1) **Data collection.** Have any data been collected for this study already?

No data have been collected for this study yet.

2) Hypothesis

Research question:
How can we behaviourally measure naturalistic speech comprehension in real-time (during listening) and in a time-resolved manner (with ms resolution)?
The research proposed here represents the first experiment in a series aimed at answering the question: "What are the causal neural mechanisms of speech comprehension?"

We will introduce a measure allowing participants to provide continuous comprehension ratings while listening to naturalistic speech segments at varying speech rates. This measure will be validated against traditional post-hoc assessments.

Hypotheses:

Real-time comprehension measure:

[1] The novel real-time comprehension measure scores will significantly correlate with the comprehension scores of the traditional post-hoc tests.

[2] The scores of our novel real-time comprehension measure will be less correlated with nuisance variables like working memory memory capacity, than the post-hoc measures.

Behavioral performance: Comprehension will be worse for higher speed speech segments compared to lower ones.

3) Dependent variable

Speech Comprehension

Measures:

> Real-time Speech Comprehension Measure: Participants will report continuous comprehension ratings using a slider while listening to speech segments. They will express their comprehension level by adjusting the distance between their index finger and thumb of their dominant hand — maximizing the distance indicates complete comprehension (a maximum value of 255), while minimizing signifies no comprehension (a minimum value of 0).

> Post -hoc Comprehension Measures:

> Post-hoc Comprehension Rating: After each segment is finished, participants will rate their comprehension on a 10 point scale, with 0 indicating no comprehension and 10 indicating full comprehension. The purpose of this task is to have a post-hoc self-report comprehension measure in addition to the real-time self-report comprehension measure.

> Summary Task: Participants will be asked to give a written summary of the segment. The semantic similarity of the segment and the summary will be evaluated using a Large Language Models. We will get word embeddings from GLoVe (Pennington et al., 2014) and BERT (Reimers & Gurevych, 2019), and compare the vector cosine distance between what is heard and what is written. The order of recall is not important and will not be taken into account.

> Multiple Choice Question: A four option multiple choice question will be administered after each segment, to assess whether participants understood the content delivered in the segments.

Currently, the standard method of assessing comprehension is using these post-hoc measures. For example, by asking listeners to estimate how much they understood on a scale (Gillis et al., 2023; Schiavetti, 1992). Another established speech intelligibility metric includes “word identification tasks” in which listeners are asked to repeat the words or sentences they heard as accurately as possible, and the number of true recalls constitutes the comprehension score (Beukelman & Yorkston, 1979; Davis et al., 2005; Lubinus et al., 2023). Multiple choice questions are frequently used as well, both in the experimental context (Moody et al., 1987; Wester & Watts, 2016) and for language skill evaluation purposes such as the Test of English as a Foreign Language (TOEFL), and IELTS (International English Language Testing System).

4) Conditions

Stimuli

An audiobook, *Someday, Someday, Maybe* by Lauren Graham (2013), served as the source for our original recording, from which we extracted 30-second segments that start and end with sentence beginnings and endings. These segments were sped up to create stimuli at five different speech rates, ranging from a minimum speeding up factor of 1 (the same as the original recording rate) to a maximum factor of 5 (five times faster than the original recording rate). The speech rate conditions are composed of 1, 2, 3, 4 and 5 times faster than the original recording rate. Segments were compressed in PRAAT (Boersma & Weenink, 2024) by factor while keeping the pitch unchanged. To avoid potential differences in context or word salience, and to ensure uniformity in the difficulty of post-hoc comprehension questions, we randomized segment-to-speed allocation across participants. In total, there are 125 segments, consisting of 25 instances for each of the five speech rates.

Design and Procedure

Firstly, participants will complete the control tasks, which include the working memory task using Digit Span (Richardson, 2007; Wechsler, 1981), and the speech-perception-in-noise task using the Digit-In-Noise test (WHO, 2022).Then, prior to the experimental trials, participants will have 2 training blocks with the slowest and fastest speech rates respectively will be administered to familiarize participants with the range of speech rate conditions and the protocol. The experiment will employ a block design. Each of the 125 blocks will start with listening to a speech segment while simultaneously rating speech comprehension using the real-time comprehension measure (slider), and will end with post-hoc comprehension measures, i.e., comprehension rating, a summary and a multiple choice question. Each block consists of a presentation of a single speech rate. The 125 blocks are randomly ordered.

5) Analyses

We will use a Mixed Effects Linear Regression Analysis to model the real-time comprehension scores. Comprehension scores of the post-hoc tests, which are the multiple-choice questions, 10-scale comprehension ratings, and written summaries, will be utilized as factors in the regression. Digit Span and Digit-In-Noise scores will be assigned as random effects, accounting for individual subject variations. This approach allows us to explore to what degree the scores of our novel real-time comprehension measure can be explained by comprehension, above and beyond the contribution of individual differences in working memory and speech perception in noise capacities.
We will investigate how comprehension varies across speech rates by employing a Mixed Effects Linear Regression Analysis. The model will show how comprehension measures predict speech rate.

