JOURNAL OF MEMORY AND LANGUAGE 24, 679–698 (1985)

The Phonological and Metaphonological Representation of Speech: Evidence from Fluent Backward Talkers

NELSON COWAN

Department of Psychology, University of Massachusetts, Amherst

MARTIN D. S. BRAINE

Department of Psychology, New York University

AND

LEWIS A. LEAVITT

Department of Pediatrics, University of Wisconsin, Madison

The psychological representation of phonemes and syllables was examined with a special group of subjects who voluntarily and rapidly rearrange speech units (i.e., "talk backward"). Each subject clearly used a primarily sound- or spelling-based representation to talk backward, and the present work focused on the sound-based skill. Backward speech differed from a total acoustic reversal: 12 subjects reordered phonemic units, and one reordered syllables. These speech units proved to be abstract to some degree, and hierarchically organized. However, the representation used in backward speech differed from the primary phonological system. It appeared to be a metaphonological system based on phonology but occasionally influenced also by orthography. Phonological principles seem to set lower limits for the size of units, and orthographic principles seem to set upper limits. A model of speech processing that includes both a primary, phonological, and a secondary, metaphonological level of representation is proposed.

The present paper provides evidence about speakers' access to phonological structure. The evidence comes from adults and children with the unusual ability to voluntarily "talk backward" by rapidly reversing the order of speech units they hear within normal utterances. We use this evidence to characterize the phonological analysis that subjects perform, and we consider the implications for models of speech representation. Some of these issues were previously mentioned by Cowan, Leavitt, Massaro, and Kent (1982) and Cowan and Leavitt (1981, 1982), who reported four case studies in all. However, the areas of agreement across subjects are clarified in the present research with a larger subject sample, and many more aspects of phonological representation are addressed. We are particularly concerned with the degree of abstraction of the phonemic and syllabic units used by backward talkers, and the levels of speech representation involved.

It seems remarkable that humans perceive discrete units within speech even though speech generally appears continuous when examined spectrographically (e.g., Liberman, 1982; Linell, 1979). As Sapir (1949) first demonstrated, the perceived units of speech are somewhat abstract. Still, it is unclear exactly what the units are in many instances, and it is un-

We thank Charles Clifton, Lyn Frazier, Priscilla Roth, Jola Jakimik, and an anonymous reviewer for commenting on an earlier draft of the manuscript. Address reprint requests to Nelson Cowan, Department of Psychology, University of Missouri, 210 McAlester Hall, Columbia, MO 65211.

0749-596X/85 $3.00

clear how these units are identified and organized. Listeners probably perceive several basic types of units: phonemes, syllables, morphemes, and words. However, each of these types of unit is associated with theoretical uncertainties and difficulties, despite the familiarity of some units to the layman. (For a theoretical introduction to units of speech, see Ladefoged, 1982; Makkai, 1972; and Sloat, Taylor, & Hoard, 1978). One issue is the location of boundary points between adjacent units, and consequently, the size of units. A second issue is the degree to which a particular unit is specific to the acoustic or phonetic form of an utterance, versus abstract and applicable across a family of acoustic/phonetic forms.

The present evidence is relevant to these theoretical issues for phonemes and syllables, because of the manner in which the subjects perform the backward speech task. Rather than producing something resembling a tape recording played in reverse, subjects appear to segment speech into phonemic or syllabic units and reverse the order of units, leaving intact the sound sequence within each unit. By transcribing backward speech, it is possible to identify the units the subject has used. For example, if the word *basket* were to be reversed as /kɛtbæs/, it would be clear that syllables were used, whereas if the reversal were /tɛksæb/, it would be clear that smaller units such as phonemes were used.

In backward speech the boundary points between units become clear. For example, the data are relevant to the phonemic status of diphthongs (two vowels conjoined) and affricates (a stop consonant and fricative conjoined). In the Trager–Smith (1951) phonology, each diphthong was considered to be biphonemic; for Pike (1947), the prominent diphthongs (/aɪ/ as in *fine,* /ɔɪ/ as in *choice,* and /aʊ/ as in *mouse*) were considered to be biphonemic, and less prominent diphthongs were considered monophonemic; and for Jones (1950) as well as Chomsky and Halle (1968), each English diphthong was considered to be monophonemic. The phonemic status of affricates in English (/dʒ/ as in *judge* and /tʃ/ as in *church*) has been disputed similarly (Hyman, 1975; Martinet, 1964). However, a backward talker must decide between the possibilities. For example, it would be possible to reverse the word *choice* (/tʃɔɪs/) as /sɔɪtʃ/, which would imply that both the affricate and the diphthong were considered monophonemic; as /sɔɪʃt/ (only the diphthong monophonemic); as /sɪɔtʃ/ (only the affricate monophonemic); or as /sɪɔʃt/ (both diphthong and affricate biphonemic).

The backward speech data also are relevant to the issue of the abstractness of speech units, inasmuch as the physical realization of a unit varies with the sequence of units. For example, the words *metal* and *medal* often are pronounced identically, with the alveolar flap [ɾ] as the medial consonant. These two words might be said to contain the same phonemes. Alternatively, the words might contain the more abstract phonemes /t/ and /d/, both of which would be converted to [ɾ] by phonetic realization rules. Backward talkers must decide between these possibilities (e.g., when reversing *metal,* they could say [lətɛm] or [ləɾɛm]).

Based on a description of the speech units that backward talkers produce, it is possible to address three general questions concerning speech processing: the source of the speech representation used in talking backward, the nature of this representation, and finally, the relationship of backward speech to ordinary spoken language. These topics will be introduced below.

