Online Supplemental Materials

Features in Random Item Generation (RIG)

For each sequence of characters, we obtain the frequency histograms for each character (*unigraph*) and for each pair of adjacent characters (*bigraph*) and normalize these, giving discrete probabilities distributions P_1^{uni} and P_1^{bi}, respectively. The means of these distributions are μ_1^{uni} and μ_1^{bi}, respectively, and their standard deviations are σ_1^{uni} and σ_1^{bi}, respectively. These parameters are necessary to measure the shapes of these distributions, as mentioned in the main article text. The specific formulae for those measures are:

$Entropy = \sum_i P_1(i) \log 1/P_1(i)$),
$Skewness = E\left[\left(\frac{X-\mu_1}{\sigma_1}\right)^3\right]$, and
$Kurtosis = \frac{E[(X-\mu_1)^4]}{(E[(X-\mu_1)^2])^2}$)

where i is either a character (unigraph, e.g., "A") or a pair of adjacent characters (bigraph, e.g., "AB"). We also compare the distributions P_1^{uni} and P_1^{bi} against the respective unigraph and biograph distributions for all *other* speakers to examine deviation from group means. Therefore, we compute group probability distributions P_m (with mean μ_m and standard deviation σ_m) and measure the amount of deviation between distributions according to:

Kullback–Leibler divergence $KL = \sum_i \ln\left(\frac{P_m(i)}{P_1(i)}\right) P_m(i)$),
$Cohen's\ d = \mu_1 - \mu_m/\sigma_m$),

and the *F* test statistic, where *i* retains its earlier meaning.

The recurrence quantification analysis consists of several features, namely, given a recurrence plot $\boldsymbol{R}_{i,j}^{m,\epsilon} = \Theta(\epsilon_i - \|\vec{x_i} - \vec{x_j}\|)$, where ϵ is an empirical threshold and $\vec{x_i}$ and $\vec{x_j}$ are phase–space trajectories in an *m*-dimensional space,

$$Recurrence\ rate\ RR = \frac{1}{N^2} \sum_{i,j=1}^{N} \boldsymbol{R}_{i,j}^{m,\epsilon}$$

$$Determinism\ DET = \frac{\sum_{l=l_{\min}}^{N} l \cdot P^{\epsilon}(l)}{\sum_{i,j=1}^{N} \boldsymbol{R}_{i,j}^{m,\epsilon}}$$

where $p^{\epsilon}(l)$ is the frequency distribution of the lengths *l* of diagonal structures and *N* is the number of diagonal lines. Similarly,

$$L_{\max} = \max_{l_i} l_i \quad i = 1 \dots N$$

$$Entropy = -\sum_{l=l_{\min}}^{N} p(l) \ln p(l)$$

where $p(l) = \frac{P^{\epsilon}(l)}{} \sum_{l = l_{\min}}^{N} P^{\epsilon}(l)$ and

$$Laminarity = \frac{\sum_{v = v_{\min}}^{N} vP^{\epsilon}(v)}{\sum_{v = 1}^{N} vP^{\epsilon}(v)},$$

$$Trapping\ time = \frac{\sum_{v = v_{\min}}^{N} vP^{\epsilon}(v)}{\sum_{v = v_{\min}}^{N} vP^{\epsilon}(v)}$$

Note that *trapping time* is identical to *laminarity* except it is normalized relative to shorter periods—this does not relate to the rate of symbol generation in RIG. All recurrence quantification analysis (RQA) measures are derived from the Cross-Recurrence Plot Toolbox, Version 5.18 (R29.3), available for free online at http://tocsy.pik-potsdam.de/CRPtoolbox/, which includes additional documentation on these features.

