

ASPECTS OF THE SYNTAX OF BEHAVIOR: SYLLABLE STRUCTURE AND SPEECH RATE

DONALD G. MACKAY

Psychology Department, University of California, Los Angeles, U.S.A.

This paper shows that maximal rate of speech varies as a function of syllable structure. For example, CCV syllables such as [sku] and CVC syllables such as [kus] are produced faster than VCC syllables such as [usk] when subjects repeat these syllables as fast as possible. Spectrographic analyses indicated that this difference in syllable duration was not confined to any one portion of the syllables: the vowel, the consonants and even the interval between syllable repetitions was longer for VCC syllables than for CVC and CCV syllables. These and other findings could not be explained in terms of word frequency, transition frequency of adjacent phonemes, or coarticulation between segments. Moreover, number of phonemes was a poor predictor of maximal rate for a wide variety of syllable structures, since VCC structures such as [ulk] were produced slower than phonemically longer CCCV structures such as [sklu], and V structures such as [a] were produced no faster than phonemically longer CV structures such as [ga]. These findings could not be explained by traditional models of speech production or articulatory difficulty but supported a complexity metric derived from a recently proposed theory of the serial production of syllables. This theory was also shown to be consistent with the special status of CV syllables suggested by Jakobson as well as certain aspects of speech errors, tongue-twisters and word games such as Double Dutch.

Introduction

A syntax of behaviors such as playing the piano or speaking must contain two basically different types of rules, syntagmatic rules for specifying the order of the behavioral components, and paradigmatic rules for specifying the form or simultaneous configuration of the components. Syntagmatic rules sequence the phonemes in speech production as well as the notes or chords in playing the piano, while paradigmatic rules specify the articulatory features for each phoneme as well as the locus and in motoric configuration for each note or chord.

The present study examines some properties of the syntagmatic rules underlying syllable production. Syllables have received little theoretical attention beyond rather rudimentary categorization, cf. Stetson (1947).^{*} Bolinger (1968) attributes this neglect to acoustic fuzziness (since phonetic records often fail to demarcate syllabic boundaries) and to formal difficulties in defining the syllable. But such

^{*} Linguists nonetheless recognize the importance of syllables, the smallest unit pronounced by itself, the unit taking sentential stress (Bolinger, 1968, p. 4) and the unit within which phonological constraints operate (e.g. SB occurs between syllables as in CROSSBOW but not within as in SBILL).

objections seem bankrupt. All linguistic units have fuzzy acoustic borders and could be disregarded on this basis. And although the syllable has no formal definition, behavioral definitions of syllables seem both valid and reliable. For example, speakers instructed to do so can pause consistently at the syllable boundaries within a word. And acoustically equivalent pauses inserted within a word are perceived as more natural when situated at a syllable boundary, e.g. RE-VEAL than within a syllable, e.g. REV-EAL (unpublished data).

Recent studies using behavioral definitions (MacKay, 1972) suggest that syllables play an important role in speech production and that producing syllables involves a complex set of syntagmatic and paradigmatic recoding rules as part of a more general system which proceeds hierarchically from general to more detailed output specifications. The syntagmatic rules involve a serial or one-to-more-than-one expansion, recoding a unit into two or more subunits if and only if these subunits are serially ordered in the output. The first and most general syntagmatic rule at the syllable level, i.e. (1), is universal; that is, it applies in the production of any syllable and expands S into an Initial Consonant Group (ICG) and a Vowel Group (VG). Then the "leftmost" subunit, in this case the ICG is expanded as one or more consonants until a paradigmatic rule is reached. Unlike syntagmatic rules, paradigmatic rules always involve a 1-1 expansion, e.g. ICG → (F) where (F) stands for a simultaneously activated set of distinctive features (cf. Table I).*

(1) Syllable (S) → Initial Consonant Group (ICG) + Vowel Group (VG).

The present study hinges on the fact that paradigmatically equivalent syllables can differ in syntagmatic complexity in this model. For example, CCV syllables

TABLE I

Syllabic recoding rules for CCV syllables such as [spae], CVC syllables such as [paes], and VCC syllables such as [aesp]

CCV Syllables	CVC Syllables	VCC Syllables
1. Syllable (S) → Initial Consonant group (ICG) + Vowel group (VG)	1(a) S → ICG + VG 2(a) ICG → [p]	1(b) S → ICG + VG 2(b) ICG → ∅
2. ICG → Consonant 1 (C ₁) + Consonant 2 (C ₂)	3(a) VG → V + Final Consonant group [FCG] 4(a) V → [ae]	3(b) VG → V + FCG 4(b) V → [ae]
3. C ₁ → [s]	5(a) FCG → [s]	5(b) FCG → C ₁ + C ₂ 6(b) C ₁ → [s]
4. C ₂ → [p]		7(b) C ₂ → [p]
5. VG → [ae]		

Abbreviations are represented in parenthesis, and terminal elements (at this level) in square brackets.

* Of course, segments are not the lowest level units in speech production and the recoding model can be readily adapted to generate a hierarchy of distinctive features, the highest level feature being C or V as in rules 3 and 6 for example (cf. Table I). And the differences between recoding rules and phrase structure rules should be stressed. For example, the symbols in recoding rules are syllable-specific since S stands for a particular syllable rather than all of the possible syllables in English. And the recoding model cannot be viewed as a grammar of syllables: it does not generate (describe) all and only all of the syllables in English or any other language. The recoding model, being based on behavioral data such as speech errors, and rules for abbreviation, poetic rhyme and Pig Latin (cf. MacKay, 1972) is clearly a model of performance rather than competence.

such as [spae] and CVC syllables such as [paes] are syntagmatically simpler than VCC syllables such as [aesp], partly because the ICG in [aesp] is expanded as ϕ , a null consonant without surface manifestation (cf. Table I). This aspect of the model is known as the Null Consonant Hypothesis. We first examined evidence for the Null Consonant Hypothesis in speech errors and tongue twisters and then devised an experiment to determine whether syllabic complexity plays a role in the maximal rate of speech.