The Source of the Representation in Backward Speech

The subjects are able to reverse even long words quickly and without rehearsal. We assume that in order to accomplish this, subjects deposit into working memory some representation of each word as a sequence of units of some type, and then scan this sequence in reverse or else reorder the units in working memory. However, it is necessary to determine not only what units are placed into working memory, but also what processes help to put them there. We refer to these processes as the "source" of representation. One possible source is that subjects might transfer a lexical representation of each word from long-term memory to working memory and then reverse that stored representation. Alternatively, subjects might have a set of analytic procedures that map an auditory image of a word into a string of phonological units. It would then be the output of these procedures that is put into working memory and reversed. The source of representation could become clear from the transcriptions (e.g., if subjects used only prestored lexical items they would not be able to reverse nonsense words).

The Nature of the Representation in Backward Speech

In about half of the backward talkers, the representation reversed is orthographic (Cowan & Leavitt, 1981). However, in the remaining subjects, it is clear from the responses to homophones and other diagnostic words that a basically phonological representation is used, and it is only these subjects with whom we are concerned. However, even in this phonological group, the exact nature of the representation has to be determined. Are the units consistently drawn from the same phonological system used in the perception and production of ordinary speech? If not, do the representations suggest subtle influences of orthography, indicating that subjects combine information from the primary system and from orthography to construct a secondary representation that can be reversed? Favoring this possibility, written language seems capable of making some form of segmental analysis available for a variety of uses. The relationship between reading and access to phonological units has been discussed by a number of investigators (Bradley & Bryant, 1983; Gleitman & Rozin, 1977; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Read, 1978). Access to linguistic structure has been referred to as "metalinguistic ability" (Hakes, 1980), of which "metaphonological ability" is one variety. Metaphonological ability would then be the basis of the secondary level of representation used in backward speech.

Relationship to Ordinary Language Use

The implications of backward speech for ordinary language depend largely on the representations used in ordinary versus backward speech. If backward speech were based on the primary phonological representation used in ordinary speech, it might provide extensive insight into the nature of this phonological system. Conversely, when only an orthographic representation is used, no insight into ordinary phonology is gained. However, if a subject's backward speech is based on phonology with orthographic influences (i.e., a secondary level of representation), there might be two types of benefit. First, because the primary level of representation might be observable through constraints it places on the secondary level, backward speech could clarify aspects of the primary representation. Second, a more direct contribution of backward speech would be to aid in a description of the secondary level of representation.

It is helpful to compare the potential contribution of backward speech to other methods that have been used to examine speech representation. The logic is similar for all of these types of evidence, but in some respects, the present type of evidence is complementary to the others. One of the oldest and best known types of evidence is the interpretation of speech by language informants and transcribers, discussed by Sapir (1949). Informants typically superimposed interpretations on speech that could not have come from the surface form alone. For example, a native speaker of the American Indian language Southern Paiute was asked to repeat the word [pa β a] ("at the water") with its syllables spoken separately. The speaker insisted on the pronunciation [pa.pa], even though he used [β] in the intact pronunciation of the word. Linguistic analyses of the language provided converging evidence that [β] was a phonetic expression of the underlying phoneme /p/ in particular phonetic contexts. Sapir concluded that aspects of phonological structure such as this one are psychologically real to native speakers. Speakers' perceptions deviated from the surface forms in a systematic rather than a capricious fashion. Similarly, in the present research, it is found that subjects superimpose an interpretation onto the speech stream (e.g., they pronounce differently the sound [ɾ] in the words *medal* versus *metal,* distinguished by sentence context).

Another type of evidence for the psychological reality of phonemes is that of transposition errors in speech (Fromkin, 1971). When speakers reverse the order of units composed of more than one phonetic segment, this provides evidence for the psychological reality of the unit. In these transposition errors there is variability in the types of reordering that occur. Nevertheless, the recurrence of a particular unit (e.g., a diphthong or a syllable) argues for the psychological reality of the unit even though it is not observed in every instance. However, speech errors generally cannot be recorded for later transcription, and the same diagnostic errors cannot be obtained uniformly across subjects. The backward speech task helps to eliminate these problems.

Previous investigators also have used language games to examine speech representation. Halle (1962) examined Pig Latin to aid in the assessment of phonological theories, and Treiman (in press) used an experimentally elicited speech game to examine syllable structure. Sherzer (1970) reported a speech game played by the Cuna Indians of Panama, in which the first syllable of each word is moved to the end. By examining the form of reordering, aspects of phonological representation were clarified. For example, all subjects reversed the word [biriga] ("year") as [gabir], not as [rigabi]. This demonstrated that the underlying representation was /bir.ga/, and that an epenthetic [i] was inserted in the surface expression of the word. In some types of linguistic example, though, Sherzer's subjects differed in the form of backward speech. It was evident that the data reflected certain uniform rules of the language, but also areas of individual variation in perceptual analysis (findings confirmed in the present study). However, in Sherzer's data the necessary phonological analyses may have been culturally transmitted from one player to another, whereas the present study reveals subjects' self-generated phonological analyses. Study 1 examines the division of speech into phonemes by some of the backward talkers in our sample of 50, and Study 2 examines one subject's division of speech into syllables.

Study 1: Phonemes

Our first subject was a 31-year-old philosophy professor (Cowan et al., 1982). Following a conference presentation of this man's abilities, he gained public media attention. To our surprise, 50 other people contacted us, claiming to have similar talents with backward speech. (For autobiographic details, see the Appendix in Cowan, Braine, & Leavitt, in press). Most of the backward talkers were adults, but there also were five children, 8 to 11 years of age, and four adolescents. Most adult subjects thought they began talking backward in late childhood (7–11 years), and the remainder in early adolescence. Most of the subjects could reverse speech so rapidly that a "simultaneous translation" was possible, although several subjects (including the children) were somewhat slower.[1] Most could reverse words with up

[1] It would appear to require a tremendous working-memory capacity to retain words of up to 10 phonemic units (e.g., *philanthropy*) or more, while manipulating the units. After all, most people cannot perform these

to about 10 phonemes. A few talked backward starting with the last word in each sentence, but most left the words in forward order.

