How’s what sound?
Perceptual shortcomings in current hearing aid personalization
William M. Whiting1,2, David McShefferty1, Benjamin Caswell-Midwinter2 & Graham Naylor1
1 Hearing Sciences – Scottish Section, University of Nottingham, Glasgow, UK
2 Institute of Health and Wellbeing, University of Glasgow, UK
3 Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA

Rationale
In the personalisation of hearing aids, it is common practice to adjust gain away from prescription based on feedback from the patient. Underlying this practice are two key assumptions about the perception of those adjustments:
1. The adjustments are adequately large enough and the stimuli adequately long enough to elicit a reliable preference, and
2. The patient’s feedback is reliable, and there is a common language for gain adjustments across patients.

In two psychological studies we separately explore the viability of these commonly held assumptions.

STUDY 1: Speak longer & fit better? Spoiler: Yes, but limited improvement

Methods
29 adults with hearing aid experience listened to 2, 4 & 6-s segments of a continuous monologue (story, reader) presented successively over
• 1st at each individual’s RNID gain
• 2nd with a 0, 6 or 12 dB adjustment in 1 of 5 broad frequency bands
• Minimal interstimulus interval (=50ms), fixed interval order
Participants judged if 2nd segment was better, worse or no different from 1st

Raw results – Similar pattern of preference for increased LF gain across durations

Preference JNDS = gain adjustments necessary to elicit better or worse preferences

JNDS decrease & reliability increases with duration
• Providing longer stimuli – taking longer – improves likelihood of patients providing reliable judgments of gain adjustments
But duration has limited & diminishing effect
• 6-s segments still required at least 3 dB gain adjustments to elicit preference
• Decreasing effect slope -0.8 dB increase/100 ms

STUDY 2: Are gain descriptors useful? Spoiler: No, too little agreement

Methods
28 adults with thresholds ≥10 dB HL @ 3-5 freqs. of median online audiogram
• 1st at standard gain
• 2nd at 412 dB adjustment in 1 of 5 broad frequency bands
• Participants could listen repeatedly to segments before two tasks per trial
1. Discrimination – Do you hear a difference in quality between 2 segments
2. Description – If you heard a difference, describe difference in 12 words

Discrimination results – Do you hear a difference?
Decrease in discriminability with increasing centre frequency the same as Discriminating decrements more difficult than increments previous

No difference in discriminability between ≥3 of voices

Description results – Describe the difference
Conversion of all responses to refer to adjusted standard segment
Condensation of phrases (e.g., “farther away” → “distant”)

Many participants showed signs of difficulty in describing judgments

• Very little agreement among participants (≤25%) when applying a 50% interrater reliability criterion (i.e., descriptor was used for half of the trials for any given adjustment by a participant)
• Ambiguous gain patterns for common terms (e.g., “clear” vs “unclear”)
• Descriptors aren’t useful

Participants could listen repeatedly to segments before two tasks per trial
1. Discrimination – Do you hear a difference in quality between 2 segments
2. Description – If you heard a difference, describe difference in 12 words
Discrimination results – Do you hear a difference?
Decrease in discriminability with increasing centre frequency the same as Discriminating decrements more difficult than increments previous

No difference in discriminability between ≥3 of voices

Description results – Describe the difference
Conversion of all responses to refer to adjusted standard segment
Condensation of phrases (e.g., “farther away” → “distant”)

Many participants showed signs of difficulty in describing judgments

• Very little agreement among participants (≤25%) when applying a 50% interrater reliability criterion (i.e., descriptor was used for half of the trials for any given adjustment by a participant)
• Ambiguous gain patterns for common terms (e.g., “clear” vs “unclear”)
• Descriptors aren’t useful