

SURVEY OF AFFECT REPRESENTATION SCHEMES FOR SPEECH EMOTION RECOGNITION

ADITI RAWAT

a.rawat-4@student.tudelft.nl

1. BACKGROUND

- Affect refers to feelings experienced; subjective in nature
- Automatic affect prediction systems have underlying **dimensional** or **categorical** affect representation schemes (ARS)
- Emotionally-receptive Alexa or Siri
- Human-Robot Interaction
- Market research
- Monitoring mental health

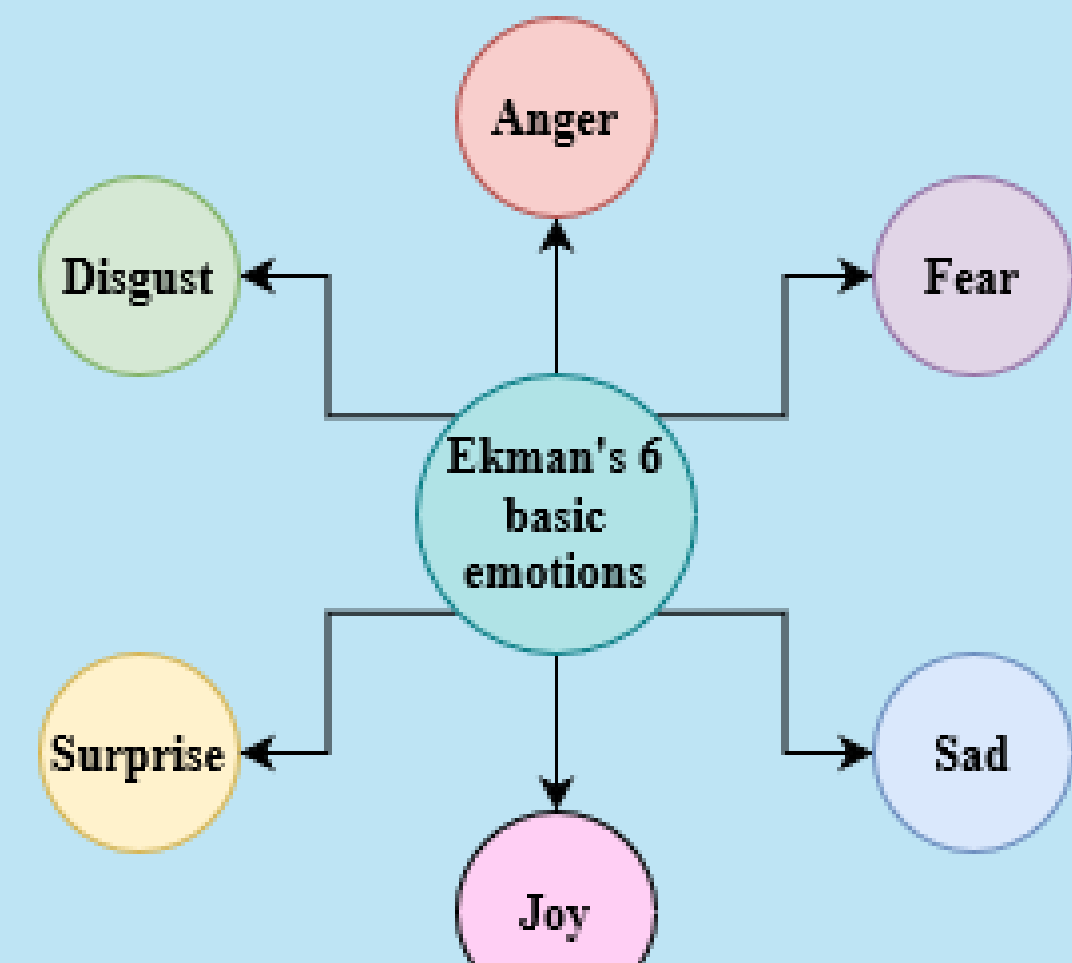


Fig. 1: Ekman's 6 basic emotions [1]

