icassp 2022 Singapore

Exploiting Annotators' Typed Description of Emotion Perception to Maximize Utilization of Ratings for Speech Emotion Recognition

Huang-Cheng Chou[1,2] , Wei-Cheng Lin[1], Chi-Chun Lee[2], Carlos Busso[1]

[1] Multimodal Signal Processing (MSP) lab,
Department of Electrical and Computer Engineering, University of Texas at Dallas (UTD), USA

[2] Behavioral Informatics & Interaction Computation (BIIC) lab,
Department of Electrical Engineering, National Tsing Hua University (NTHU), Taiwan

THE UNIVERSITY OF TEXAS AT DALLAS

UT Dallas MSP Multimodal Signal Processing Laboratory

TSING HUA UNIVERSITY

biic 人類行爲訊息暨互動計算研究室 Behavioral Informatics and Interaction Computation Lab

Emotion Perception

Emotional stimulus **Emotion perception** **Emotion decoding** **Annotation**

Speech Emotion Recognition

Example of Annotation in the MSP-Podcast corpus

Annotations for primary emotion (single-choice):

Disagreement

MSP-PODCAST_0004_0073.wav
Rater 1: Neutral
Rater 2: Neutral
Rater 3: ~~Happy~~
Rater 4: ~~Other~~ (accusatory)
Rater 5: ~~Other~~ (Pleased)

Discarded

Never used

Consensus label: **Neutral**

R. Lotfian and C. Busso, "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings," IEEE Trans. Affect. Comput., vol. 10, no. 4, pp. 471-483, October-December 2019.

Where Typed Words Come From

Is any of these emotions the primary emotion in the audio? If not, select **Other** and specify the emotion.

○ Angry ○ Sad ○ Happy ○ Surprise ○ Fear ○ Disgust ○ Contempt ○ Neutral ○ Other

Single-choice

(d) Primary emotion

Please pick all the emotional classes that you perceived in the audio(Include the primary emotions selected in previous question)

