Perception of Prosody, Emotional Speech, and Indirect Speech in Sensorineural Hearing Loss

Frédéric Marmel[1], Haiping Huang[2], Elizabeth Davis[2], Eric Branda[3], Dina Lelic[4], Erin Picou[2]

[1, 4]ORCA Labs (Stockholm, Sweden & Lynge, Denmark), [2]Vanderbilt University Medical Center - Nashville (TN, USA), [3]WS Audiology - Iselin (NJ, USA)

Introduction: Most studies of the communication difficulties experienced by hearing aid wearers (HAWs) focus on speech understanding. Studying communication beyond speech understanding may shed light on hearing difficulties that are detrimental to quality of life but are currently overlooked. This ongoing study focuses on the perception of speech prosody — the aspect of communication that creates emotional tone and subtle grades of meaning. We are investigating whether HAWs have difficulties perceiving several types of prosody and whether HAs provide benefit for prosody perception. We investigate emotional prosody conveying basic emotions (anger, disgust, fear, happiness, neutral, sadness) or more nuanced emotions (adoration, amusement, anger, awe, confusion, contempt, desire, disappointment, distress, fear, interest, sadness), as well as prosody conveying intentions (sarcasm, teasing, white lie, sincere, questions and indirect requests).
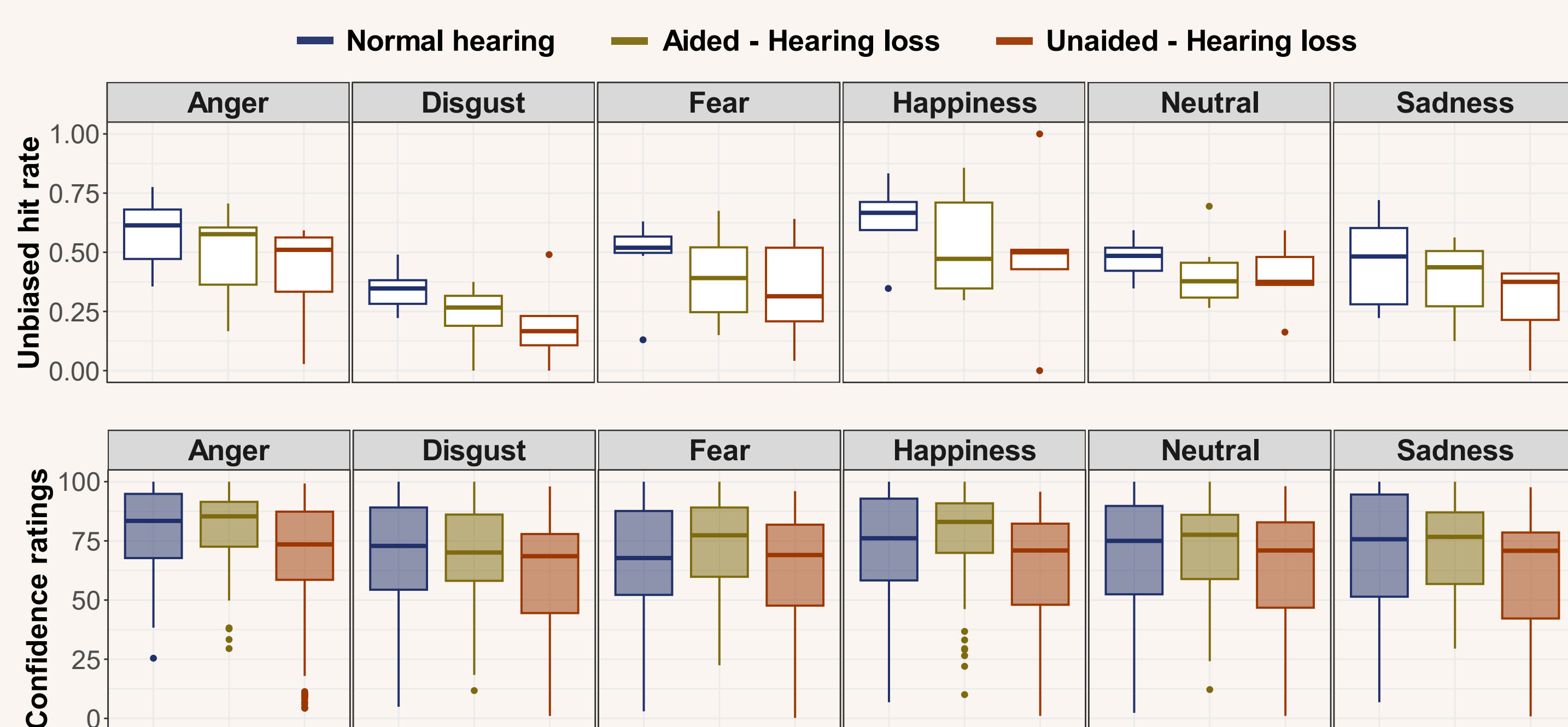
General method

