FACIAL EXPRESSIONS AS AN INDICATOR OF 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

HAAR 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).
- Our long-term goal is to determine the feasibility of using real-time facial-expression recognition algorithms to quantify listening difficulty in-situ. We selected facial expressions because: 1) It may be recorded more naturally and is an easier 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) With advancements in technology and if they could be analyzed in real-time, facial expressions may be useful for tele-audiology.

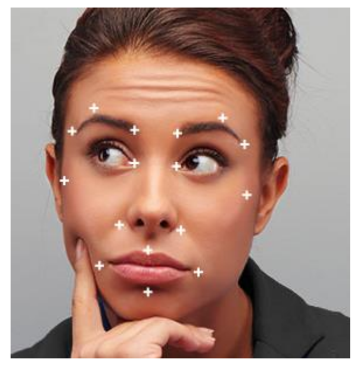
Questionnaires	* Recall Bias
Subjective Rating	* Reliability issues[1]
Pupillometry	* Controlled Situations[2]
Dual-Task Paradigm	*Complex task
Skin Conductance	*Questionable validity[3,4]

- 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.

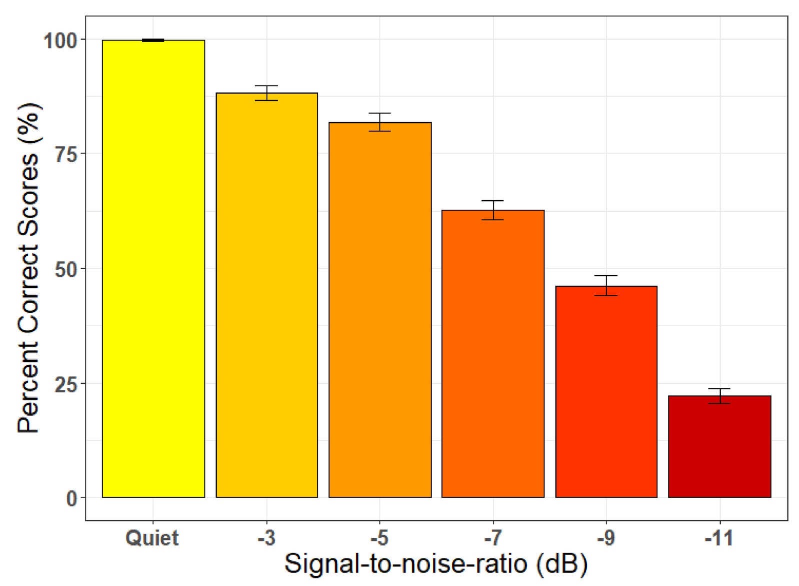
METHODS

- Participants: 20 adults, aged 22 to 37 (Mean = 27.45, SD = 4.92) with normal hearing.
- Stimuli used: Speech perception testing using IEEE[6] 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 various signal-to-noise ratios (SNRs: -3,-5,-7,-9,-11 dB: presented randomly), and in quiet. Participants were also asked to subjectively report their listening effort for each condition.
- The iMotions software assesses the movement, texture, and shape of the face and defines facial expressions as a combination of action units. Different action units work synergistically to produce a particular expression, which is identified by the software. It then computes the evidence level, which is the probability of the presence of a given facial expression. Some of the many emotions analyzed include: joy, anger, fear, contempt, frustration, sadness, confusion, negative 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 in individuals with hearing loss in difficult listening situations[7,8].

