Multimodal Signal Processing (MSP) lab
The University of Texas at Dallas
Erik Jonsson School of Engineering and Computer Science

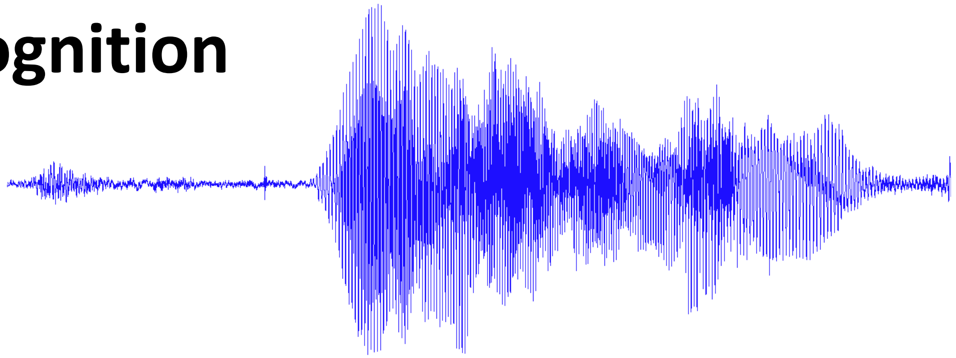
Predicting Categorical Emotions by Jointly Learning Primary and Secondary Emotions Through Multitask Learning

Reza Lotfian
Carlos Busso

THE UNIVERSITY OF TEXAS AT DALLAS

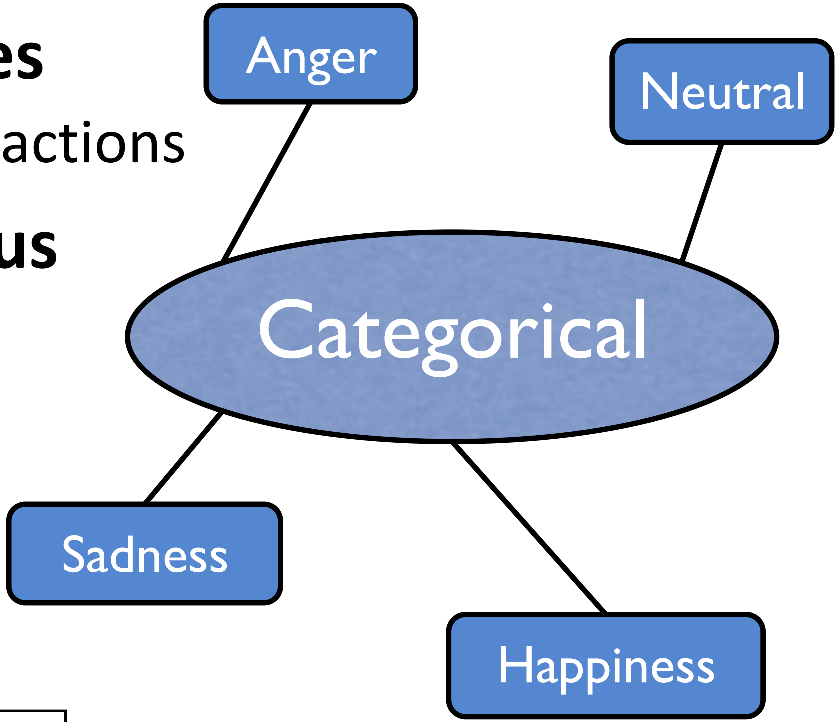
Introduction

- **Increasing interest in speech emotion recognition**
- **Emotion recognition from speech**
  - Call centers
  - Healthcare
  - Education
  - Entertainment
  - Creating emotions aware human computer interaction

