# CAN FACIAL EXPRESSIONS MEASURE LISTENING DIFFICULTY?

**Soumya Venkitakrishnan[1], Nicholas Giuliani[2], Yu-Hsiang Wu[1]**

[1]Department of Communication Sciences & Disorders, The University of Iowa; [2]University of Iowa Hospitals and Clinics

Hearing Aid and Aging Research Laboratory

## INTRODUCTION

- Various behavioral and physiological measures have been used in research settings to assess listening difficulty (e.g. the dual-task paradigm, subjective rating scales, pupillometry). However, all these measures have some drawbacks (as listed alongside).

| | |
|---|---|
| Questionnaires | * Recall Bias |
| Subjective Rating | * Reliability issues[1] |
| Pupillometry | * Controlled Situations[2] |
| Dual-Task Paradigm | *Complex task |
| Skin Conductance | *Questionable validity[3,4] |

- Our long-term goal is to determine the feasibility of using real-time facial-expression recognition algorithms to quantify listening difficulty. We selected facial expressions because: 1) It may be recorded more naturally & can be a passive task, 2) it may be recorded in real-world listening situations by the camera in mobile devices, while a person is filling out in-situ surveys (i.e. Ecological Momentary Assessment[5]), 3) If facial expression correlates with pupil response, then it could be used to measure listening effort with less constraints.
- The goal of the present study was to explore how listeners' facial expressions changed as a function of speech listening difficulty. We hypothesized that with increasing difficulty in speech listening, listeners would be more likely to generate facial expressions that reflect negative emotions such as confusion and frustration.
- Our secondary goal was to compare the results of facial expressions to pupil response which is known to increase with listening difficulty[6,7]. We hypothesized that both these responses would correlate because they are both based on the arousal of the autonomic nervous system.

## METHODS

- Participants: 20 adults, aged 22 to 37 (Mean = 27.45, SD = 4.92 ) with normal hearing.
- Stimuli used: Speech perception testing using IEEE[8] sentences.
- The facial expressions of individuals were recorded using a camera (Logitech HD Pro Webcam C920) and the Emotient FACET software (v6.3.6973.6; iMotions) at signal-to-noise ratios (SNRs) of -5, and -11 dB (representing a high or ~80% and low or ~25% speech perception accuracy), and in quiet.
- The pupil response was simultaneously recorded using Tobii Pro X2 screen-based eye tracker.
- The iMotions software assesses the movement of the different muscles of the face. It defines facial expressions as a combination of action units. These movements and the resulting facial expressions are interpreted based on the Facial Action Coding System[9]. The algorithm then computes the evidence level, which is the probability of the presence of a given facial expression. The emotions analyzed by the software include joy, anger, fear, contempt, frustration, sadness, confusion and negative emotions.

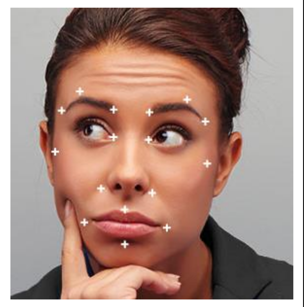
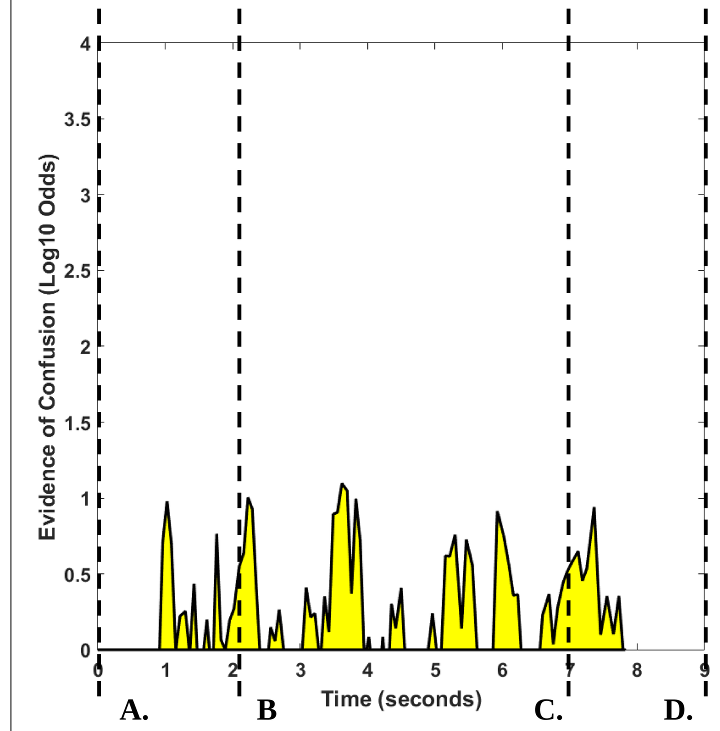
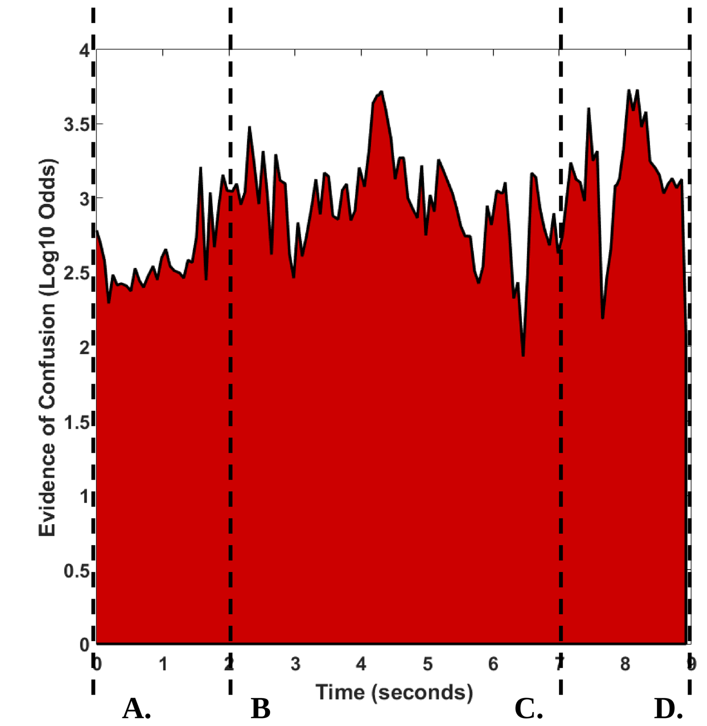


