Laura Winther Balling[1] & Frédéric Marmel[2] · ORCA Labs ([1]Lynge, Denmark & [2]Stockholm, Sweden)

Exploring listening efficiency in a lexical decision task as a measure of hearing-aid outcomes at realistic signal-to-noise ratios

Introduction

- Positive SNRs are prevalent in everyday situations [1], but performance in standard speech tests is often near ceiling at these realistic SNRs
- There is a need for outcome measures sensitive to listeners' difficulties at realistic SNRs with a focus on capturing how much effort they need to exert, a crucial factor for living with hearing loss
- Listening efficiency integrates accuracy and effort (as indexed by response time) into a unified measure that is better able to capture hearing ability differences between listeners in realistic situations [2]
- The lexical decision task (LDT) taps into many of the same cognitive processes as the speech tests more typically used in audiology. Participants are presented with a mixture of words and nonwords and are asked to respond as quickly and as accurately as possible whether each is a real word. Response times may be taken to reflect their underlying capabilities and the effort they exert

Objective & Research questions

To explore the potential of the lexical decision task and of listening efficiency metrics to reveal effects of SNR and hearing-aid (HA) signal processing on the performance of HA wearers, we asked the following **research questions**:

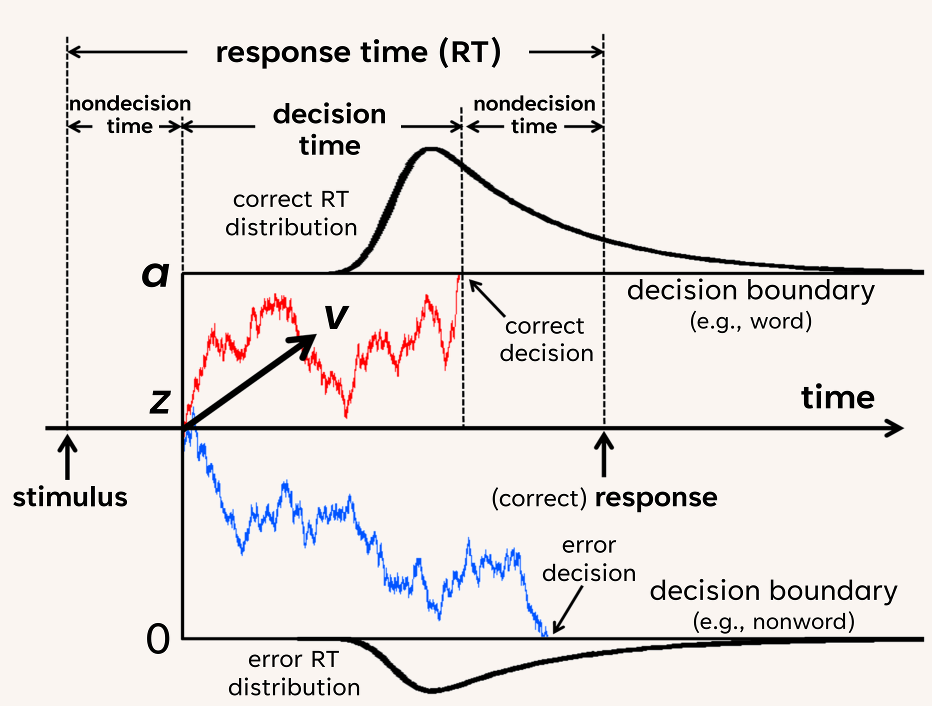
- Is the LDT suitable for evaluating the listening difficulties of HA wearers at realistic SNRs?
- Do listening efficiency metrics uncover larger differences between experimental conditions (SNR, HA program) than accuracy or response time alone?
- Do listening efficiency metrics correlate with subjective measures of listening effort/fatigue better than accuracy or response time?