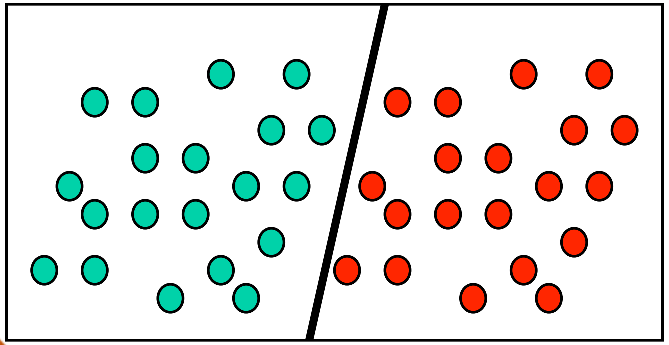


Motivation

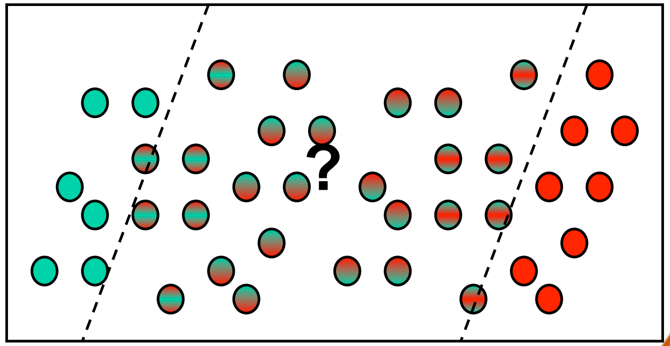
- **Interest in the recognition of discrete categories**
  - Useful in human-human and human-computer interactions
- **Spontaneous human interactions are ambiguous**
  - The boundary between categories are not clear
  - Difficult machine learning problem



Conventional machine learning problem



Emotion recognition



Motivation: Annotation of Emotions

- **Spontaneous corpora**
  - Emotions are not predetermined during recording
  - Need to be emotionally annotated
- **Emotional labels often come from perceptual evaluations from multiple evaluators**
  - Compensate for outlier and individual variations
- **Aggregating annotators' votes (consensus label)**
  - Majority vote

- Hap
- Hap
- Sur
- Ang
- Neu

Noise

Hap

Motivation: Annotation of Emotions

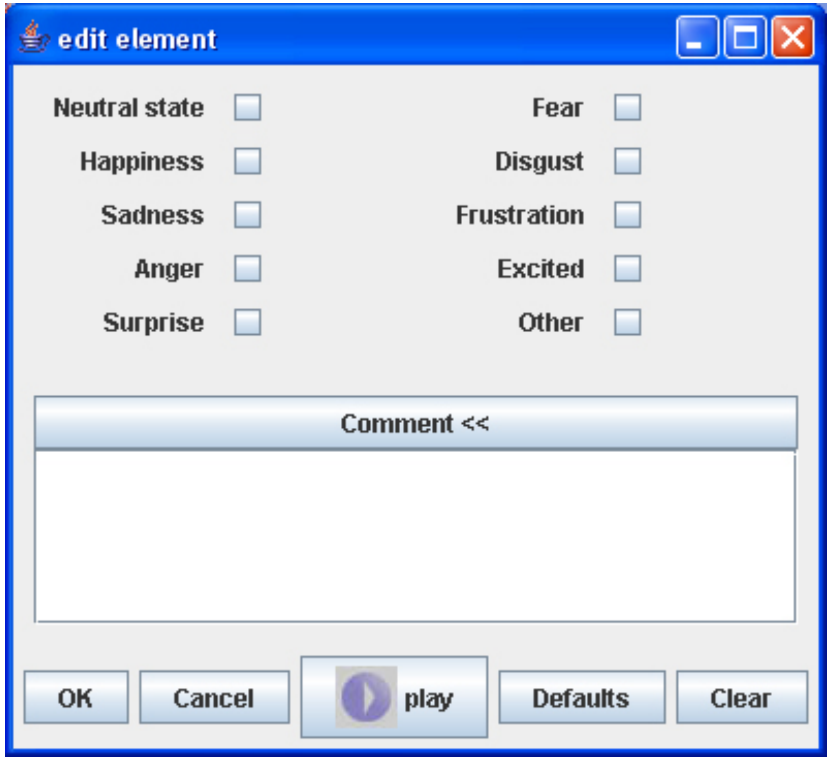
- **Evaluators often disagree on the perceived emotion**
  - Noise or information?
- **Assigning a single emotion per sentence oversimplifies the subjectivity in emotion perception**
  - More than one label can be relevant
- **Evaluator should identify as many emotions as they perceived**
  - Concept of major emotion versus minor emotion [Devillers et al., 2005]

Expression of Emotion

- **Mock subjective evaluation**

✦Sample 1: [fru; ()] [ang; ()] [neu; ()]