The subjects were intelligent and generally interested in reading and verbal games. However, only two had any formal linguistic training. (These two were not tested further.) The subjects had no dyslexias or other verbal problems, and never talked backward accidentally. A few had other special abilities: One was able to multiply and divide large numbers quickly, and another was able to rapidly alphabetize the letters within a word or phrase. All except three were native speakers of English whose backward speech abilities preceded fluency in any foreign language. The remaining three were native German speakers who later learned English.

Because we were interested in studying speakers' processing of phonemes, we separated subjects into two groups: those who based their backward speech upon sound, and those who based it upon spelling. We judged the spelling-based skill to be irrelevant to our analysis of the phonological system, and concentrated on the sound-based skill.

Method

Subjects. It was possible to contact 20 of the English backward talkers for further interviews and testing: 10 who reported using spelling-based methods of reversal, and 10 who reported using sound-based methods. The subjective reports were later found to be correct in all cases. The sound-based backward talkers (6 males and 4 females, including an 8-year-old boy, a 16-year-old girl, and 8 adults ranging from 18 to 54 years of age) were tested in greater detail. Three German backward talkers (females ages 27, 64, and 66) also were interviewed.

Procedure. A variety of words, phrases, and sentences were presented to subjects in random order for conversion to backward speech, which was recorded for later transcription. Each subject was told to rearrange the stimuli in his or her usual fashion. Recordings were made in person or by telephone using a magnetic loop that fit over the earpiece. The recordings were consistently clear with either method. Broad phonetic transcription was used; reliability was high and disagreements primarily regarded details of vowel quality (cf. Cowan et al., 1982; Cowan & Leavitt, 1982). Subjects found to use a sound-based method also were given the digit span portion of the WAIS (Wechsler, 1955) or of the McCarthy Scale for children (McCarthy, 1970).

In the first phase of testing, subjects were interviewed and given a list of words (e.g., *judge, xerox, bomb, island*) to determine whether sound or spelling was used as a basis of backward speech. The 10 English-speaking subjects who could use a sound-based method of reversal were given 30 to 53 additional stimulus words, as well as 8 to 10 sentences. The words, which appear in Table 1, were selected to determine if the subjects use phonemes as units of speech reversal, and if so, to provide a detailed description of the units used. The sentences were included to examine stress cues, homophonic word pairs, and the fluency of reversal. The native German backward talkers received portions of the same English list as well as German words and sentences.

Results and Discussion

Subjects using sound versus spelling methods. Subjects could be separated quite

---

reversals even slowly, let alone rapidly. However, the subjects did not display truly extraordinary memory in conventional tasks (e.g., in digit span). It seems likely that the subjects recode sequences of units into higher order chunks whose reversals have been memorized. For example, the subject might have memorized the reversed pronunciation of such frequent sequences as [kw], [ɛst], [ɪŋ], and [pl]. This could reduce the load in working memory to a reasonable limit. The power of such recoding has been demonstrated in other memory skills (e.g., Ericsson, Chase, & Faloon, 1980).