☐ Angry ☐ Sad ☐ Happy ☐ Amused ☐ Neutral

☐ Frustreated ☐ Depressed ☐ Surprise ☐ Concerned

☐ Disgust ☐ Disappointed ☐ Excited ☐ Confused

☐ Annoyed ☐ Fear ☐ Contempt ☐ Other

Multi-choice

(e) Secondary emotion

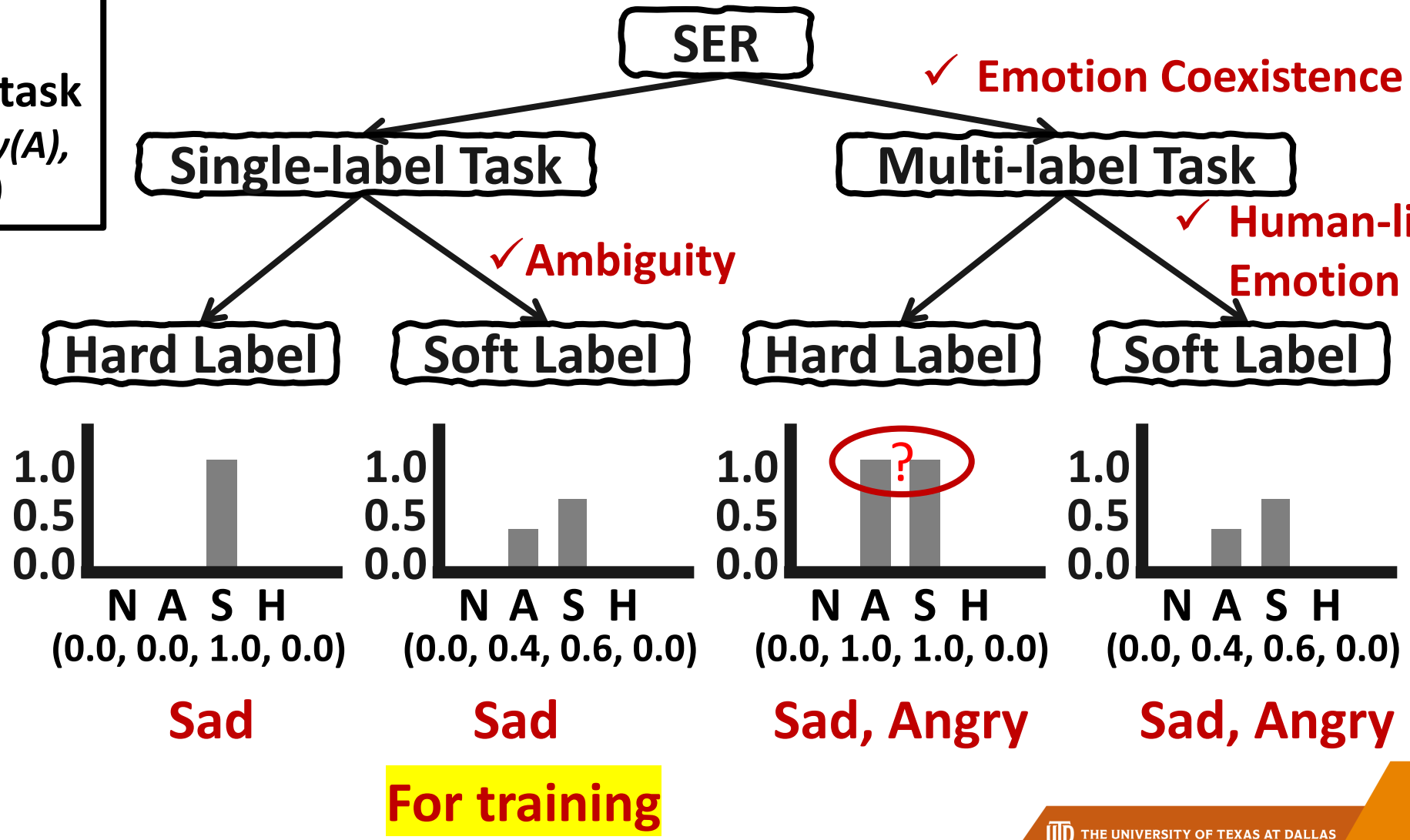
Primary emotion example:
MSP-PODCAST_0004_0073.wav

1. **W0002117; Other (Pleased)**
2. **W0000060; Neutral**
3. **W0003012; Other (accusatory)**
4. **W0002999; Neutral**
5. **W0003011; Happy**

R. Lotfian and C. Busso, "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings," IEEE Trans. Affect. Comput., vol. 10, no. 4, pp. 471-483, October-December 2019.

Decision of Labels for Speech Emotion Recognition (SER)

Example:
Four-class SER task
Neutral (N), Angry(A), Sad (S), Happy (H)



For training

Motivation

We aim to utilize all emotional annotations to improve the prediction of primary and secondary emotions!

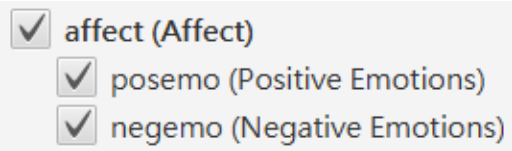
Methodology

Purpose:

- Explore the benefits of using the typed words provided by evaluators when they selected the class ``**other**'' in the primary or secondary emotions for improving performance of SER systems

Method:

- Propose a **three-dimensional (3D) polarity label** (positive, negative, and ambiguous emotion words) with all emotional annotations
  - Include all typed words
  - Include primary and secondary emotions
  - Polarity obtained with Linguistic Inquiry and Word Count (**LIWC**) 2015



Pennebaker, J. W., Booth, R., Boyd, R., & Francis, M. Linguistic Inquiry and Word Count: LIWC2015. 2015. Austin, TX: Pennebaker Conglomerates (www. LIWC. net).

