The subjective estimation of relative syllable frequency

DAVID C. RUBIN*

622 William James Hall, Harvard University, Cambridge, Massachusetts 02138

Ss are able to judge the relative frequency of occurrence in English of nonmorphemic syllables independent of phoneme frequency. The results support a theory of speech perception based on the syllable as a unit as opposed to the phoneme.

A common assumption of most studies of speech perception is that the best candidate for a perceptual unit is the phoneme (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The syllable has theoretical advantages as a perceptual unit, but has met with the objection that there are too many possible syllables in English (Liberman et al, 1967, pp. 451-452). Before providing experimental evidence against this objection, the advantages of using the syllabic level will be reviewed.

There are three main lines of evidence for using the syllable as opposed to the phoneme. First, it is compatible with the psychological evidence. Second, it allows for easier utilization of contextual information. Third, it allows for a simpler model of speech perception.

Two recent reviews consider the diverse psychological evidence for favoring the syllable as a perceptual unit. Lenneberg (1967, pp. 107-119) summarized the evidence from delayed auditory feedback, signal-switching between right and left ear, rate of interruption of speech, EEG, and paced speech during thalamic stimulation studies, which all point to a rhythmic time constant equal to the 6/sec rate of production of syllables. Massaro (1972) reviewed the literature on preperceptual auditory memory. Considering evidence from many areas, including the perception of repeating white noise segments, recognition under several masking conditions, and judgments of temporal order, he concludes that there exists a preperceptual auditory image of sufficient duration to allow for the 100-300 msec necessary to process incoming sounds. Both reviewers interpret their findings as evidence for the syllable as a perceptual unit.

The syllable allows for easier utilization of syntactic and semantic information which are of utmost importance in normal speech perception (International Joint Council on Artificial Intelligence, 1973, pp. 173-222; Wanner, 1973). In this respect, the syllable level has a distinct advantage over the phonetic. Phonemes are, with a few exceptions, totally neutral with respect to syntax and semantics; syllables are not.

Syllables are often morphemes, and all common words are monosyllables. For example, the 57 words which occur most frequently in English (Kučera & Francis, 1967) are all monosyllables. These 57 words alone accounted for 42% of the sampled text and probably more of actual speech. As most function words and inflections, which can be viewed as providing the syntactic frames of utterances, are monosyllables, the syllabic level allows for an interaction with syntactic structure. Thus, the syllable allows for a direct mix of auditory information from a distinctive feature-type level and semantic and syntactic information from "higher" levels.

The syllable also allows for a simpler theory of speech perception. In the field of perception, passive theories which extract all necessary information from each segment of the stimulus are distinguished from active theories which go beyond the information given in each segment and make use of information external to the stimulus. As has been argued in the previous paragraph, a totally passive theory of speech perception is not feasible, but the degree to which a theory is passive can be increased if the size of the unit of analysis is increased. For example, the sensitivity of vowels to their immediate context can be viewed as evidence for a highly active system of phoneme perception or as evidence for a more passive system using units which combine the vowel with its immediate context. Increasing the size of the perceptual unit from the phoneme to the syllable does, in fact, increase in the stimulus invariance. Syllables are less sensitive to their contex than are phonemes. For instance, while one cannot synthesize speech by splicing together phoneme-length segments, attempts with syllable-length segments have been more successful (Liberman et al, 1967, pp. 441 and 451).

There is, however, a limit on the size of a perceptual unit. As the size of the units increases indefinitely, the number of units becomes too large. Consider, for example, a theory using all well-formed sentences of English as units. The purpose of this paper is to demonstrate that, in fact, there are not too many syllables to allow for a reasonable theory of speech perception.

Using a 42-sound basis (41 phonemes plus "the"), Dewey found that a 100,000-word sample of written

English contained a highly skewed distribution of about 4,400 different syllables. Since a modest vocabularly contains twice this number of words, the pure number of syllables itself does not seem to present any storage problems. A further objection is that the high decision rate required of the listener when he makes choices among the 4,400 syllables is beyond his limited capacity to process information. If, however, the listener can employ information about the relative frequency of occurrence of the syllables by using either a biased passive (Broadbent, 1973) or an active hypothesis-forming strategy, then this would not be the case.

Based on Dewey's (1923) count, a syllable contains 8.5 bits of information, or the equivalent of a syllable in a language comprising 373 equiprobable syllables. A phoneme contains 5.2 bits of information. [As Dewey did not list the syllables with frequencies of 10 or less, it was necessary to assume a log-normal distribution (see Carroll, 1967) to estimate the number of different syllables at each frequency between 1 and 10 in order to calculate the information of the distribution.] Using Dewey's (1923) figures of 1.43 syllables per word and 3.65 phonemes per word, the channel capacity necessary for transmitting speech at the normal rate of 180 words/min is calculated as 38 bits/sec for syllables vs 57 bits/sec for phonemes.

