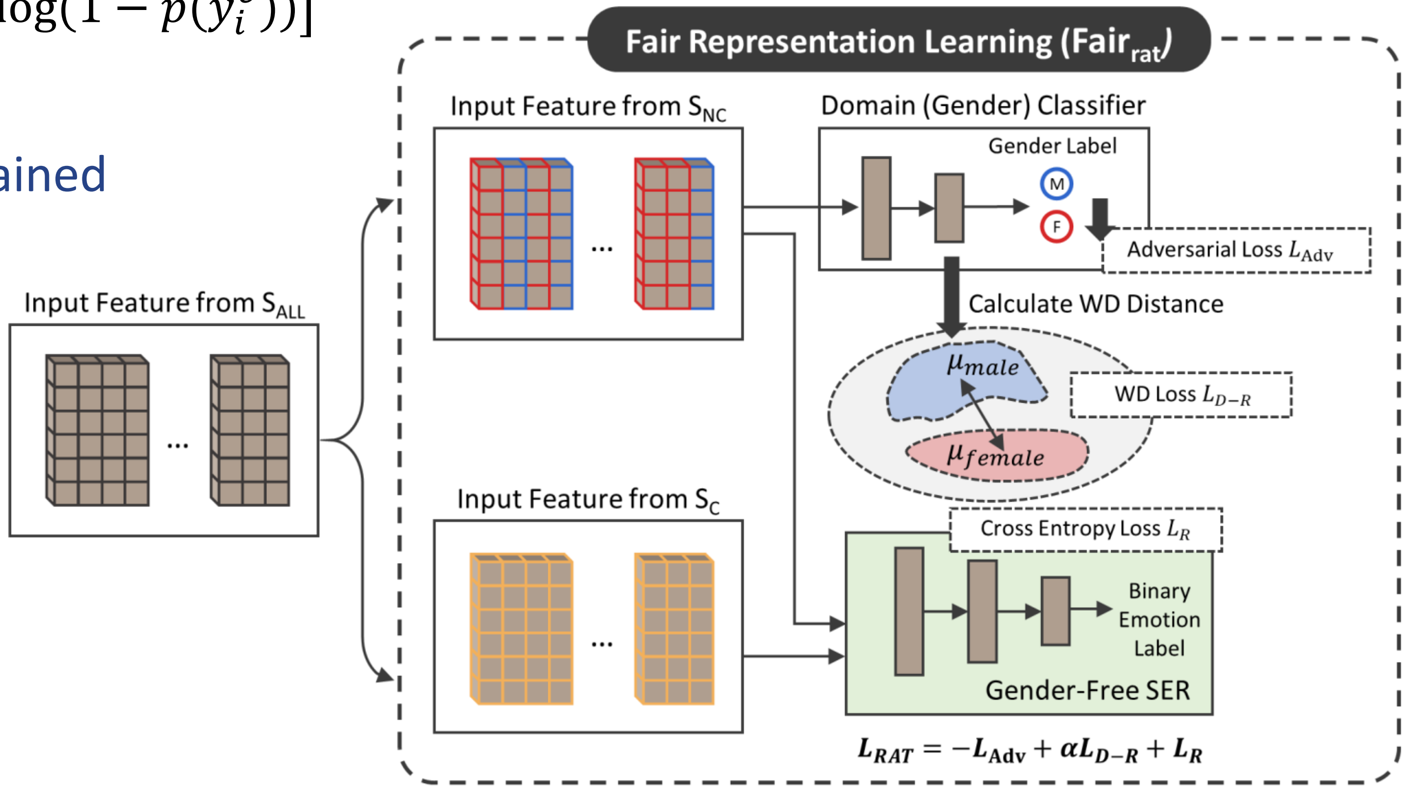
Rater-sided Fair Representation Learning (Fair$_{rat}$)

- Cross entropy loss for binary emotion classification

$$L_R = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i^e \log\left(p(y_i^e)\right) + (1 - y_i^e)\log(1 - p(y_i^e))\right]$$

- The parameters of this network are trained by minimizing the loss function

$$L_{\mathrm{RAT}} = L_R - L_A + \alpha L_{D-R}$$



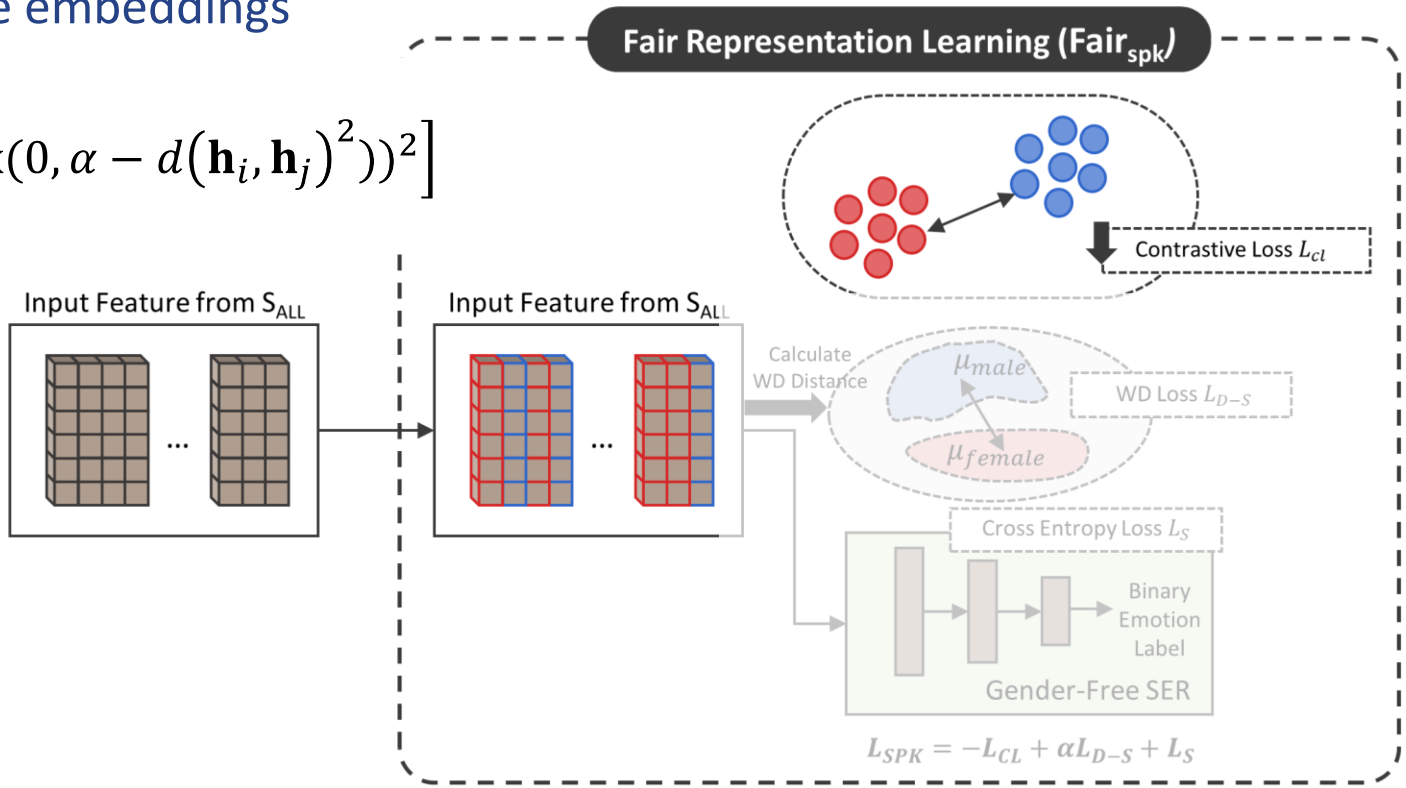
Bias Mitigation Methods

Woan-Shiuan Chien, Shreya G. Upadhyay and Chi-Chun Lee, "**Balancing Speaker-Rater Fairness for Gender-Neutral Speech Emotion Recognition.**" in *Proceeding of the 49th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '24), 2024*.