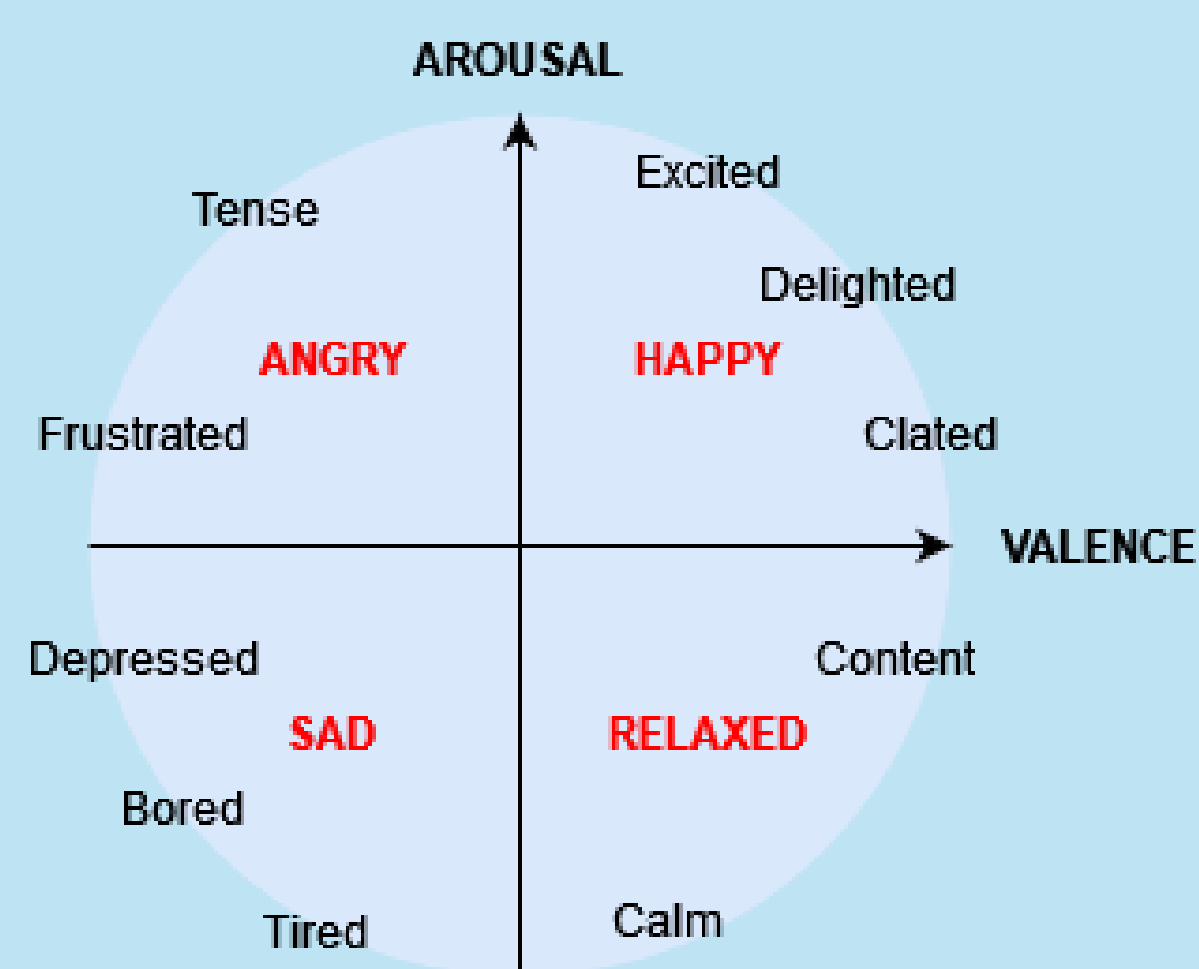


Fig. 2: Russell's Valence and Arousal ARS [2]

2. RESEARCH QUESTION

How are different affect representation schemes utilised in the context of speech emotion recognition?