Emotion Corpus: MSP-PODCAST version 1.9

Audio sentences:

- Train set: 55,283
- Validation set: 9,546
- Test set: 16,570

Emotional Annotations:

- Crowdsourcing platform: **Amazon Mechanical Turk**
- Every sentence has more than **5 annotators**
- 8-class Primary emotion (**P**) (**Single-choice**):
  - *anger, sadness, happiness, surprise, fear, disgust, contempt, neutral, and* ***other***
- 16-class Secondary emotion (**S**) (**Multi-choice**):
  - **Primary emotions**
  - *amusement, frustration, depression, concern, disappointment, excitement, confusion, and annoyance and* ***other***

R. Lotfian and C. Busso, "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings," IEEE Trans. Affect. Comput., vol. 10, no. 4, pp. 471-483, October-December 2019.

Polarity Label Processing

Primary emotion (P):
(**W1**) *Other(Excited)*, (**W2**) *Happy*, (**W3**) *Other(Pleased)*, (**W4**) *Neutral*, (**W5**) *Angry*
Secondary emotion (S):
(**W1**) *Other(Excited)*, (**W2**) *Happy*, (**W3**) *Other(Pleased)*, (**W4**) *Neutral*, (**W5**) *Excitement, Other(interesteede, CURIOUSITY, EnERgetic), Neutral*

Step 1: Pre-processing

- Lowercase and spell correction
- Check if secondary emotions (S) includes primary emotions (P) based on the rater-level

P: *Other(excited), happy, Other(pleased), neutral, angry*

S: *Other(excited), happy, Other(pleased), neutral, + angry, Excitement, Other(interested, curiosity, energetic), neutral*

Step 2: Check variants of options

- Check if typed emotions are variants of list of emotions

S: *Other(excited)* → *excitement*, *happy, Other(pleased), neutral, angry, excitement, Other(interested, curiosity, energetic), neutral*

Step 3: Classify polarity of emotional terms

- Linguistic Inquiry and Word Count (LIWC)
- **Ambiguous emotion**: LIWC does not provide a class

Positive emotion: *happy, pleased, interested, curiosity, energetic, excitement, excitement*
Ambiguous emotion: *neutral, neutral*
Negative emotion: *angry*

Step 4: Generate the final polarity label (Po)

Po = (Neg, Amb, Pos) = (0.1,0.2,0.7)

Example of Sentiment Words in the MSP-Podcast Corpus

UT Dallas MSP
Multimodal Signal Processing Laboratory

Negative Emotion Words | **Ambiguous Emotion Words** | **Positive Emotion Words**