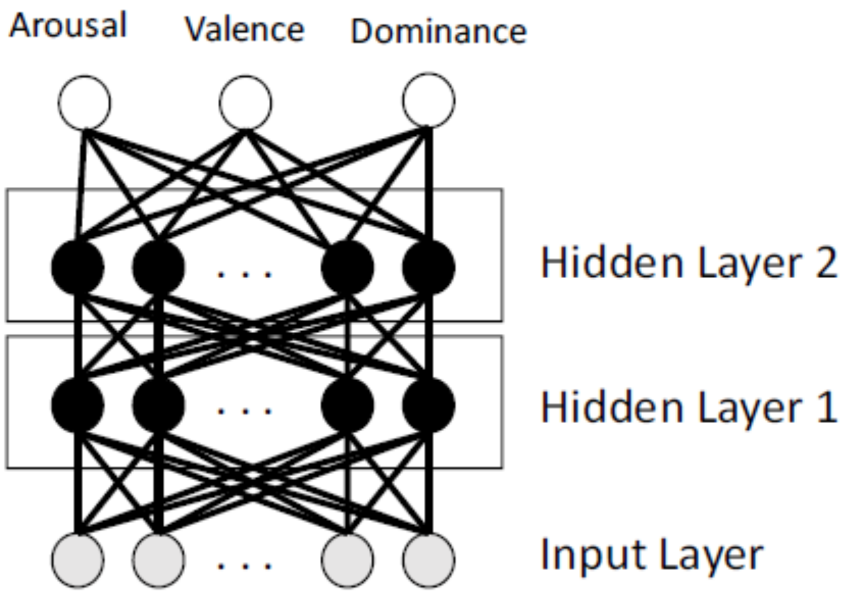
☑ Angry	☐ Sad	☐ Happy	☐ Amused	☑ Neutral
☑ Frustrated	☐ Depressed	☐ Surprise	☐ Concerned	
☑ Disgust	☑ Disappointed	☐ Excited	☐ Confused	
☐ Annoyed	☐ Fear	☐ Contempt	☐ Other	



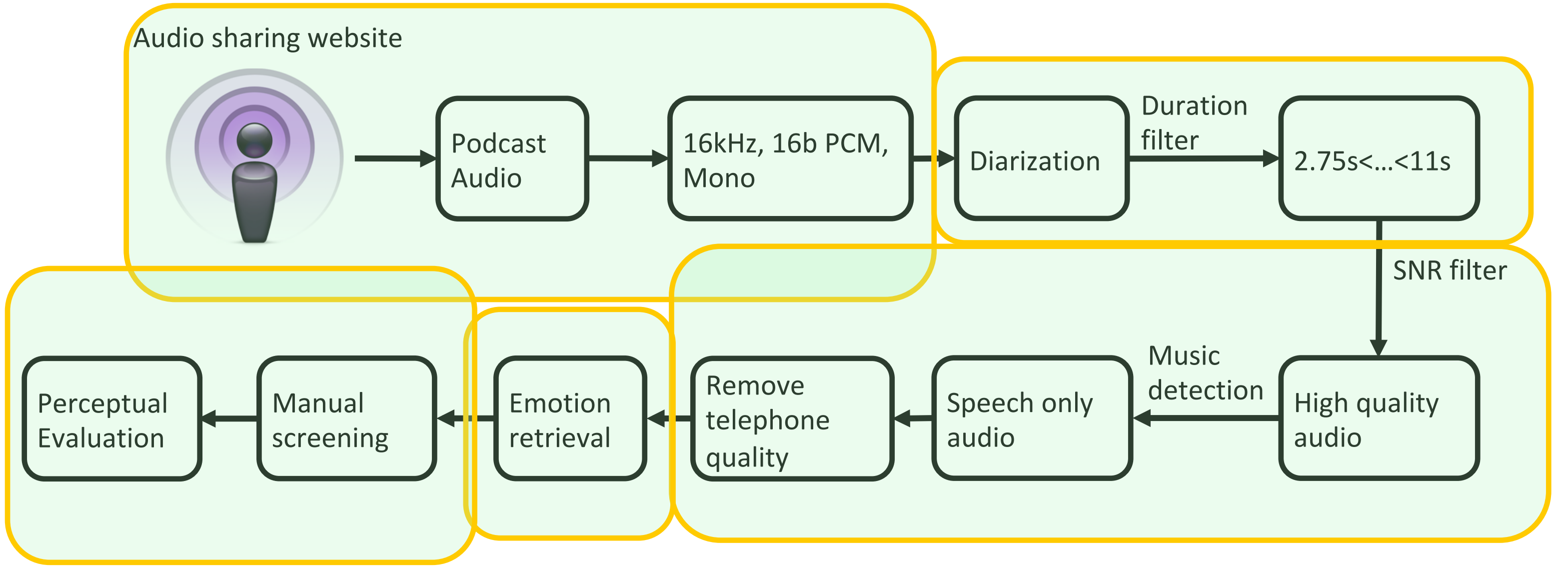
We hypothesize that secondary emotions provide useful information, even to predict the dominant emotion

Related Work

- **Better use of emotional annotations collected from multiple raters**
  - Consider the disagreement between multiple annotators as measure of difficulty [Lotfian and Busso, 2018a]
  - Soft label: instead of 1-hot ground truth [Fayek et al., 2016]
  - Ensembles: Train multiple classifiers, aggregate outcomes [Lotfian and Busso, 2018b]
- **Multitask learning in emotion recognition**
  - Use of multiple emotional attributes (arousal, valence, dominance) [Parthasarathy and Busso, 2017]
  - Gender and emotion [Ververidis 2004, Vogt 2006]
  - Attributes and emotional classes [Xia & Liu, 2016]