Dewey's (1923) count appears to be a carefully executed tabulation of a representative sample of American English prose. The differences in the relative frequency values of the syllables caused by using a count based on written as opposed to spoken English is difficult to assess, but it is fairly certain that the information calculation for the syllables is slightly higher than it would be for speech. It is probably that a syllable count using a larger sample size than Dewey's (1923) might increase the number of distinct syllables. The increase in the storage requirements would not be critical, however, and the information rate would change little since the additional syllables would all be of low probability.

To summarize, there are surely not too many syllables for long-term memory capacity, and if the listener could make use of the relative frequency of the individual syllables to aid in his perception, there would not be too many syllables in decision-rate terms either.

The purpose of the present experiment is to demonstrate that the listener does, in fact, have access to relative frequency information. Magnitude estimation (Stevens, 1971) was chosen to measure this information. The method requires little training on the part of the S and is easy to implement. The Ss are simply asked to assign each stimulus a number proportional to the variable of interest. If the Ss could demonstrate knowledge of the relative frequency of syllables in English in this artificial manner, then it seems reasonable to assume that the human listener could use such information in speech perception. Such a demonstration

is not intended as direct evidence for speech perception using syllables, but rather as evidence that the major empirical argument against the use of syllables as a basic unit of speech perception is untenable. Independent evidence for using the syllabic level has been given.

METHOD

Subjects

Thirteen native speakers of English enrolled in a psychology course at Harvard University volunteered to be Ss.

Procedure

On two separate occasions, the 13 Ss were presented with a tape-recorded list of 10 syllables and asked to "assign a number proportional to the relative frequency of occurrence of each syllable in English." The Ss were given prior training in the magnitude estimation of the size of circles and examples of clusters of phonemes which are and are not considered to be syllables. Following Stevens's (1971) suggestion, no standard was assigned, leaving each S free to choose his own modulus.

Materials

The two lists of 10 syllables were chosen from the Dewey (1923) count. Since Ss can magnitude-estimate the frequency of occurrence of words in English (Shapiro, 1969), an attempt was made to use nonmorphemic syllables where possible. The first list consisted of open syllables formed by a consonant followed by the vowel /I/. As only the initial consonant is varied, this list provides a good control for many possible cues to frequency, including stress and the length of the syllables in phonemes. If Ss were estimating syllable frequency from phoneme frequency, this list should clearly reveal that fact. The particular syllables were chosen to cover the log-frequency range in as close to equal intervals as possible, while providing as much contrast as possible to a strategy based on predicting the syllable frequency from the independent phoneme frequencies. That is, where a choice was possible, syllables were chosen to provide maximum contrast between actual syllable frequency and the syllable frequency predicted on the basis of the product of the independent probability of the component phonemes. The test stimuli, written in the Simplified Spelling Board notation used by Dewey (1923) are these: /si/, /di/, /i/, /li/, /bi/, /fi/, /mi/, /vi/, /ħi/, and /zi/. In the International Phonetic Association transcription, they are these: /sI/, /dI/, /I/, /lI/, /bI/, /fI/, /mI/, /vI/, /θI/, and /zI/.

The second list was chosen by first finding the 10 log-frequency values, including the 2 extreme values, which evenly cover the total log-frequency range. The first nonmorphemic, closed syllable encountered at those 10 values in the Dewey (1923) count were chosen. If there was no such syllable at a given value, a 10% deviation was allowed. If still no syllable met the requirements, the closed preference was dropped, and then, if necessary, the nonmorphemic requirement was dropped. The resulting list in the Simplified Spelling Board notation is /pʋr/, /tên/, /zes/, /han/, /mem/, /ʃən/, /bl/, /ri/, /ħi/, and /i/. In the International Phonetic Association transcription, it is /pə/, /ten/, /zɛs/, /hæn/, /mɛm/, /ʃn̩/, /bl̩/, /rI/, /ðə/, and /I/. Dewey (1923) combined all forms of the word "the," making it the most common syllable in English. Under the conditions used, most Ss did not consider the utterance /ʒə/ as representative of the word "the," but just as the particular syllable /ðə/. For this reason, this syllable has been excluded from the analysis.

The stimulus tape was recorded in a sound-deadening chamber by a trained phonetician at the rate of one syllable every 5 sec. The pseudorandom order of presentation of the syllables in the list was identical to that given above.

RESULTS AND DISCUSSION

The results of the magnitude estimation for the two lists are given in Fig. 1. With magnitude estimation, the Os' judgments are typically described by a power function of the stimulus scale. Logarithmic coordinates are used to produce a linear plot. The power function exponent was .22 for the first list and .15 for the second. These exponents are lower than the range of .51 to .83 obtained by Shapiro (1969, p, 251) for words. The correlation between the objective and subjective scales in log-log coordinates was .768, p < .01, for the first list, and .729, p < .05, for the second.

Thus, the Ss can estimate the relative frequency of occurrence of syllables in English. If, however, the Ss performed this task by using estimates based on phonemes, the theoretical argument would be vacuous. The Ss could have information about the number of phonemes in each syllable, and about the relative frequency of those phonemes (Carroll & Lamendella, in press).