TABLE 1
RESPONSES OF ENGLISH-SPEAKING SOUND-BASED SUBJECTS TO WORDS RELEVANT TO VARIOUS ISSUES

Issue	Example	Frequency	
1. Silent letters pronounced?			
bomb [bam]	[mab] vs [bəmab]	10 vs 0	
island [aIlənd]	[dənəlaI] vs [dənəlsaI]	10 vs 0	
plague [pleg]	[gelp] vs [jugelp]	10 vs 0	
weigh [we]	[ew] vs [həgew]	7 vs 0	
ghost [gost]	[tsog] vs [tsohəg]	9 vs 0	
though [ɟo]	[oɟ] vs [həguat]	8 vs 0	
thought [θat]	[taθ] vs [təhəguat]	7 vs 0	
judge [dʒʌdʒ]	[dʒʌdʒ] vs [ɛgdʌdʒ]	7 vs 0[a]	
2. Homographic sequences distinguished?			
g's in garage ([g], [ʒ])	[ʒarəg] vs [garəg]	10 vs 0	
c's in cycle ([s], [k])	[ləkaIs] vs [ləsaIs]	10 vs 0	
ough in though, thought ([-o], [-at])	[o-, ta-] vs [həgua-,həgua-]	7 vs 0	
ct in lecture, dictionary ([lɛktʃɚ], [dIkʃənɛri])	[-tʃk-, -ʃk-] vs both [-tʃk-] or [-ʃk-]	7 vs 3	
M. Begin ([e]) vs begin ([i])	[e, i] vs both [i] or [e]	10 vs 0	
use ([z]) vs use ([s])	[z, s] vs both [z] or [s]	10 vs 0	
3. Diphthongs /aI/, /ɔI/, and /aʊ/ preserved?[b]		$n = 8$	$n = 2$
fine [faIn]	[naIf] vs [nIaf]	7 vs 0	0 vs 2
eye [aI]	[aI] vs [Ia]	7 vs 0	0 vs 2
sky [skaI]	[aIks] vs [Iaks]	7 vs 0	0 vs 2
buy [baI]	[aIb] vs [Iab]	6 vs 0	0 vs 2
island [aɪlənd]	[dənəlaɪ] vs [dənəlɪa]	8 vs 0	1 vs 1
cycle [saɪkəl]	[əlkaIs] vs [əlkIas]	8 vs 0	1 vs 1
join [dʒɔɪn]	[nɔIdʒ] vs [nIɔdʒ]	8 vs 0	0 vs 2
boy [bɔɪ]	[ɔIb] vs [Iɔb]	8 vs 0	0 vs 2
house, mouse [-aʊs]	[saʊ-] vs [sʊa-]	8 vs 0	1 vs 1
now [naʊ]	[aʊn] vs [ʊan]	6 vs 1	0 vs 2
4. Minor diphthongs preserved?			
Words with [e^{I}], e.g., *weigh*	[e^{I}w] vs [Iew]	10 vs 0	
Words with [o^{u}], e.g., *ghost*	[tso^{u}g] vs [ts^{u}og]	10 vs 0	
5. /ju/ preserved?			
use (v) [juz]	[zju] vs [zui] or [zujə]	3 vs 4	
use (n) [jus]	[sju] vs [sui] or [sujə]	3 vs 4	
youth [juθ]	[θju] vs [θui] or [θujə]	2 vs 6	
6. Affricates preserved?			
judge [dʒʌdʒ]	[dʒʌdʒ] vs [ʒdʌʒd]	10 vs 0	
join [dʒɔIn]	[nɔIdʒ] vs [nɔIʒd]	8 vs 0	
giraffe [dʒəræf]	[færədʒ] vs [færəʒd]	8 vs 0	
church [tʃɝtʃ]	[tʃɝtʃ] vs [ʃtɝʃt]	8 vs 0[c]	
lecture [lɛktʃɚ]	[rətʃkɛl] vs [rəʃtkɛl]	8 vs 0	
fetuccini [fɛtətʃini]	[initʃətɛf] vs [iniʃtətɛf]	7 vs 0	
7. Sound sequences [ks], [gz]			
xerox [ziraks]	[skariz] vs [ksariz]	6 vs 4	
examine [ɛgzæmIn]	[-zgɛ] vs [-gzɛ]	6 vs 2	
locks, tacks, strikes	[-sk-] vs [-ks-]	7 vs 0	
8. Syllabic /l/ sound			
castle [kæsəl]	[ləsæk] vs [əlsæk]	4 vs 3	
subtle [sʌɾəl]	[lətʌs] vs [əltʌs]	3 vs 3	
cycle [saIkəl]	[ləsaIk] vs [əlsaIk]	5 vs 5	

TABLE 1—*Continued*

medal [mɛɾəl]	[lədɛm] vs [əldɛm]	3 vs 1
metal [mɛɾəl]	[lətɛm] vs [əltɛm]	4 vs 2
9. [r]-colored vowels		
burn, turn [-ɝn]	[nrət] vs [nɝt]	7 vs 2
dollars [dalɚz]	[srəlad] vs [sɝlad]	6 vs 1
lecture [lɛktʃɚ]	[rətʃkɛl] vs [ɝtʃkɛl]	8 vs 1
ladder [læɾɚ]	[rədæl] vs [ɝdæl]	6 vs 1
latter [læɾɚ]	[rətæl] vs [ɝtæl]	6 vs 1
finger [fIŋgɚ]	[rəgnIf] vs [ɝgnIf]	4 vs 1
10. Homophonic medial letters "d" and "t"		
ladder, latter ([d, t])	[d, t] vs both [ɾ]	7 vs 0
medal, metal ([d, t])	[d, t] vs both [ɾ]	3 vs 1
11. The sound [ŋ]		
bank [baŋk]	[kŋæb] vs [knæb] or [kənæb]	1 vs 6
finger [fIŋgɚ]	[-gŋIf] vs [-gnIf] or [-gənIf]	1 vs 4
ring [rIŋ]	[ŋIr] vs [gnIr] or [gənIr]	4 vs 6
aching [ekIŋ]	[ŋI-] vs [gnI-] or [gənI-]	5 vs 5
bang [bæŋ]	[ŋæb] vs [gnæb] or [gənæb]	3 vs 4
12. Phonemic alternation		
ser*e*ne ([i]), ser*e*nity ([ɛ])	[-i-, -ɛ-] vs both [i] or [ɛ]	6 vs 1
educa*t*e ([t]), educa*t*ion ([ʃ])	[-t-, -ʃ-] vs both [t] or [ʃ]	6 vs 0
13. Stress differences		
*con*trast (n), con*trast* (v)	(a) ['tsærtnak], [tsært'nək]	0
	(b) [tsært'nak], ['tsærtnək]	1
	(c) both ['tsærtnak]	4
	(d) both [tsært'nak]	1
*pre*sent (adj), pre*sent* (v)	Like (a), (b), (c), or (d) above	4, 2, 3, 1
*con*tent (n), con*tent* (adj)	Like (a), (b), (c), or (d) above	1, 1, 6, 0
*per*mit (n), per*mit* (v)	Like (a), (b), (c), or (d) above	5, 1, 2, 0
(In all, stress contrast was maintained in 15 instances and omitted in 17 instances.)		
14. Stress-related vowel quality differences		
*con*trast ([ka-]), con*trast* ([kə-])	[-ak, -ək] vs both [-ak]	2 vs 4
*pre*sent ([prɛ-]), pre*sent* ([prə-])	[-ɛrp, -irp] vs both [-ərp]	6 vs 3
*con*tent ([ka-]), con*tent* ([kə-])	[-ak, -ək] vs both [-ak]	4 vs 4

Note. Row sums differ because it was impossible to give some subjects all of the words. Examples should not be considered exhaustive except for the feature under consideration.

