Supporting Information

Stilp and Kluender 10.1073/pnas.0913625107

SI Methods

Procedure. Sentences were upsampled (44.1 kHz) and presented diotically at 72 dBA via circumaural headphones (Beyer-Dynamic DT-150). Listeners participated individually in single-subject soundproof booths. After informed consent, listeners read instructions on a computer screen explaining the nature of the experiment, stating some sentences were expected to be difficult to understand, and that guessing was encouraged because every word of their responses counted.

Experiments lasted 30–40 min. Each sentence was played once, after which listeners were prompted to type any words they understood. Participants first completed practice sentences arranged to progressively increase in predicted difficulty (Exp. 1: 12 sentences increasing in percentage replaced; Exp. 2: 14 sentences increasing in duration replaced and entropy level). Next, each listener heard experimental sentences (Exp. 1: 120, Exp. 2: 70), one per trial, without hearing any sentence more than once. Although all listeners in each experiment heard sentences in the same order, the order of experimental conditions was pseudorandomized. In every block of trials (Exp. 1: 5 blocks; Exp. 2: 10 blocks), listeners heard one sentence from each experimental condition in random order. Each sentence in each condition was presented multiple times across each group of listeners (Exp. 1: twice; Exp. 2: three times).

Intelligibility was scored as the percentage of words in each sentence correctly identified. Three raters, blind to experimental conditions and purpose, independently scored responses offline, ignoring minor errors in spelling, verb tense, and number for regular nouns and verbs that did not change pronunciation of the word. When stricter criteria are adopted by not permitting these errors, percentages correct change by <1% and all statistical comparisons remain the same. Scoring was highly consistent (interrater reliability, Exp. 1: $r_{\text{intraclass}} = 0.99$; Exp. 2: $r_{\text{intraclass}} = 0.97$). Percentages of words correctly identified across all trials in each condition were used for data analyses.

Statistical Analyses. ***Experiment 1.*** Results from experiment 1 were analyzed in a 4 (speech segment) by 6 (nominal level of segment replacement) repeated-measures analysis of variance (ANOVA). Nominal level of noise replacement was a significant factor ($F_{5,235} = 73.15$, $P < 0.001$, $\eta_p^2 = 0.61$), as was speech segment replaced ($F_{3,141} = 33.22$, $P < 0.001$, $\eta_p^2 = 0.41$). Paired t tests Bonferroni-corrected for multiple comparisons reveal that replacement of consonants (mean intelligibility: 76.63%, SEM = 1.86) or vowels (mean = 79.46%, SEM = 1.86) results in worse intelligibility compared with replacing CVs (mean = 84.39%, s.e.m. = 1.55) or VCs (mean = 83.21%, SEM = 1.64) ($\alpha = 0.01$). Critically, replacing consonants with noise resulted in worse performance than replacing vowels ($\alpha = 0.05$). This result is also observed when analyzing performance at maximum replacement (i.e., matching the experimental design of related studies (1–3); 100%-consonants-replaced intelligibility: 56.36%, SEM = 2.06; 100%-vowels-replaced: 63.34%, SEM = 2.36). Finally, the interaction between factors was also significant ($F_{15,705} = 6.92$, $P < 0.001$).

The relationship between duration of sentence replaced and intelligibility was assessed via linear regression. Mean proportion of sentence replaced (predictors; one value for each of 24 experimental conditions) were derived by multiplying mean segment duration (C duration: 77.2 ms, V = 104.0 ms, CV = 167.6 ms, VC = 177.4 ms) by percentage of segment replaced (50–100% in 10% steps for Cs and Vs, 25–50% in 5% steps for CVs and VCs) by mean sentence proportion occupied by that segment (Cs = 0.62, Vs = 0.38, CVs = 0.30, VCs = 0.31) by mean sentence duration (2,112.4 ms). Proportion of sentence replaced exhibited a strong linear relationship with intelligibility ($r^2 = 0.65$, $P < 0.001$).

