Visual Effects from Facial Information to Voice Perception in Normal Hearing Individuals and Cochlear Implant Users

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA
Institute of Psychology
Department of General Psychology and Cognitive Neuroscience

Friedrich Schiller University Jena Voice Research Unit

Celina I. von Eiff[1,2] and Stefan R. Schweinberger[1,2]

[1]Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University Jena, Germany; [2]Voice Research Unit, Friedrich Schiller University Jena, Germany

celina.isabelle.von.eiff@uni-jena.de, stefan.schweinberger@uni-jena.de

1. Introduction

Recent research has demonstrated influences of visual facial information to vocal emotion recognition (VER). Here we considered the role of visual information in normal hearers (NH) and in individuals with cochlear implants (CIs) – hearing prostheses designed to functionally replace damaged parts of the inner ear. Importantly, current studies suggest that the ability to perceive emotional communicative cues is highly relevant for CI users' quality of life [1,2].

Rationale of Studies

To contribute to a recent controversy around adaptive effects of visual information to communication with a CI, we systematically investigated audiovisual (AV) integration of emotional cues in NH and CI users. Applying state-of-the-art voice morphing [3], we examined whether CI users gained more perceptual benefit from the addition of congruent facial information (or larger interference from incongruent facial information) compared to NH.

2. Materials and Methods

Experiment 1

- **Participants**: 26 (17 female) adult CI users (M_{Age} = 55.65) and
  26 (17 female) individuals with NH abilities (M_{Age} = 55.50).
- **Task**: Surprise-anger, two-alternative forced choice discrimination task on vocal emotions
- **Stimuli**: Phonetically balanced two-syllable pseudowords, either auditory-only, or with time-synchronized videos with either congruent or incongruent facial emotional expressions **(Fig.1).** Altogether, 256 stimuli (2 emotions x 8 speakers x 4 pseudowords x 4 conditions) were presented in the experiment, in fully randomized order. Surprised and angry stimuli were selected based on classification rates in a pilot study **(Fig. 2)**.
- **Instructions**: Judge emotion in the voice, ignore emotion in the face while looking at the face.

Experiment 2

- **Participants**: 25 (15 female) adult CI users (M_{Age} = 50.36) and
  25 (15 female) individuals with NH abilities (M_{Age} = 50.32).
- **Task**: Surprise-anger, two-alternative forced choice discrimination task on vocal emotions, with an adaptive testing procedure (employing a voice morphing approach that varied the degree of emotion-diagnostic information in the voice).
- **Stimuli**: Identical to Experiment 1. The experiment included 640 stimuli (2 emotions x 8 speakers x 2 pseudowords x 4 conditions x 5 morph levels (MLs)). 384 stimuli (96 trials/condition) were presented to each participant, with conditions in fully randomized order.
- **Instructions**: Identical to Experiment 1.
- **Dependent Variable**: According to the adaptive testing protocol, the task became easier (higher MLs) or more difficult (lower MLs) depending on accuracy in recent trials, in order to eventually arrive at a constant mean performance of .75 correct. Accordingly, it should be noted that lower MLs indicate better performance.

3. Results

Experiment 1 (Fig. 3)

- CI users were substantially impaired in VER, with or without AV information (main effect LGroup: $F(1, 50) = 60.485$, $p < .001$, $\eta_p^2 = .547$).
- We quantified benefits and costs from AV congruent and incongruent facial information, respectively: We calculated differences for congruent AV minus auditory-only, and for incongruent AV minus auditory-only (per stimulus and subject). CI users exhibited stronger benefits to VER if congruent facial information was available, interaction LGroup x Difference, $F(1, 50) = 15.536$, $p < .001$, $\eta_p^2 = .237$.
- CI users benefitted more from congruent faces, $t(32.031) = 4.580$, $p < .001$, Welch test.
- CI users exhibited marginally larger costs from incongruent faces, $t(50) = -1.735$, $p = .089$.
- Performance was positively correlated to quality of life (environmental health domain, WHOQOL-BREF, $rs = .41$, $p = .043$, $n = 25$).

