

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	We used custom experiment scripts written in JavaScript (version 1.5) and HTML (HTML5) to collect the data in this study. This code is publicly available at https://osf.io/pybwd/ .
Data analysis	We used custom code written in MATLAB (version R2022a) to analyze the behavioral data in this study. We also used the audio software Voicesauce (version 1.37) to analyze the acoustic features of our stimuli.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data analyzed in this study is available via the Open Science Framework at <https://osf.io/pybwd/>. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus used in this study as stimuli is available for download at <https://academictorrents.com/details/34e2b78745138186976cbc27939b1b34d18bd5b3>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Gender of participants was determined based on self-report. In total, 2054 participants reported their gender as "Female", 1866 reported their gender as "Male", and 345 answered "Other/Prefer not to answer". These values are reported in the manuscript. Gender was not considered in the study design, nor were any gender-based analyses performed, as we hypothesized that our results would be consistent across participants and would not vary based on gender.

Reporting on race, ethnicity, or other socially relevant groupings

Race and ethnicity of participants was determined based on self-report. Participants reported their race from the options "American Indian/Native American", "Asian", "Native Hawaiian/Pacific Islander", "Black/African American", "White", "More than one race", or "Prefer not to answer". They reported their ethnicity as "Hispanic/Latinx" or "Not Hispanic/Latinx". Race and ethnicity were not considered in the study design, as we hypothesized that our results would be consistent across participants and would not vary based on these factors.

Population characteristics

Age of participants was determined based on self-report; all participants were over the age of 18. Mean age in the memory experiment was 36.9 (SD = 11.7) and mean age in the attribute rating experiment was 38.1 (SD = 11.1).

Recruitment

Participants were recruited through the online platform Amazon Mechanical Turk (AMT). Participants were required to have previously completed ≥ 50 AMT Human Intelligence Tasks ("HIT"s), have an AMT approval rating of $\geq 95\%$, and reside in the U.S. The demographics of Amazon Mechanical Turk workers in the U.S. have been shown to roughly match that of the U.S. population; therefore, self-selection bias in our sample is not likely to significantly impact results.

Ethics oversight

University of Chicago Social & Behavioral Sciences Institutional Review Board (Study number IRB19-1395)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences

☒ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Quantitative experimental study type

Research sample

The research sample consisted of 4255 Amazon Mechanical Turk (AMT) workers living in the U.S. 48% were female, 44% were male, and 8% marked "Other/Prefer not to answer". All participants were over 18 years old. The demographics of AMT workers in the U.S. have been shown to roughly match that of the U.S. population, more so than in most in-person studies; therefore, the sample is relatively representative. We chose this study sample following the methodology of many previous online behavioral studies on memorability/ memory conducted by our lab and others. Running our study on AMT allows us to collect large amounts of representative data quickly.

Sampling strategy

The sampling procedure used in this study is convenience sampling, as participants signed up to take studies on AMT, and chose to participate specifically in our study. However, as we collected data from thousands of participants whose demographics roughly match that of the U.S. population, sampling can be considered relatively random. The total sample size for the three memory experiments ($N = 3352$) was chosen so that each target stimulus was heard by an average of at least 40 participants, as previous work finds that a stable measure of memorability is reached at 40 participant ratings/ stimulus (Isola et al., 2014). Thus, we used the calculations: $378 \text{ stimuli} * (40 \text{ participants} / 17 \text{ target stimuli})$ in Exp. 1-2, and $378 \text{ stimuli} * (40 \text{ participants} / 22 \text{ target stimuli})$ in Exp. 3. We rounded the resulting values up to account for the high number of participants who fail our internal vigilance checks and who do not completely finish the experiment. The sample size for the attribute rating experiment ($N = 903$) was chosen to collect 30 ratings of all attributes for each stimulus, following the procedure in Bainbridge et al., 2013. As participants were able to rate more than one voice clip, data collection was stopped once we reached 30 ratings/voice clip.

Data collection

As this experiment was run online, participants performed the experiment on their own computer. Following their completion of the experiment, data was saved and accessed by the researchers through the AMT requestor site. No one was involved besides the participants and the researchers. Researchers were blinded to the experimental conditions, but not the study hypothesis.

Timing

Data was collected between May 2022 and November 2023. Specifically, data in Exp. 1 was collected from May 2022-Jan 2023, data in Exp. 2 was collected from April-July 2023, data in Exp. 3 was collected from Oct-Nov 2023, and data in the attribute rating experiment was collected from Feb-March 2023.

Data exclusions

In the memory experiments, 4389 participants were excluded from the analyses. These included participants who missed more than 70% of filler repeats or had false alarms on more than 70% of the initial presentation of fillers, as this suggested that they did not

provide effortful performance on the task. An additional 3229 participants were excluded and replaced with new participants for failing one of five exclusion criteria: failing the common sense question check, failing the noise level check, failing the sentence transcription check, responding to > 95% of trials, or not starting the main task. In the attribute rating experiment, 244 participants who did not correctly answer a catch question were replaced with new participants. All exclusion criteria were pre-established.

Non-participation

No participants dropped out/ declined participation.

Randomization

Participants were not allocated into experimental groups. Each participant heard a randomized sequence of a pseudo-randomly selected subset of all stimuli. Stimuli in each participant's sequence were pseudo-randomly selected as either target or filler stimuli, to ensure that each stimulus was heard as a target by about 40 participants.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Included in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Included in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.