The Null Consonant Hypothesis: evidence from speech errors and tongue twisters

The Null Consonant Hypothesis was based originally not on empirical evidence but on theoretical considerations. The phenomena discussed below provide empirical support for the hypothesis. The first two phenomena, anticipatory omissions and null transpositions were discovered originally in Meringer's (1906 and 1895) corpus of German speech errors, although similar phenomena have since been noted in English errors.

Anticipatory omissions

The 97 omission errors in Meringer's corpus had the following general characteristics (cf. MacKay, 1969): they usually occurred in unstressed syllables and were followed by an identical phoneme in identical syllabic position in an adjacent syllable, but rarely involved word-initial phonemes. The English omission in 2(a) exemplifies all of these characteristics.

- 2(a) Intern→inter (omission);
- (b) Stieglitz ist kein Esel→Ieglitz ist kein Esel (anticipatory omission);
- (c) temporal difference→temporal (anticipation);
- (d) Emser Kränchen→Kremser Änchen (null transposition);
- (e) throat cutting→coat thrutting (transposition);

Curiously though, some omissions ($n = 13$) violated every aspect of this general pattern [cf. 2(b)], but fit all of the characteristics of anticipation errors such as 2(c), which usually involve word-initial consonants in stressed syllables. Under the Null Consonant Hypothesis these otherwise inexplicable omissions reflect misexpansions of the rule $ICG \rightarrow \phi$, just as anticipations reflect misexpansions of syllabic recoding rules (cf. MacKay, 1972). In anticipation 2(c) for example, the term $ICG \rightarrow (T)$ was misexpanded as $ICG \rightarrow (D)$ (thereby capturing the fact that these segments are in identical syllabic position since misexpansions can only occur when the same term appears twice in the hierarchy). And similarly in anticipatory omissions 2(b), the term $ICG \rightarrow (ST)$ in STIEGLITZ was misexpanded as $ICG \rightarrow \phi$ under the Null Consonant Hypothesis, thereby capturing the fact that anomalous omissions always occurred in syllable-initial position and were followed by a vowel-initial syllable.

Null transpositions

A special class of transposition errors also supports the Null Consonant Hypothesis. Transposed consonants or consonant clusters almost invariably originate in identical syllabic positions as in 2(e). Sometimes, however, only one element is misordered as in 2(d), where the ICG of KRÄN is transposed to the beginning of EMSER. In this and nine similar cases an ICG was transposed to the beginning of a vowel-initial syllable. Under the Null Consonant Hypothesis, the only difference between these cases and standard transpositions is that one of the transposed elements is ϕ . The explanation for both null transpositions is identical: 2(d) reflects misexpansion of the rule $ICG \rightarrow \phi$ for EMSER and $ICG \rightarrow (KR)$ for KRÄN, while 2(e) reflects misexpansion of $ICG \rightarrow (THR)$ and $ICG \rightarrow (C)$.

Tongue twisters

The Null Consonant Hypothesis also finds support in the nature of British tongue twisters, as seen in extensive collections such as Parkin (1969). One of the major factors appearing over and over in tongue twisters is the alternation of initial consonants. For example, the difficulty in 3(a) seems to lie in the alternation of [t] and [θ], initial consonants differing in a single distinctive feature (manner of articulation). However, syllable-initial vowels do not seem to participate in this "feature alternation" principle. Parkin (1969) found no tongue twisters similar to 3(b), where initial vowels differ in a single distinctive feature, but found 28 similar to 3(c-d) where H alternates with an initial vowel. These tongue twisters are consistent with the Null Consonant Hypothesis and suggest that H and ϕ may have similar underlying features.

- 3(a) Thelma twisted ten tinted thistles in her teeth.
- (b) All ill engineers allow errors. (Hypothetical example.)
- (c) He ate hot apples and halibut hastily.
- (d) Hath Hazel asthma?

Experiment I*Syllable structure and speech rate*

Experiment I was designed to determine the relation between maximal speech rate and the syntagmatic complexity of CCV, CVC and VCC syllables containing identical phonemes. We reasoned that syntagmatic complexity may place an upper bound on speech rate and predicted that subjects would produce CCV and CVC syllables faster than paradigmatically identical but structurally more complex VCC syllables (cf. Table I).

Alternative predictions were based on more traditional notions of articulatory difficulty. Consonant clusters are inherently difficult according to one hypothesis since children commonly have problems in producing consonant clusters (Falk, 1973). The Consonant Cluster Hypothesis predicts slower rates for CCV and VCC than CVC structures (which lack a consonant cluster). Another hypothesis held that maximal speech rate will depend on the degree of coarticulation between the segments of a syllable. Greatest coarticulation apparently occurs between adjacent consonants (Kozhevnikov and Chistovich, 1965), less between vowels and preceding consonants, and least between vowels and subsequent consonants (MacNeilage and DeClerk, 1969). The Coarticulation Hypothesis thus predicts faster rates for CCV than CVC structures. A third hypothesis held that speech rate depends on phoneme transition frequencies. The Transition Frequency Hypothesis predicted faster rates the higher the average transition frequency of phonemes within the syllable.