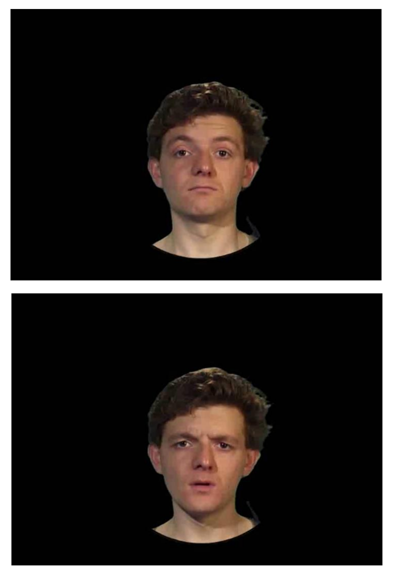

Figure 1. Video frames from AV Stimuli.

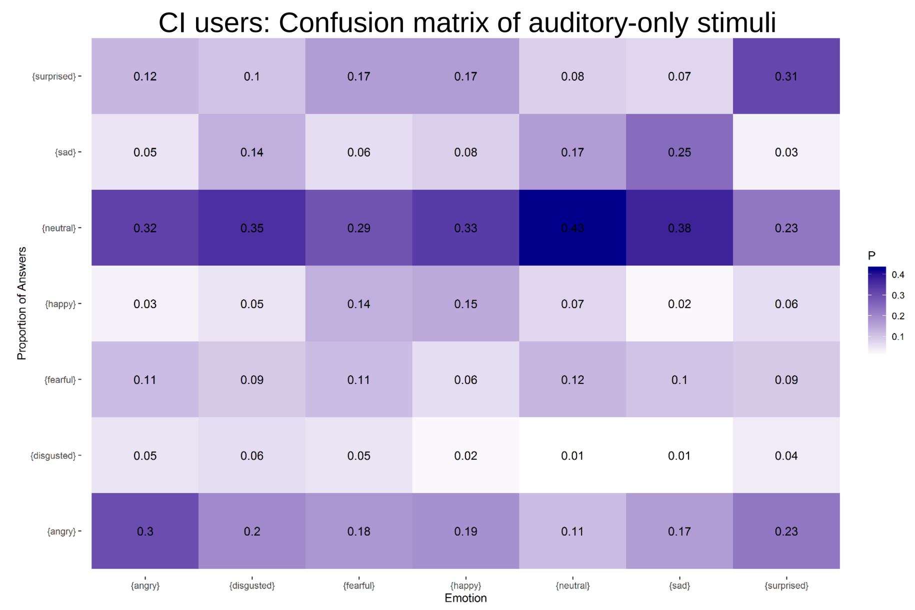

Figure 2. Classification rates in a pilot study. NH classified each emotion ≥ .38 correctly (angry: .51, surprised: .68). *Note:* Chance level is at .14.

Experiment 2 (Fig. 4)

- CI users showed substantially lower VER abilities, with or without AV information (main effect LGroup: $F(1, 48) = 45.294$, $p < .001$, $\eta_p^2 = .485$).
- We quantified benefits and costs from AV congruent and incongruent facial information, respectively: We calculated differences between MLs in each trial round for congruent AV minus auditory-only, and for incongruent AV minus auditory-only (per subject).
- CI users benefitted more from congruent faces, $t(38.501) = -2.909$, $p = .006$, Welch test, but did not exhibit larger costs from incongruent faces, $t(48) = -1.481$, $p = .145$.
- We created subgroups of CI users and NH (n = 10 per subgroup) with equal auditory-only performance levels, $t(18) = 0.391$, $p = .700$, by excluding CI users with lowest and NH with highest auditory-only performance. These subgroups performed on different MLs for AV congruent stimuli, $t(18) = -1.808$, $p = .087$, indicating that CI users benefitted marginally more from congruent faces than NH, with no differences for incongruent trials, $t(18) = 0.182$, $p = .858$.