The results of correcting for these factors are shown in Table 1. It should be noted that in the first list there was only one syllable that was not two phonemes long. Thus, for the first list, the statistics based on the number of phonemes per syllable are difficult to interpret. For

Table 1
Partial Correlation Coefficient of the Actual and Estimated Syllable Frequencies Holding Two Factors Constant

Correlations	Factors	
	N†	Log††
	First List	
With Log–Estimates	−.576†††	.519
With Log–Actual Frequency	−.536†††	.736**
Partial Correlation	.665†††	.667*
	Second List	
With Log–Estimates	−.346	.306
With Log–Actual Frequency	−.744*	.690*
Partial Correlation	.752*	.752*

†Number of phonemes per syllable
††Log of the product of the independent phoneme probabilities
†††Significance figures would be misleading, as all syllables but one are two phonemes long.
*p < .05   **p < .01

the same reason, however, there is little need to isolate this factor statistically. From the figures in Table 1 and the correlation of the number of phonemes per syllable with the log of the product of the independent phoneme probabilities (−.856 and −.972 for the first and second lists, respectively), the second-order partial correlation coefficients were calculated as .812 and .747, p < .05. Again, significance figures would be misleading for the first list.

Thus, the strength of the relationship is not greatly diminished when we control for these factors. Considering the limited range of the stimuli, less than two-and-a-half orders of magnitude, these figures are quite high. It should be noted that this limited range is not a fact of nature, but is due fo the limits of Dewey's (1923) count of syllable frequency. In addition to an increased range, the inclusion of morphemic syllables to the list would also almost definitely increase the correlation (Shapiro, 1969).

In this paper, evidence was presented for increasing the size of the basic perceptual unit in speech perception from the phoneme to the syllable, and a major empirical objection to this increase was demonstrated to be untenable.
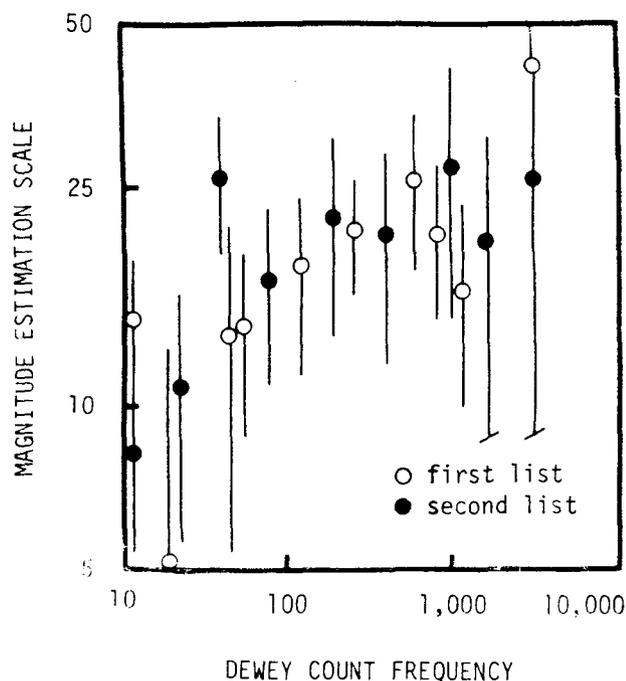

Fig. 1. The relationship between the geometric mean of the subjective estimates of relative syllable frequency and the actual syllable frequency plotted on log-log coordinates. The distance from the ends of the vertical lines to the mean value is equal to the standard deviation of the individual estimates. In order to correct for variations due solely to differences in the moduli used, each estimate was divided by the geometric mean of its S's estimates and multiplied by the grand geometric mean before the standard deviations were calculated.

REFERENCES

Broadbent, D. E. In defense of empirical psychology. London: Methuen, 1973.

Carroll, J. B. On sampling from a lognormal model of word frequency distribution. In H. Kučera and W. Francis (Eds.), Computational analysis of present-day American English. Providence, R.I: Brown University Press, 1967.

Carroll, J. B., & Lamendella, J. T. Subjective estimates of consonant phoneme frequencies. Language & Speech, in press.

Dewey, G. Relativ frequency of English speech sounds. Cambridge: Harvard University Press, 1923.

International Joint Council on Artificial Intelligence. Third International Joint Conference on Artificial Intelligence: Advance papers of the conference. Menlo Park, Calif: Publications Department, Stanford Research Institute, 1973.

Kučera, H., & Francis, W. N. Computational analysis of present-day American English. Providence, R.I: Brown University Press, 1967.

Lenneberg, E. H. *Biological foundations of language.* New York: Wiley, 1967.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Massaro, D. W. Perceptual images, processing time, and perceptual units in auditory perception. Psychological Review, 1972, 79, 124-145.

Shapiro, B. J. The subjective estimation of relative word frequency. Journal of Verbal Learning & Verbal Behavior, 1969, 8, 248-251.

Stevens, S. S. Issues in psychophysical measurement. Psychological Review, 1971, 78, 426-450.

Wanner, E. Do we understand sentences from the outside-in or from the inside-out? Daedalus, 1973, 102, 163-184.