- Targeted types of affective states
- Motivations for different ARS
- ARS used in combinations
- Popularity of ARS overall and over time
- Psychological basis for ARS

3. METHODOLOGY

- PRISMA guidelines for systematic review [3]
- Databases: **Scopus & Web of Science**
- Included:** Automatic affect prediction via speech
- Excluded:** Multimodal, Non-English papers, Reviews
- Benchmarking datasets for feasibility

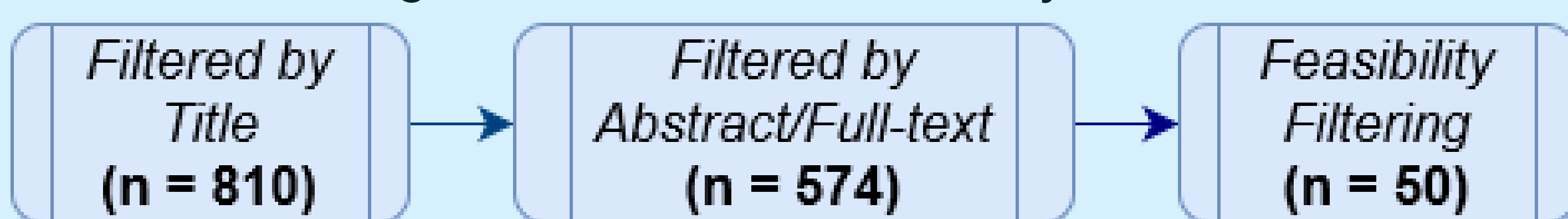


Fig. 3: Selection of papers

- 50 papers selected (10 each from years 2019-2023)
- 15 supplementary dataset papers

4. RESULTS

Type of Affective State Targeted	Corresponding ARS
Emotion	All except D.5f, D.5a
Emotion-related state / Affect / Behaviour	D.5f
Behaviour	D.5a
Attitude	D.7r, D.6t

Table 1: Types of affective states targeted

Behaviors: aggressive, cheerful, intoxicate, nervous, tired
Emotion-related states: anger, emphatic, positive, rest

MOTIVATION

- Context-specific:**
 - Labels based on interaction with robot
 - drop happiness (COVID-related)
- Database-specific:**
 - to be consistent with other datasets
 - convenient annotation of dataset (motivation for categorical schemes)
 - imbalance in dataset
- Better discriminability**
- Categories being too limited**
- Not enough positively-valenced emotions:** use of calm as baseline
- Time-specific:** dimensional ARS can recognize rise and fall in emotions

POPULARITY

- Categorical ARS much more popular
- Most popular ARS are also from most popular datasets
- C.6e:** Ekman's basic 6
- C.6ed:** EMO-DB (anger, boredom, disgust, fear, happiness, sadness)
- C.3i:** IEMOCAP (anger, happiness, sadness)