Speaker-sided Fair Representation Learning (Fair$_{spk}$)

- A similar framework as Fair$_{rat}$ by using a fairness constraint contrastive framework to train the gender debiasing model
- Eliminate gender information from the embeddings

$$L_{cl} = \frac{1}{N}\sum_{i,j}\left[y_{ij}^{e} \cdot d(\mathbf{h}_i, \mathbf{h}_j)^2 + (1 - y_{ij}^{e}) \cdot (\max(0, \alpha - d(\mathbf{h}_i, \mathbf{h}_j)^2))^2\right]$$



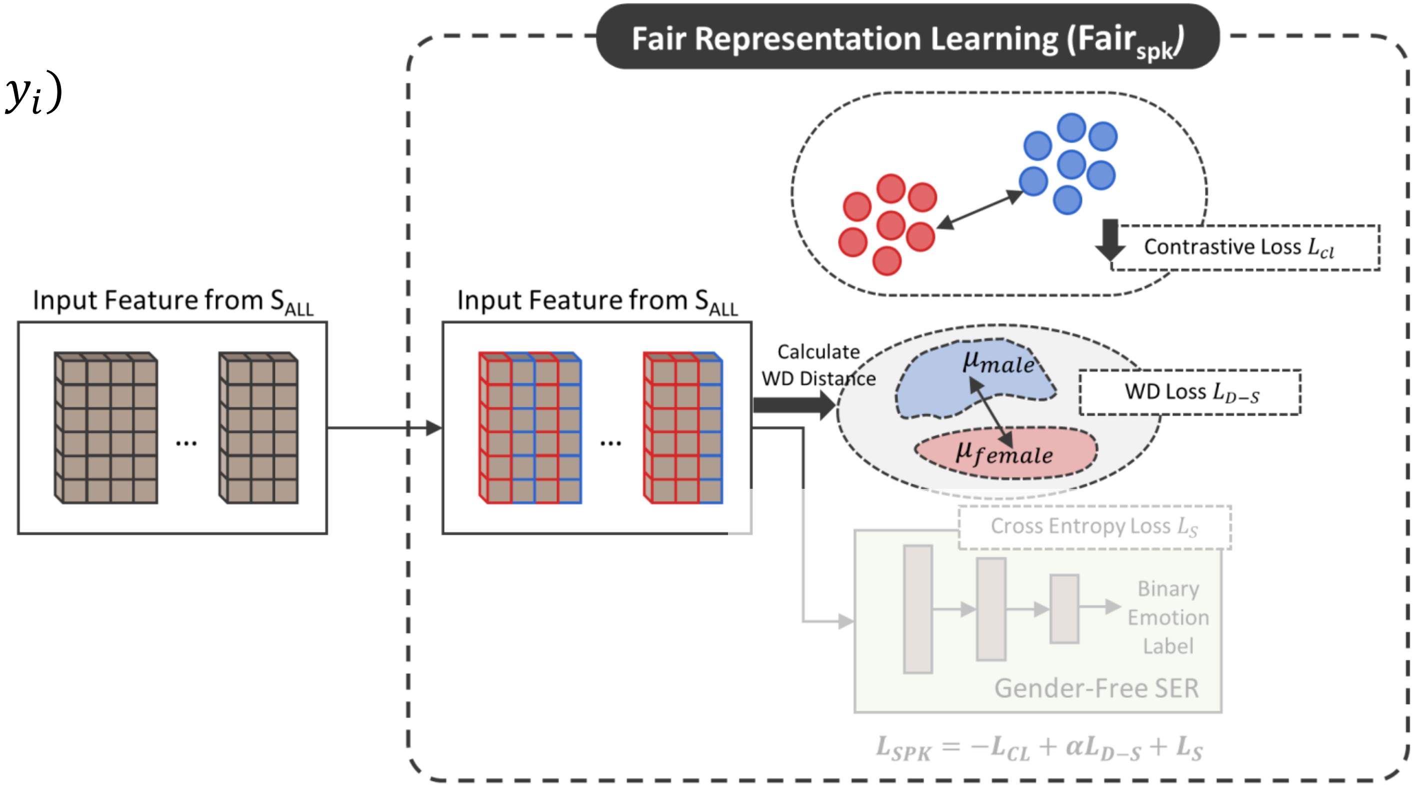
Bias Mitigation Methods

Woan-Shiuan Chien, Shreya G. Upadhyay and Chi-Chun Lee, "**Balancing Speaker-Rater Fairness for Gender-Neutral Speech Emotion Recognition.**" in *Proceeding of the 49th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '24), 2024*.

Speaker-sided Fair Representation Learning ($Fair_{spk}$)

- Impose fairness constraints on the distribution of instances in the feature space

$$L_{D-S} = W_1(\mu_{male}, \mu_{female})$$
$$= \min_{\pi \in \Pi(\mu_{male}, \mu_{female})} \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{i,j}\, d(x_i, y_i)$$



Bias Mitigation Methods

Woan-Shiuan Chien, Shreya G. Upadhyay and Chi-Chun Lee, "**Balancing Speaker-Rater Fairness for Gender-Neutral Speech Emotion Recognition.**" in *Proceeding of the 49th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '24), 2024*.