Figure 1[9] and 2[10]. Action units and their use in identifying emotions



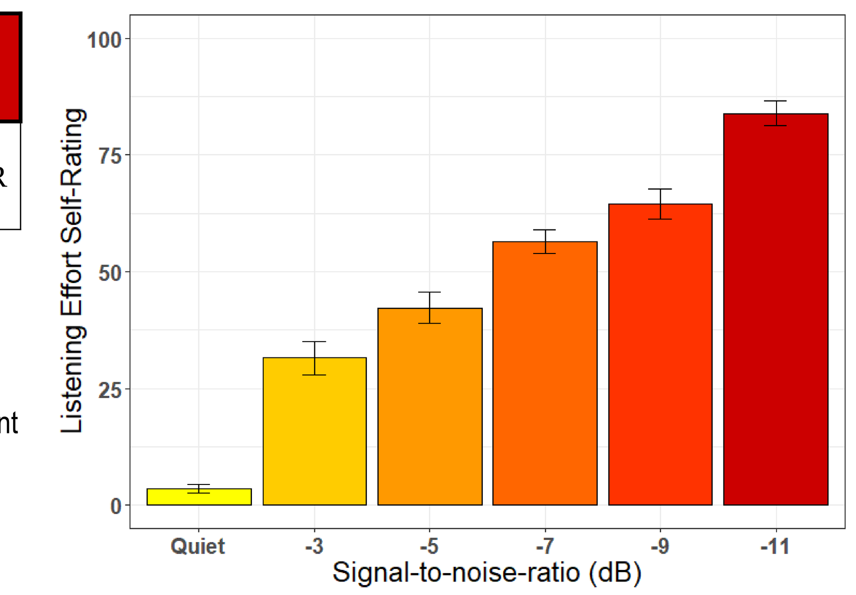
RESULTS

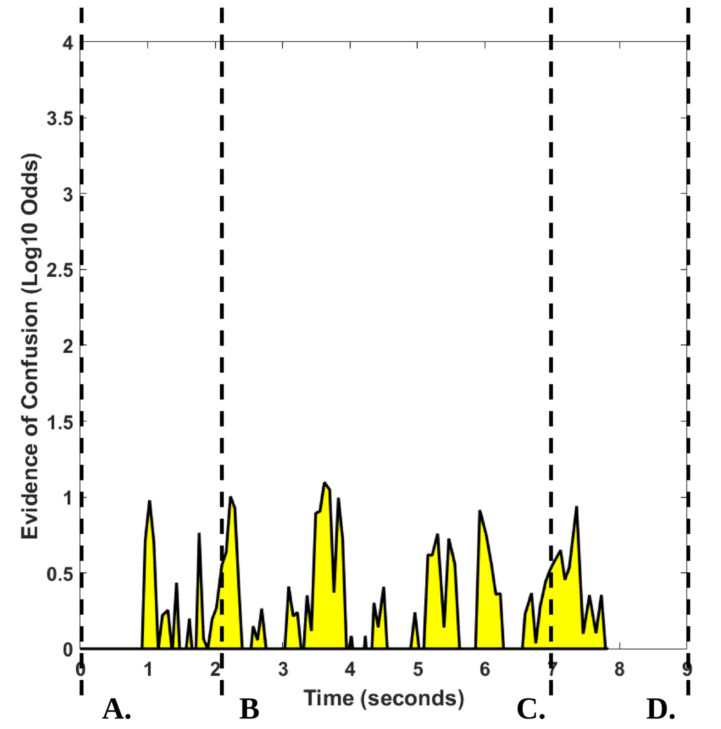


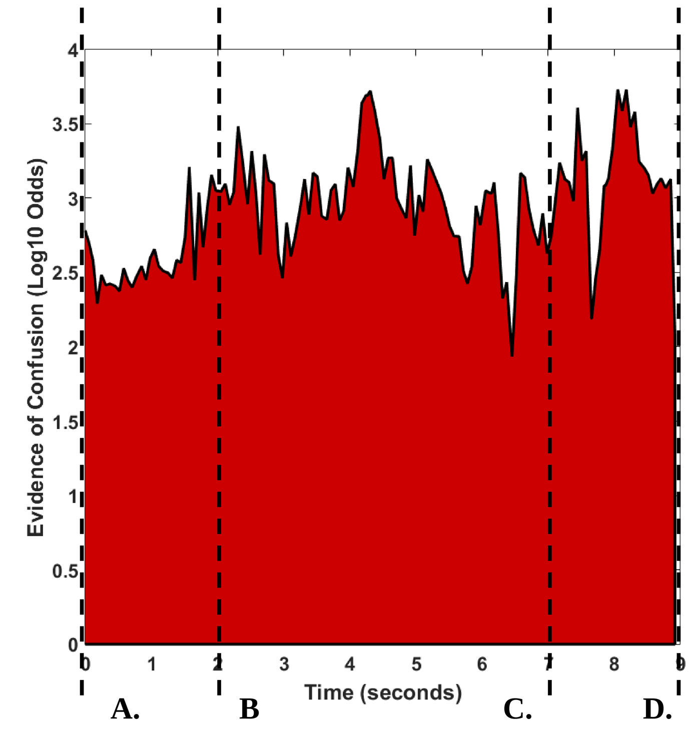
Quiet	-3 SNR	-5 SNR	-7 SNR	-9 SNR	-11SNR