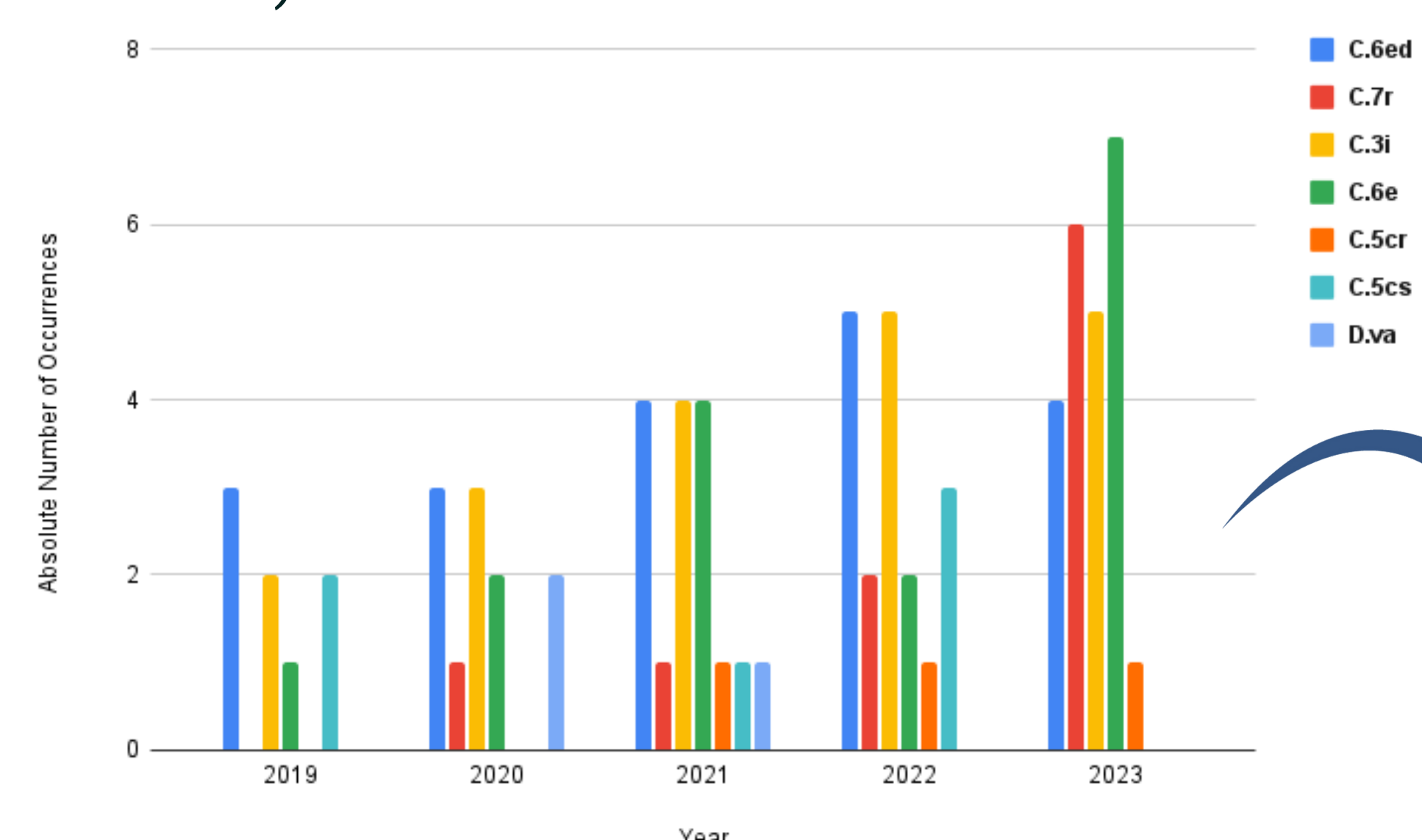


Fig. 5: ARS Popularity 2019-2023

ARS AND PSYCHOLOGY

- Ekman's basic 6 emotion models and modifications
- Russell's Valence and Arousal
- No psychological basis mentioned for behavior and attitude ARS