Materials

The materials (cf. Table II) consisted of 60 syllables; 20 with VCC structure, 20 with CVC structure, and 20 with CCV structure, where C stands for a consonantal segment and V for a simple or complex vowel. Half the materials were monosyllabic English words and half were nonwords. The nonwords were phonologically possible but nonoccurring monosyllabic words of English. To control for paradigmatic or phonemic complexity in the nonwords, we permuted the same three segments across the three syllabic structures, e.g. [skow], [kows], [owsk]. This control proved infeasible with the words: lexical gaps in English made it

impossible to maintain the same three segments across all three syllable structures, although we tried to keep as many segments identical as possible, e.g. SPA, PASS, ASP.

TABLE II
The median syllable duration and semi-interquartile range for the 60 syllables in Experiment I

	Non-words (phonemic representation)			Words (conventional orthography)		
	CCV	CV C	VCC	CCV	CVC	VCC
	[sta]	[tas]	[ast]	spa	pass	asp
	[ska]	[kas]	[ask]	ski	case	ask
	[gra]	[rag]	[arg]	sky	kiss	axe
	[bli]	[lib]	[ilb]	spy	puss	ups
	[spuw]	[puws]	[uwsp]	stay	toss	art
	[spow]	[pows]	[owsp]	fro	roof	orb
	[skuw]	[kuws]	[uwsk]	crew	lake	Earp
	[klow]	[lowk]	[owlk]	blow	lab	old
	[frow]	[rowf]	[owrf]	grey	rag	ark
	[pruw]	[ruwp]	[uwrp]	pro	rope	orb
Number of segments	3	3	3	3	3	3
Number of recoding operations	5	5	7	5	5	7
Median syllable duration (s)	0.275	0.278	0.333	0.263	0.278	0.294
Semi-interquartile range	0.0445	0.0445	0.0751	0.0441	0.0425	0.0445

Complex vowels e.g. [ow] are treated as single segments in the syllable structure.

Procedures

The subjects were 10 native speakers of English recruited from the UCLA subject pool (6 males, 4 females, mean age 18.5). Each subject produced the 60 syllables in a different randomized order. The instructions* were as follows:

This is a study in speech production. We are interested in how quickly you can produce speech sounds. Some of the sounds will be nonsense syllables such as [ruk] and others will be words. We want you to produce the syllables as fast as you can for about 5 s. Make sure you continue to say the same word or syllable throughout. Remember though that we want your maximum rate.

The experimenter pronounced a syllable which the subject repeated once. Then on the signal "go", the subject repeated the syllable at maximum rate for approximately 6 s as determined by a Standard electric stop-clock, model S-1. This interval was chosen as being well within the breath span for subjects in our pilot studies. A Sony model TC-355 recorded the subject's output, which was then transformed into oscillographic tapes using a Siemens oscillogram with a paper speed of 50 mm/s. The number of syllables produced during the first 5 s was determined from the oscillographic tapes and translated into milliseconds per syllable. Finer measurements of consonant and vowel durations were obtained with a Kay electric type B sonograph.

* Although sophisticated methods for measuring articulatory durations have been in use for some time, little attention has been paid to subject instructions. Maximal rate instructions seem essential in studies of articulatory duration since syllables can in principle be prolonged indefinitely. Having the subject speak at "normal" or any other voluntary rate is essentially an undefined experimental procedure, giving the subject complete control over the dependent variable.

Results

Since preliminary analyses indicated that the data failed many of the criteria for parametric analyses, e.g. variance across syllable structures was significantly non-homogeneous, we employed nonparametric statistics throughout, specifically two-tailed sign tests with subjects as the unit of analysis (unless otherwise indicated). Table II contains the main results. Syllable durations for words vs. nonwords did not differ statistically ($P > 0.178$) and were combined in subsequent analyses. And within the word materials there was no significant effect of word frequency: the Spearman correlation between word frequency (as determined from Carroll, Davies and Richman, 1971) and syllable duration was low ($r = -0.25$) and nonsignificant at the 0.20 level. But syllable duration varied as a function of syllabic structure. The median duration for VCC syllables was significantly longer than for CCV syllables ($P < 0.001$ using either subjects or syllables as the unit of analysis) and CVC syllables ($P < 0.001$, same tests) but did not differ for CCV vs. CVC syllables ($P > 0.824$). Variance for CVC and CCV syllables was identical (semi-interquartile range 44 ms) and considerably less than for VCC syllables (semi-interquartile range 60 ms). Subject differences accounted for most of the variability: median syllable duration was 223 and 335 ms for the fastest and slowest subjects.

The peak amplitude of the vowels in the 30 nonwords was determined for two subjects chosen because of the clarity and measurability of their oscillographic records. Relative vowel amplitudes were significantly greater for VCC syllables than for either CVC or CCV syllables ($P < 0.01$, syllables as unit of analysis) but did not differ for CCV and CVC syllables ($P > 0.824$, same test). Rank correlation between vowel amplitude and syllable duration was 0.63, significant at the 0.05 level. Using the sonograph, we also determined the average duration of the vowels, consonants and the time between syllable repetitions for nine syllables produced by these subjects: [spow], [owsp], [pows], [skuw], [uwsk], [kuws], [spuw], [uwsp], and [puws], also chosen for reasons of measurability. All three measurements were longer for VCC syllables than for CVC and CCV syllables. The vowel was 24 ms longer, the two consonants 54 ms longer and the interval between syllable repetitions 18 ms longer in VCC syllables than in CVC and CCV syllables (cf. Table III).