6) Outliers and Exclusions

For the post-hoc comprehension measures, to ensure that we are validating the novel measure against actual comprehension:
-Absolute threshold for comprehension: 75% correctness on the multiple choice questions for the slowest (easiest to comprehend) speech rate.
-Ratings on the 10-point scale should be significantly different for the slowest and fastest speech rates.
- If a participant didn’t provide any summaries (summary left blank for all conditions), their data will be excluded.

To ensure that participants actually used the physical slider actively:
- Distribution of the amount of slider movements across trials and participants- If a participant's slider movements exceeds ±3.5 standard deviations from the mean, we will remove those participants' data.

7) Sample Size

We are aiming to recruit 30 participants as was done in a previous study on comprehension of naturalistic speech at various rates (Lubinus et al., 2023).

8) Other

Criteria for participants:
-Native English speakers
-Normal hearing
-Neurotypical (not diagnosed neurological or psychiatric disorders)

Control Measures

Digit Span: We will measure working memory capacity using the forward and backward Digit Span Test (Richardson, 2007; Wechsler, 1981). The forward and backward tasks will consist of different digit sets, and the forward test will be applied before the backward test. Each of them will include seven levels (3 to 9 digits) with 2 different spans (sequences) for each length level. The procedure will start with the shortest digit spans (3 digits) and will stop as soon as participants fail to correctly repeat both spans belonging to one level or when the two longest digit spans (9 digits) are successfully completed. Digit spans will be presented auditorily and participants will verbally report their responses.

Digit-In-Noise Test: We will measure speech comprehension ability in noise using “hearWho”, a mobile and web-based software application for hearing screening developed by the World Health Organization (WHO). HearWho utilizes a Digit-In-Noise Test by presenting 23 spoken sets of US English digit triplets over progressively increasing white noise (WHO, 2022).

Note: We collected pilot data to be informed about our design choice. The pilot utilized 6 speech rates (the original multiplied by 1, 2, 2.5, 3, 3.5, and 4), had the summary task for only 50% of the segments, and had 30 segments for each 6 conditions. To obtain more summaries and keep the experiment time no more than 2 hours, we decided to change the design: We will obtain summaries for 60% of the segments and have 5 speech rates (multiplied by 1,2,3,4,5) with 25 segments for each. The maximum speech rate is changed from multiplied by 4 to 5 to make sure we have a condition that is fully non-comprehendible. We did not collect any data with the new design.

References

Beukelman D.R, Yorkston K.M. (1979). The relationship between information transfer and speech intelligibility of dysarthric speakers. J. Commun. Disord. 12, 189–196. doi:10.1016/00219924(79)90040-6

Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.10, retrieved 21 April 2024 from http://www.praat.org/

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. Journal of Experimental Psychology: General, 134(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

Deafness and hearing loss: hearing checks and the hearWHO app . (2022). Accessed: March 9, 2022:

https://www.who.int/news-room/questions-and-answers/item/deafness-and-hearing-loss-hea ring-checks-and-the-hearwho-app

Gillis, M., Vanthornhout, J., & Francart, T. (2023). Heard or Understood? Neural Tracking of Language Features in a Comprehensible Story, an Incomprehensible Story and a Word List. Eneuro, 10(7), ENEURO.0075-23.2023. https://doi.org/10.1523/ENEURO.0075-23.2023

Graham, L. (2013). Someday, someday, maybe. Random House Publishing Group

Lubinus, C., Keitel, A., Obleser, J., Poeppel, D., & Rimmele, J. M. (2023). Explaining flexible continuous speech comprehension from individual motor rhythms. Proceedings of the Royal Society B: Biological Sciences, 290(1994), 20222410. https://doi.org/10.1098/rspb.2022.2410

Moody, T., Joost, M., & Rodman, R. (1987). The effects of various types of speech output on listener comprehension rates. Human–Computer Interaction–INTERACT '87, 573–579. https://doi.org/10.1016/b978-0-444-70304-0.50094-7

Pennington, J., Socher, R., Manning, C. D. (2014). GloVe: Global Vectors for Word Representation.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

Richardson, J. T. E. (2007). Measures of Short-Term Memory: A Historical Review. Cortex, 43(5), 635–650. https://doi.org/10.1016/S0010-9452(08)70493-3

Schiavetti N. 1992 Scaling procedures for the measurement of speech intelligibility. In Studies in speech pathology and clinical linguistics (ed. RD Kent), p. 11. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale-Revised. Psychological Corporation, New York.

Wester, M., Watts, O., & Henter, G. E. (2016). Evaluating comprehension of natural and synthetic conversational speech. Speech Prosody 2016, 766–770. https://doi.org/10.21437/SpeechProsody.2016-157