Experiment 2. Control sentences were included primarily as reference for performance in experimental conditions; therefore, results were analyzed in a 2 (segment duration) by 3 (entropy) repeated measures ANOVA omitting control data. As expected, replacing longer segments (112-ms) with noise compromised performance more than replacing shorter segments (80-ms) ($F_{1,20} = 48.07$, $P < 0.001$, $\eta_p^2 = 0.71$). Most important is the inverse relationship between sentence intelligibility and amount of cochlea-scaled spectral entropy replaced ($F_{2,40} = 17.78$, $P < 0.001$, $\eta_p^2 = 0.47$). Paired t tests with Bonferroni correction for multiple comparisons reveal that intelligibility when low-entropy portions were replaced exceeded performance in medium-entropy ($\alpha = 0.05$) and high-entropy ($\alpha = 0.01$) conditions. The interaction between segment duration and entropy was significant ($F_{2,40} = 5.05$, $P < 0.05$). For 80-ms segments, intelligibility in the low-entropy condition was higher than in other conditions. For 112-ms segments, performance in the high-entropy condition was poorer than in other conditions. (Bonferroni-corrected $\alpha = 0.01$ for all contrasts). Interestingly, performance in the high-entropy, 80-ms condition (mean intelligibility: 69.02%, SEM = 2.79) and the low-entropy, 112-ms condition (mean = 71.08%, SEM = 2.23) was equivalent despite replacing less of the signal but more of overall sentence entropy in the former case (proportion sentence entropy replaced: 0.57, SEM = 0.01; mean duration sentence replaced = 718.86 ms, SEM = 16.32) and vice versa in the latter (entropy: 0.19, SEM = 0.01; duration = 800.00 ms, SEM = 19.78).

The correlation between intelligibility and proportion of cochlea-scaled spectral entropy replaced in sentences was assessed. Proportion of sentence entropy replaced by noise was calculated as the sum of Euclidean distances in replaced segments divided by the sum of Euclidean distances between all adjacent slices in the sentence. Proportions were averaged for all sentences in each condition, then entered as predictors in linear regression with intelligibility as the outcome variable. Intelligibility closely follows measures of entropy replaced ($r^2 = 0.80$, $P < 0.01$). Conversely, a regression analysis examining intelligibility as predicted by duration of signal replaced did not reach significance ($r^2 = 0.54$, n.s.).

Relative (mean-centered) proportions of replaced signals marked as consonants and vowels in TIMIT are shown in Fig. 5*A*. Using TIMIT demarcations, raw numbers of samples replaced were summed for each speech sound in each entropy condition, then divided by total number of samples for that speech sound across all sentences. Results are mean-centered to emphasize relative changes between entropy conditions. With each increase in spectral entropy, significantly more vowel and fewer consonant intervals were replaced. More consonants (0.69) than vowels (0.28) are replaced in low-entropy conditions, and more vowels (0.61) than consonants (0.38) are replaced in high-entropy conditions (proportions do not sum to 1 because segments not classified as consonants or vowels in TIMIT, e.g., pauses and epenthetic silence, were excluded from analyses).

1. Cole R, Yan Y, Mak B, Fanty M, BaileyT (1996) The contribution of consonants versus vowels to word recognition in fluent speech. *Proc Internatl Conf Acoustics Speech Signal Processing,* 853–856.
2. Kewley-Port D, Burkle TZ, Lee JH (2007) Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing impaired listeners. *J Acoust Soc Am* 122:2365–2375.
3. Fogerty D, Kewley-Port D (2009) Perceptual contributions of the consonant–vowel boundary to sentence intelligibility. *J Acoust Soc Am* 126:847–857.