TABLE III

The mean duration of the vowels, consonants, and the interval between syllable repetitions for VCC structures, and CVC and CCV structures (ms)

Syllable structure	Vowel duration	Consonant duration	Interval between repetitions
VCC	150	188	95
CVC and CCV	126	136	77
Difference	24	54	18

To test for fatigue effects, syllable duration for each of the 5 s was determined from the oscillographic records of all 10 subjects. Syllable duration increased as a

monotonic function of time for all three syllable structures, with a significant difference between first (263 ms) and fifth seconds (290 ms), 0.01 level, either subjects or syllables as unit of analysis. But although this finding can be viewed as a fatigue effect, other explanations are possible. For example, volume and relaxation pressure of the lungs is decreasing over the 5 s interval, so that later syllables are probably produced with increasing amounts of internal intercostal activity (Whitteridge, Ladefoged and Draper, 1958), an increase in effort which might reduce speech rate independently of neural fatigue.

To test the Transition Frequency Hypothesis we calculated the mean first order transitional frequencies of segments in the 30 nonwords, using Hultzen, Allen and Miron (1964). For example, #BLI# has transition frequency 197 for #B, 7 for BL, 36 for LI, and 8 for I#, giving a mean of 110 per transition. But transition frequency cannot explain our main results since Spearman correlation between speech rate and transition frequency was non-significant, $P > 0.20$, $r_s = -0.27$. To check this conclusion we examined the syllable durations of 5 VCC syllables with transitional frequencies greater on the average than those of corresponding CCV and VCC syllables. In line with our main results the median times in this partial analysis were significantly longer for VCC (0.401 s) than CCV and CVC (0.264 s), $P < 0.01$.

Finally, about half the nonword syllables contained liquids (R or L), e.g. [klow] and half did not, e.g. [spow]. Varying the syllable structure introduced a subtle difference between these two sets of materials: due to distributional constraints on English liquids the temporal order of the consonants had to be permuted in syllables containing liquids, e.g. [klow] vs. [owlk] but not in the remaining syllables, e.g. [skow] vs. [owsk]. But the relative speech rate for the three syllable structures was identical in both sets of materials although materials containing liquids were produced significantly slower than the remaining materials ($P < 0.001$), an effect due to paradigmatic or phonemic complexity.

Discussion

The data showed that maximal speech rate varied as a function of syllable structure: VCC syllables took longer to produce than CVC and CCV syllables. This difference was not due to transitional frequency, phonemic complexity, or the relative order of segments (which is identical for SKOSKOSKO, KOSKOSKOS, and OSKOSKOSK). Nor did the data support either the Consonant Cluster or Coarticulation Hypotheses. The Consonant Cluster Hypothesis incorrectly predicted faster rates for CVC than CCV and VCC syllables, while the Coarticulation Hypothesis incorrectly predicted faster rates for CCV than CVC syllables. The slower rate for VCC syllables was also not due to special efforts in enunciating final consonants, e.g. the release of syllable-final stops so as to make VCC syllables more distinct. In that the same process should also occur in CVC syllables, this hypothesis fails to explain the difference between CVC and VCC syllables, and the lack of difference between CCV and CVC syllables. It is also difficult to explain our results in terms of vowel laxing or degradation in acoustic output as a function of syllable structure. If such degradation occurred it was not apparent from examining the spectrographic records or from listening to the tapes: indeed, it was

difficult to identify syllable structure from the spectrographic records of noninitial portions of SKOSKOSKO, KOSKOSKOS, and OSKOSKOSK. However, our spectrographic measurements showed that the consonantal portion, the vocalic portion and even the time between syllable repetitions was longer for VCC than CCV or CVC syllables. So although these syllable structures could sound and look identical, there must be some syntagmatic device which accounts for the difference in output times for VCC vs. CVC and CCV syllables, a syntagmatic device similar to that postulated in the Syllabic Recoding Theory.

However, the present results were not unambiguous. For example vowel amplitude was confounded with syllabic complexity in the data. Since VCC syllables had greater vowel amplitude, and since the distance traversed by the jaw varies with vowel amplitude, the slow VCC rates may reflect increased vowel amplitude rather than increased syllabic complexity. Glottal or preglottal activity might also account for the slow VCC rates. Glottalization preceding vowel-initial syllables would increase the subglottal pressure, explaining the increased vowel amplitude and intersyllabic interval for VCC syllables. But although vowel-initial syllables are frequently released by glottal stops in some languages, this is apparently not the case in English. Acoustic and electromyographic data indicate that ϕ is rarely realized as a glottal stop in the natural production of English words (P. Ladefoged, personal communication). But glottalization may have occurred in the present task, and we have no evidence to the contrary. However glottalization fails to explain the lengthening of the consonantal portion of VCC syllables and in any case does not conflict with the Null Consonant Hypothesis, since ϕ may have optional overt realization involving closure of the vocal bands so as to increase subglottal pressure and thereby release the vowel with a synchronized and vigorous explosion.

The main problem with the present data concerns the number of syllable structures: only three of the 20 possible syllable structures (see Table IV). True these three structures were particularly interesting since they contained identical segments but different complexity. However structures with differing numbers of segments and identical complexity are just as interesting for the theory. For example, CV and V syllables are equally complex under the Recoding model, while CCCV syllables are simpler than VCC syllables. The theory thus generates the clearly counterintuitive predictions that V syllables will be produced no faster than CV syllables, despite the extra segment, and that CCCV syllables will be produced *faster* than VCC syllables which have *fewer* segments.