Figure 1 and 2. Action units and their use in identifying emotions


- Initially, each participant was asked to maintain a neutral expression, which became their baseline. We analyzed the expressions of confusion, frustration and negative emotions as these are seen when individuals encounter cognitive disequilibrium or gaps in knowledge[10,11].

## RESULTS




*A: Start of noise B: Start of sentence C: End of sentence D: Response cue*

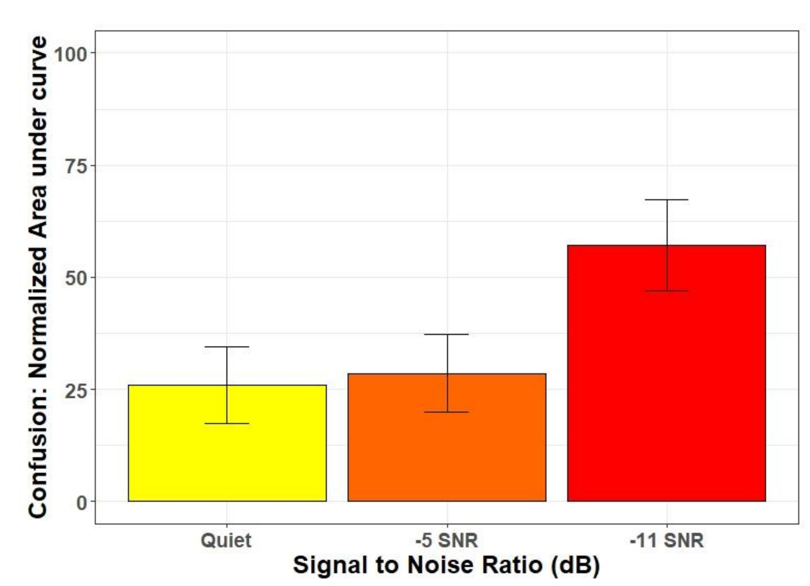
**Figure 3.** Time course of confusion during **quiet** condition for the subject 20

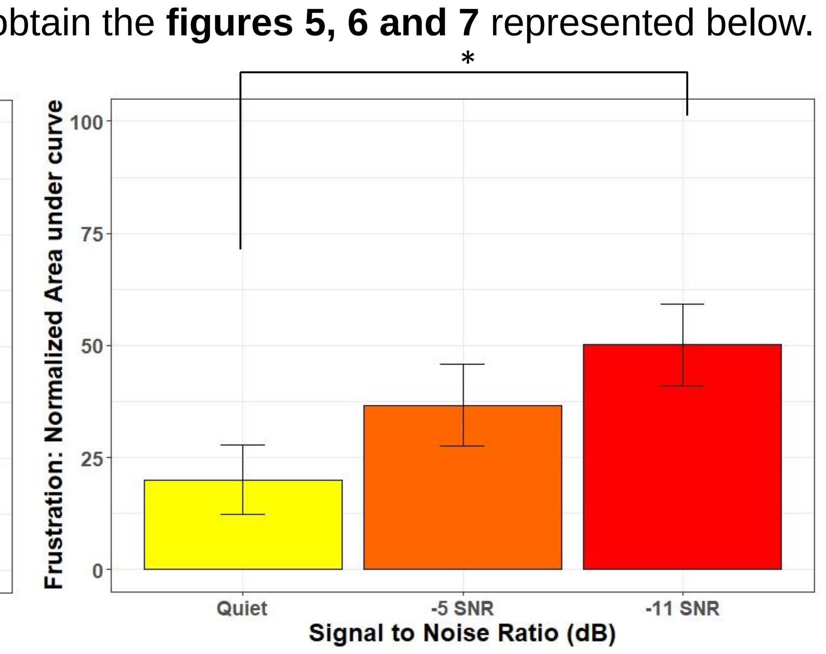
**Figure 4.** Time course of confusion at **SNR of -11** for the subject 20