ARS IDENTIFIERS EXAMPLES

C.4i
C: Categorical
4: Number of emotions
i: Dataset Initial

D.vad
D: Dimensional
vad: valence, arousal, dominance

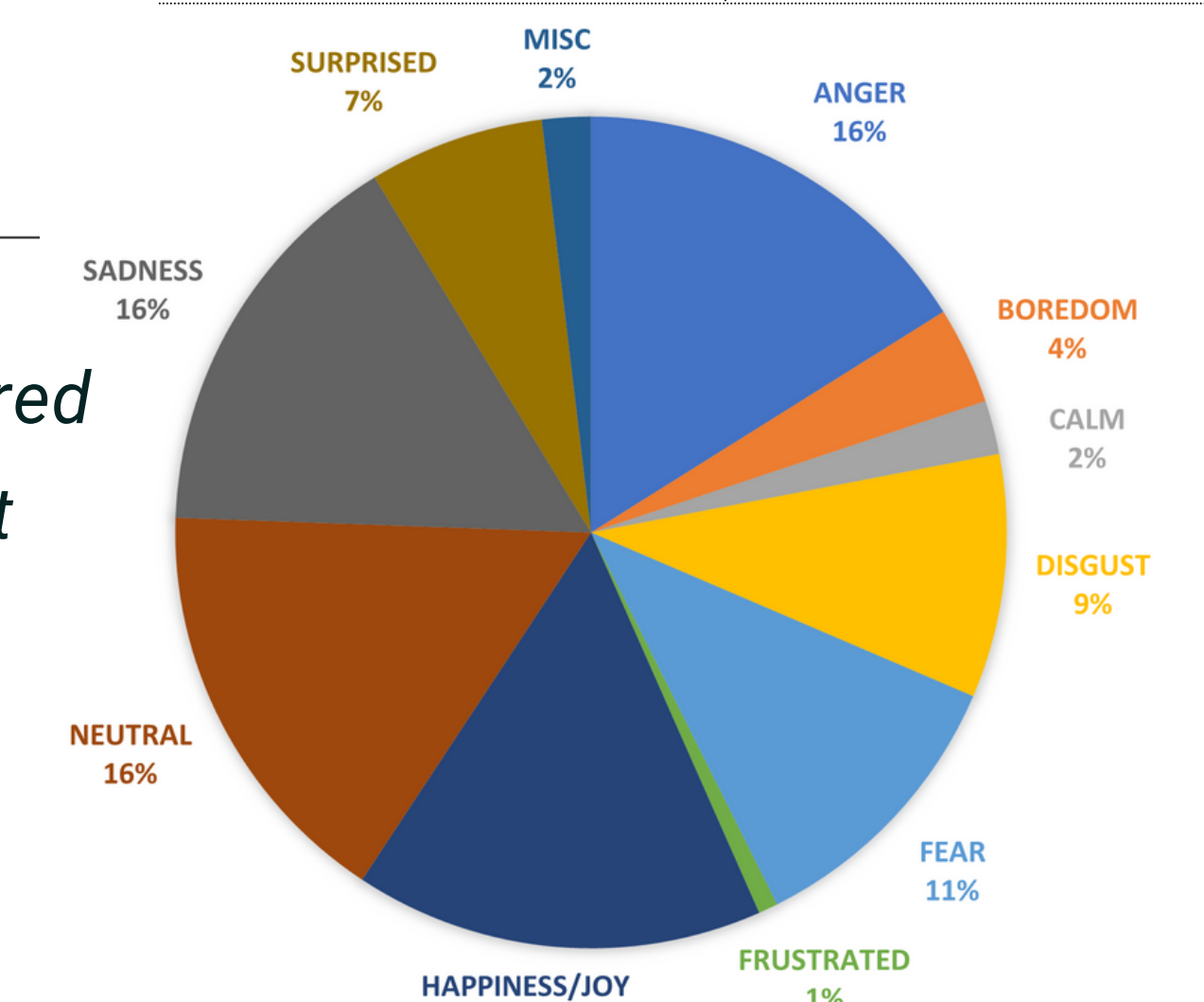


Fig. 3: Emotions targeted by categorical ARS

COMBINATIONS OF ARS

- 56% of the systems use multiple categorical schemes together
- 1 paper uses coarse to fine approach
 - division of valence-arousal space into quadrants
 - quadrants mapped to categories