Results

Relative error for predictions is computed according to the actual (X_{actual}) and predicted (X_{pred}) values of a variable:

$$RelError = \frac{E[X_{actual} - X_{est}]^2}{E[X_{actual}]^2} \times 100$$

Among the RIG features, several features varied significantly across age. Specifically, where d is Cohen’s effect size: *NumChar* ($t(29) = -4.1, p < .001, \mathrm{CI} = [-94, -31],\ d = -1.58$), *determinism* ($t(29) = -5.1, p < .001, \mathrm{CI} = [-0.014, -0.006],\ d = -1.98$), *maximal vertical line length* ($t(29) = -3.9, p < .001, \mathrm{CI} = [-94, -30], d = -1.25$), *transitivity* ($t(29) = -3.9, p < .001, \mathrm{CI} = [-0.022, -0.0069], d = -1.51$), *unigraph Cohen* ($t(29) = -4.1, p < .001, \mathrm{CI} = [-1.2, -0.4], d = -1.57$), *bigraph entropy* ($t(29) = -4.6, p < .001, \mathrm{CI} = [-1.2, -0.44], d = -1.78$), *bigraph Cohen* ($t(29) = -4.1, p < .001, \mathrm{CI} = [-0.4, -0.13], d = -1.59$), *bigraph skewness* ($t(29) = 4.5, p < .001, \mathrm{CI} = [0.51, 1.4], d = 1.74$), and *bigraph kurtosis* ($t(29) = 4, p < .001, \mathrm{CI} = [3.6, 11], d = 1.54$) all were significantly different across age groups, with several other features having $p < .05$.

In our experiments, the mixture density network is trained (over 200 iterations, determined empirically) with vectors of RIG features (one per subject, of dimensionality $NR = 23$) and outputs a single Gaussian, given a hidden layer of 34 ($= \frac{NR+3}{2}$, determined empirically) neural units. In this case, the mean and variance of the Gaussian indicates a likelihood function over possible values of the predicted variable.

As mentioned in the main article text, we evaluate which features are affected by age using a heteroscedastic two-tailed *t* test across younger and older speakers. In order to emphasize the validity of these tests, we also compute the Kolmogorov–Smirnov test of normality, since the *t* test assumes Gaussian distributions. Those data are presented in Supplemental Table 1, below.

Because root-mean-square error (RMSE) and relative error are our primary objective evaluation metrics in estimating measures in different tests given RIG measures as input, those results are presented in Supplemental Table 2, below, across all predicted features. Although the results from standard multilinear regression are often relatively good (i.e., < 10% relative error), it is important to note that the neural network model achieves much higher accuracy, often by an order of magnitude.

Supplemental Table 1. One-sample Kolmogorov–Smirnov (KS) test of normality (with *p*-value and *k*-statistic) and two-tailed *t* test between age groups with unequal variances, over all features grouped by test.