Experiment II

Syllable complexity and the Single Order Hypothesis

Experiment II was designed to overcome the limitations of Experiment I, to examine a larger range of syllable structures and to test the Single Order Hypothesis. The Single Order Hypothesis hinges on the fact that consonants in syllable-initial clusters can occur in one and only one possible order in most languages (Sigurd, 1955). For example, given that the segments [s] and [k] occur in an ICG in English, they must occur in the order [sk]. However, this is not true for the

TABLE IV

The syllable structures of monosyllabic words, with the number of coding rules postulated in the Syllabic Recoding Theory

Syllable structure	Exemplar	Number of segments	Number of coding rules
1. V	a	1	3
2. CV	saw	2	3
3. VC	as	2	5
4. CVC	cat	3	5
5. CCV	ski	3	5
6. VCC	ask	3	7
7. CVCC	tart	4	7
8. CCVC	star	4	7
9. CCCV	stray	4	6
10. VCCC	amps	4	8
11. CCVCC	stink	5	8
12. CVCCC	burst	5	8
13. CCCVC	straight	5	8
14. VCCC*	ursts*	5	9*
15. CCCVCC	sprint	6	10
16. CCVCCC	prints	6	10
17. CVCCCC	bursts	6	10
18. CCVCCCC*	glursts*	7	11*
19. CCCVCCC	sprints	7	11
20. CCCVCCCC*	splursts*	8	12*

Asterisks designate syllable structures which are possible under the phonological rules of English but lack exemplars in the language. The exemplars given are possible but non-occurring examples of these structures.

FCG: given that [s] and [k] occur in an FCG, they can appear in either order: [sk] as in ASK or [ks] as in AKS (axe). The reason for this distributional constraint is not known at present, nor is it known whether the constraint plays any role in speech production. But under the Single Order Hypothesis, syllable-initial consonant clusters are easier to produce than syllable-final consonant clusters, since phonemic order is determinate in the one case but not in the other.

Experiment II included all 10 syllable structures with segment length four or less (see Table IV). The Single Order Hypothesis generates two predictions above and beyond those shown in Table IV for the Syllabic Recoding Theory. According to the Single Order Hypothesis the maximal rate for CCVC should be faster than for CVCC syllables, despite their equivalence in syllabic complexity and number of segments. Thus a CCVC syllable such as [klis] with a syllable-initial consonant cluster having determinate order should be produced faster than a CVCC syllable such as [lisk] with a syllable-final cluster having indeterminate order. The Single Order Hypothesis also predicts that adding a consonant to an ICG is easier than adding the same consonant to an FCG, since the syllable position of the added consonant is determinate in the ICG but indeterminate in the FCG. Consequently, the difference in maximal rate between CV, CCV, and CCCV syllables

should be rather small, and definitely smaller than the difference between VC, VCC, and VCCC syllables, under the Single Order Hypothesis.

Instructions and procedures were identical to those in Experiment I: the subjects (10 native speakers from the UCLA subject pool who had not participated in Experiment I) repeated a syllable at maximum rate. The syllables ($n = 140$) were mainly nonwords constructed as in Experiment I except for an additional rule prohibiting segment repetition, thereby excluding syllables such as [siks] due to possible interactions between the repeated Ss. There were 30 VC structures, 30 CV structures, and 10 each for the remaining structures: V, CV, CCV, CVC, VCC, CCCV, CCVC, CVCC, and VCCC (cf. Table V).

TABLE V

The median syllable duration and semi-interquartile range for the syllables in Experiment II (phonemic representation with diphthongs treated as single vowels)

	V	VC	CV				
	[ey]	[eyr]	[ses]	[eyk]	[rey]	[sae]	[key]
	[iy]	[iyl]	[el]	[iyp]	[liy]	[le]	[piy]
	[ay]	[ayl]	[il]	[ayk]	[lay]	[li]	[kay]
	[ow]	[owr]	[ar]	[owt]	[row]	[ra]	[tow]
	[uw]	[uwr]	[ar]	[uwp]	[ruw]	[ra]	[puw]
	[ae]	[eys]	[aes]	[aek]	[sey]	[sae]	[kae]
	[e]	[iys]	[es]	[ep]	[siy]	[se]	[pe]
	[i]	[ays]	[is]	[ik]	[say]	[si]	[ki]
	[a]	[ows]	[as]	[at]	[sow]	[sa]	[ta]
	[a]	[uws]	[as]	[ap]	[suw]	[sa]	[pa]
Number of segments	1	2	2				
Number of recoding operations	3	5	3				
Median Syllable duration (s)	0.227	0.238	0.208				
Semi-interquartile range	0.035	0.043	0.032				
	CCV	CVC	VCC	CCCV	CCVC	CVCC	VCCC
	[krey]	[reyk]	[eyrk]	[skrey]	[kreys]	[reysk]	[eyrks]
	[piy]	[liyp]	[iykp]	[spliy]	[pliys]	[liysp]	[iykps]
	[klay]	[layk]	[aylk]	[sklay]	[klays]	[laysk]	[aylks]
	[trow]	[rowt]	[owrt]	[strow]	[trows]	[rowst]	[owrts]
	[pruw]	[ruwp]	[uwrp]	[spruw]	[pruws]	[ruwsp]	[uwrps]
	[krae]	[raek]	[aerk]	[skrae]	[kraes]	[raesk]	[aerks]
	[ple]	[lep]	[elp]	[sple]	[ples]	[lesp]	[elps]
	[kli]	[lik]	[ilk]	[skli]	[klis]	[lisk]	[ilkp]
	[tra]	[rat]	[art]	[stra]	[tras]	[rast]	[arts]
	[pra]	[rap]	[arp]	[spra]	[pras]	[rasp]	[arps]
Number of segments	3	3	3	4	4	4	4
Number of recoding operations	5	5	7	6	7	7	8
Median syllable duration	0.238	0.217	0.294	0.278	0.278	0.313	0.333
Semi-interquartile range	0.030	0.032	0.057	0.065	0.031	0.045	0.078

Results

Table V contains the main results. Both predictions of the Single Order Hypothesis were confirmed: CCVC syllables were produced significantly faster than CVCC syllables ($P < 0.002$) and consonants added to an FCG increased syllable duration relatively more than consonants added to an ICG. The average difference between CV, CCV, and CCCV syllables was 35 ms whereas the average difference between VC, VCC, and VCCC syllables was 48 ms. This difference in differences was significant at the 0.05 level, subjects as unit of analysis.