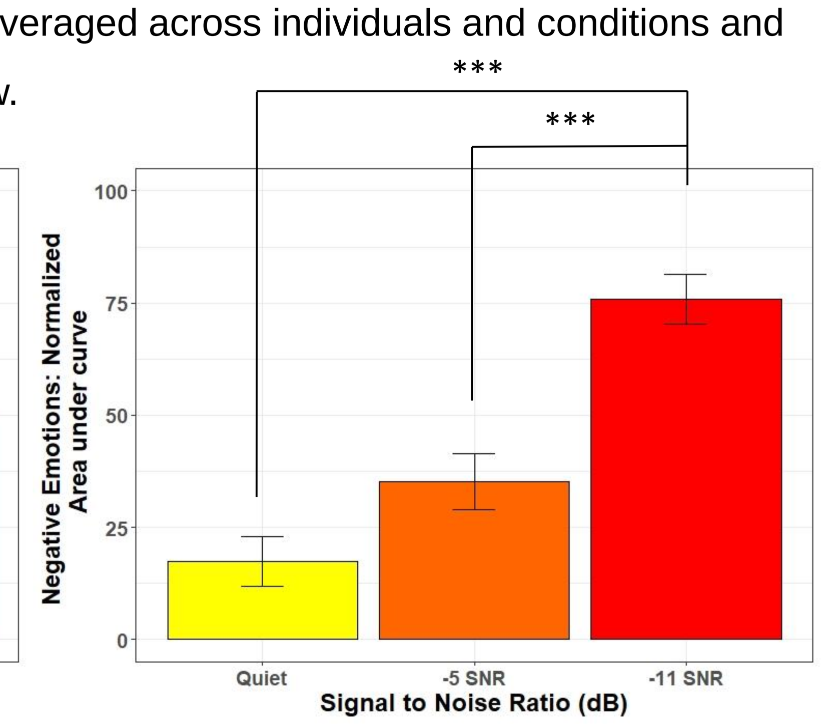
The **figure 3** and **4** represent the evidence level of the emotion 'confusion'. The evidence value represents the odds in logarithmic (base 10) units of an emotion being present. E.g. An evidence value of 1 for confusion indicates that the observed expression is 10 times more likely to be categorized by an expert human coder as confused than not confused, and a value of 2 indicates that the observed expression is 100 times more likely to be coded as confused than not confused by an expert human coder.

For analysis, the evidence levels obtained were first baseline corrected. Following this, the area under the curve was obtained for each sentence.

The peaks of these graphs (**Figure 3 and 4**) denote a higher probability of presence of the emotion. The positive area under the graph (integrated value) was obtained for each participant for different conditions. This was then averaged across individuals and conditions and then normalized to account for variability to obtain the **figures 5, 6 and 7** represented below.




*Significance key: *: p< 0.05 **: p<0.01 ***: p< 0.001*

**Figure 5.** The evidence for **confusion** at different signal-to-noise ratios with standard errors of mean

**Figure 6.** The evidence for **frustration** at different signal-to-noise ratios with standard errors of mean

**Figure 7.** The evidence for **negative** emotions at different signal-to-noise ratios with standard errors of mean

Linear mixed effects model with the fixed effect of SNR and random subject effect was used for analysis. With SNR as the independent variable and evidence level of emotions as the dependent variable, the main effect of SNRs was found to be non-significant for the emotion of confusion ($F(2,38)= 2.68$, $p=0.082$) and was significant for the emotions of frustration ($F(2,38)= 4.423$, $p=0.019$) and negative emotions ($F(2,38)= 26.820$, $p<0.0001$).

Pair-wise follow-up testing was conducted using Tukey test with adjustments for multiple comparisons. For frustration (**Figure 6**), we found a significant difference between the quiet and the -11 SNR condition ($p=0.015$). For negative emotions (**Figure 7**),we found a significant difference between quiet and -11 dB SNR condition and the -5 and the -11 dB SNR condition ($p<0.0001$).