MSP-Podcast corpus



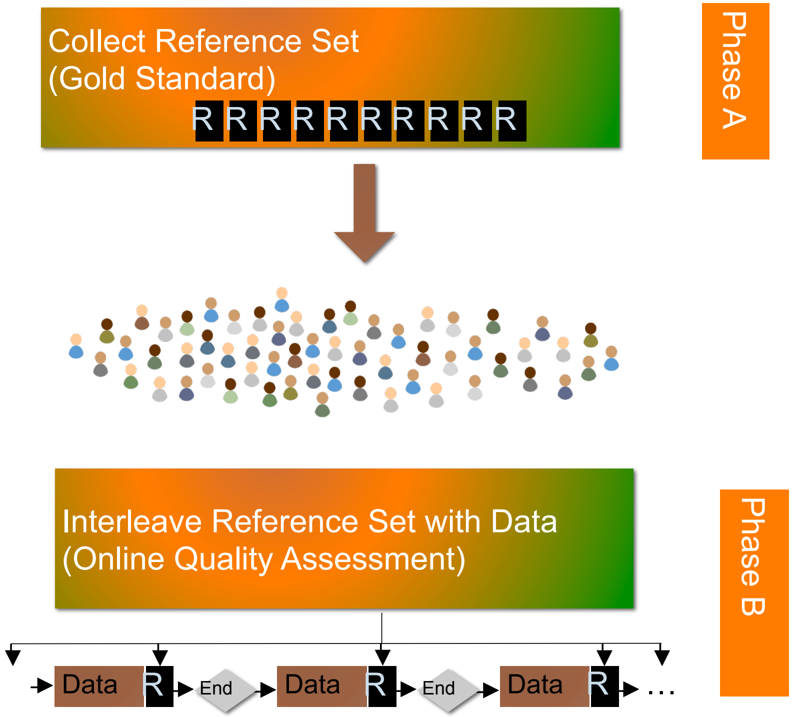
- **Collection of audio recordings[1] (Podcasts)**
  - Naturalness and the diversity of emotions
  - Creative Commons copyright licenses
  - Duration between 2.75s – 11s
  - Perceptive evaluation of emotional content

[1] Reza Lotfian and Carlos Busso, "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings," IEEE Transactions on Affective Computing

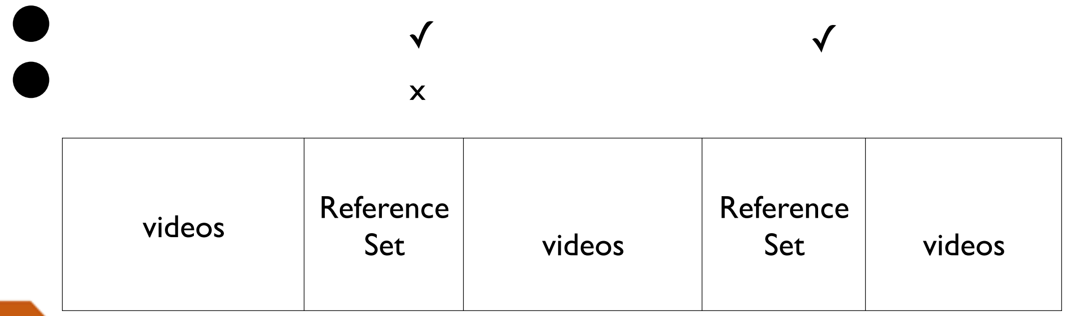
MSP-Podcast corpus

- **Study uses version V1.1 (22,630 sentences – 39hrs,12min)**
  - Test: 7,181 sentences (50 speakers)
  - Development: 2,614 (20 speakers)
  - Train: 12,835 (rest of the speakers)
- **Evaluated through Amazon Mechanical Turk**
  - At least 5 evaluations per sentence