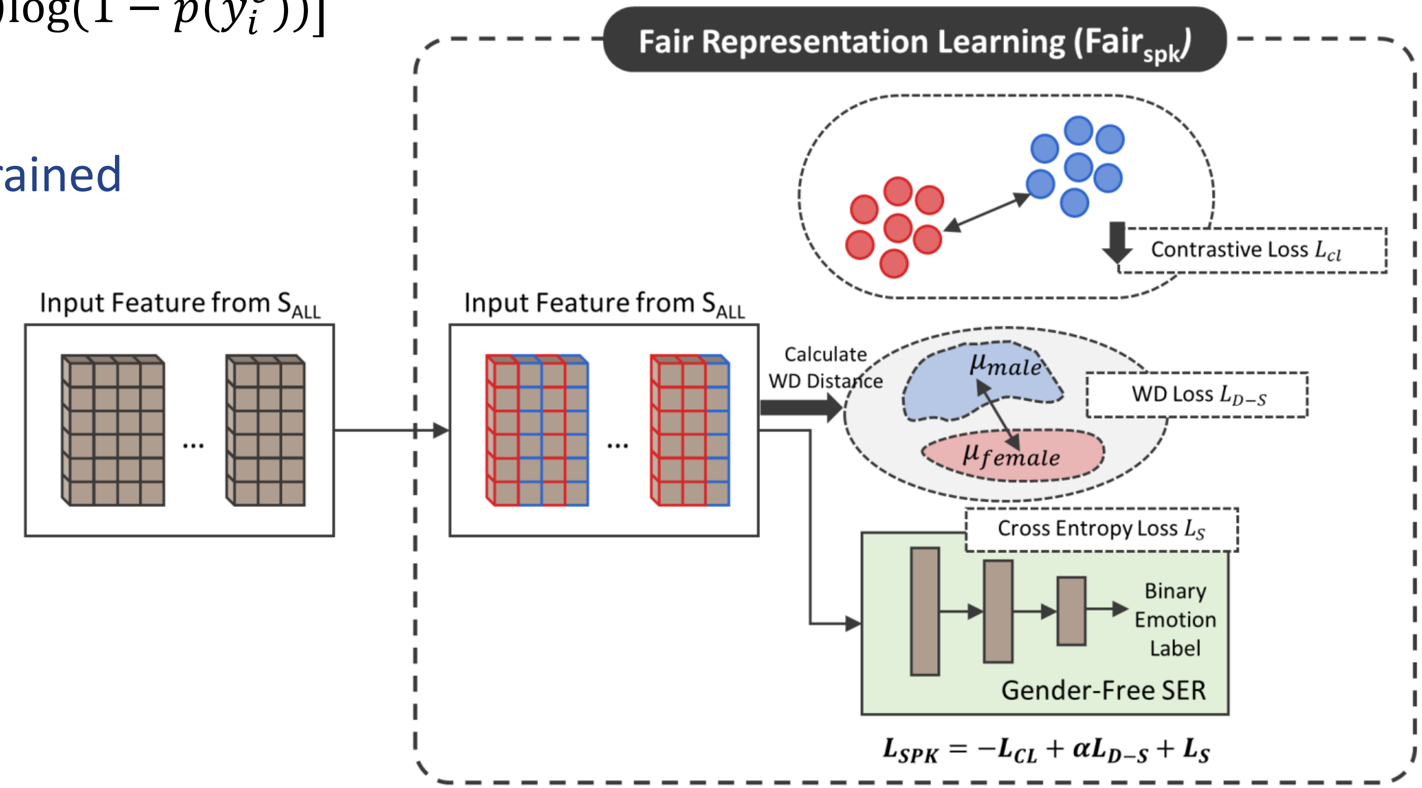
Speaker-sided Fair Representation Learning (Fair_{spk})

- Cross entropy loss for binary emotion classification

$$L_S = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i^e \log\left(p(y_i^e)\right) + (1 - y_i^e)\log(1 - p(y_i^e))\right]$$

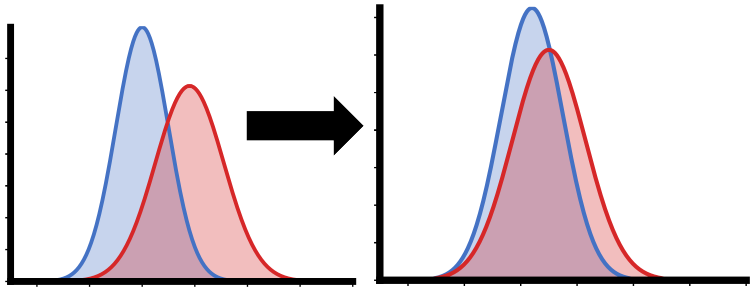
- The parameters of this network are trained by minimizing the loss function

$$L_{\text{SPK}} = L_S - L_{cl} + \alpha L_{D-S}$$



Experiments

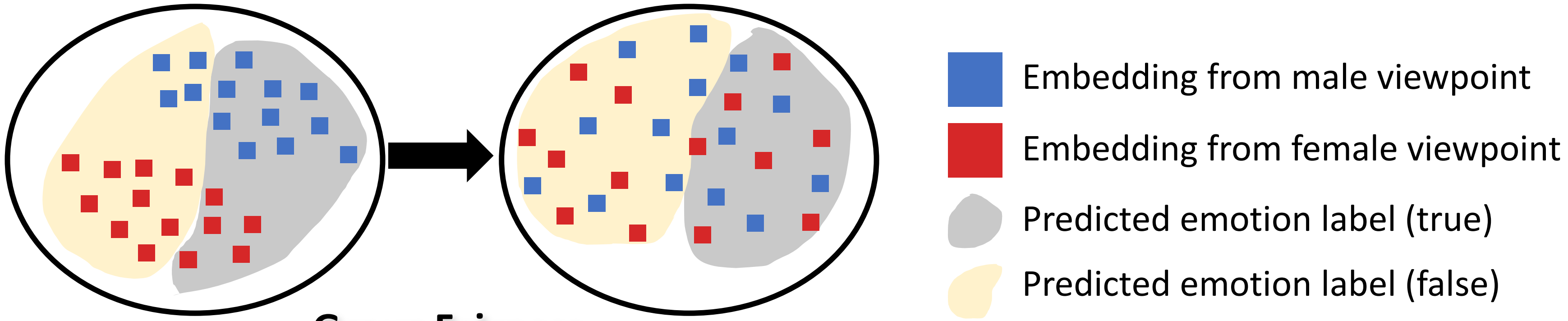
Experimental Setups and Evaluations



- Features: vq-wav2vec representation
- Target emotion label: voted ground truth
- Emotion recognition performance: weighted F1-score on S_{ALL} dataset
- Fairness metric: **statistical parity score** ΔSP (ideal value=0)

$$\Delta SP = \left| P(\hat{Y} = \textbf{emotion label} \mid A = male) - P(\hat{Y} = \textbf{emotion label} \mid \bar{A} = female) \right|$$

- Evaluate on $\mathbf{S_{NC}}$ dataset between different **rater's gender** and our predictions
- Evaluate on $\mathbf{S_{ALL}}$ dataset between different **speaker's gender** and our predictions



Group Fairness

Experiments

Woan-Shiuan Chien, Shreya G. Upadhyay and Chi-Chun Lee, "**Balancing Speaker-Rater Fairness for Gender-Neutral Speech Emotion Recognition.**" in *Proceeding of the 49th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '24), 2024*.