[a] Two subjects produced /dʒdʌdʒ/.

[b] As these responses demonstrate, eight subjects regarded diphthongs as single units, and two regarded them as two units each.

[c] Two subjects produced /həsrəs/.

clearly into two groups. In the responses of subjects with a sound-based method of backward speech, silent letters were not pronounced. In contrast, subjects with an orthographic method always pronounced silent letters. In illustration, the word *bomb* was reversed as /mab/ only by sound-based subjects, but as /bəmab/ only by orthographic subjects. Orthographic subjects rarely took into account auditory aspects of language (e.g., although the letter *g* represents both /g/ and /ʒ/ in the word *garage,* it was typically reversed by orthographic subjects as /ɛgəræ g/). In contrast, sound-based subjects preserved sound distinctions in homographs (e.g., *garage* was typically reversed as /ʒarəg/). Ten of the eleven subjects who used a spelling-based method also maintained that they visualized the words before reversing them. The exception was one subject who was able to use either method: He did not know if he visualized. The subjects who used a sound-based method denied that they visualized, with the exception of R.B., the subject of Study 2.

Treatment of diphthongs and affricates. For each word presented to the 10 English speaking subjects who used a sound-based method of reversal, Table 1 contains the subjects' responses along with an example of each type of theoretically possible response. One major option occurred with diphthongs and affricates. Eight subjects rapidly and automatically preserved the structure within the diphthongs /aɪ/, /ɔɪ/, and /aʊ/ 98.6% of the time. The other two subjects attempted to reverse the order of sounds within these diphthongs, but they did this slowly and with 15% failures to reverse them. For example, the words *fine* and *join* were reversed quickly as /naɪf/ and /nɔɪdʒ/ by eight subjects, but more slowly as /nɪaf/ and /nɪɔdʒ/ by the other two subjects. On the other hand, all 10 preserved rather than reversed the structure within the minor diphthongs /o^{w}/ as in *though* and /e^{I}/ as in *weigh*. Further, all 10 preserved rather than reversed the order of sounds within the affricates /dʒ/ as in *judge* and /tʃ/ as in *church*.

The two subjects who generally reversed elements within diphthongs were not fluent: They were not able to talk backward at a normal speech rate and could not carry out a "simultaneous" translation to backward speech. In contrast, all seven of the adults who preserved diphthongs were consistent and could carry out simultaneous translation, even when long words such as *automobile* and *pneumonia* were introduced. The difference in fluency between the two groups was not due simply to a memory difference: the two adults who reversed elements within diphthongs, although less fluent in backward speech, had higher mean forward and backward digit spans (8.5 and 7.5, respectively) than the other seven adults (7.4 and 6.3). Thus, the style of talking backward in which the major diphthongs are reversed may not serve as a suitable substrate for rapid, automatic backward speech. It is possible that in these two subjects the same monophonemic representation of diphthongs exists as in the other subjects, but that it has been overcome through conscious reflection to afford a more accurate reversal of speech.

Other aspects of segmental representation. In contrast to the diphthongs /aɪ/, /ɔɪ/, and /aʊ/, Table 1 shows that there is a good deal of inconsistency in subjects' representation of /ju/. Some subjects analyzed /ju/ as two units and reversed them, whereas other subjects analyzed it as a single unit that was preserved. Subjects' representation of /ju/ also seems to have been influenced by orthography. When it was represented by a single letter, as in the word *use,* about half of the subjects treated it as a single unit. However, when it was represented by several letters in the word *youth,* more subjects treated it as two units. Notice that unlike /ju/, the representation of /aɪ/, /ɔɪ/, and /aʊ/ was entirely independent of orthography. It may be that /ju/ is more likely to be treated as two units when /j/ is thought to be a consonant rather than a vowel (i.e., when it is represented by the letter *y*).

In spite of a basically phonemic method of speech reversal, subjects' knowledge of orthography sometimes influenced their phonological representation. An interesting example is the letter *x,* which typically represents two phonemic units (/ks/, as in *fox,* or /gz/, as in *examine*). About one-third of the time *x* was taken to represent a single unit. Nevertheless, *x* was more often analyzed as /ks/, reversed as [sk]. Other examples that might suggest an influence of orthography are the unstressed syllable /-l/, which was reversed as either [lə-] or [əl-], and the "r-colored" vowels, which were generally perceived as a vowel plus /r/. Thus, although subjects' choice of units was not dictated by the orthography (which could not account for their treatment of silent letters, homographs, major diphthongs and affricates, or words with *ng*), they did prefer a representation that could easily be reconciled with the orthography.

Some features illustrated in Table 1 help to determine the degree of abstractness of phonological units. One such case is the word-medial, alveolar flapped [ɾ] sound, which can be orthographically represented by *d* (e.g., *medal*) or by *t* (e.g., *metal*). Subjects' reversals of these words usually contained either [d] or [t] rather than [ɾ], and therefore suggest that [ɾ] is the surface realization of two more abstract forms, /t/ and /d/. Another important case is the sound [ŋ], in words containing *ng* or *nk*. Interestingly, the treatment of this speech sound depended upon the phonological context in which it occurred. In two words, *finger* and *bank*, [ŋ] was followed by a velar stop. For these words there was a strong tendency to reproduce [ŋ] abstractly, as /ng/. However, in words without a following velar stop (e.g., *bang*), subjects more accurately reproduced [ŋ]. In words like *finger* and *bank*, subjects may think that the [ŋ] sound is a phonetic variation of /n/. Note that the presence versus absence of a velar stop is not signaled in the orthography.