Figure 3. (left) Speech Recognition performance at different signal-to-noise ratios with standard errors of mean

Figure 4. (right) Perceived (subjective) listening effort at different signal-to-noise ratios with standard errors of mean







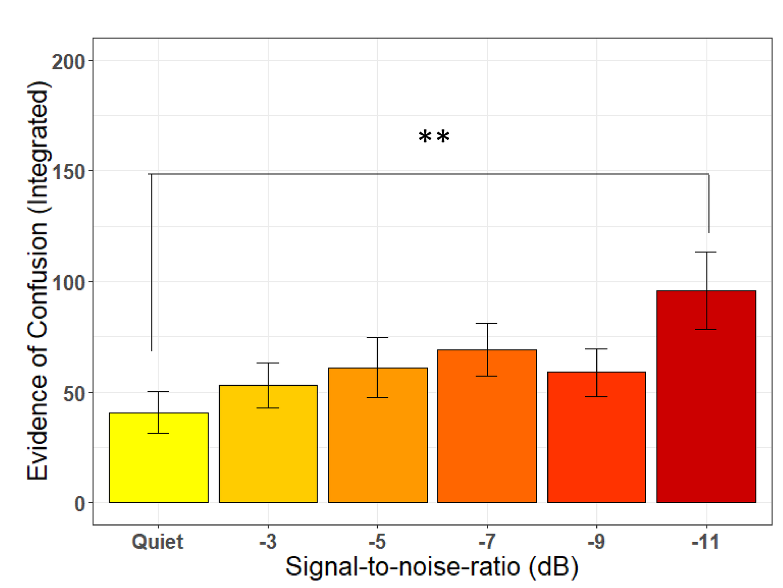
A: Start of noise B: Start of sentence C: End of sentence D: Response cue

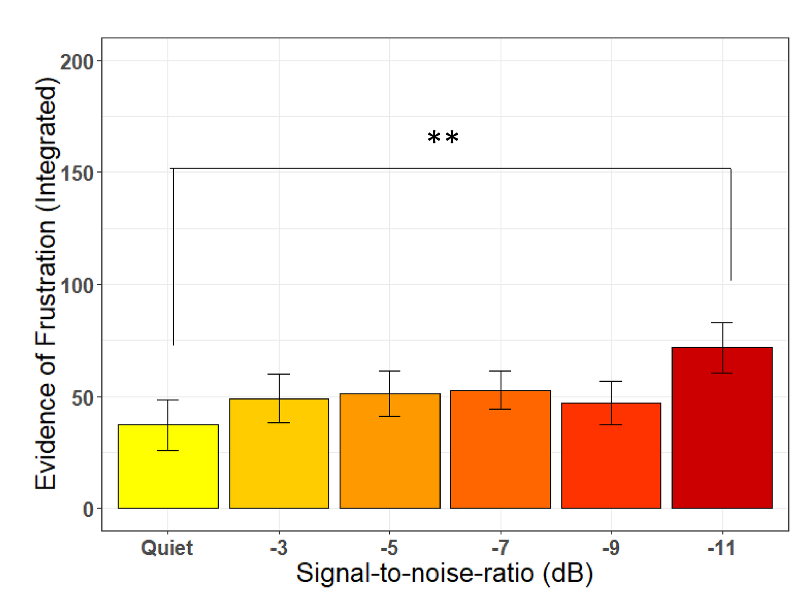
Figure 5. Time course of confusion during **quiet** condition for the subject 20