Fairness Evaluation Scheme

- Fairness metric: statistical parity score (ideal value=0)
  - **Intra-Fairness**: evaluate the one-sided gender-neutral fairness in their own corresponding gender viewpoint, i.e., using ΔSP_{spk} for Fair_{spk} and ΔSP_{rat} for Fair_{rat}
  - **Inter-Fairness**: evaluate the fairness metric of one-sided using the model of the other. This means using ΔSP_{spk} for Fair_{rat} and ΔSP_{rat} for Fair_{spk}

		Gender Viewpoint	
		$\boldsymbol{\Delta SP_{spk}}$	$\boldsymbol{\Delta SP_{rat}}$
Gender-Neutral Model	$\textbf{Fair}_{\textbf{spk}}$	V	V
	$\textbf{Fair}_{\textbf{rat}}$	V	V

Results and Analyses

Woan-Shiuan Chien, Shreya G. Upadhyay and Chi-Chun Lee, "**Balancing Speaker-Rater Fairness for Gender-Neutral Speech Emotion Recognition.**" in *Proceeding of the 49th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '24), 2024*.

Intra-Fairness

- It suffers the least performance drop on the recognition performance
- It better satisfies statistical parity metrics than methods without consideration of fairness

Inter-Fairness

- Fair_{spk} exhibits a substantial increase in ΔSP_{rat}
- Fair_{rat} exhibits a substantial increase in ΔSP_{spk}

	Neutral			Happiness			Anger			Sadness		
	F1(%)	ΔSP_{spk}	ΔSP_{rat}	F1(%)	ΔSP_{spk}	ΔSP_{rat}	F1(%)	ΔSP_{spk}	ΔSP_{rat}	F1(%)	ΔSP_{spk}	ΔSP_{rat}
DNN	77.73	0.452	0.649	70.00	0.511	0.428	76.44	0.378	0.389	82.28	0.359	0.169
Fair_{spk}	70.68	0.226	0.488	65.80	0.380	0.366	73.26	0.234	0.379	75.50	0.260	0.208
Fair_{rat}	68.80	0.403	0.352	65.14	0.691	0.126	75.68	0.372	0.189	76.84	0.291	0.088

The one-sided fair SER model does not generalize well across different viewpoints.

Question!!

- Can we make both sides fair at the same time?
  - Two-sided Fairness?

Guiding Question!!

Group Fairness

versus

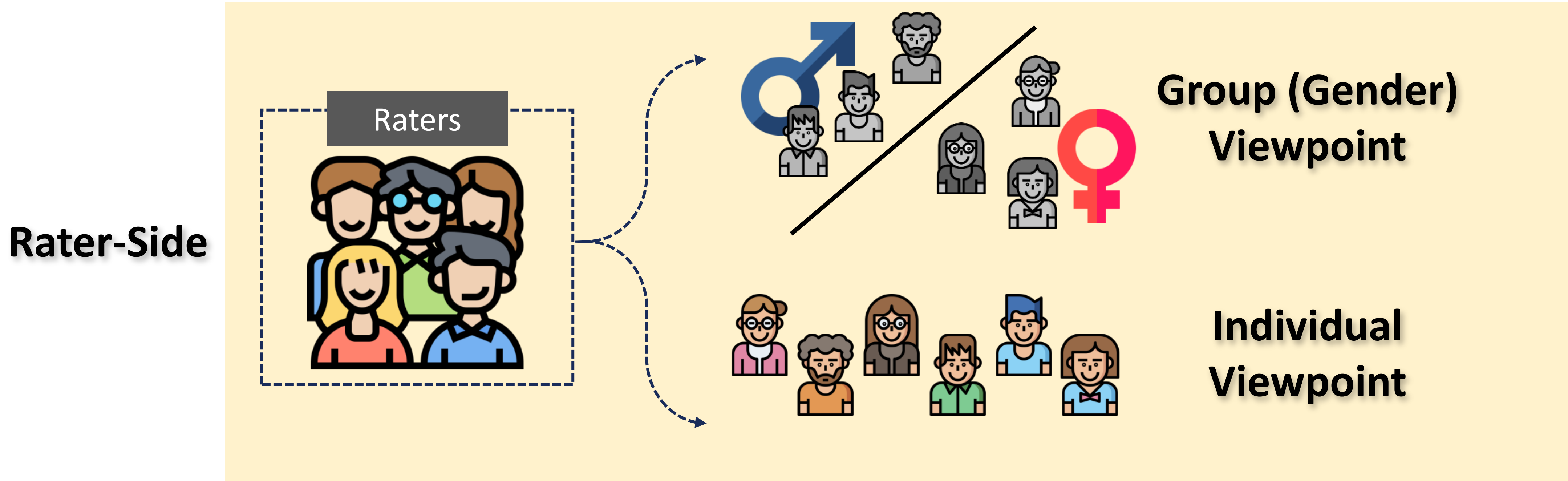
Individual Fairness

Evaluation Frameworks

Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024*.

Group Fairness versus Individual Fairness

- Achieve either **group** or **individual** fairness alone may not be sufficient for comprehensive fairness due to the distinct nature of these fairness concepts. → Conflicts between the two fairness paradigms
  - Group Fairness: Achieve equitable outcomes across groups (predefined attributes)
  - Individual Fairness: Ensure that individuals with similar representations would receive similar predictions from the model (system).

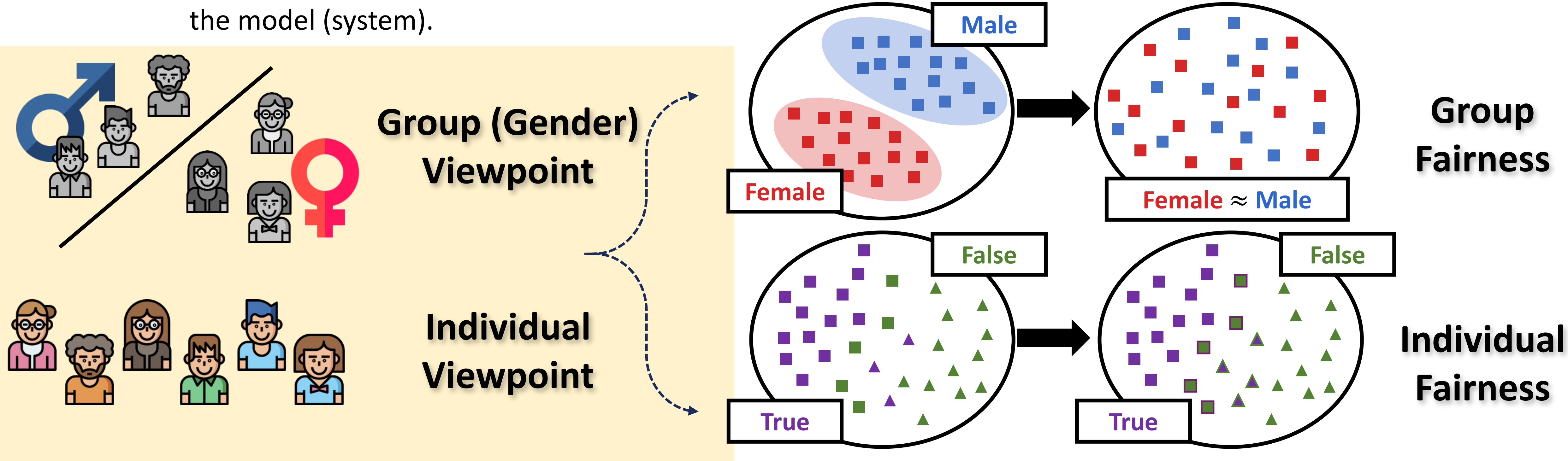


Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024.*

Evaluation Frameworks

Trade-off Between Group and Individual Fairness

- Achieve either **group** or **individual** fairness alone may not be sufficient for comprehensive fairness due to the distinct nature of these fairness concepts. → Conflicts between the two fairness paradigms
  - Group Fairness: Achieve equitable outcomes across groups (predefined attributes)
  - Individual Fairness: Ensure that individuals with similar representations would receive similar predictions from the model (system).