Anxious, Argue, Aggression

Trust, Joyful, Relax

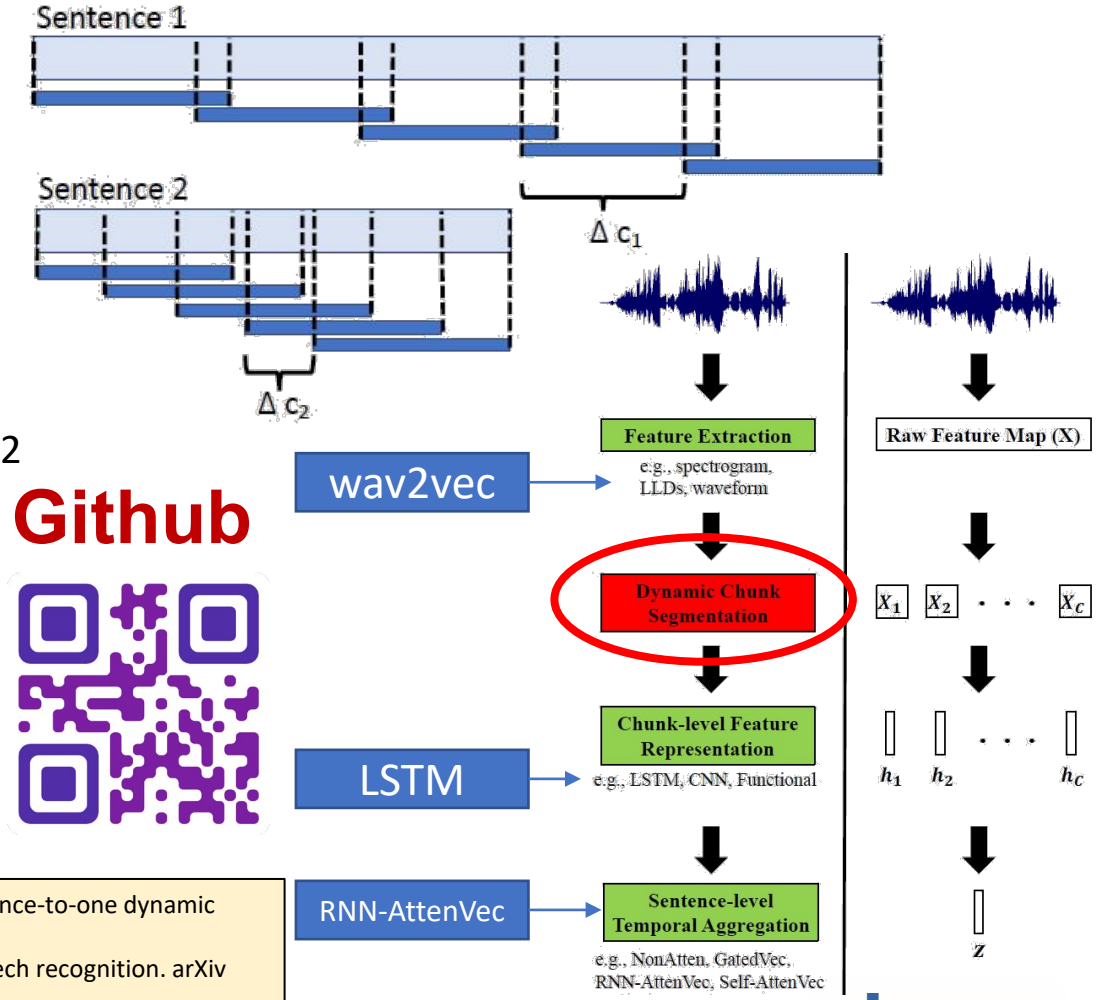
Experiment Setup (Model and Feature)

Speech Emotion Classification (SEC) Model:

- Chunk-level SER model with the RNN-AttenVec chunk-level attention[1]
- Same hyperparameters as the original paper[1]

Acoustic feature extraction:

- Extract 512-dimensional wav2vec feature vector[2] inspired by the analysis of Keesing et al. [2021][3]
- Features are z-normalized:
  - The parameters for the mean and standard deviation are estimated from the **train set**



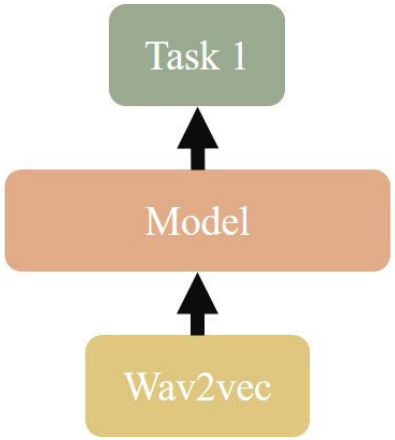
Github

[1]Lin, W. C., & Busso, C. (2021). Chunk-level speech emotion recognition: A general framework of sequence-to-one dynamic temporal modeling. IEEE Transactions on Affective Computing.
[2]Schneider, S., Baevski, A., Collobert, R., & Auli, M. (2019). wav2vec: Unsupervised pre-training for speech recognition. arXiv preprint arXiv:1904.05862.
[3]Keesing, A., Koh, Y. S., & Witbrock, M. (2021, August). Acoustic Features and Neural Representations for Categorical Emotion Recognition from Speech. In Proceedings of the 22nd Annual Conference of the International Speech Communication Association, Brno, Czech Republic (pp. 3415-3419).