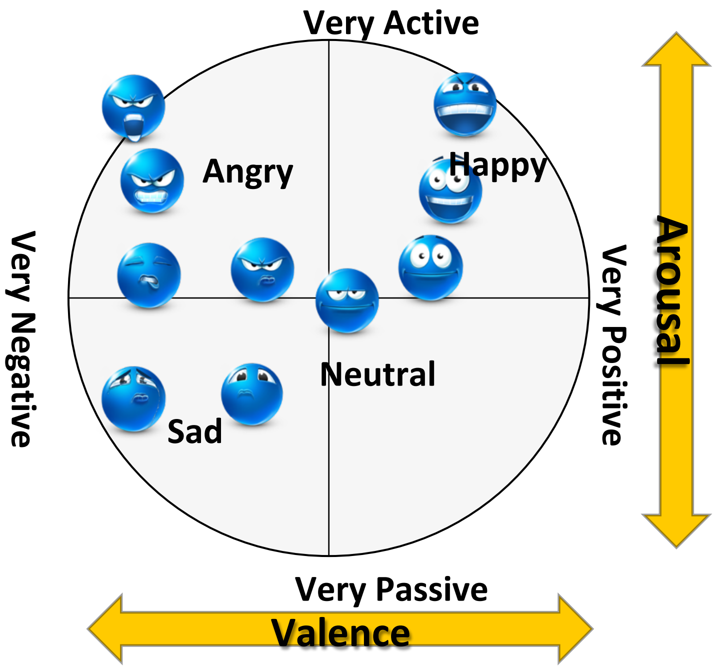
Collect reference set



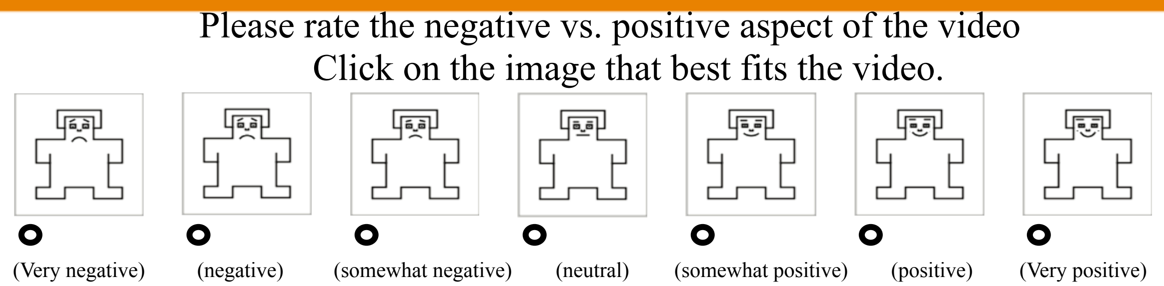
Trace performance in real time



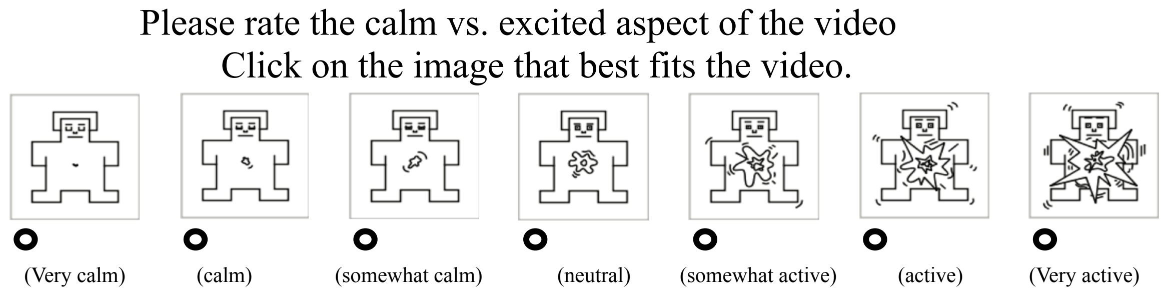
MSP-Podcast corpus



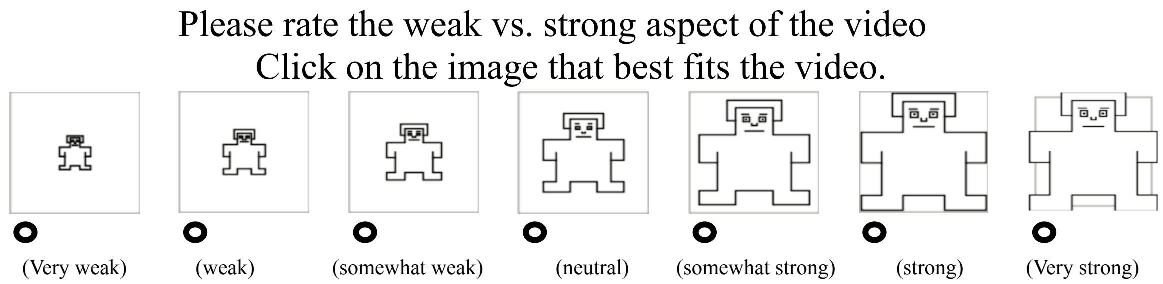
Valence



Arousal



Dominance



MSP-Podcast corpus

Primary emotion

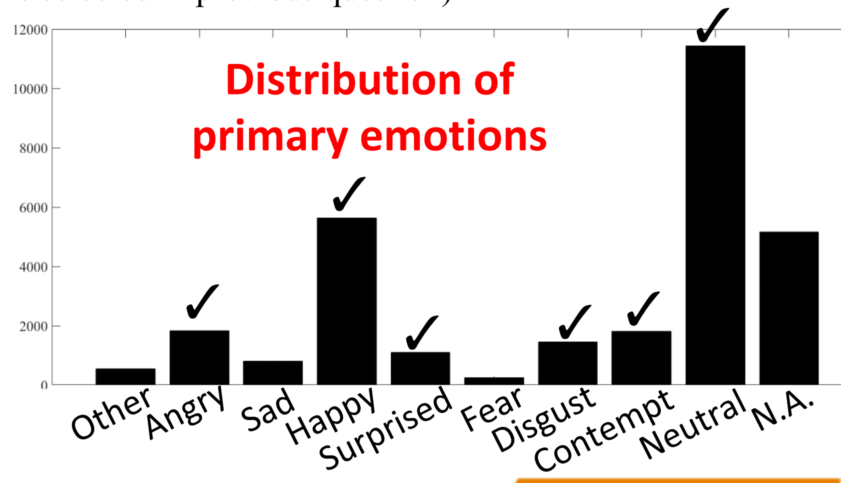
Is any of these emotions the primary emotion in the audio? If not, select **Other** and specify the emotion.