Evaluation Frameworks

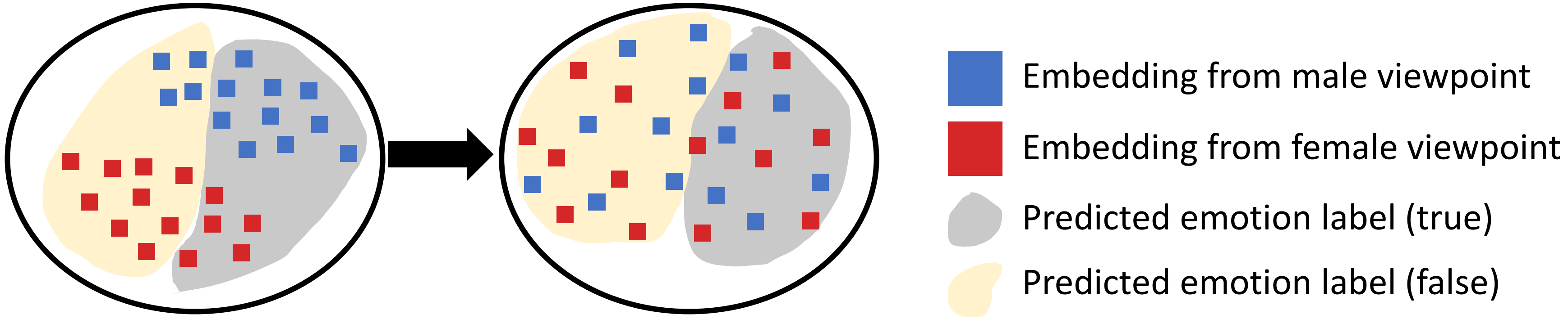
Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024*.

Evaluations

- Group Fairness: **statistical parity score** ΔSP (ideal value=0)

$$\Delta SP = \left|P(\hat{Y} = \textbf{emotion label} \,|A = male) - P(\hat{Y} = \textbf{emotion label} \,|\bar{A} = female)\right|$$

  - Evaluate on $\mathbf{S_{NC}}$ dataset between different **rater's gender** and our predictions



Group Fairness

Evaluation Frameworks

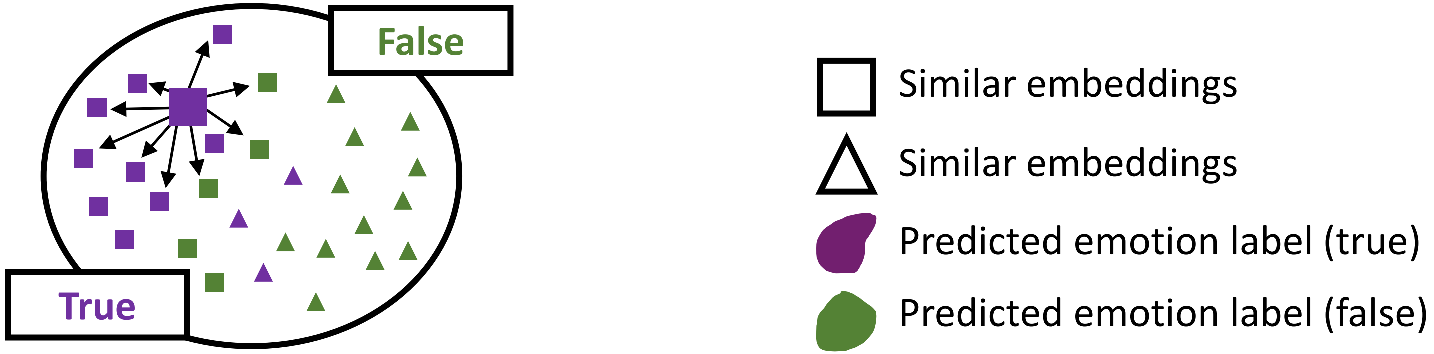
Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024*.

Evaluations

- Individual Fairness: **consistency score** ΔC (ideal value=1)

$$\Delta C = 1 - \frac{1}{k}\sum_{i=1}^{k}\left|\hat{y}_i - \frac{1}{k_{\text{neighbors}}}\sum_{j\in\mathcal{K}_{\text{neighbors}}(i)}\hat{y}_j\right|$$

  - Evaluate on $\mathbf{S}_{\mathbf{ALL}}$ dataset between different **rater's gender** and our predictions (k=20)



Individual Fairness

Guiding Question!!

Can a model be fair to groups but unfair to individuals?

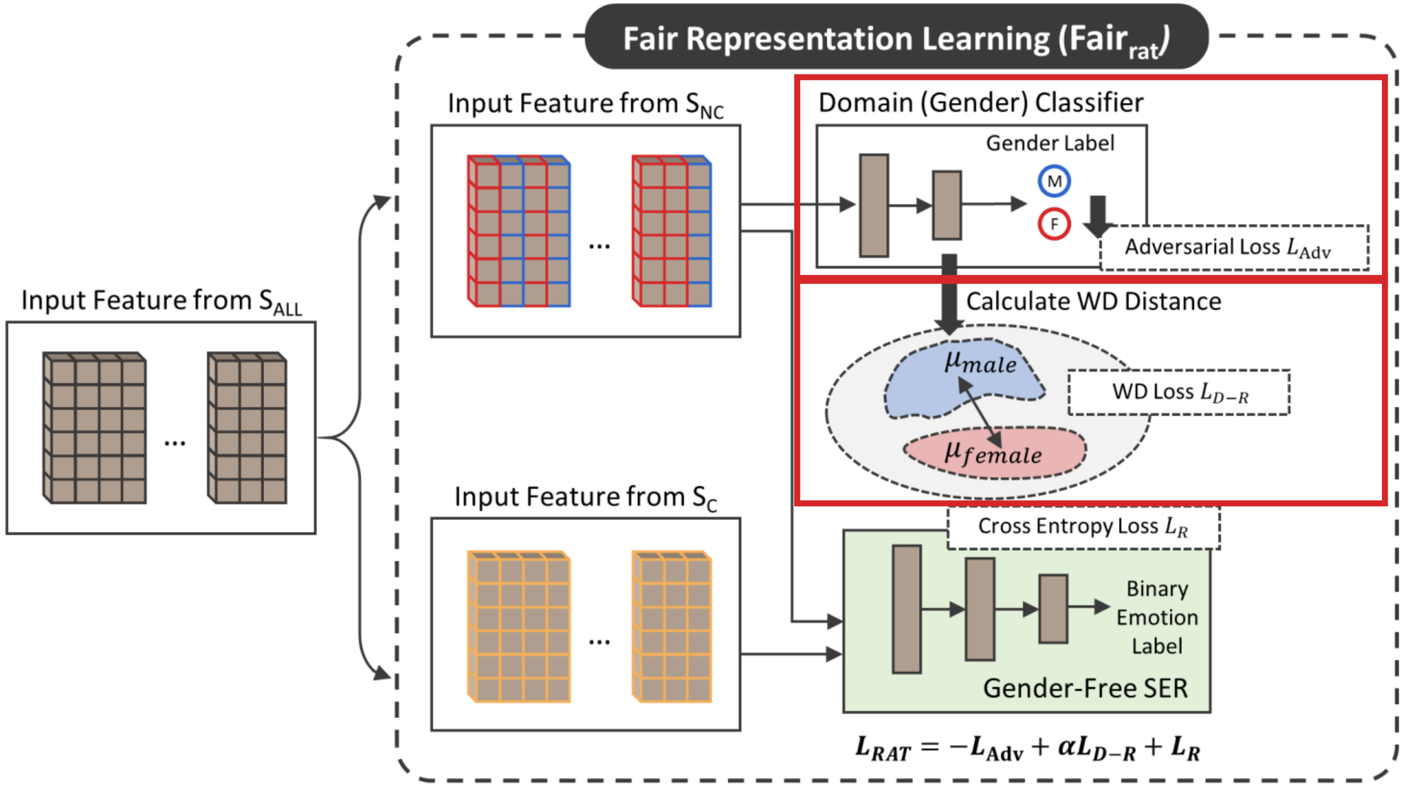
How would individual fairness be affected when we improve group fairness?