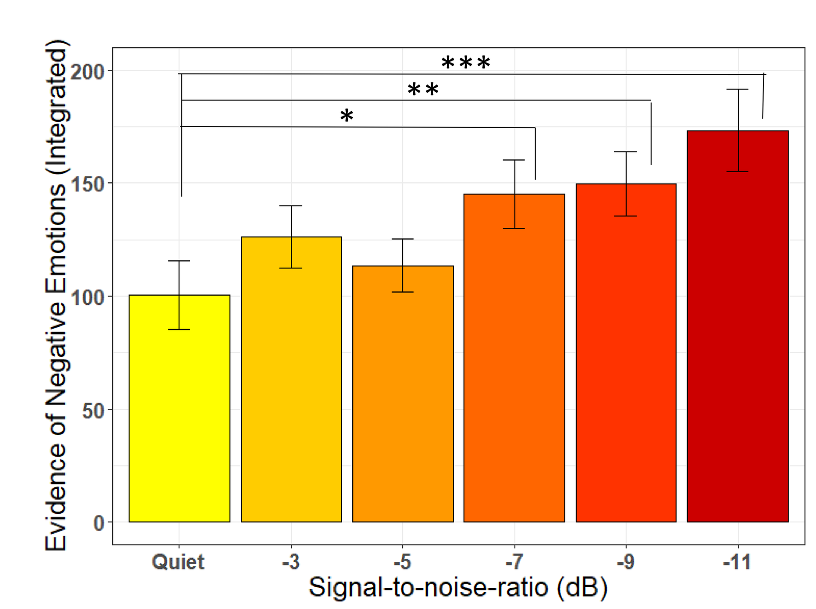
Figure 6. Time course of confusion at **SNR of -11** for the subject 20

- The **figure 5** and **6** (left) represent the evidence level of 'confusion'. The evidence value represents the odds in logarithmic (base 10) units of a particular 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 graph above was obtained for each sentence. The peaks of this graph denote a higher probability of presence of the emotion. The positive area under the graph (integrated value) was obtained for each individual for different conditions. This was then averaged across individuals and conditions to obtain the **figures 7, 8 and 9** represented below.







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

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

Figure 8. The evidence for **frustration** at different signal-to-noise ratios with standard errors of mean

Figure 9. 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 significant for the emotions of confusion ($F(5,95)= 2.757$, $p=0.0228$), frustration ($F(5,95)= 2.683$, $p=0.0260$) and negative emotions ($F(5,95)= 6.515$, $p<0.0001$).
- Pair-wise follow-up testing was conducted. When confusion (**figure 7**) and frustration (**figure 8**) were analyzed separately, it was found that there was a significant difference only between the quiet and the -11 dB SNR conditions (confusion: $p=0.0094$, frustration: $p=0.0080$). The evidence level for negative emotions (**figure 9**) was significantly higher for the SNR conditions of -7 ($p=0.0134$), -9 ($p=0.0021$) and -11 ($p<0.0001$) as compared to speech in quiet.

DISCUSSION

- Although there is no graded change in facial expression for each condition, the significant main effect of SNR suggests that the evidence level for confusion, frustration and negative emotions increased monotonically as SNR decreased. These findings support the feasibility of using facial expression to assess listening difficulty, at least in controlled environments.
- Our next step is to determine the relationship between facial expressions and pupillometry, an established method of measuring listening effort, during a more narrow time window centered around the offset of the stimulus.

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. If these can be transmitted along with information about the listening environment logged by the hearing aids, then audiologists may be able to program hearing aids in real-time and implement telerehabilitation.

ACKNOWLEDGEMENTS

National Institute on Disability, Independent Living and Rehabilitation Research (NIDILRR 90RE5020-01-00)

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. 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.
7. Cacchione, P. Z., Culp, K., Dyck, M. J., & Laing, J. (2003). Risk for acute confusion in sensory-impaired, rural, long-term-care elders. *Clinical Nursing Research*, *12*(4), 340-355.
8. Ciorba, A., Bianchini, C., Pelucchi, S., & Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clinical interventions in aging*, *7*, 159.
9. How do we detect emotion? Retrieved from https://www.spacebetween.co.uk/facial-expression-analysis
10. Avoid the hassle of facial coding. Retrieved from https://imotions.com/emotient/