- 4 closed-set categorization tasks implemented in PsychoPy [1], each using a different corpus of emotional / intentional speech [2–3, 5–6]
- Tasks: judge emotions / intentions expressed and rate confidence in judgments
- Free-field sound presentation at an average of 60 dB SPL
- Questionnaires: cognition, anxiety and depression, social isolation, HA satisfaction, perceived hearing ability, perceived emotion recognition, visual emotion recognition

- Experimental group: experienced ($\geq$ 1 year) HAWs, fit to loss with Signia Pure Charge & Go, and given 3-week acclimatization period
- Control group: Age-matched participants with near-normal hearing [7]
- HAWs tested aided and unaided (counterbalanced order), NHs tested once
- Testing ongoing: 8 HA wearers (73 ± 6 years, 6 F / 2M) and 7 NH participants (66 ± 8 years, 6 F / 1 M) took part in the study
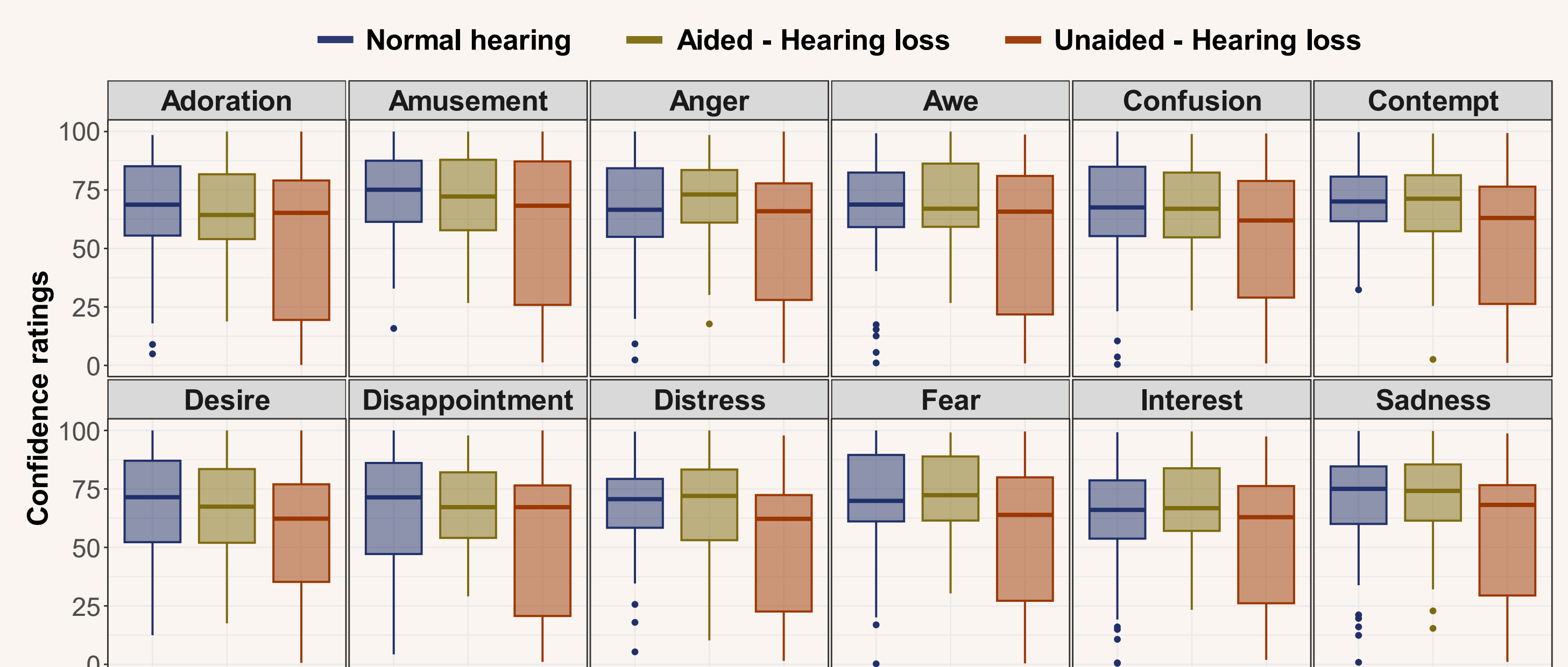
Basic emotions