Evaluation Frameworks: Trade-off

Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024*.

In-processing Learning for Achieving Group Fairness

- Effects of removing group information on fairness metrics
- Influence when satisfying group fairness through Wasserstein Distance (WD) measures

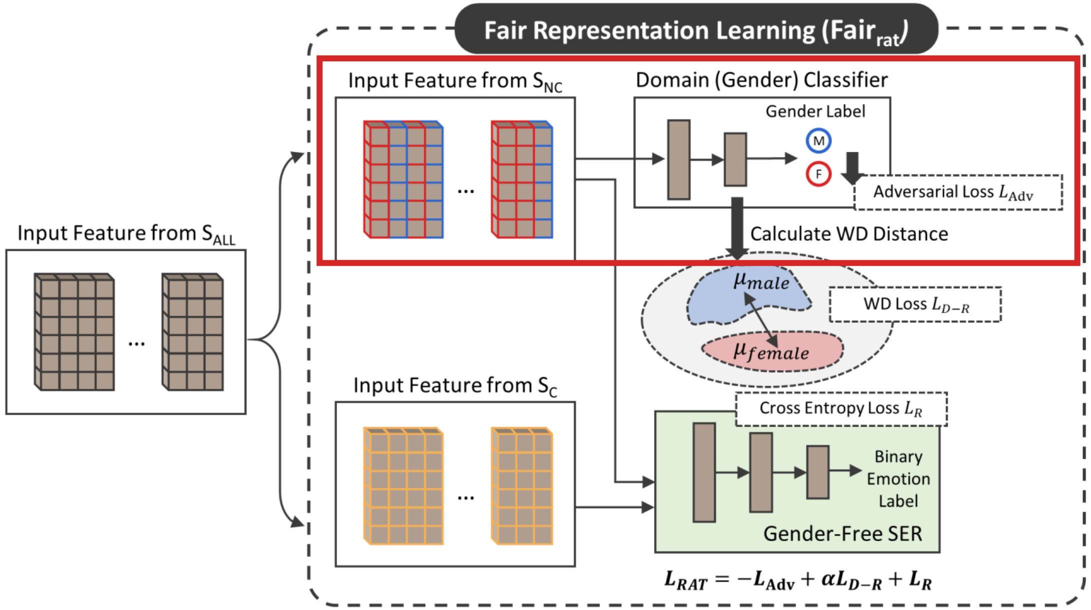


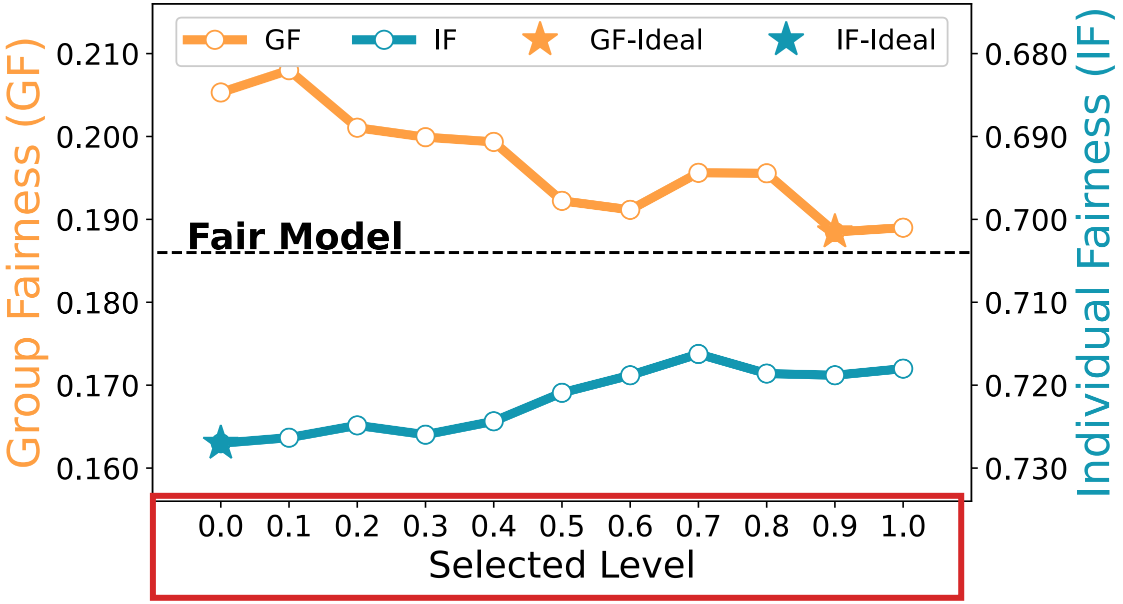
Evaluation Frameworks: Trade-off

Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024.*

Effects of Partial Group Information Elimination

- Randomly remove gender information from the S_{NC} data to weaken the domain-invariant classifier
- Train the domain-invariant classifier using **N%** of S_{NC} data, where **N** varies from 0 to 100 in increments of 10



