The redundancy of phonemes in sentential context^{a)}

Christian E. Stilp

*Department of Psychology, University of Wisconsin–Madison,
1202 West Johnson Street, Madison, Wisconsin 53706
christian.stilp@gmail.com*

Abstract: Printed English is highly redundant as demonstrated by readers' facility at guessing which letter comes next in text. However, such findings have been generalized to perception of connected speech without any direct assessment of phonemic redundancy. Here, participants guessed which phoneme or printed character came next throughout each of four unrelated sentences. Phonemes displayed significantly lower redundancy than letters, and possible contributing factors (task difficulty, experience, context) are discussed. Of three models tested, phonemic guessing was best approximated by word-initial and transitional probabilities between phonemes. Implications for information-theoretic accounts of speech perception are considered.

© 2011 Acoustical Society of America

PACS numbers: 43.71.Sy [SGS]

Date Received: July 28, 2011 Date Accepted: September 9, 2011

1. Introduction

Language is highly redundant. In a classic example, Shannon^{1,2} measured the redundancy of English text. Incorporating probabilities of occurrence and higher-order transitional probabilities between characters yields redundancy estimates around 75% (≈ 1 bit/character). This approach has generated widespread interest in measuring redundancy of texts on different topics and in different languages.^{3–5} However, while stimulus redundancy is easily calculated, do perceivers exploit this redundancy? Shannon's² "guessing game" had participants guess which character came next throughout a passage of text. Following an incorrect guess, the correct answer was provided (single-guess procedure) or guessing continued until reaching the correct answer (multiple-guess). Consistency in participants' responses was taken to reflect stimulus redundancy. Response entropy was initially high (4.03 bits/character, 15% redundancy) but decreased with increasing context, reaching 1.30 bits/character (73% redundancy) when the preceding 100 characters were known.

Van Rooij and Plomp⁶ used the guessing game paradigm to measure linguistic entropy in written sentences. They next measured speech reception thresholds as a function of linguistic entropy across varying signal-to-noise ratios, reporting a positive correlation between thresholds and sentence entropy. Van Wijngaarden *et al.*⁷ showed that differences in speech reception thresholds across native versus non-native language speakers were well-captured by linguistic entropy of the written materials. Similarly, Müsch and Buus⁸ incorporated linguistic entropy as a key stage in their speech recognition sensitivity model.

While correlation between printed linguistic entropy and speech reception thresholds is notable, discrepancies between orthography and phonology (inventory size, statistical properties, etc.) complicate comparison between measures of perceptual sensitivity to redundancy in spoken language. Much like letters, phonemes in English

^{a)}Portions of this work were presented at the 160th Meeting of the Acoustical Society of America (November 2010, Cancún, Mexico).

occur with varying probability. Second-order redundancy in speech is conveyed by phonotactic probabilities, which play an important role in language acquisition^{9–12} and processing.^{13,14} Quantifying the degree to which listeners use this information, and how that compares to estimates from written language, has never been demonstrated to our knowledge. The present experiments investigate the extent to which participants exploit redundancy when asked to select the next speech sound or character in a sentence. Sentences are presented auditorily (Expt. 1a) or visually (Expts. 1b and 1c) to allow direct comparison of phonemic versus orthographic redundancy estimates for the same test items.

2. Methods

2.1 Participants

Thirty-nine undergraduate students from the University of Wisconsin participated in the experiment (19 in Expt. 1a, 10 each in Expts. 1b and 1c) with no one in multiple experiments. Nineteen listeners (Expt. 1a) guessed at auditorily presented sentences while remaining participants (Expts. 1b and 1c) guessed at visually presented (written) sentences. All reported being native speakers of English. Participants in Expt. 1a reported no known hearing impairments. While participants in Expts. 1b and 1c were not screened for normal/corrected-to-normal vision, none reported any difficulty viewing the experimental materials. All were compensated for their time with extra credit in an introductory psychology course.

2.2 Stimuli

Stimuli were drawn from the TIMIT sentence database.¹⁵ Auditory task demands imposed some constraints on sentence selection: no glottal stops (to avoid transcription errors), no sounds repeated across successive words (e.g., this sport, where only one /s/ is transcribed), and, citation form for clear pronunciation. Four sentences were selected: “The clerk’s eyes flickered” (S1), “Serve the coleslaw after I add the oil” (S2), “Only lawyers love millionaires” (S3), and “When all else fails use force” (S4). Each was spoken by a different male talker from the midwestern United States. Acoustic sentences had mean duration of 1771 ms (range = 1340–2430).