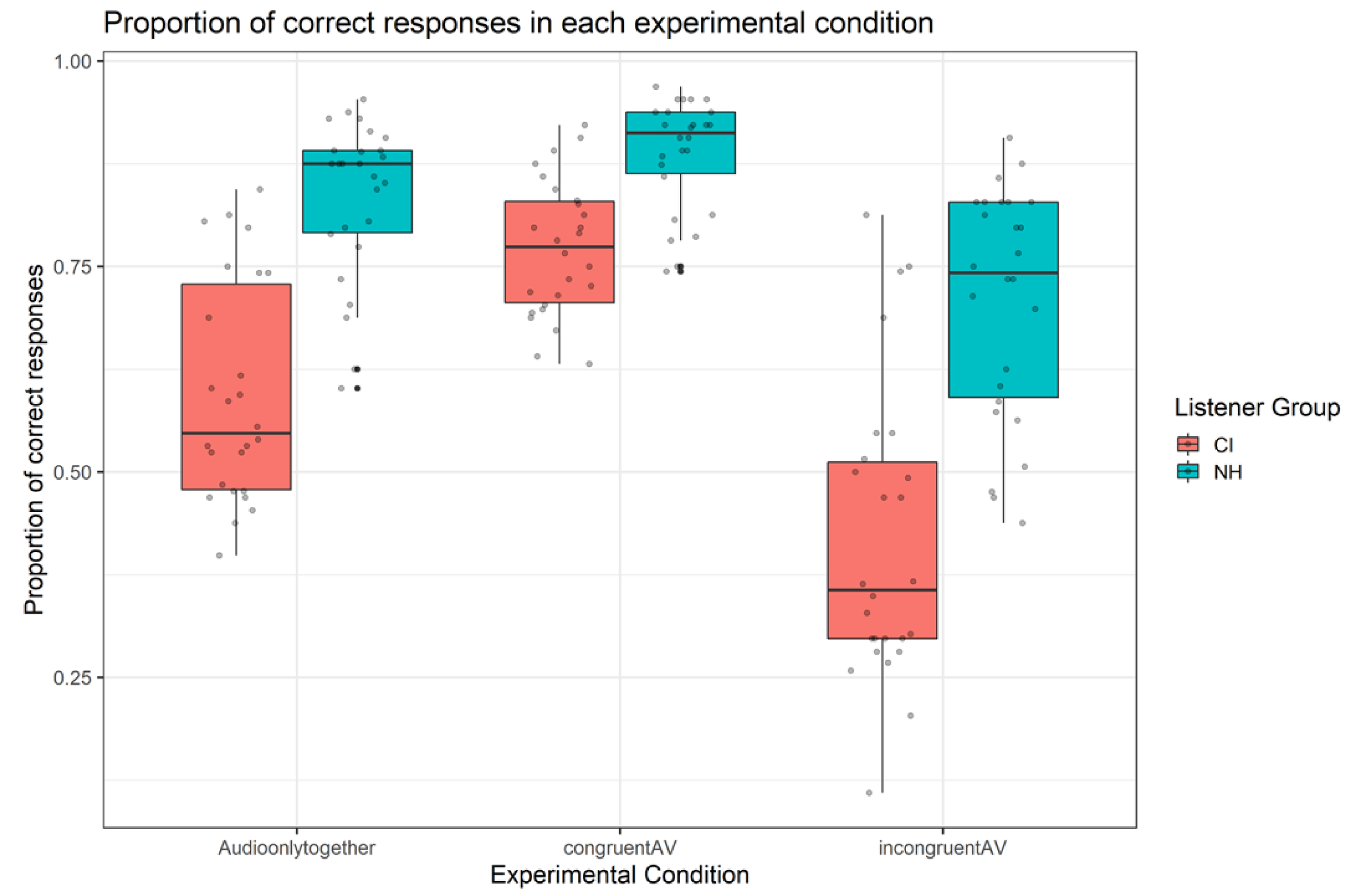

Figure 3. E1. Comparison of vocal emotion recognition in CI users and NH.