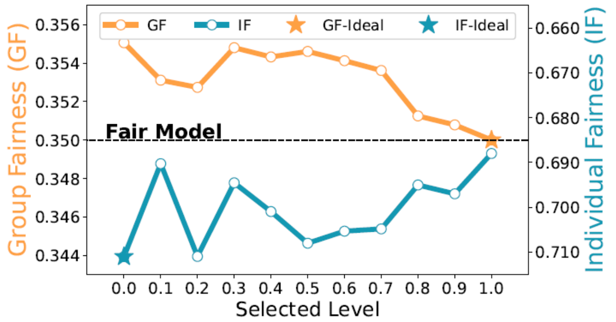


Evaluation Frameworks: Trade-off

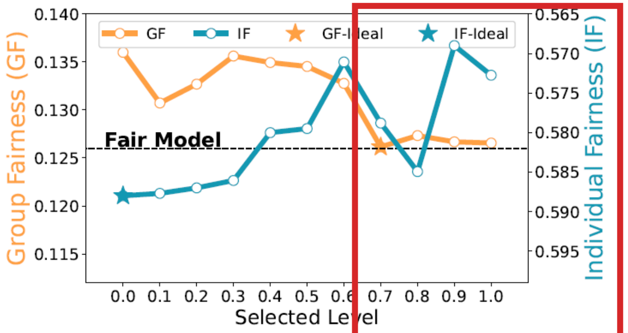
Woan-Shiuan Chien and Chi-Chun Lee, "**An Investigation of Group versus Individual Fairness in Perceptually Fair Speech Emotion Recognition.**" in *Proceeding of Conference of the International Speech Communication Association (Interspeech '24), 2024.*

Effects of Partial Group Information Elimination

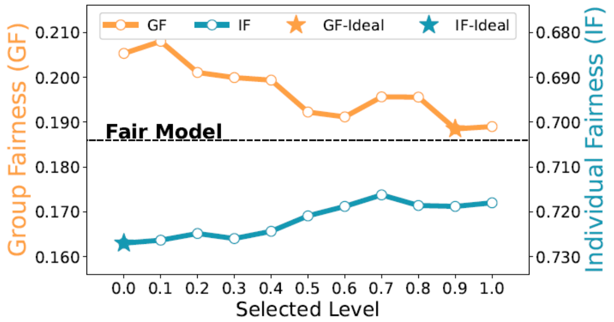
- Significant reduction in individual fairness when over 70% of data was de-gendered
- Differences in individual fairness effects were pronounced between IEMOCAP (less than 4% discrepancy) and BIIC-Podcast (up to 20% discrepancy)