Last, subjects consistently maintained surface distinctions in word alternations such as the /i/ versus /ɛ/ in *serene* versus *serenity*. Thus, reversals were not based upon the abstract phonemes proposed by Chomsky and Halle (1968), despite the fact that the orthography is consistent with Chomsky and Halle in these examples. In general, then, the units for which there is evidence can be described as intermediate in their level of abstraction. They are more abstract than surface phonetic or "taxonomic" phonemic categories, but less abstract than the systematic phonemes of Chomsky and Halle (1968). We cannot claim that the type of representation Chomsky and Halle proposed does not exist; proofs of nonexistence are rare. However, according to Chomsky and Halle, the level of representation that was observed would not have been expected.

The data also suggest that subjects' speech representations tended to be relatively stable. For some aspects of representation (e.g., representation of homographic words, diphthongs, and affricates), this stability was obvious because of the extent of intersubject agreement. Even when subjects disagreed, though, they displayed substantial internal consistency. For example, of eight subjects who received the two words with *x* shown in Table 1, five put the stop consonant before the fricative in both words, two put the fricative first both times, and one subject was inconsistent. Second, an r-colored vowel was treated as a single unit rather than /r/ + vowel in 7 out of 44 instances, but 6 of these 7 were produced by a single subject. Third, each subject consistently used /ŋ/ or /ng/ as the representation for *ng*, except that two subjects used /ŋ/ only for *ng* in the word-final position. Similarly, three subjects were inconsistent for words with /ju/, but this inconsistency was predictable via the orthography (/ju/ was treated as a single unit unless represented with the letter y as in *youth*). Lastly, seven out of eight subjects who received two or more words with a syllabic /l/ were consistent across at least 80% of the words they received. Thus, subjects were relatively consistent and apparent inconsistencies often could be related to orthographic differences between stimuli.

Suprasegmental properties. A final issue illustrated in Table 1 is how subjects map English stress and intonation onto their backward speech productions. Subjects varied in this regard. Sentence-length intonations were never produced in reverse: Most subjects preserved the forward sentence intonation contour, superimposing it on their backward speech. A few subjects used a more monotonic, "list" contour that did not resemble a normal speech contour either forward or backward. To examine word stress, word pairs differing only in stress and meaning were examined (e.g., the noun "*con*trast" vs. the verb "con*trast*"). Subjects sometimes reversed the stress pattern within words. An example would be the noun *con*trast reversed as /tsært 'nak/ and the verb con*trast* reversed as /'tsærtnək/. However, Table 1 indicates that subjects more frequently superimposed the forward stress pattern on the reversed word (e.g., *con*trast reversed as /'tsærtnak/) or else failed to use stress distinctively. The data suggest, therefore, that stress and intonation both are separate from the sequence of phonemes within most subjects' mental representation of language. A similar conclusion can be drawn on the basis of speech errors (Cutler, 1980).

German backward talkers. The three native speakers of German reported having spoken German backward as a childhood game. Because one backward talker used a bidirectional tape recorder for many years to attain a backward speech style more like the acoustic signal, we discuss only the remaining two. They were tested with the English words in Table 1, and with a set of German words. Both subjects clearly used sound- rather than spelling-based methods of speech reversal for English words. For example, they omitted silent letters and consistently distinguished homographs. The variety of evidence is smaller in the German corpus, because German has a closer letter-to-sound correspondence than English does. Nevertheless, whenever a homograph did occur the two pronunciations were maintained. For example, the letter *s* may be pronounced /s/ (as in *aus*) or /ʃ/ (as in *Stein*), and this distinction was maintained. There were many vowel homographs in the sample, and these distinctions also were consistently maintained. The German backward talkers *differed* from most of the English backward talkers in their treatment of diphthongs: they consistently reversed the order of elements within the major diphthongs /aɪ/, /ɔɪ/, and /aʊ/ in both languages. For example, they reversed the word *fine* as /nɪaf/. A possible reason why the German speakers were able to analyze diphthongs into smaller elements is that in German, unlike English, each of the major diphthongs always is represented by two or more letters. Because a reversal of elements within diphthongs generally would result if the subject were to read German from right to left, there could be a contribution of orthography.

Summary and Conclusions

Although both sound and spelling were used as bases for backward speech, it was clearly possible to classify each subject's skill as basically sound- or spelling-based. The main conclusions of this study came from 12 subjects who used sound-based methods of speech reversal. For them, the goal of reordering speech typically was to match as closely as possible a recording played in reverse. This was sometimes stated explicitly by the subjects. However, their reorderings were further from a reversed recording than would be motorically possible. The most likely explanation is that their analysis of speech into units constrained their productions (e.g., their analysis was not fine-grained enough to perceive separately the vowels within diphthongs or consonants within affricates). Some aspects of the data also offer clues to the available speech representation. Most subjects used segments that seem to reflect a moderately abstract representation. Examples are the use of /ng/ for [ŋ], as in *ring*, and the use of /d/ versus /t/ to replace [ɾ], as in *metal* versus *medal*. The representations subjects used were influenced by orthography in certain instances, but many aspects of the representation could not have come from the orthography (e.g., units represented by two or more letters). No subject used speech sounds that would directly represent the level of "systematic phonemics'' described by Chomsky and Halle (1968). Instead, subjects seemed to be using a less abstract level of representation, closer to the level described by Sapir (1949).