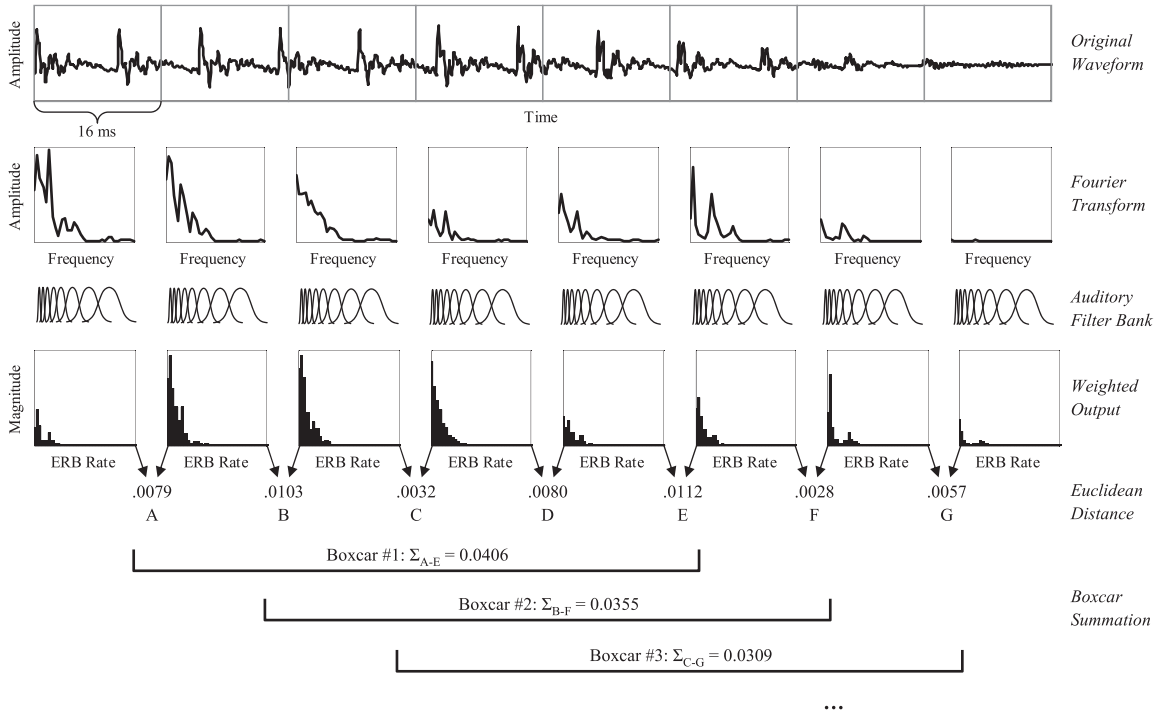

Fig. S1. Method by which measures of cochlea-scaled spectral entropy were calculated. See *Methods* for details.

Table S1. Table of speech sound classification for experiment 2 and VCV analyses according to manner of articulation (consonants) or vocal tract configuration (vowels)

	Phonetic symbols (TIMIT symbols)
Consonants	
Closure	(bcl, dcl, gcl, pcl, tcl, kcl, q)
Stops	/b/ (b), /d/ (d, dx), /g/ (g), /p/ (p), /t/ (t, q), /k/ (k)
Affricates	/č/ (ch), /j/ (jh)
Fricatives	/s/ (s), /z/ (z), /f/ (f), /v/ (v), /š/ (sh), /ž/ (zh), /θ/ (th), /ð/ (dh)
Laterals/ glides	/r/ (r), /l/ (l), /w/ (w), /y/ (y), /h/ (hh, hv), /ʔ/ (el)
Nasals	/m/ (m, em), /n/ (n, en, nx), /ŋ/ (ng, eng)
Vowels	
High	/i/ (iy), /u/ (uw, ux), /I/ (ih, ix), /U/ (uh), /ɜ^/ (er, axr), /ɛ/ (eh)
Low	/ʌ/ (ah, ax, ax-h), /c/ (ao), /æ/ (ae), /ɑ/ (aa)
Front	/i/ (iy), /I/ (ih, ix), /ɛ/ (eh), /æ/ (ae)
Mid	/ɜ^/ (er, axr), /ʌ/ (ah, ax, ax-h), /ɑ/ (aa)
Back	/u/ (uw, ux), /U/ (uh), /c/ (ao)
Diphthongs	/e^{y}/ (ey), /a^{y}/ (ay), /o^{y}/ (oy), /a^{u}/ (aw), /o^{u}/ (ow)

Phonetic symbols are denoted between slashes, and TIMIT symbols are denoted in parentheses.