Results pertaining to syllabic complexity were mixed since average predictions of the model were confirmed while some detailed predictions were not. As can be seen in Figure 1, the predicted overall relationship between speech rate and

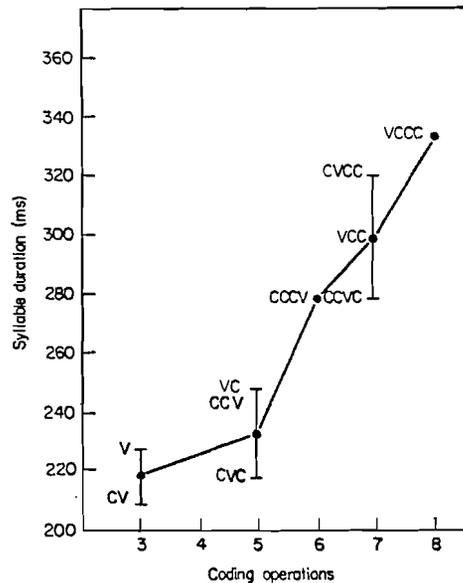


FIG. 1. Syllable duration in milliseconds, as a function of syllable structure and number of coding operations.

syntagmatic complexity was confirmed: maximal rate decreased systematically as a function of syllabic complexity. Monosyllabic V structures were produced no faster than CV structures ($P > 0.11$), and VCC structures were produced no faster than CVCC structures ($P > 0.764$). And as predicted, CCVC, CVCC, and VCC syllables (averaged together) were produced faster than VCCC syllables ($P < 0.02$), but slower than CCCV syllables ($P < 0.02$). But Experiment I was not completely replicated. Although CCV and CVC syllables were produced faster than VCC syllables ($P < 0.002$), CVC syllables were produced faster than either CCV or VC syllables ($P < 0.05$, but only with subjects as the unit of analysis), despite their equivalent complexity. And CV syllables were produced no faster than CVC syllables ($P > 0.20$), despite their difference in complexity. However these discrepancies probably reflect an error in constructing CVC materials, which contained

mo:
vari:
And
($P >$
[s],
spe:
sigr
in r
eith
T
was
osci
amp

Sylli

1. S
2. I
3. V

CC

1. S
2. IC
3. C
4. C
5. C
6. V

Abbr

little
vow
and
dura
rate
T
vow
were

more actual or occurring English syllables than V, CV, VC and CCV materials, a variable having significant effects on syllable duration (see subsidiary results). Another error may account for the statistically equivalent syllable durations ($P > 0.20$) of CVCC and VCCC structures. None of the CVCC syllables ended in [s], unlike virtually all of the VCCC syllables. The final [s] probably facilitated speech rate since VCCC syllables with short vowels ending in [s] were produced significantly faster than those not ending in [s] e.g. [ilsk]. However the equivalence in rate for CCCV and CCVC syllables is both unpredicted and unexplained under either the Single Order or Syllabic Complexity Hypotheses.

The first set of subsidiary analyses concerned effects of vowel amplitude, which was determined for two subjects chosen for the clarity and measurability of their oscillographic records. Vowel-initial structures had significantly greater vowel amplitude than all other structures ($P < 0.01$). But vowel amplitude accounted for

TABLE VI

Syllabic recoding rules for V, VC, CV, VCC, CCCV, CCVC, CVCC and VCCC syllables

V Syllables [i]	VC Syllables [is]	CV Syllables [si]	VCC Syllables [ist]
1. S→ICG+VG	1. S→ICG+VG	1. S→ICG+VG	1. S→ICG+VG
2. ICG→ϕ	2. ICG→ϕ	2. ICG→[s]	2. ICG→ϕ
3. VG→[i]	3. VG→V+FCG	3. VG→[i]	3. VG→V+FCG
	4. V→[i]		4. V→[i]
	5. FCG→[s]		5. FCG→C ₁ +C ₂
			6. C ₁ →[s]
			7. C ₂ →[t]

CCCV Syllables [stri]	CCVC Syllables [star]	CVCC Syllables [sart]	VCCC Syllables [arts]
1. S→ICG+VG	1. S→ICG+VG	1. S→ICG+VG	1. S→ICG+VG
2. ICG→C ₁ +C ₂ +C ₃	2. ICG→C ₁ +C ₂	2. ICG→[s]	2. ICG→ϕ
3. C ₁ →[s]	3. C ₁ →[s]	3. VG→V+FCG	3. VG→V+FCG
4. C ₂ →[t]	4. C ₂ →[t]	4. V→[a]	4. V→[a]
5. C ₃ →[r]	5. VG→V+FCG	5. FCG→C ₁ +C ₂	5. FCG→C ₁ +C ₂ +C ₃
6. VG→[i]	6. V→[a]	6. C ₁ →[r]	6. C ₁ →[r]
	7. FCG→[r]	7. C ₂ →[t]	7. C ₂ →[t]
			8. C ₃ →[s]

Abbreviations after Table I, terminal elements in brackets.

little of the overall temporal variability: no other syllable structures differed in vowel amplitude ($P > 0.10$) despite the large rate differences between say CVCC and CCVC syllables. And vowel amplitude correlated weakly with both syllable duration ($rs = 0.43$) and structural complexity ($rs = 0.39$) but highly ($rs = 0.98$) with rate variability (semi-interquartile range for the 10 syllable structures).