STUDY 2: SYLLABLES

The syllable is a paradoxical unit, in that it seems to be intuitively available to untrained and even illiterate speakers (Morais, Carey, Alegria, & Bertelson, 1979) but is poorly understood linguistically. Syllables may be necessary to describe rules of phoneme sequencing within various languages. In English, for example, a word can begin with three consonants only if the first consonant is /s/, the second is /p/, /t/, or /k/, and the third is /l/ or /r/ (Sloat et al., 1978, p. 64). It seems reasonable that such rules should apply to every syllable rather than just word onsets. However, there is a problem: the rules for determining syllable boundaries are in dispute (cf. Hooper, 1976; Kahn, 1976; Pulgram, 1970). These rules logically must precede any syllabic restrictions upon phoneme sequencing.

The subject of the second study (R.B.) may be relevant to this theoretical concern, because she reordered speech in such a way that we could determine the syllable boundaries used. A previous study of syllable boundaries by Fallows (1981) provided a useful framework for this task, with a discussion of possible influences on syllabification. The most important of these influences were (a) phonotactic constraints, (b) the principle of maximal syllabic onset, (c) the effect of word stress, and (d) the principle of ambisyllabicity. Phonotactic constraints are limitations in the allowable sequence of phonemes within a syllable. For example, the word *empty* could not be divided /ɛm.pti/ because a syllable of English cannot begin with /pt/. The word *basket* probably cannot be divided /bæ.skɛt/, because English syllables rarely end in a stressed lax vowel. The principle of maximal onset states that each syllable begins with the maximal cluster of phonemes allowed by the phonotactic constraints (e.g., in the word *basket,* phonotactic constraints prevent /bæ.skɛt/, and of the remaining two possibilities, maximal onset favors /bas.kɛt/ over /bæsk.ɛt/. The third principle states that stressed syllables attract consonants in both initial and final position (e.g., in conflict with the maximal onset principle, the stress principle favors /bæsk.ɛt/ over /bæs.kɛt/). Finally, ambisyllabicity states that a syllabic boundary may fall within a single consonant, in effect making the consonant part of two syllables. This could help to resolve conflicts among the other syllabification rules. Fallows tested these principles in a task in which children heard 71 bisyllabic words with critical features. They were required to say each word with one of the syllables repeated twice (e.g., "chipchipmunk"). Fallows' words were included in the present stimulus list, but we also included many other words with additional diagnostic significance (e.g., words with three or more syllables).

Method

Subject. Only one subject (R.B., a 29-year-old female who said that she began her special skill at 8 years of age) reordered speech in a way that required segmentation into syllables (as well as phonemes). She fluently transformed speech in three ways that will be described in the results.

Procedure. A new set of 230 stimulus words and 37 sentences and phrases was selected in order to examine issues of syllabification. The words included all 71 from Fallows (1981). R.B. was instructed to reorder each utterance in her own way. These data were supplemented with a session in which many of the words were readministered, and in which R.B. was asked to explain to the best of her ability the rationale or basis of many of her syllabic divisions.

Results and Discussion

It is first necessary to describe the unusual skill of R.B. Upon hearing a word or phrase for the first time, she transformed each utterance in three ways, in rapid succession: (a) first, the order of syllables within the utterance was reversed; (b) next, the order of phonemes within each syllable was reversed, but the syllables themselves were put in their normal forward order; and (c) finally, the order of phonemes in the utterance was completely reversed. Alternatively, if requested to do so R.B. could produce any one of the three reorderings in isolation. In each reordering, a normal forward stress and intonation pattern seemed to be superimposed on the reordered phoneme string. Two one-word utterances and a multiword utterance to which R.B.'s response speeds are typical appear in Table 2.

R.B.'s skill requires her to make rapid decisions about the syllabification of words. For example, consider the word "basket" (/bæskɛt/). It might be divided as /bæ.skɛt/, /bæs.kɛt/, or /bæsk.ɛ t/. If R.B. selected /bæ.skɛt/, she would be expected to produce [skɛtbæ, æbtɛks, tɛksæb]. However, if she selected /bæs.kɛt/ she should produce [kɛtbæs, sæbtɛk, tɛksæb]. Finally, the syllabification /bæsk.ɛt/ should result in the production of [ɛtbæsk, ksæbtɛ, tɛksæb]. Thus, given R.B.'s reorderings of a particular utterance, one can determine her syllabification. Of course, it is theoretically possible for R.B. to use one syllabic division for the first reordering and a different division for the second one. However, such inconsistencies occurred in very few (<2%) of her responses.

Transcriptions of R.B.'s responses to Fallows' stimuli suggest that they were quite similar to Fallows' subjects. Consonant sequencing constraints (e.g., the division *eve.ning* rather than /i.vnɪŋ/ or /ivn.ɪŋ/) were preserved 98% of the time by Fallows' subjects and 99% by R.B. For words in which consonant sequencing rules permitted either an open or a closed first syllable, both vowel quality and stress played a role. Fallows reported that when a stressed first syllable contained a lax vowel (e.g., in *father*) subjects closed the first syllable 85% of the time, but in unstressed first

TABLE 2
THREE STIMULUS UTTERANCES REORDERED BY SUBJECT R.B.