2.3 Procedure

Following acquisition of informed consent, participants read a set of instructions on a computer screen. Participants in Expt. 1a were also provided printed instructions highlighting discrepancies between orthography and phonology (e.g., differences in the number of letters per phoneme and vice versa, two vowel letters/sounds making up a diphthongal vowel, the letter “r” acting as a consonant /r/ or vowel /ɜ/). Experiments followed the multiple-guess procedure of Shannon’s² guessing game: on each trial, participants made guesses by clicking labeled buttons on the computer screen until the correct answer was selected. Feedback was provided after each guess: when incorrect, that menu button was replaced with dashes (“---”) to indicate it was unavailable for future guesses on that trial; when correct, the sentence was presented up to that guess, all menu options were restored, and the next trial began. Thus the first trial for each sentence offered no context and guesses approached chance performance. In Expt. 1a, menu buttons were labeled with orthographic approximations of phonemes (39 in all), and the final button, labeled “play sentence,” could be clicked by the listener at any point to hear the sentence played up to the most recent correct guess [as indicated by phonetic boundaries in TIMIT (Ref. 16)]. In Expts. 1b and 1c, buttons corresponded to letters of the alphabet with (1b; 27 characters total) and without (1c; 26 letters) the space between words as a response option.¹⁷ Sentence text up to the most recent correct guess was presented on the screen. Punctuation was removed with the exception of the apostrophe in S1 appearing only after the space (1b) or letter (1c) following “clerk’s” was correctly guessed. Participants were informed when a sentence was completed and the experiment was proceeding to the next randomly selected sentence.

All participants completed the experiment individually in a double-wall sound-proof booth. Speech sounds in Expt. 1a were presented diotically at 72 dB SPL via circumaural headphones (Beyer-Dynamic DT-150); headphones were unnecessary for Expts. 1b and 1c. Auditory and visual stimulus presentation and response recording were done in MATLAB. Each correct response in Expts. 1b and 1c subtended a mean visual angle of 0.63° . Experiment 1a lasted approximately 70 min, and other experiments around 30 min.

3. Results

Given that participants in Expt. 1a were guessing which sound came next and not assessing articulatory precision, scoring allowed for acceptable variants of correct answers (as listed in TIMIT transcription) where appropriate (e.g., /ð/ for /θ/; /s/ for /z/; /i/ or /ɪ/ for /ɪ/ following /ð/). Shannon² revealed a logarithmic relationship between performance (response uncertainty) and context (known letters), with large benefits of added context when little was available but asymptotic performance when larger amounts were available. Figure 1 depicts mean guesses per trial averaged across sentences. The relationship between context and the log-transformed number of guesses is well-captured by linear regression with a negative slope (Expt. 1a: $r = -0.55$, $P < 0.05$; 1b: $r = -0.56$, $P < 0.01$; 1c: $r = -0.41$, $P = 0.055$). Beyond some item-specific deviations, these functions are fairly well-defined.

3.1 Upper bound of entropy

Following Shannon,² upper bounds of entropy were calculated using Eq. (1):

$$H = -\sum q_i^N \log_2 q_i^N \quad (1)$$

where q_i^N is the proportion of trials where the participant was correct on the i th guess. Sentences in Expt. 1a conveyed mean entropy of 3.00 bits/phoneme (43% redundancy). However, Eq. (1) may underestimate entropy when the number of trials in a sentence is less than the number of speech sounds available for guessing. As a result, some q_i^N with zero proportion of correct guesses may be due to undersampling and not redundancy. To address this point, for each listener, guesses for all trials in the experiment (76 total, across all four sentences) were concatenated into a single vector and analyzed using Eq. (1). The average upper bound of entropy for the entire session was 3.75 bits/phoneme (29% redundancy).

Sentences in Expt. 1b conveyed mean entropy of 2.42 bits/character (49% redundancy) across sentences, and 2.86 bits/character (40% redundancy) across the session. Sentences in Expt. 1c conveyed mean entropy of 2.63 bits/letter (44% redundancy) across sentences, and 3.07 bits/letter (35% redundancy) across the session.

Fig. 1. Performance as a function of context for Experiments 1a (left), 1b (middle), and 1c (right). Mean number of guesses is on the ordinate (maximum = chance guessing for each experiment), and trial number is on the abscissa (maximum = length of shortest sentence for each experiment, so guessing is averaged across all four sentences). Dashed lines denote logarithmic fits to the data. Marked deviations from these fits are generally item-specific. Error bars are standard error of the mean.

Independent samples t -tests reveal that the upper bound of entropy in Expt. 1a is significantly higher than those observed in Expts. 1b (across sentences: $t_{27} = 7.15$, $P < 0.001$; across session: $t_{27} = 9.57$, $P < 0.001$) and 1c (across sentences: $t_{27} = 4.84$, $P < 0.001$; across session: $t_{27} = 7.91$, $P < 0.001$) despite testing the same four sentences.