Feature		**KS test**		H_0 = **ages equal?**
		p	***k***	
PPVT-4	Raw scores	.2	1	*t*(29) = 1.3, *p* = .2, CI = [–2.8, 13]
	Standard scores	.72	1	*t*(29) = 0.36, *p* = .72, CI = [–7.7, 11]
Stroop	**Interference**	**< .001**	**1**	***t*(29) = 5.1, *p* < .001, CI = [178, 412]**
	Facilitation	.12	.52	*t*(29) = –1.6, *p* = .12, CI = [–1.4e+02, 17]
	Total Stroop Effect (IC-C)	**< .001**	**.97**	***t*(29) = 4.6, *p* < .001, CI = [199, 512]**
RBANS	Immediate Memory Index Score	.69	1	*t*(29) = –0.4, *p* = .69, CI = [–12, 8.1]
	Visuospatial/Constructional Index Scores	.032	1	*t*(29) = 2.3, *p* = .032, CI = [1.3, 28]
	Language Index Scores	.051	1	*t*(29) = –2, *p* = .051, CI = [–17, 0.054]
	Attention Index Scores	.24	1	*t*(29) = 1.2, *p* = .24, CI = [–4.5, 17]
	Delayed Memory Index Scores	.62	1	*t*(28) = 0.5, *p* = .62, CI = [–7.8, 13]
	Sum of Index Scores	.54	1	*t*(28) = 0.62, *p* = .54, CI = [–26, 49]
	Total Scale	.53	1	*t*(28) = 0.63, *p* = .53, CI = [–7.3, 14]
RBANS sub scores	**List Learning**	**< .001**	**1**	***t*(29) = –3.7, *p* < .001, CI = [–8.4, –2.5]**
	Story Memory	.022	1	*t*(29) = –2.4, *p* = .022, CI = [–5.3, –0.44]
	Figure Copy	.23	1	*t*(29) = 1.2, *p* = .23, CI = [–0.85, 3.3]
	Line Orientation	.52	1	*t*(29) = –0.65, *p* = .52, CI = [–2.8, 1.5]
	Picture Naming	.65	1	*t*(29) = –0.46, *p* = .65, CI = [–0.87, 0.55]
	Semantic Fluency	.0045	1	*t*(29) = –3.1, *p* = .0045, CI = [–9.9, –2]
	Digit Span	.074	1	*t*(29) = 1.9, *p* = .074, CI = [–0.17, 3.6]
	Coding	< .001	1	*t*(29) = –5.9, *p* < .001, CI = [–24, –12]
	List Recall	.0031	.98	*t*(29) = –3.2, *p* = .0031, CI = [–4.3, –0.96]
	List Recognition	.0095	1	*t*(29) = –2.8, *p* = .0095, CI = [–2.2, –0.33]
	Story Recall	.048	.98	*t*(29) = –2.1, *p* = .048, CI = [–3.2, –0.017]
	Figure Recall	.4	1	*t*(28) = 0.86, *p* = .4, CI = [–1.7, 4.1]
RIG	**NumChar**	**< .001**	**1**	***t*(29) = –4.1, *p* < .001, CI = [–94, –31]**
	Recurrence rate	.078	.79	*t*(29) = –1.8, *p* = .078, CI = [–0.017, 0.00094]
	Determinism	**< .001**	**.83**	***t*(29) = –5.1, *p* < .001, CI = [–0.014, –0.006]**
	<L>	.014	1	*t*(29) = –2.6, *p* = .014, CI = [–1.1, –0.14]
	LMAX	.0031	1	*t*(29) = –3.2, *p* = .0031, CI = [–18, –4]
	Entr	.0039	1	*t*(29) = –3.1, *p* = .0039, CI = [–0.39, –0.082]
	Lam	.037	.83	*t*(29) = –2.2, *p* = .037, CI = [–0.017, –0.00058]

Feature	**KS test**		H_0 **= ages equal?**
	p	***k***	
TT	.015	1	*t*(29) = –2.6, *p* = .015, CI = [–1.8, –0.21]
Max. vertical line length	**< .001**	**1**	***t*(29) = –3.9, *p* < .001, CI = [–94, –30]**
T1	.033	.88	*t*(29) = 2.2, *p* = .033, CI = [0.0013, 0.029]
T2	.029	1	*t*(29) = –2.3, *p* = .029, CI = [–1.8, –0.1]
RTE	.043	.73	*t*(29) = 2.1, *p* = .043, CI = [0.0044, 0.26]
Clust	.17	.82	*t*(29) = –1.4, *p* = .17, CI = [–0.0055, 0.001]
Transitivity	**< .001**	**.8**	***t*(29) = –3.9, *p* < .001, CI = [–0.022, –0.0069]**
RIGentropy	.1	1	*t*(29) = –1.7, *p* = .1, CI = [–0.15, 0.014]
RIGKL	.007	.52	*t*(29) = 2.9, *p* = .007, CI = [0.031, 0.18]
Unigraph Cohen	**< .001**	**.15**	***t*(29) = –4.1, *p* < .001, CI = [–1.2, –0.4]**
RIGFtestP	.41	.5	*t*(29) = –0.83, *p* = .41, CI = [–7.6e-20, 3.2e-20]
RIGskewness	.43	.39	*t*(29) = 0.81, *p* = .43, CI = [–0.19, 0.43]
RIGkurtosis	.16	.98	*t*(29) = 1.4, *p* = .16, CI = [–0.14, 0.82]
Bigraph entropy	< .001	1	*t*(29) = –4.6, *p* < .001, CI = [–1.2, –0.44]
Bigraph Cohen	< .001	.38	*t*(29) = –4.1, *p* < .001, CI = [–0.4, –0.13]
Bigraph skewness	< .001	.98	*t*(29) = 4.5, *p* < .001, CI = [0.51, 1.4]
Bigraph kurtosis	< .001	1	*t*(29) = 4, *p* < .001, CI = [3.6, 11]