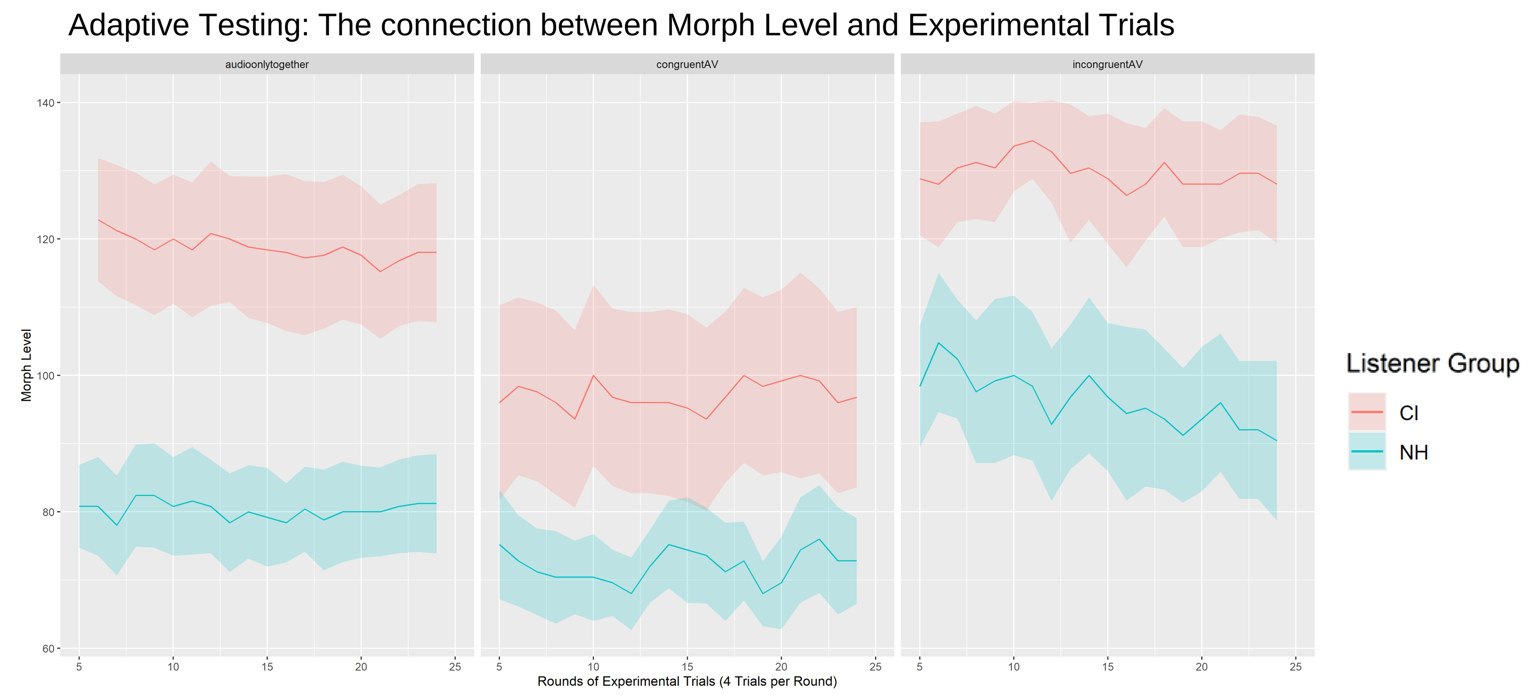

Figure 4. E2. Comparison of morph levels in the adaptive testing procedure in CI users and NH.

4. Conclusions

Performance in both groups benefitted from congruent facial expressions, but this visual benefit was stronger in CI users than NH. Importantly, we saw evidence for a larger influence of visual information to VER in CI users even when accounting for auditory-only performance differences – suggesting effects of deafness-related compensations rather than degraded and unequal acoustic representations between the groups. Echoing previous research, we also demonstrated that better VER was associated with better quality of life ratings in CI users. Overall, our results are in line with adaptive benefits from visual facial information to socio-emotional communication with a CI.

References

1. Luo, X., Kern, A., & Pulling, K. R. (2018). Vocal emotion recognition performance predicts the quality of life in adult cochlear implant users. *Journal of the Acoustical Society of America, 144(5)*, EL429.
2. Schorr, E. A., Roth, F. P., & Fox, N. A. (2009). Quality of Life for Children With Cochlear Implants: Perceived Benefits and Problems and the Perception of Single Words and Emotional Sounds. *Journal of Speech, Language, and Hearing Research, 52(1)*, 141–152.
3. Kawahara, H., & Skuk, V. G. (2019). Voice morphing. In: S. Frühholz and P. Belin (Eds.): *The Oxford Handbook of Voice Perception*. Oxford: Oxford University Press (pp. 685–706).