Our secondary aim was also to compare the facial expression responses to an already established method of measuring listening effort, i.e. pupil responses. For this purpose, we used change in measures of facial expressions and pupil response between the easiest (Quiet) and the most difficult (-11 dB SNR) conditions. We calculated correlation using Spearman's correlation co-efficient as the change in facial expressions were not normally distributed. We saw a weak positive correlation between pupil responses and facial expressions of frustration ($r = 0.29$, $p=0.23$) and of negative emotions ($r = 0.28$, $p=0.25$: **Figure 8**). However, the correlations were not statistically significant.

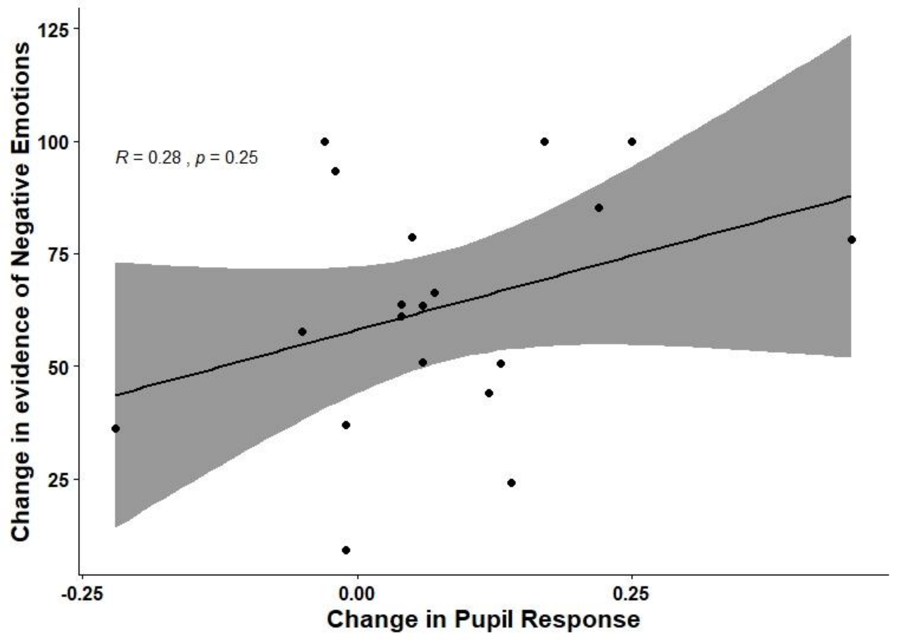



**Figure 8:** Correlation between change in facial expressions (of negative emotions) and pupil response between the easiest and the most difficult conditions.

## DISCUSSION

- Although there was no graded change in facial expression for each condition, the increasing trend of confusion, frustration and negative emotions suggests that the evidence level of these emotions increased monotonically as SNR decreased. These findings support the feasibility of using facial expression to assess listening difficulty, at least in controlled environments.
- The weak positive correlation between facial expressions and pupil response could be explored and indicates that facial expression shows promise in the area of measuring listening difficulty.

## CONCLUSIONS AND IMPLICATIONS

- Facial expressions could be explored further as an easier objective method of measuring listening effort.
- With advancements in technology, facial expressions may be recordable at a remote location along with self-reported survey results and help with tele-rehabilitation.
- This measure could also provide an unbiased method of measuring the interaction of human emotions, environment and hearing aids.

## ACKNOWLEDGEMENTS

National Institute on Disability, Independent Living and Rehabilitation Research (NIDILRR 90REGE0013).

## CONTACT

Contact soumya-venkitakrishnan@uiowa.edu for further information.

## REFERENCES

1. Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, *34*(5), 523-534.
2. Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*(4), 498-510.
3. Mackersie, C. L., & Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: Heart rate variability and skin conductance. *Ear and Hearing*, *37*, 118S-125S.
4. Mackersie, C. L., MacPhee, I. X., & Heldt, E. W. (2015). Effects of hearing loss on heart-rate variability and skin conductance measured during sentence recognition in noise. *Ear and hearing*, *36*(1), 145.
5. Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1-32.
6. Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277-284.
7. Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and hearing*, *31*(4), 480-490.
8. Institute of Electrical and Electronics Engineers (IEEE). (1969). IEEE recommended practice for speech quality measurements. IEEE Transactions on Audio and Electroacoustics AU-17, 225–246.
9. Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology*, *30*(1), 527-554.
10. Di Leo, I., Muis, K. R., Singh, C. A., & Psaradellis, C. (2019). Curiosity… Confusion? Frustration! The role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology*, *58*, 121-137.
11. D'Mello, S. K., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, *20*(4), 361-389.