Experimental design

- 2*2 factorial design: SNR (+5 vs. +10 dB) x HA program (omnidirectional microphone mode vs. directional microphone mode)
- 19 experienced HA wearers (1+ years), 60–95 years old
- Fixed 60-dB HINT noise from 3 back loudspeakers (speech from 1 front loudspeaker)
- LDT: 4 blocks (1 per condition), 60 words & 60 nonwords (from [3, 4]) per block
- Listening effort questionnaires (after each block): selected questions from [5–8]
- Modified HINT (at the end of session): same fixed noise as for LDT, one training list & one test list for each of the two SNRs, hearing aids in omnidirectional mode

Listening efficiency

- Listening efficiency should be more sensitive than accuracy or response times alone because it deconfounds performance from the speed-accuracy trade-off
- Listening efficiency can be computed simply as the ratio of accuracy to response times [8], *but* more statistical power can be gained by computing listening efficiency as the drift rate parameters in cognitive models of decision making [9]
- The DDM (drift-diffusion model, [10]) is a popular model cognitive model of decision making for speeded two-choice tasks:



Decision process in the DDM

*Response time on each trial is the sum of a non-decision time (perception, movement initiation, execution) and a decision time, which is determined by a process of noisy evidence accumulation. Evidence is accumulated (**drift rate: v**) towards one of the two decision boundaries, and a decision is taken when a boundary is reached. Other parameters are the decision bias (z) and the evidence threshold (a).*

Hierarchical Bayesian modelling

Full parameter distributions for the population, for specific groups, and for individual participants

Data analysis

- Effects of SNR & HA program on LDT performance assessed for four metrics: **Correctness** (at trial level, binomial data), **Response time** (at trial level, log transformed), **Correctness-to-response-time ratio** (at block level, log transformed), and **Drift rate** (at block level, log transformed)
- Drift rates estimated for each participant by fitting a hierarchical DDM using [11] (*a*, *z*, and *t* estimated as group-level parameters)
- Covariates: flow-related (block number, list), subject-related (gender, age, PTAs), HINT scores, effort questionnaires
- Random effect structure: intercepts for participants for all metrics, intercepts for individual items for Correctness and Response times, random slopes only when improving the model

Results (mixed regression models)

Results summary

- SNR & HA-program manipulations only had small effects
- SNR contributed significantly only to Correctness-to-response-time ratio model
- HA program contributed significantly only to Correctness model (but adding random effects for HA program improved the Correctness-to-response-time ratio model)
- Best fit (larger R^2) for Correctness-to-response-time ratio model (model based on averages, so substantial individual variation already taken out)
- No conclusive indication that combining accuracy and response times heightens sensitivity to experimental manipulations
- HINT performance did not contribute significantly to the models, even though it was retained in the final model for Correctness
- Most questionnaires did not contribute to the models

Discussion

- Contrary to our hypotheses and in contrast to [2], no evidence that listening efficiency metrics uncover larger differences between experimental conditions or that they are stronger correlates of subjective measures of listening effort/fatigue
- Among the effort ratings, only participants' ratings of how likely they are to change the listening situation was significant, in line with [5]
- Performance in the modified HINT and in the LDT did not align. The LDT is probably not **suitable to evaluate the listening difficulties of HA wearers at realistic SNRs**
- More work using the hierarchical Bayesian DDM seems worthwhile given its potential for increased sensitivity to listeners' difficulty [2, 9] and for disentangling perceptual and decision-making processes in perceptual tasks [12]

References

[1] Smeds, Wolters, Rung, M. (2015). J Am Acad Audiol, 26(02), 183–196.
[2] Perea Pérez, Hartley, Kitterick, Zekveld, Naylor, Wiggins (2023). Front Hum Neurosci, 17.
[3] Balling, Baayen (2008). Lang Cogn Proc, 23(7-8), 1159–1190.
[4] Balling, Baayen (2012). Cognition, 125(1), 80–106.
[5] Picou, Ricketts (2018). Int J Audiol, 57(6), 457–467.
[6] Meister, Rahlmann, Lemke, Besser (2018). Trends Hear, 22.
[7] Group (1986). Task Load Index (NASA-TLX) v 1.0.
[8] Prodi, Visentin, Farnetani (2010). J Acoust Soc Am, 128, 172–181.
[9] Stafford, Pirrone, Croucher, Krystalli (2020). Behav Res Methods, 52, 2142–2155.
[10] Ratcliff, Rouder (1998). Psychol sci, 9(5), 347–356.
[11] Fengler, Xu, Bera, Omar, Frank (in preparation, https://github.com/lnccbrown/HSSM)
[12] Oganian, Froehlich, Schlickeiser, Hofmann, Heekeren, Jacobs (2016). Front Psychol, 6, 1958.

The Scientific Institute of WS Audiology