Note. PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); CI = confidence interval; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status (Randolph, 1998, 2008, 2012); NumChar = number of characters; <L> = mean diagonal-line length; LMAX = maximal diagonal-line length; RIG = random item generation; Entr = entropy; Lam = laminarity; TT = trapping time; Max. = maximum; T1 = recurrence time of first type; T2 = recurrence time of second type; Clust = clustering coefficients.

References

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. San Antonio, TX: Pearson.

Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status*. San Antonio, TX: Pearson.

Randolph, C. (2008). *Repeatable Battery for the Assessment of Neuropsychological Status*. Toronto, Ontario, Canada: Psychological Corporation.

Randolph, C. (2012). *Repeatable Battery for the Assessment of Neuropsychological Status*. San Antonio, TX: Pearson.

Supplemental Table 2. Root-mean-square error (RMSE) and relative error (Rel. error) in predicting the indicated feature from random item generation (RIG) measurements across a standard linear regression and a mixture-density network.

Feature		**Regression**		**Mixture-density network**	
		RMSE	**Rel. error (%)**	**RMSE**	**Rel. error (%)**
PPVT	Raw Scores	4.085822	1.932477	1.134743	0.536701
	Standard Scores	4.621449	4.193036	1.654792	1.501391
Stroop	**Interference**	81.35883	26.31997	2.922117	0.945319
	Facilitation	17.27275	16.82317	0.778355	0.758096
	Total Stroop Effect (IC-C)	64.08611	17.85691	3.652135	1.017628
RBANS	Immediate Memory Index Score	8.58571	8.463277	0.076863	0.075766
	Visuospatial/Constructional Index Scores	6.552457	6.851812	0.050156	0.052448
	Language Index Scores	3.10128	3.10935	0.006363	0.00638
	Attention Index Scores	5.066363	4.446502	0.154897	0.135946
	Delayed Memory Index Scores	3.498282	3.55242	0.136897	0.139016
	Sum of Index Scores	27.7323	5.562196	0.019479	0.003907
	Total Scale	1.138312	1.117996	0.034316	0.033703
RBANS sub scores	List Learning	2.391059	7.913112	0.248867	0.823616
	Story Memory	0.855803	4.616994	0.079013	0.426272
	Figure Copy	0.413054	2.422306	0.038035	0.223052
	Line Orientation	2.580822	14.32507	0.061148	0.339407
	Picture Naming	0.535071	5.792674	0.126956	1.374421
	Semantic Fluency	3.255901	13.93974	0.078436	0.335815
	Digit Span	0.408282	3.04643	0.123305	0.920048
	Coding	2.778818	4.808992	0.036596	0.063332
	List Recall	0.340807	4.615788	0.131944	1.787005
	List Recognition	0.049491	0.25512	0.02466	0.12712
	Story Recall	0.437389	4.305006	0.162874	1.603091
	Figure Recall	2.662915	18.25359	0.037356	0.256062

Note. PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); RBANS = Repeatable Battery for the Assessment of Neuropsychological Status (Randolph, 1998, 2008, 2012).

References

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. San Antonio, TX: Pearson.

Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status*. San Antonio, TX: Pearson.

Randolph, C. (2008). *Repeatable Battery for the Assessment of Neuropsychological Status*. Toronto, Ontario, Canada: Psychological Corporation.

Randolph, C. (2012). *Repeatable Battery for the Assessment of Neuropsychological Status*. San Antonio, TX: Pearson.