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

Secondary emotion

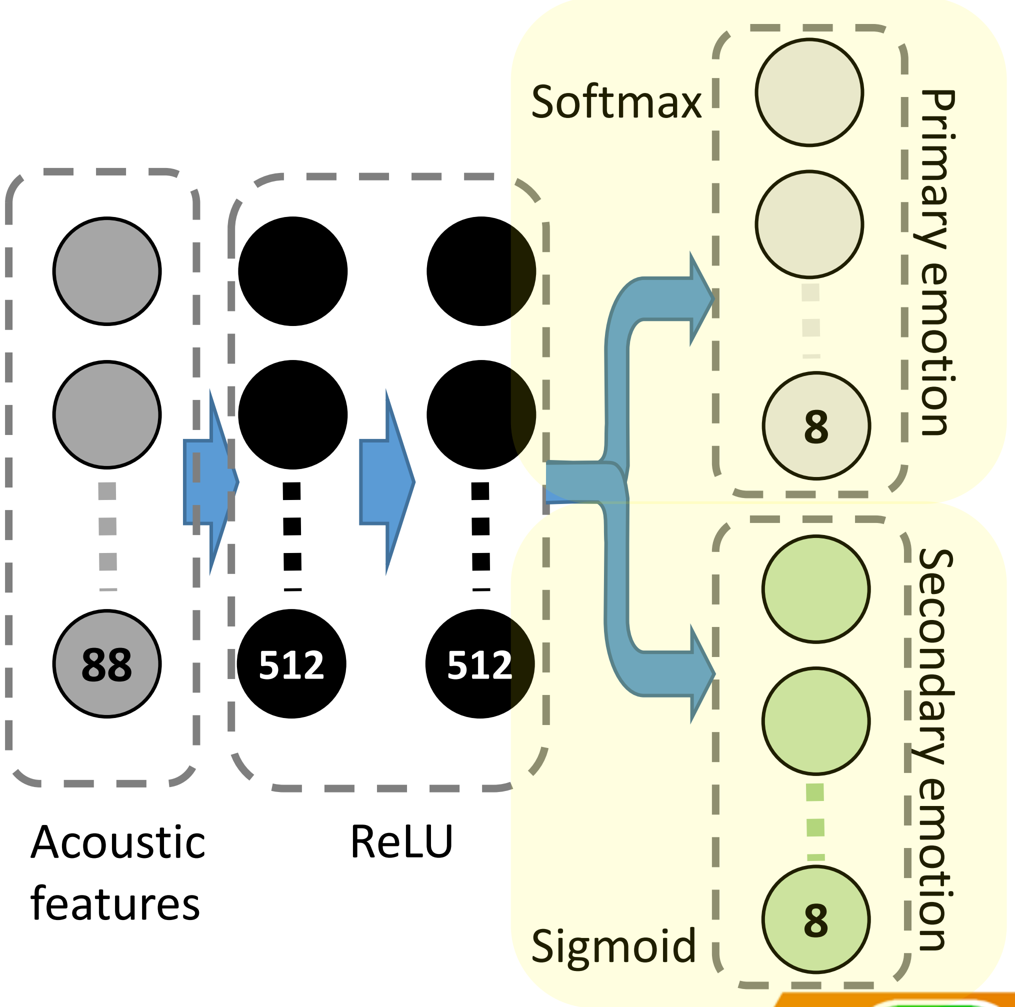
Please pick all the emotional classes that you perceived in the audio(Include the primary emotions selected in previous question)

☐ Angry	☐ Sad	☐ Happy	☐ Amused	☐ Neutral
☐ Frustrated	☐ Depressed	☐ Surprise	☐ Concerned	
☐ Disgust	☐ Disappointed	☐ Excited	☐ Confused	
☐ Annoyed	☐ Fear	☐ Contempt	☐ Other ☐	