3.2 Lower bound of entropy

Shannon² proposed a method for calculating lower bounds of entropy, but assumptions of ideal guessing performance and the monotonicity of q_i^N have drawn criticism.^{3,4,18} Instead, three psycholinguistic models were explored to assess how each compared to listener performance in Expt. 1a. For each model, on each trial, phoneme probabilities were derived from Vitevitch and Luce's¹⁹ Phonotactic Probability Calculator, ranked, and then "guesses" were made and entropy calculated using Eq. (1). Simple phoneme probability and position-specific phoneme probability were poor predictors of listener performance. A third model, combining word-initial and transitional probabilities between phonemes, was significantly correlated with listener performance on three of the four sentences ($r \geq 0.45$) and across the entire session ($r = 0.50$, $P < 0.005$). However, all models failed to achieve true lower bounds of entropy, producing estimates that exceeded listener performance at both sentence (≥ 3.39 bits/phoneme) and session levels (≥ 4.09 bits/phoneme). It is clear that participants employed higher-level linguistic knowledge (e.g., word-level and sentence-level cues, semantics, syntax) beyond that captured by phonemic probabilities, particularly when more context was available (Fig. 1), but formal analyses at this level are beyond the scope of the present effort.

4. Discussion

Phonemic versus orthographic redundancy in English was investigated using Shannon's² guessing game paradigm, revealing significantly lower redundancy estimates for heard sentences. While some of this difference may be attributable to the number of response options presented (worse performance with more response options²⁰), discrepancies in phonemic and orthographic segmentation and redundancy as well as participant experience are expected to be primary factors affecting performance. The discrete nature of printed text conveys all cues to the identity of a letter or character on a given trial. The same cannot be said for continuous speech, in which the acoustic cues to phonemic identity are spread across time (and potentially across multiple trials) due to coarticulation, increasing task difficulty. Further, experience breaking words into constituent letters, a skill developed through literacy training, far exceeds that of breaking words into individual phonemes. While this is expected to be particularly pronounced for a subject population with no assumed experience in linguistics or phonetics, such experience was not formally assessed. The degree to which redundancy estimates increase for participants with significant experience in English phonetics and phonology, and whether this redundancy approaches that of orthographic presentation, remains an intriguing question for future research.

Previous studies^{6,7} employed orthographic versions of the guessing game to measure linguistic entropy. This measure was correlated with speech reception thresholds, such that sentences with higher redundancy (predictability) corresponded to lower thresholds. These relationships are impressive given nontrivial differences between English orthography and phonology. The present results encourage investigating the degree to which phonemic redundancy correlates with other speech perception tasks and whether its predictive power matches or surpasses that of orthographic redundancy.

Contributions of contextual information to word and sentence perception are well-documented.^{21–24} In the present paradigm, higher-level linguistic knowledge such as sentence syntax or semantic plausibility contributes to higher redundancy estimates than those predicted by first- or second-order phonemic probability. This is also evident in improved performance as a function of context (Fig. 1). However, lower-level analyses of interest (phonological, orthographic) and higher-level influences on

performance cannot be isolated in the present paradigm. Testing individual or non-sense words in the guessing game paradigm would limit influences of higher-level predictability, but such tests have limited utility in understanding perception of running speech.

Item difficulty was consistent across participants in each experiment. This is evident in performance on the first trial of each sentence when no context was available. Upon identifying the first sound or character in a word, numbers of guesses generally decreased until that word was completed, as participants extrapolated information from earlier trials to aid guessing. While these four sentences do not fully represent the entire English language, consistency in redundancy estimates (all s.e. ≤ 0.12 bits) and clear trends in Fig. 1 suggest that they are sufficiently representative of its redundancy.²⁵

Redundancy estimates and listener performance share a strong relationship, but an important point merits discussion. Information theory¹ is agnostic with respect to the meaning of the message being transmitted. In the present experiments, redundancy estimates are agnostic to accuracy. By way of example, if the listener's first guess on every trial was correct, the sentence would be completely redundant. However, if the listener was always correct on the 10th (or any other) guess, Eq. (1) produces the same estimate of redundancy. Perfectly consistent responses achieve maximum information transmission even if those responses are incorrect²⁶ (e.g., always guessing incorrectly about the outcome of a coin flip still provides 100% information transmission). It is thus inappropriate to equate entropy with task difficulty or redundancy with accuracy. Sentences can possess redundancy not necessarily reflected by understanding (i.e., chance guessing), just as performance can reflect systematicity not superficially conveyed by the stimulus materials (e.g., lexical, syntactic, semantic redundancy). The present experiments encourage consideration of both information (entropy, redundancy) and meaning (accuracy) in information-theoretic approaches to speech perception.

Acknowledgments

The author wishes to thank Keith Kluender and two anonymous reviewers for comments on a previous version of this manuscript and Kyira Hauer, Raymond Kluender, Brittany Thomson, and Nora Brand for assistance in pilot studies and data collection. Funding was provided by Grant No. F31 DC009532 from the National Institute on Deafness and Other Communication Disorders.