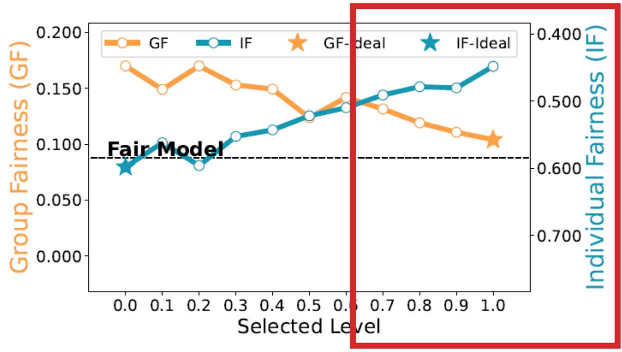
(a) *Neutral-IEMOCAP*



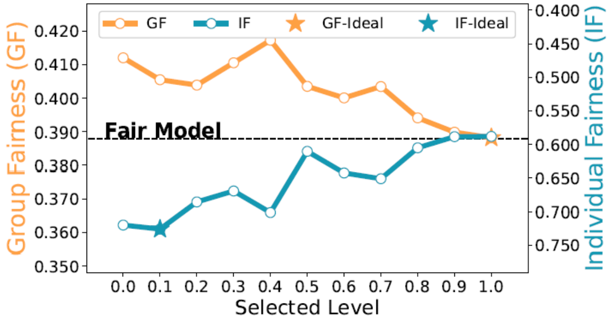
(b) *Happiness-IEMOCAP*



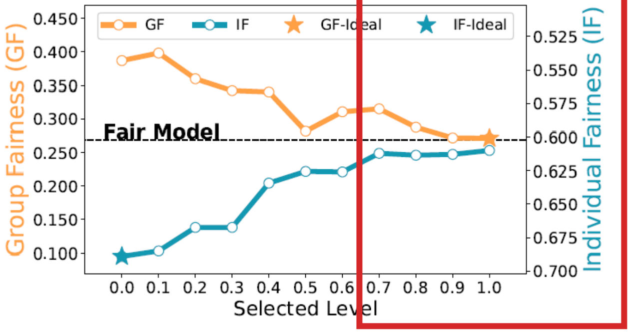
(c) *Anger-IEMOCAP*



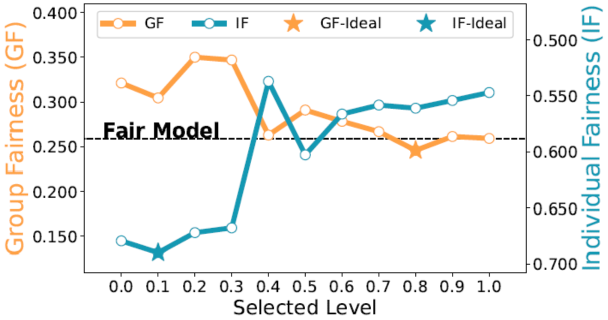
(d) *Sadness-IEMOCAP*



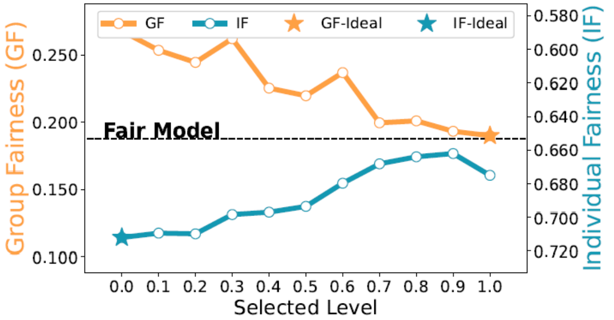
(e) *Neutral-BIIC-Podcast*



(f) *Happiness-BIIC-Podcast*



(g) *Anger-BIIC-Podcast*



(h) *Sadness-BIIC-Podcast*

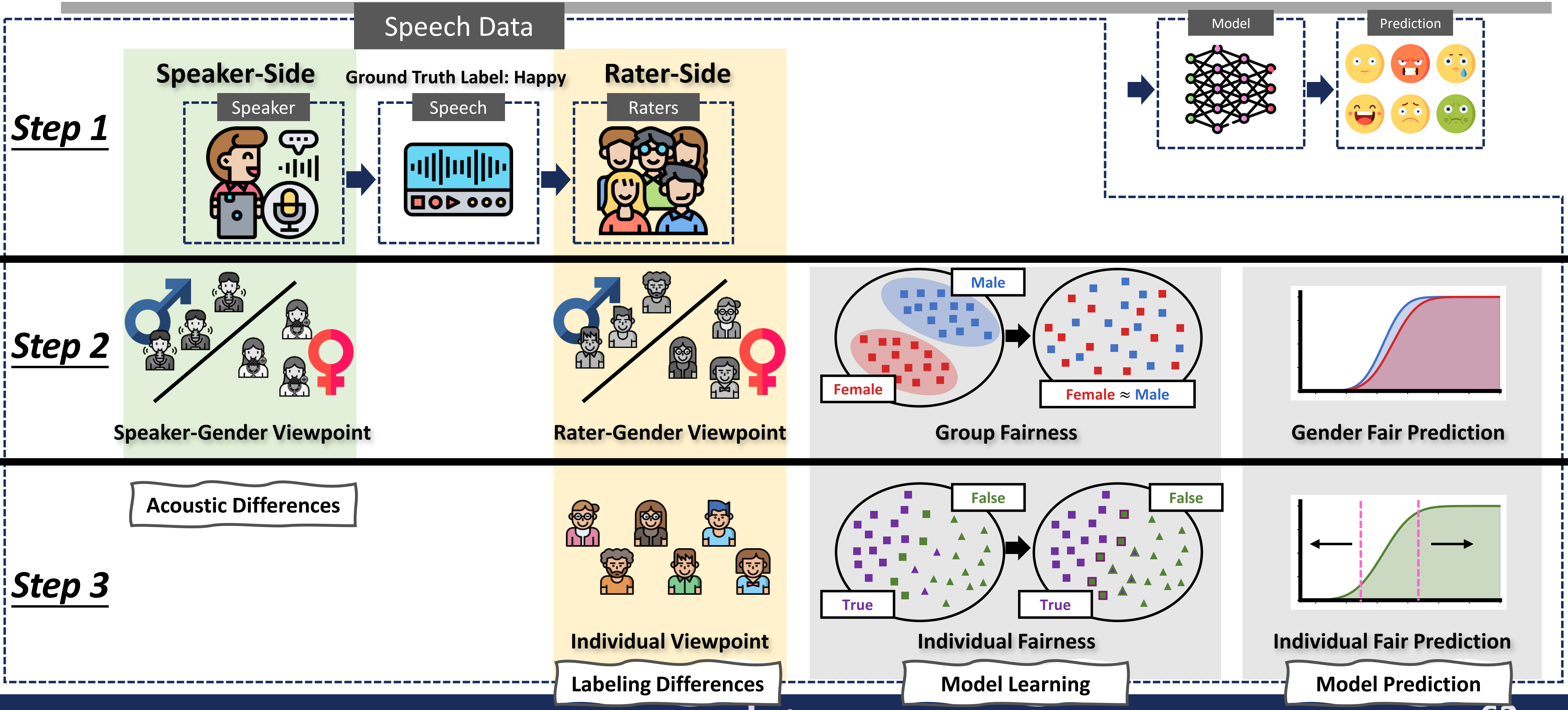
Open Reflections!!

- Fairness == Debiasing?
- Who defines what is FAIR?
  - The model, the data, or the people?
- Who matters most?
  - The speaker, the rater, or the society behind them?

Challenges And Opportunities

- No Consensus on Definition
- Transparent Debiasing and Fairness
- Fairness–Performance Relationship
- Better Evaluation
- …

Summary

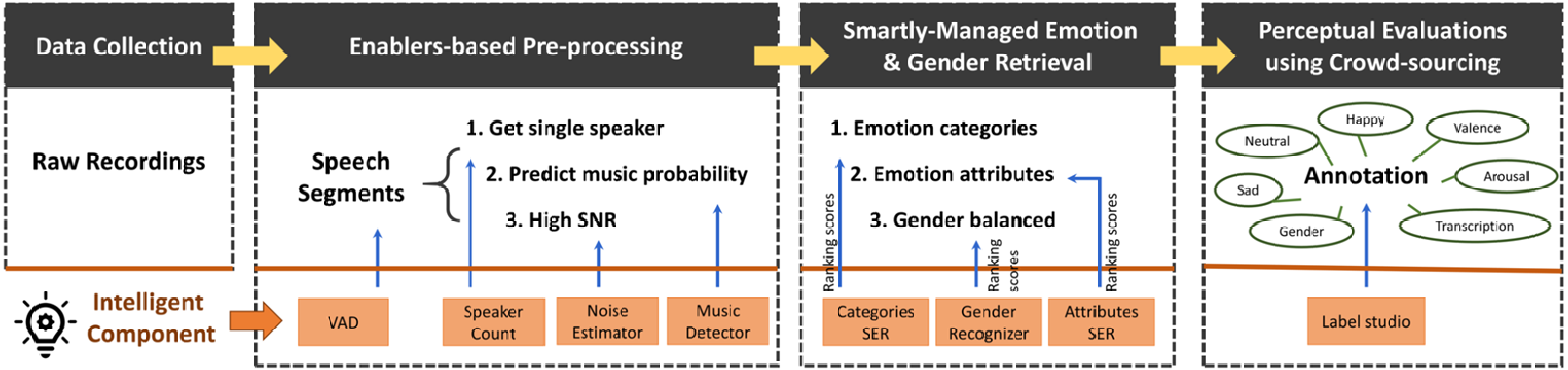


Speech Data
Model
Prediction
Step 1
Speaker-Side
Ground Truth Label: Happy
Rater-Side
Speaker
Speech
Raters
Step 2
Male
Female
Female ≈ Male
Speaker-Gender Viewpoint
Rater-Gender Viewpoint
Group Fairness
Gender Fair Prediction
Step 3
Acoustic Differences
False
True
Individual Viewpoint
Individual Fairness
Individual Fair Prediction
Labeling Differences
Model Learning
Model Prediction

Data Resources: BIIC-Podcast

We provide a centralized platform for researchers, offering a customizable-standard pipeline and access to affective speech corpora, collaborating with MSP lab at UT Dallas, USA (>200 hours, continuing…)

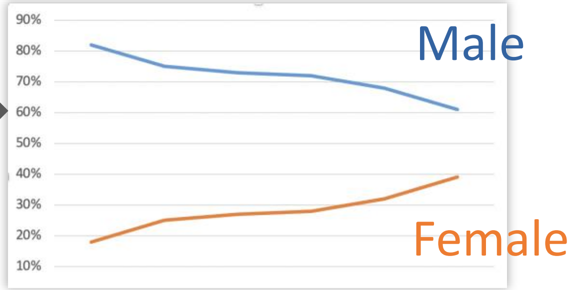
The collection of data will be optimized over time, and the process is transparent to all researchers.



Affective Naturalistic Database Consortium

http://andc.ai/

From October 2022 to October 2023. Initially, the labels released show a majority of males outnumbering females.



Shreya G. Upadhyay*, Woan-Shiuan Chien*, Bo-Hao Su, Lucas Goncalves, Ya-Tse Wu, Ali N. Salman, Carlos Busso and Chi-Chun Lee, "An Intelligent Infrastructure Toward Large Scale Naturalistic Affective Speech Corpora Collection." in Proceeding of the 11th International Conference on Affective Computing & Intelligent Interaction (ACII '23), 2023.

THANK YOU !!

biic
人本訊號運算研究室
Behavioral Informatics and Interaction Computation Lab

//Q&A

NSTC 國家科學及技術委員會
National Science and Technology Council