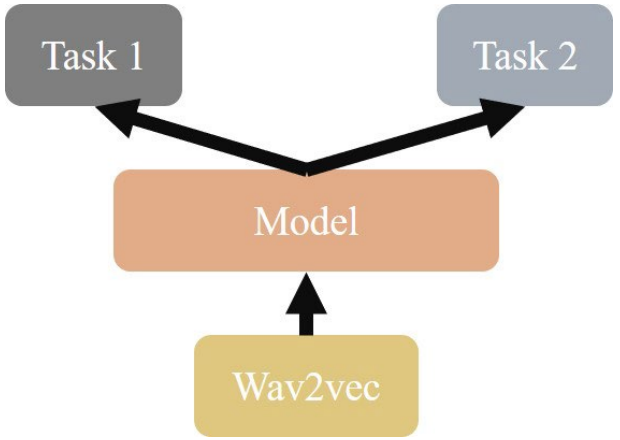
Experiment Setup (Multi-task SER)

Goal: Investigate the benefits of the proposed polarity label in the predictions of primary or secondary emotions

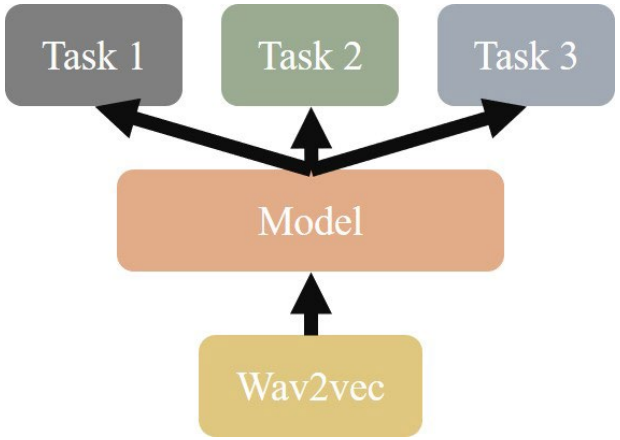
- **Single-task: Primary emotion (P), Secondary emotion (S), Polarity label (Po)**
- **Multi-task: (P+Po), (S+Po), (S+P), (S+P+Po)**



One single-task



Two multi-tasks

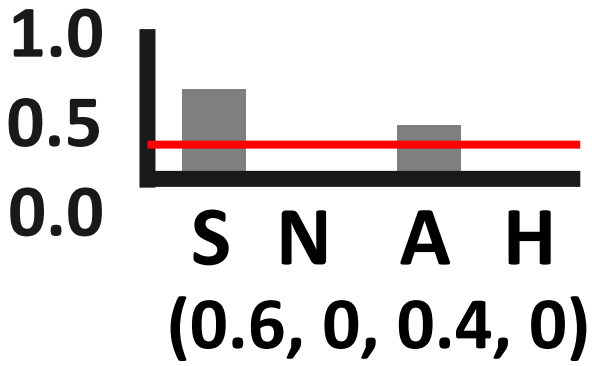


Three multi-tasks

Learning Objective Function and Evaluation Metric

Objective functions (Loss):

- Cross-entropy (**CE**) (softmax)
- Binary cross-entropy (**BCE**) (sigmoid)
- Kullback–Leibler divergence (**KLD**) (softmax)