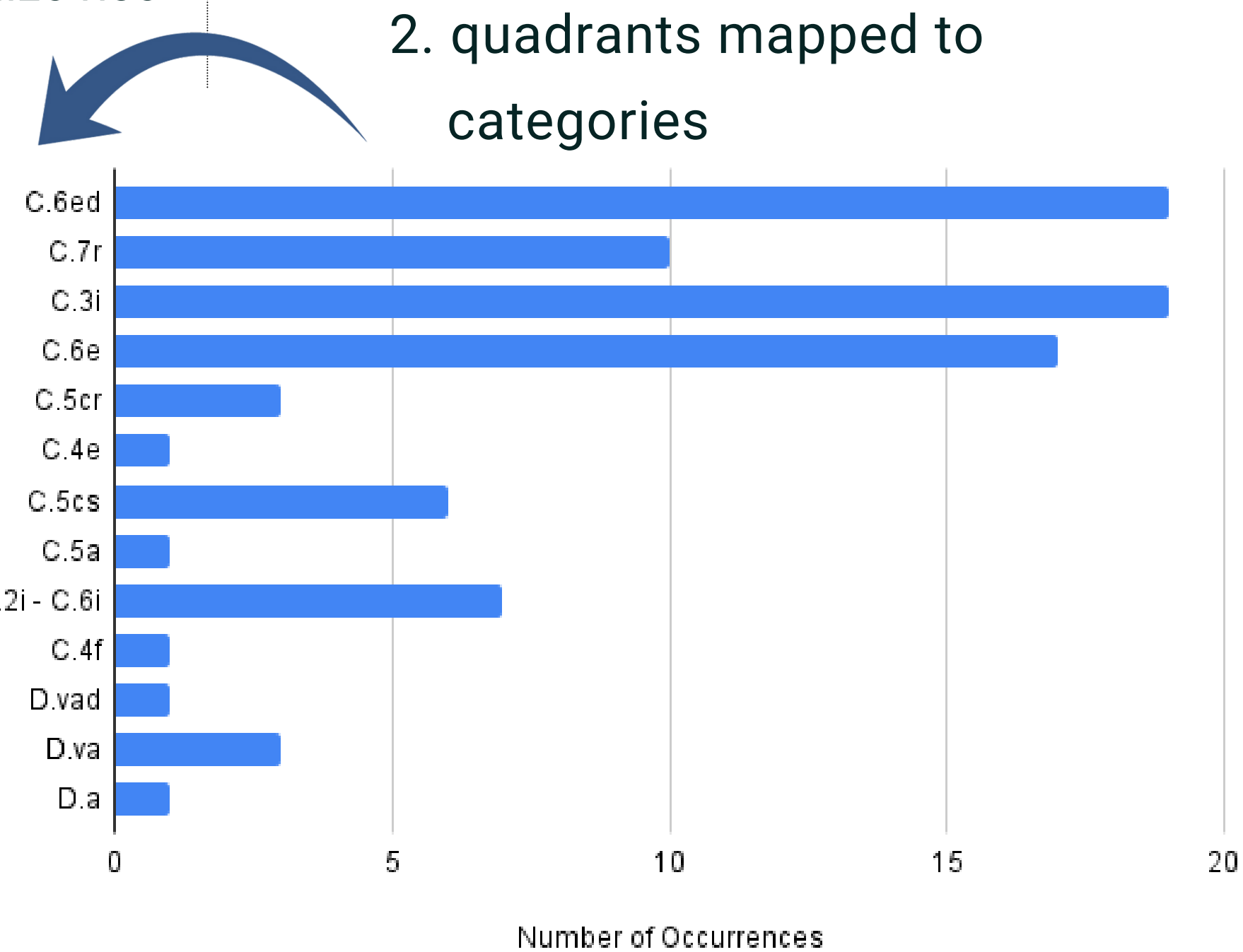


Fig. 4: ARS Popularity Overall

- C.6e:** Ekman's basic 6 (most popular in 2023) show steady growth
- C.3i:** IEMOCAP and **C.6ed:** EMO-DB grow in popularity with a slight fall in 2023
- C.7r:** RAVDESS dataset (Ekman's 6 + calm) rose in popularity significantly
- Dimensional ARS only in 2020 and 2021 (mostly valence & arousal)

5. DISCUSSION

- Papers themselves do not motivate ARS
- Supplementary papers are more informative
- Neutral as an emotion/behavior is controversial
- Confusion between types of affective states
- Too many ARS relative to number of papers
- Datasets provide opportunity to use dimensional with categorical, but systems do not use this

6. CONCLUSIONS

- Categorical schemes overwhelmingly more popular than dimensional (in combinations as well)
- Happiness, Anger, Sadness, Neutral most common
- Most ARS based on Ekman's but varying labels
- Popularity of ARS highly linked with that of dataset
- Practicality prioritized over psychological research

7. FUTURE EXTENSIONS

- More researchers; more representative sample
- Exploration into supplementary dataset papers
- Multimodal input (for example, audio-visual)
- Motivation behind lack of dimensional representation

REFERENCES

- [1] Paul Ekman and Wallace V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*. 17(2):124-129, 1971
- [2] Pekka Siirtola, Satu Tamminen, Gunjan Chandra, Anusha Ihalapathirana, and Juha Rönning. Predicting emotion with biosignals: A comparison of classification and regression models for estimating valence and arousal level using wearable sensors. *Sensors*, 23(3):1598, 2023.
- [3] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*. 88:105906, 021.