Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): Arousal and Valence Validation

Karla Kovacek[1], Steven Livingstone[2], Gurjit Singh[1,3], Frank A. Russo[1]

1. Department of Psychology, Ryerson University, Toronto, Canada, 2. Department of Computer Science, University of Otago, Dunedin, New Zealand, 3. Sonova, Stäfa, Switzerland

Background

- The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) has become a widely used tool in psychological and affective computing studies of discrete emotions.
- Discrete models, such as the basic emotion model (Ekman, 1992; Ekman, 1999), group emotions into distinct categories – the RAVDESS has been validated along discrete emotions of neutral, calm, happy, sad, angry, surprised, disgusted and fearful.
- In contrast, dimensional models organize emotions along continuums within an n-dimensional space. Russell's (1980) circumplex model of affect organizes emotions along a dimension of arousal (calm to excited), and valence (unpleasant to pleasant)
- The aim of the current study was to extend the use of the RAVDESS by validating the files along dimensions of arousal and valence.

Methods

Participants

- 24 young adults, 19-34 years old
- Normal hearing & vision, fluent in English (from age 5+)

Screening

- Subjective hearing and vision measures, emotional communication measure, demographics questionnaire

RAVDESS

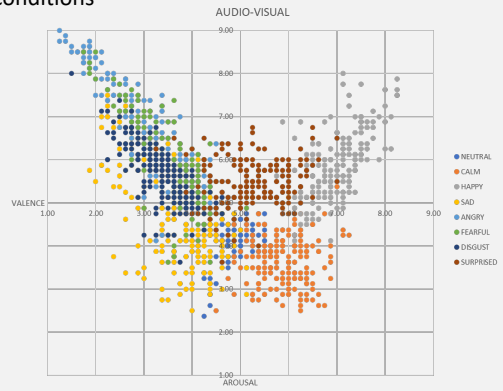
- 7,356 dynamic files: 4,330 speech files, 3,026 song files
- 24 professional actors (12 male, 12 female) speaking two lexically matched statements in North American accent
- Audio-only, visual-only, audio-visual
- Normal and strong emotional intensity levels

Procedure

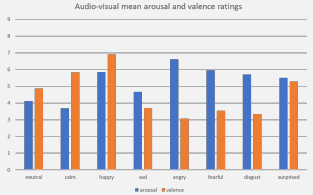
- Use provided equipment to rate 4,320 RAVDESS speech stimuli on dimensions of arousal & valence using two 9-point Self-Assessment Manikin Likert scales

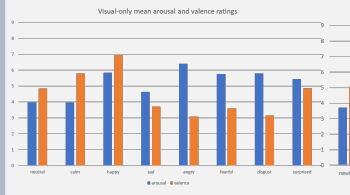
Results

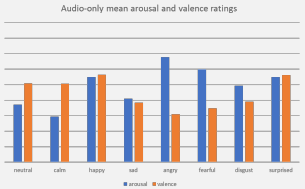
- Mean arousal and valence ratings under audio-only (n=21), visual-only (n=14), and audio-visual (n=8) conditions



- Trends in the pattern of arousal and valence ratings for stimuli drawn from different emotion categories was largely as expected
- Differentiation of stimuli was best in the audio-visual condition.







Discussion

Implications

- This validation increases the scope of uses for RAVDESS including better characterizing emotion perception in special populations. For example, there is reason to believe that hearing impairment restricts the range of dimensional ratings in audio-only conditions (Picou, 2016), while paradoxically expanding it in visual-only conditions (Glick & Sharma, 2017).
- This validation also opens up the possibility of building affective computing systems capable of characterizing speech emotion using discrete or dimensional models.

References & Acknowledgements

Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3-4), 169-200.

Ekman, P. (1999). Facial expressions. Handbook of cognition and emotion, 16(301), e320.

Glick, H., & Sharma, A. (2017). Cross-modal plasticity in developmental and age-related hearing loss: clinical implications. Hearing Research, 343, 191-201.

Picou E. M. (2016) How hearing loss and age affect emotional responses to nonspeech sounds. The Journal of Speech, Language, and Hearing Research 59: 1233–1246.

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161.