The second set of subsidiary results concerned a curious interaction between vowel length and syllable structure. As expected, syllables with simple vowels [i] were produced faster than those with complex vowels [liy] ($P < 0.002$). But this

difference varied with syllable structure. Syllables with complex vowels were only 6 ms longer in vowel-final structures (V, CV, CCV, CCCV) but 41 ms longer in consonant-final structures, a significant outcome, $P < 0.02$, subjects as unit of analysis and the difference between complex vs. simple vowels in the two sets of materials as the variables. Moreover the difference between complex vs. simple vowels varied systematically with the number of final consonants: 6 ms with 0 final consonants, 28 ms with one final consonant, 32 ms with two final consonants and 97 ms with three final consonants.

The remaining subsidiary results concern a difference between actual vs. possible but nonoccurring English syllables. Actual syllables are either monosyllabic words, e.g. [prey] or occur in some multisyllabic English word, e.g. [pruw] represents an actual syllable in PROOVING. Possible but nonoccurring syllables obey the phonological rules of English but never occur as part of any actual English word. The data showed no difference between the two types of actual syllables, but actual syllables were produced significantly faster than nonoccurring syllables ($P < 0.05$, syllable structures as the unit of analysis and actual vs. nonoccurring syllables as the variables).

Discussion

According to the Syllabic Recoding Theory, CV syllables are no more complex than monosegmental V syllables. Our data supported this hypothesis since V syllables were produced no faster than CV syllables. However, neither the theory nor the data contradict the special status of CV syllables suggested by Jakobson (1966). Jakobson argued that CV syllables represent a universal type present in all languages of the world, and that children find CV syllables particularly easy to remember and produce. These phenomena support the Syllabic Recoding Theory since CV syllables represent the simplest expansion of the one universal or obligatory rule in the model: $S \rightarrow ICG + VG$. But Jakobson's universals cannot indicate that CV structures are simpler than ϕV structures: ϕV represents a particular manifestation of the CV structure according to the theory, so that the absence of ϕV in some language is no more interesting or relevant than the absence of any other syllable-initial consonant (e.g. English η). And contrary to Jakobson, the recoding model predicts that CV and ϕV structures will be equally easy for children to learn and produce, a prediction supported in Weir (1966). The recoding model is not just consistent with linguistic universals and child language learning but explains why CV is universal and easily learned since CV constitutes the simplest possible structure in the model.

The recoding model is also consistent with most of our data for more complex syllable structures. As predicted, speech rate for syllables of equal length can differ and longer syllables can be produced faster than shorter ones, depending on syllabic complexity. But other factors play a significant role in speech rate. There are complex interactions between speech rate, vowel complexity and syllable structure. And data supporting the Single Order Hypothesis indicate that sequential constraints within initial consonant clusters somehow facilitate speech rate. Moreover the present data do not substantiate the detailed characteristics of recod-

ing rules, but only the complexity metric these rules provide. Indeed it may seem intuitively implausible that speakers specify more general aspects of words and syllables before specifying detailed aspects such as the phonemes composing the word. Other data nonetheless support this general thesis. Consider the tip-of-the-tongue phenomenon for example cf. Brown and McNeill (1966) and Yarmey (1973). When a word is on the tip of the tongue, speakers can often specify how many syllables the word contains and which syllable receives primary stress—all without knowing what phonemes make up the syllables. The generic specification of words, as seen in the tip-of-the-tongue phenomenon, favors a model with the same general characteristics as the Recoding Theory.

It might be argued that recoding rules apply prior to the motion of the articulators so that differences in speech rate cannot be due to recoding rules. Such an argument is fallacious on both logical and empirical grounds. It would be just as logical to argue that military commands precede troop movements, so that differences in the rate of advancement of armies cannot be due to the military commands. Moreover it has already been shown that organizational factors well above the articulator level (i.e. semantic and syntactic variables) play a role in maximal speech rate (cf. MacKay and Bowman, 1969).

But the present data have little relevance to the more general theory for producing words in which the recoding model is only one component. The normal production of words involves a conceptual component, a syntactic component, a selectional feature component and a word structure or morphological component, none of which are relevant to data showing no difference between words vs. nonwords. Probably the only rules of relevance to the present task are those for determining the order of abstract syllabic components and for specifying the phonetic features of these components, including nondistinctive or allophonic features such as the degree of aspiration of unvoiced stops. The child must learn or reconstruct these syllabic and phonetic rules for each syllable, except perhaps for universal rules such as $S \rightarrow ICG + VG$. And although the learning, reconstruction and setting up of syllabic rules is irrelevant to the present data, the model also predicts differences in the time to learn, and retrieve the program of rules for VCC vs. CCV syllables. Consider the learning of "secret languages" such as Double Dutch, one of a class of word games involving addition of dummy units. To speak Double Dutch, one begins with the ICG of a word, then adds a dummy vowel group (AWL), followed by a dummy ICG (F) and the VG of the original word. Thus the word DUTCH in Double Dutch is rendered DAWL FUTCH. As would be expected under the Recoding Theory, Double Dutch operates on "natural units" as defined in the model (i.e. units to the left of the arrows in the rules in Table I) rather than "unnatural units" (i.e. segments which do not arise from expansions of a single term at any level in the recoding hierarchy, e.g. the CV of a CVC syllable). The prediction of interest is that experimental word games should be easier to learn when they involve addition, substitution or transposition of natural as compared to unnatural units. Thus, the model does not predict that all existing word games, no matter how uncommon, will be equally simple or even readily describable under the Recoding Theory: some artificial word games may be more artificial and thus more difficult to learn and use than others.