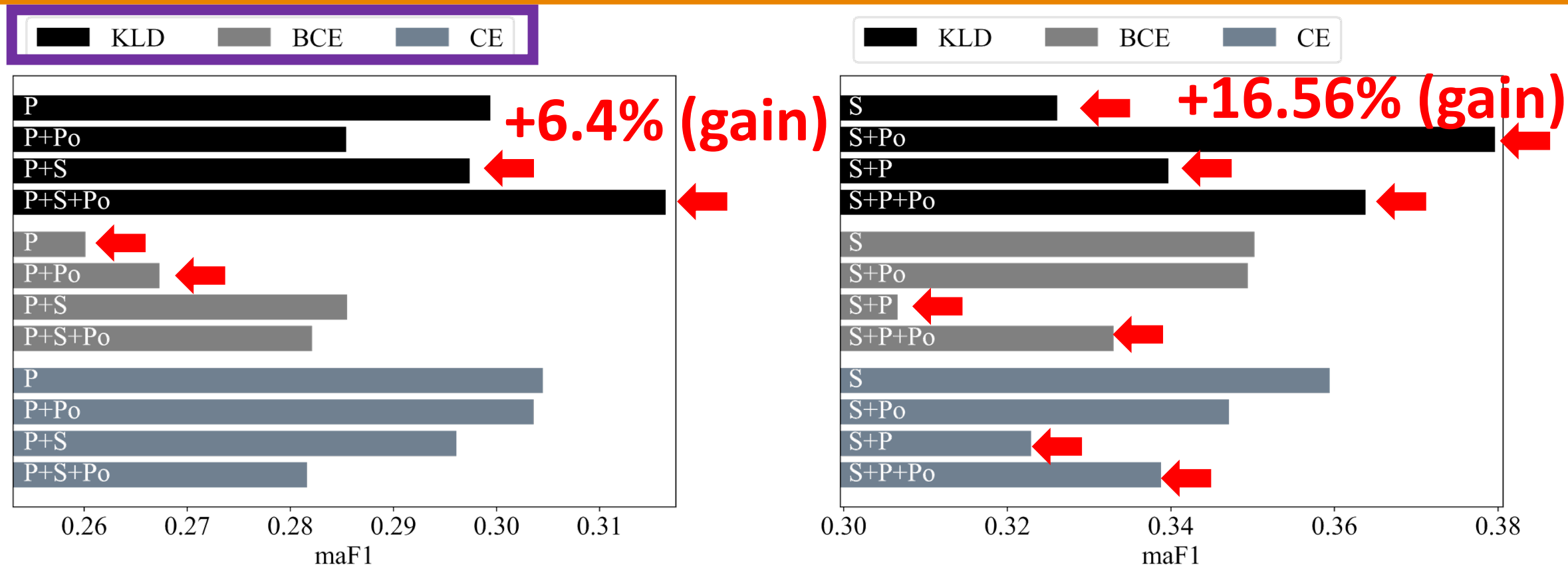
$\frac{1}{4}$

Sad, Angry

Evaluation metric:

- Macro F1-score (**maF1**)
  - Binarize threshold: 1/K, where K is the number of class in the classification task
    - **1/8** for primary emotion recognition task (**P**)
    - **1/16** for secondary emotion recognition task (**S**)

Visualization of Improvement for the Prediction of P and S



The macro-F1 scores for primary emotion recognition (P) **The macro-F1 scores for secondary emotion recognition (S)**

Po: polarity label
P: primary emotion label
S: secondary emotion label

Conclusion

Contribution:

- Utilize annotators' typed words of emotion perception to maximize the utilization of ratings for Speech Emotion Recognition (SER)

Method:

- Propose a 3D polarity label (positive/ambiguous/negative) to improve the prediction of primary and secondary emotion

Result:

- **8-class** Primary emotion classification: **+6.4%** performance gain
- **16-class** Secondary emotion classification: **+16.56%** performance gain

Findings:

- Typed words in the ``Other'' class have valuable information
- The SER task can be defined as a **multi-label task**

Thank You

110-2917-I-007-016
110-2221-E-007-067-MY3
110-2634-F-007-012

CNS-2016719

Paper Full Text

Contact **Huang-Cheng Chou**:
Email: hc.chou@gapp.nthu.edu.tw
LinkedIn: https://www.linkedin.com/in/huangchougchou/