References and links

- ¹C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).
- ²C. E. Shannon, "Prediction and entropy of printed English," *Bell Syst. Tech. J.* **30**, 50–64 (1951).
- ³T. M. Cover and R. C. King, "A convergent gambling estimate of the entropy of English," *IEEE Trans. Inf. Theory* **24**, 413–421 (1978).
- ⁴L. B. Levitin and Z. Reingold, "Entropy of natural languages: Theory and experiment," *Chaos, Solitons Fractals* **4**, 709–743 (1994).
- ⁵C. Papadimitriou, K. Karamanos, F. K. Diakonos, V. Constantoudie, and H. Papageorgiou, "Entropy analysis of natural language written texts," *Phys. A* **389**, 3260–3266 (2010).
- ⁶J. C. G. M. Van Rooij and R. Plomp, "The effect of linguistic entropy on speech perception in noise in young and elderly listeners," *J. Acoust. Soc. Am.* **90**, 2985–2991 (1991).
- ⁷S. J. Van Wijngaarden, H. J. M. Steeneken, and T. Houtgast, "Quantifying the intelligibility of speech in noise for non-native listeners," *J. Acoust. Soc. Am.* **111**, 1906–1916 (2002).
- ⁸H. Msch and S. Buus, "Using statistical decision theory to predict speech intelligibility. I. Model structure," *J. Acoust. Soc. Am.* **109**, 2896–2909 (2001).
- ⁹P. W. Jusczyk, A. D. Friederici, J. M. I. Wessels, V. Y. Svenkerud, and A. M. Jusczyk, "Infants' sensitivity to the sound patterns of native language words," *J. Mem. Lang.* **32**, 402–420 (1993).
- ¹⁰P. W. Jusczyk, P. A. Luce, and J. Charles-Luce, "Infants' sensitivity to phonotactic patterns in the native language," *J. Mem. Lang.* **33**, 630–645 (1994).
- ¹¹H. L. Storkel, "Learning new words: Phonotactic probability in language development," *J. Speech Lang. Hear. Res.* **44**, 1321–1337 (2001).

- ¹²K. Graf Estes, J. Edwards, and J. R. Saffran, "Phonotactic constraints on infant word learning," *Infancy* **16**, 180–197 (2011).
- ¹³M. S. Vitevitch and P. A. Luce, "When words compete: Levels of processing in spoken word perception," *Psychol. Sci.* **9**, 325–329 (1998).
- ¹⁴M. S. Vitevitch and P. A. Luce, "Probabilistic phonotactics and spoken word recognition," *J. Mem. Lang.* **40**, 374–408 (1999).
- ¹⁵J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, *NTIS Order No. PB91-505065: DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM* (National Institute of Standards and Technology, 1990).
- ¹⁶J. R. Glass and V. W. Zue, "Multi-level acoustic segmentation of continuous speech," in *Proceedings of the 1988 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-88)*, (April 1988, New York), pp. 429–432.
- ¹⁷Published reports using Shannon's (Ref. 2) guessing game vary as to whether the space between words is included as a response option. Both versions were tested here to ensure that this factor does not principally explain differences between phonemic and orthographic redundancy estimates.
- ¹⁸D. Kersten, "Predictability and redundancy of natural images," *J. Opt. Soc. Am. A* **4**, 2395–2400 (1987).
- ¹⁹M. S. Vitevitch and P. A. Luce, "A web-based interface to calculate phonotactic probability for words and nonwords in English," *Behav. Res. Methods Instrum. Comput.* **36**, 481–487 (2004).
- ²⁰I. Pollack, "Message uncertainty and message reception," *J. Acoust. Soc. Am.* **31**, 1500–1508 (1959).
- ²¹G. A. Miller, G. A. Heise, and W. Lichten, "The intelligibility of speech as a function of the context of the test material," *J. Exp. Psychol.* **41**, 329–335 (1951).
- ²²D. N. Kalikow, K. N. Stevens, and L. L. Elliott, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**, 1337–1351 (1977).
- ²³A. Boothroyd and S. Nittrouer, "Mathematical treatment of context effects in phoneme and word recognition," *J. Acoust. Soc. Am.* **84**, 101–114 (1988).
- ²⁴A. W. Bronkhorst, A. J. Bosman, and G. F. Smoorenburg, "A model for context effects in speech recognition," *J. Acoust. Soc. Am.* **93**, 499–509 (1993).
- ²⁵Kersten (Ref. 18) used the guessing game paradigm to measure redundancy in seven natural images. Consistent performance across stimuli was taken to indicate their representativeness of redundancy in natural images.
- ²⁶F. Attneave, *Applications of Information Theory to Psychology* (Holt, New York, 1959).