Multitask Learning Network

- **Learning two different tasks**
  - Primary emotion
  - Secondary emotion
- **Two shared layers**
- **Primary emotion**
  - Eight class problem
  - Softmax layer with cross-entropy function
- **Secondary emotion**
  - Find all the emotional categories that are relevant to the speaking turn
  - Distance between true and predicted classes
  - Kullback-Leibler divergence (KLD)

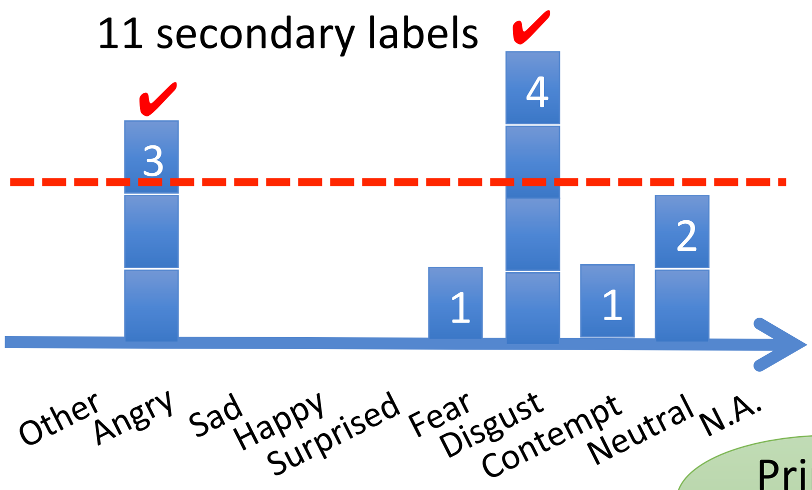


$$L_{ov} = (1-\alpha) \times L_{primary} + \alpha \times L_{secondary}$$

Labels for Secondary Emotions

- **Vector for secondary emotions**
  - We remove primary class
  - We remove the expanded list of emotions
    - Same 8 classes as primary emotion
  - *k* is the average number of secondary emotions for sentence *l*
  - A class is a secondary emotion if its votes are more than *k*
  - Add primary emotion