- CREMA-D corpus [2]: 12 sentences with neutral semantic content, 6 emotions, 3 intensities, 3 modalities (audio/audio-visual/visual), 91 actors
- Our stimulus set: audio and audio-video recordings retained if rated in [2] a) by at least 10 raters, b) so that majority of raters chose the emotional category intended by the actor, c) with a mild or medium emotional intensity
- Each participant: 6 sentences * 6 emotions * 2 intensities * 2 modalities
- Preliminary results: audio modality only, intensities averaged together

Intentions

- RISC [5] and Pros_scaled [6] corpora in two separate tasks
- RISC: short videos of question-response dyads, 4 responder intentions, 4 types of relationship (romantic, friends, colleagues, boss/employee), 4 amateur actors
  - Our stimuli: retained only colleagues and boss/employee relationships
  - Each participant: 3 scenes * 2 relationships * 4 intentions * 2 modalities
- Pros_scaled: 12 audio sentences that could be conveying indirect requests (6 modal interrogatives and 6 declarative statements), 18 speakers
  - Our stimuli: retained only 4 interrogatives associated with best discrimination between requests and questions in [6], and dropped 4 speakers whose request and question utterances could not be discriminated in [6]
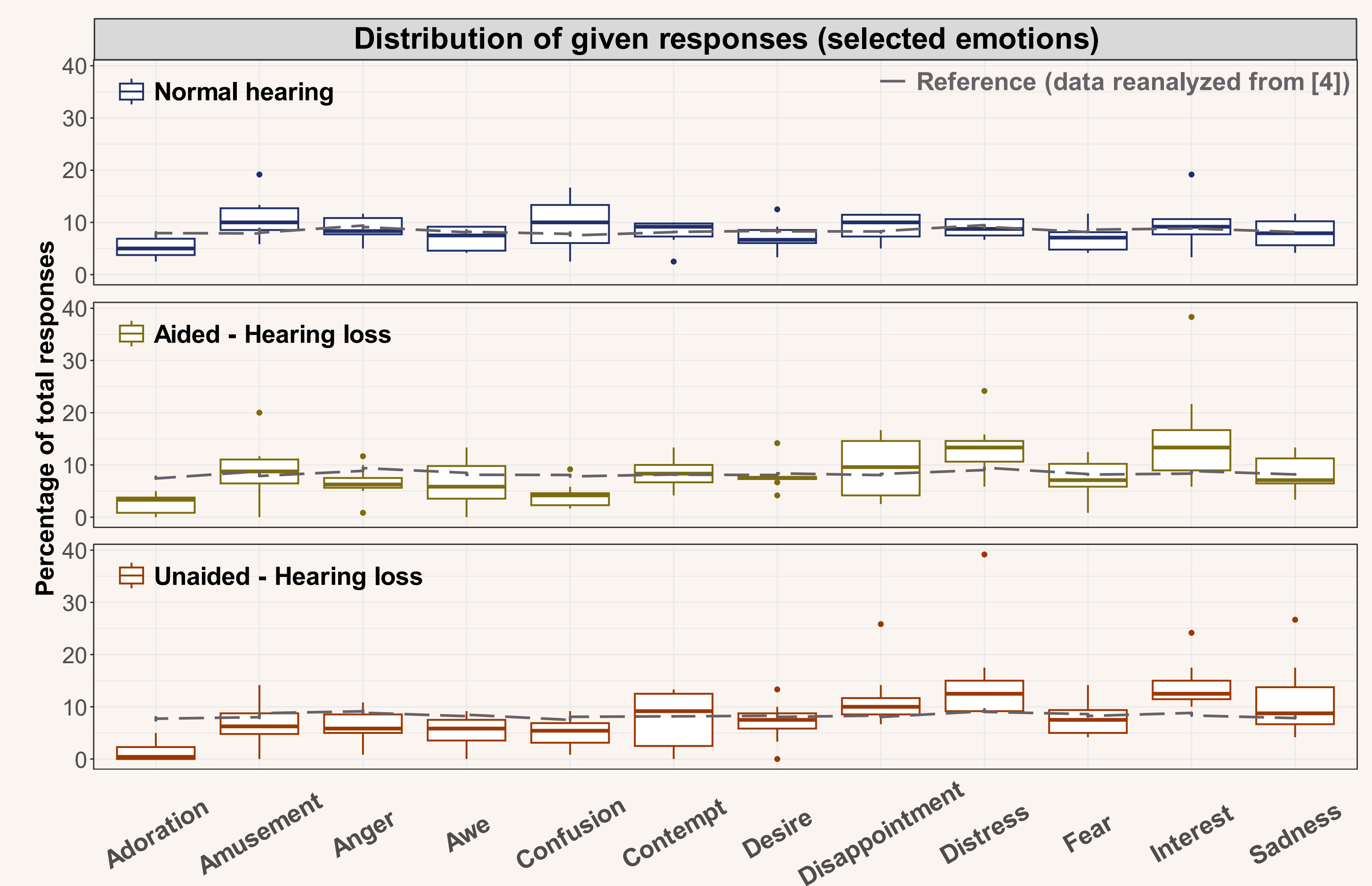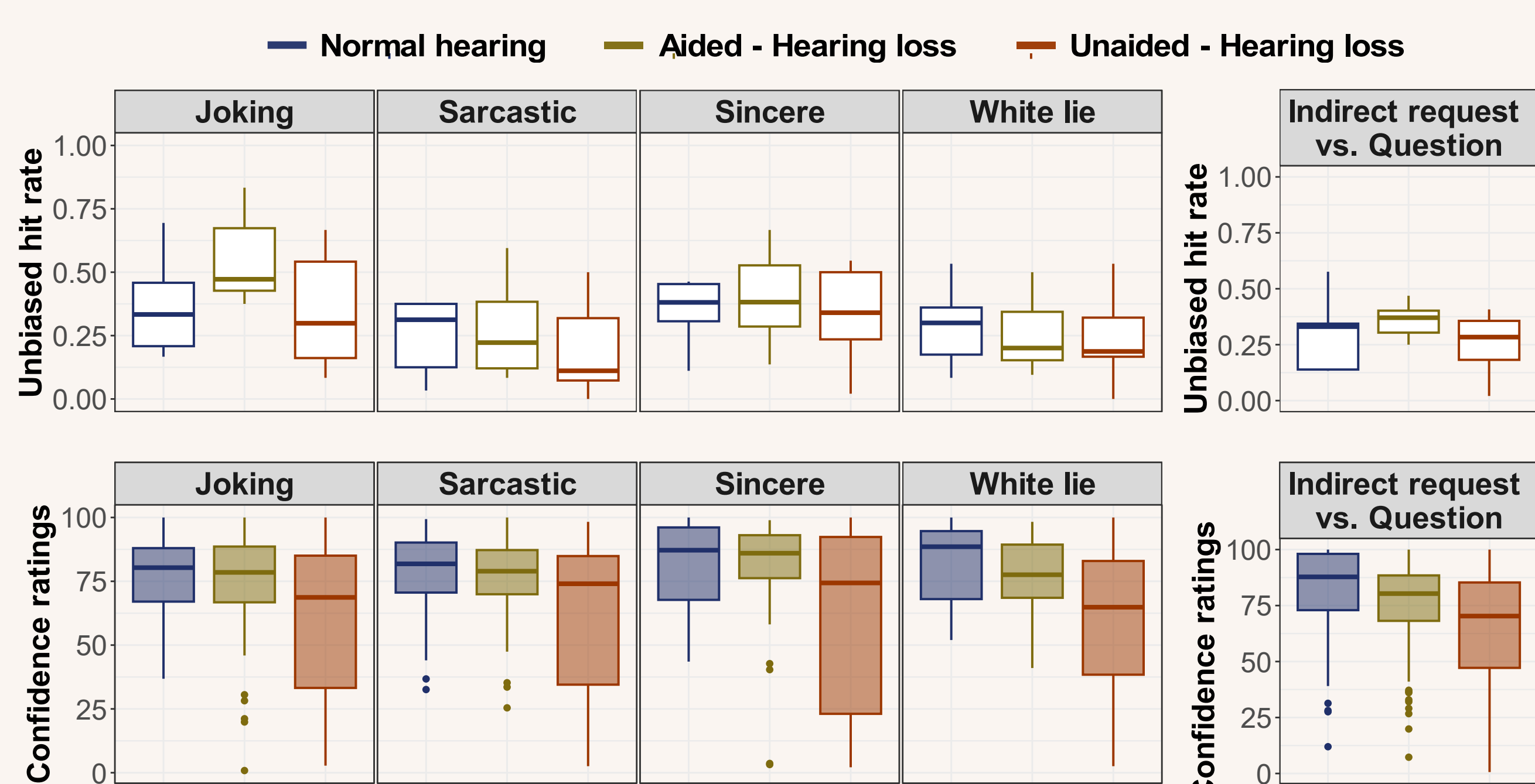 - Each participant: 4 sentences * 2 intentions * 6 speakers