The model also applies to syllable perception. For example a bisyllabic VCV input should be more readily perceived as V-CV than as VC-V, a prediction that might be tested with Warren and Gregory's (1958) verbal transformation technique. If a VCV input is replayed on a tape loop, subjects usually should hear the simpler segmentation, V-CV more often than VC-V, all other factors being equal.

The model may even apply to rapidly generated outputs in motor systems other than speech. For example a study of the time to produce patterns of finger movement with syntagmatic structures analogous to those of syllables (cf. Figure 2) would seem valuable for testing the hypothesis that similar principles underlie speech production and other syntagmatic actions. The model and task characteristics developed here may thus prove useful in exploring the acquisition, form, and consequences of syntagmatic rules for speech and perhaps other behaviors as well.

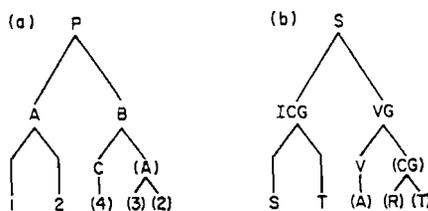


FIG. 2. The structure of the syllable START and an analogous structure for a pattern (P) of finger movements where the unit A is produced with the left hand, B with the right hand, and the numbers stand for the fingers such that fingers of the left hand are labeled 1-4 from left to right and fingers of the right hand are labeled 1-4 from right to left so that anatomically homologous fingers of the two hands have identical labels. [In this way the components of the simulated ICG and the simulated FCG can be drawn from the same set in a nonarbitrary way since MacKay and Soderberg (1970) have shown that anatomically homologous fingers are neurally interconnected and interact in the generation of finger taps.] The finger labeled 4 represents the vowel and is always stressed or tapped harder than the others.

The author gratefully acknowledges the support of NIMH Grant 19964-01 and thanks Ms B. Rosenfeld for executing the pilot studies and tabulating the data, Dr B. Burke for running subjects and Mr J. Miller for statistical analyses of the data. Part of this study was carried out at the Department of Experimental Psychology, University of Sussex with space and equipment generously provided by N. S. Sutherland.

References

- BOLINGER, D. (1968). *Aspects of language*. New York: Harcourt, Brace and World.
- BROWN, R. and MCNEILL, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325-37.
- CARROLL, J. B., DAVIES, P. and RICHMAN, B. (1971). *Word frequency book*. New York: Houghton Mifflin.
- DRAPER, M. H., LADEFOGED, P. and WHITTERIDGE, D. (1957). Expiratory muscles involved in speech. *Journal of Physiology*, 138, 17-18.
- FALK, J. S. (1973). *Linguistics and language*. Toronto: Xerox.
- HULTZEN, L. S., ALLEN, J. H. C. and MIRON, M. S. (1964). *Tables of transitional frequencies of English phonemes*. Urbana: University of Illinois Press.
- JAKOBSON, R. (1966). Implications of language universals for linguistics. In J. H. GREENBERG (Ed.), *Universals of Language*. Cambridge: MIT Press.

- KOZHEVNIKOV, V. A. and CHISTOVICH, L. A. (1965). *Speech: Articulation and Perception*. Washington: Joint Publications Research Service.
- MACKEY, D. G. and SODERBERG, G. A. (1971). Homologous intrusions: an analogue of linguistic blends. *Perceptual and Motor Skills*, 32, 645-6.
- MACKEY, D. G. (1972). The structure of words and syllables: evidence from errors in speech. *Cognitive Psychology*, 3, 210-27.
- MACKEY, D. G. (1969). Forward and backward masking in motor systems. *Kybernetik*, 6, 57-64.
- MACKEY, D. G. and BOWMAN, R. W. (1969). On producing the meaning in sentences. *American Journal of Psychology*, 82, 23-39.
- MACNEILAGE, P. F. and DECLERK, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America*, 45, 1217-33.
- MERINGER, R. (1906). *Aus dem Leben der Sprache*. Berlin: Behrs Verlag.
- MERINGER, R. and MAYER, K. (1895). *Versprechen und Verlesen*. Stuttgart: Göschensche.
- PARKIN, K. (1969). *Anthology of British Tongue-twisters*. London: Samuel French.
- SIGURD, B. (1955). Rank order of consonants established by distributional criteria. *Studia Linguistica*, 1, 8-20.
- STETSON, R. H. (1951). *Motor phonetics*. Amsterdam: North Holland.
- WARREN, R. M. and GREGORY, R. L. (1958). An auditory analogue of the visual reversible figure. *American Journal of Psychology*, 71, 612-13.
- WEIR, R. (1967). Questions on the learning of phonology. In SMITH, F. and MILLER, G. (Eds), *The Genesis of Language*. Cambridge: MIT Press.
- YARMEY, A. D. (1973). I recognize your face but I can't remember your name: further evidence on the tip of the tongue phenomenon. *Memory and Cognition*, 1, 287-90.

Revised manuscript received 4 December 1973