Example: Primary emotion: sadness
5 evaluators
11 secondary labels

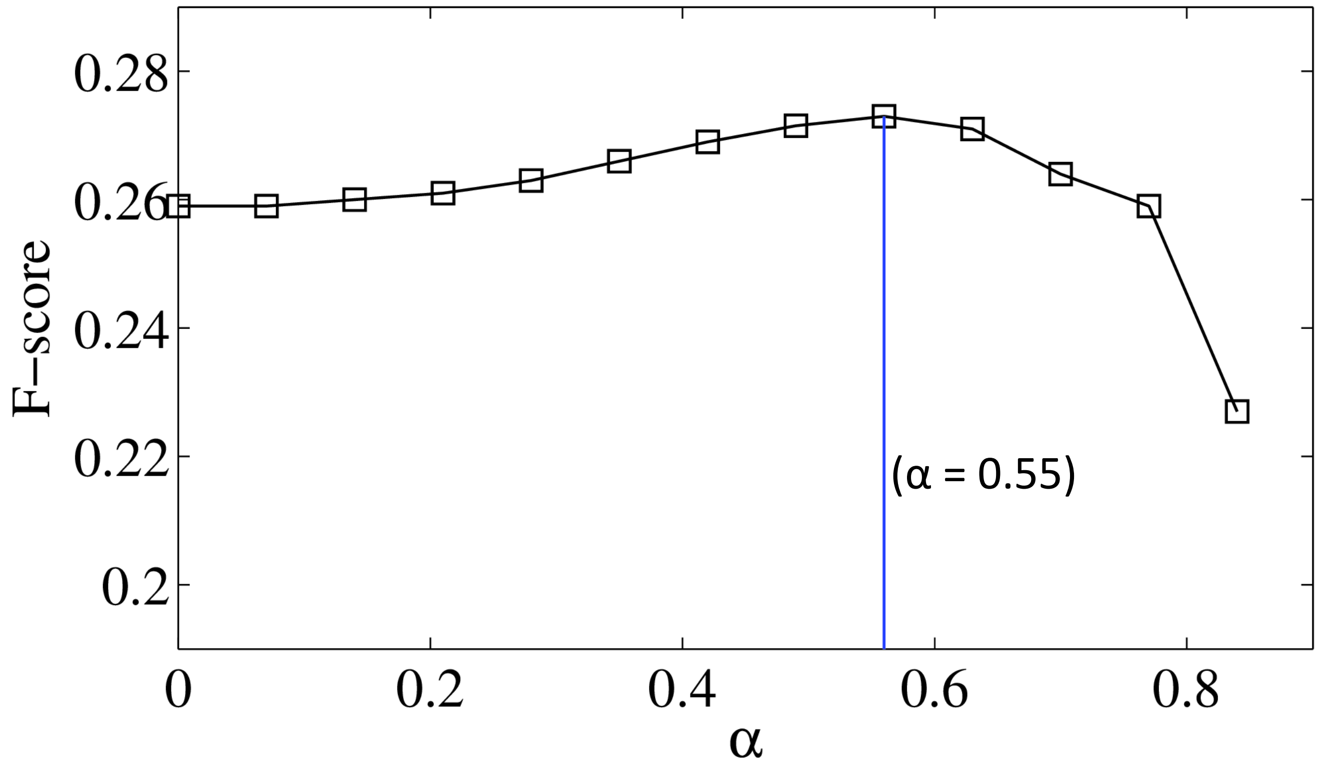


$$\begin{bmatrix} \text{Anger} \\ \text{Sadness} \\ \text{Happiness} \\ \text{Surprise} \\ \text{Fear} \\ \text{Disgust} \\ \text{Contempt} \\ \text{Neutral} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ 0 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

Primary emotion

Experimental Results

- **Acoustic features**
  - eGeMAPS set [Eyben et al., 2016]
  - 88 acoustic features
- **Hyperparameter optimization:**
  - Tradeoff between primary and secondary task in cost function (α)
- **Parameter optimization on development set**
  - More weight to secondary emotion
  - F-score of primary emotion classification increases by including secondary emotion in training



$$L_{ov} = (1 - \alpha) \times L_{primary} + \alpha \times L_{secondary}$$

Baselines

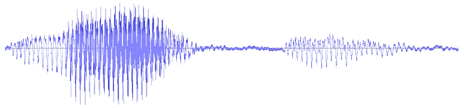
- **Hard label primary emotion (Hard label PE)**

- Hap
- Hap
- Sur
- Ang
- Neu

$\rightarrow$ Hap

Majority vote

Softmax (8 nodes)
Fully connected (512 nodes)
Fully connected (512 nodes)



- **Soft label derived from primary emotion (Soft label PE)** **[Fayek et al.2016]**

Sentence A { Hap, Hap, Sur, Ang, Neu } →

$$\begin{bmatrix} \text{Anger} \\ \text{Sadness} \\ \text{Happiness} \\ \text{Surprise} \\ \text{Fear} \\ \text{Disgust} \\ \text{Contempt} \\ \text{Neutral} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.0 \\ 0.4 \\ 0.2 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.2 \end{bmatrix}$$

Results: Cross-entropy loss

- **Average cross-entropy loss on test set**
  - Use of auxiliary task helps reducing the cross-entropy loss
  - Considering secondary emotions lead to better generalization

	Cross-entropy Loss
Hard label PE	1.391
Soft label PE	1.350
MTL (PE+SE)	**1.339**

Results: Primary Emotions

- **Detecting primary emotion**
  - Human performance is only F1-score=38.9
    - Compare labels from one rater with consensus labels from rest of the raters
    - Difficult task (spontaneous speech)
  - Chances performance is 12.5%
  - Proposed approach achieves 2.3% absolute improvements (9.6% relative gain)

	Precision	Recall	F1-score
Hard label PE	23.1%	24.9%	24.4
Soft label PE	24.9%	25.8%	25.3*
MTL (PE+SE)	26.4%	26.1%	26.3**
Human performance	40.8%	37.2%	38.9

(*) approach outperforms the Hard-label PE baseline
(**) approach outperforms other alternative methods

Results: Secondary emotions

- **Results on detecting secondary emotions**
  - MTL framework is optimized to maximize the classification performance of the primary task
  - Binary classification tasks
    - Does the sentence convey the detected emotional class?
    - multiple emotions are possible
  - Baseline: single-task learning that recognizes secondary emotions (*Hard label SE*)
  - Proposed method outperforms baseline by 5.1%

	Accuracy
Hard label SE	61.7%
MTL (PE+SE)	66.8%*

(*) approach outperforms the Hard-label SE baseline

Shared representation learned by MTL model is discriminative for both tasks

Final Remarks

- **Categorical emotions are more convenient but prototypical classes can be ambiguous**
- **Secondary emotion labels convey complementary and useful information that a classifier should leverage**
- **Multitask (Primary + Secondary emotion) improves the classification performance**
  - Efficient framework to leverage annotation of secondary labels
- **Future directions**
  - Attribute based emotions (arousal-valence) as auxiliary task
  - Investigate the optimum criteria to accept a class as a secondary emotion

Questions?

This work was funded by NSF CAREER award IIS-1453781

[MSP Lab UT Dallas]