Nuanced emotions

- VENEC corpus [3]: 2 sentences with neutral semantic content, 19 emotions, 3 intensities, audio modality only, 100 actors
- Our stimulus set: the 12 emotions consistently recognized across cultures in [4]. For each emotion, recordings retained if rated in [4] as target emotion by a) at least 20% of raters, b) more raters than any other emotion category and c) $\leq$ 50% of raters (to avoid the task being too easy)
- Each participant: 10 sentences * 12 emotions

Preliminary results & Discussion

- **Basic**: trend for better recognition performance for NHs than for HAWs, and for better recognition aided than unaided. Trend for higher confidence for NHs than for unaided HAWs
- **Nuanced**: trend for higher confidence for NHs than for unaided HAWs. Trend for a more skewed distribution of responses for HAWs than NHs and reference data
- **Intentions**: no overall trend for differences in recognition performance between groups. Trend for higher confidence for NHs and for HAWs aided than for HAWs unaided
- **State of the art**: literature points to speech emotion recognition impairment with hearing loss and there is no evidence that HAs help. There is a lack of studies looking beyond basic speech emotions
  - Questions for further analyses: potential confound of age and other factors assessed with the questionnaires, influence of visual cues

References

[1] Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK (2019) PsychoPy2: Experiments in behavior made easy. Behav Res Methods, 51(1):195–203.

[2] Cao H, Cooper DG, Keutmann MK, Gur RC, Nenkova A, Verma R (2014) Crema-d: Crowd-sourced emotional multimodal actors dataset. IEEE Trans Affect Comput 5(4):377–390.

[3] Laukka P, Elfenbein HA, Thingujam NS, Rockstuhl T, Iraki FK, Chui W, Althoff J (2016) The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. J Pers Soc Psychol 111(5):686.

[4] Cowen AS, Laukka P, Elfenbein HA, Liu R, Keltner D (2019) The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. Nat Hum Behav 3(4): 369–382.

[5] Rothermich K, Pell MD (2015) Introducing RISC: A new video inventory for testing social perception. PloS one 10(7).

[6] Trott S, Reed S, Kaliblotzky D, Ferreira V, Bergen B (2023) The role of prosody in disambiguating English indirect requests. Lang Speech 66(1): 118–142.

[7] Moore BCJ, Creeke S, Glasberg BR, Stone, MA, Sek A (2012) A version of the TEN test for use with ER-3A insert earphones. Ear Hear 33(4): 554e557.

VANDERBILT UNIVERSITY MEDICAL CENTER

ORCAlabs
The Scientific Institute of WS Audiology