Event	Time (ms)	Utterance
End of forward model	0	interesting
Beginning of Reversal 1	1200	ˈɪŋɪstɚɪn
Beginning of Reversal 2	2100	ˈnɪrətsɪŋɪ
Beginning of Reversal 3	3050	ˈŋɪtsɚətnɪ
End of Reversal 3	4050	. /ˈIn.tɚ.Ist.Iŋ/[a]
End of forward model	0	elephantitis
Beginning of Reversal 1	800	tɪstaɪfənt'ʔɪʔɛl
Beginning of Reversal 2	2350	lɛʔɪtnɪf'ʔaɪtsɪt
Beginning of Reversal 3	4000	sɪʃaɪtnɪ'fɪlɛ
End of Reversal 3	5200	. /ˌɛl.I.fənt.ˈtaI.tIs/[a]
End of forward model	0	urban and rural cultures
Beginning of Reversal 1	900	ˌtʃɝzkʌlʔəlˌrurʔænd'bɪnʔɚ
Beginning of Reversal 2	3200	ˌrʊnɪbdnɛˌ rurlə'lʌkzrʊtʃ
Beginning of Reversal 3	5600	ˌzrʊtʃlʌkləˌ rurdnɛ'nɪbrʌ
End of Reversal 3	7900	. /ˌɝ.bɪn.ænd.ˌrur.əl.ˈkʌl.tʃɚz/[a]

Note. The times are running counts measured from the end of the forward model.
[a] Syllabic representation of the stimulus, which can be determined from Reversals 1 and 2.

syllables (e.g., in *machine*) closure occurred only 35% of the time. R.B. similarly closed stressed syllables containing lax vowels 77% of the time and unstressed syllables 33% of the time. Fallows also examined cases in which the stress and maximal onset principles worked together or in opposition. When stress and maximal onset worked together, Fallows' subjects made the appropriate division (e.g., *e.nough*) 94% of the time, and R.B. did so 83% of the time. When stress and maximal onset were in conflict (e.g., *sof.a* vs. *so.fa*) maximal onset was obeyed 66% of the time by Fallows' subjects and 69% of the time by R.B. Finally, ambisyllabic responses were made 22% of the time by Fallows' subjects and 14% of the time by R.B. on the same words. Thus, there is no apparent conflict between the syllabification methods of R.B. versus the children studied by Fallows.

Other analyses. Additional aspects of R.B.'s method of syllabification are summarized in Table 3. It indicates where she placed the division between the first two syllables, separately for words with different combinations of stress, vowel quality, and number of intervocalic consonants. The responses are expressed as proportions of the available opportunities to make each type of division, where an "opportunity" is defined as a stimulus for which consonant sequencing constraints do not prevent the division. The top row of the table suggests that vowel quality and stress may affect syllabification of words with one intervocalic consonant. In the second row of the table, words in which the syllabification could be affected by a morpheme boundary or by geminate spelling were omitted, and the outcome is much clearer. Specifically, in words with a stressed first syllable, the syllable was left open (i.e., ended in a vowel) much more often when the vowel was tense rather than lax, χ^2 (1) = 9.63, $p < .005$. However, the first syllable always was left open when it was unstressed (in which case the tense/lax distinction does not apply), which was significantly more often than in stressed syllables with a lax vowel, χ^2 (1) = 19.15, $p < .001$. The difference between tense and unstressed vowels was not significant. In sum, syllable closure is most important for lax vowels in stressed syllables and least important for unstressed syllables.

In other situations, orthography or cluster division were important considerations (see Table 3). First, when there was a geminate consonant spelling between the first two syllables, ambisyllabicity always was used, but it rarely was used otherwise, χ^2 (1) = 134.7, $p < .001$ (with items divided two ways excluded). Geminate spelling between syllables other than the first two also generally produced ambisyllabicity. Second, when there were two or more intervocalic consonants, R.B. preferred to split the cluster after the first consonant (i.e., CVC.C(C)) rather than after two or more consonants (i.e., CVCC.(C)), $p <$ .001 with a sign test. This preference was maintained in each vowel condition.

There was an interesting effect of stress within words with three or more syllables. With them, it was possible to present word pairs in which a suffix resulted in a stress change. These pairs demonstrated quite clearly that stress affects syllabification. For example, R.B.'s reordered speech indicated that *photograph* versus *photography* was represented as /fot.ə.græf/ versus /fə.ta.grə.fi/ (/t/ shifted), and *telegraph* versus *telegraphy* as /tɛl.ə.græf/ versus /tə.lɛ.grə.fi/ (/l/ shifted).

In R.B.'s division of sentences and phrases, it was found that no syllable spanned more than one word (i.e., that syllables were restricted by word boundaries). R.B.'s responses also were strongly influenced by morphology. In 6 words presented to her, it would be necessary to leave a syllable open to preserve a morphene boundary (e.g., in *a.sleep* or *free.dom*), and she consistently did so even though two syllabifications were possible. Conversely, in 25 words, two syllabifications were pos-

TABLE 3
PROPORTION OF FIRST SYLLABIC DIVISIONS HAVING VARIOUS PROPERTIES

	Vowel of first syllable					
	Stressed					
Structural aspect	Tense	n^b	Lax	n	Unstressed	n
One intervocalic consonant						
Proportion open						
All	0.59	(37)	0.16	(57)	0.65	(26)
Morphs and geminates excluded	0.75	(28)	0.33	(27)	1.00	(17)
Proportion ambisyllabic[a]						
Nongeminate spelling	0.01	(39)	0.04	(35)	0.00	(18)
Geminate spelling	—	(0)	1.00	(26)	1.00	(8)
Intervocalic clusters						
CV.C(C)(C) vs	0.00		0.00		0.50	
CVC.C(C)	1.00	(2)	1.00	(3)	0.50	(4)
CVC.C(C) vs	0.66		1.00		1.00	
CVCC.(C)	0.33	(3)	0.00	(16)	0.00	(8)

Note. Because R.B. always obeyed English consonant sequencing constraints, illegal sequences were not included in the proportions (e.g., *Sa.nta* is impossible, so this word was excluded from the computation of the proportion of CV.C(C)(C), but *San.ta* and *Sant.a* both are possible, so this word did enter into the second comparison under intervocalic clusters). Words with r-colored vowels in the first intersyllabic position also were excluded. If a word was left open on one occasion and closed on another, it was omitted from the computation of percentage open, and words with consonant clusters divided two ways were excluded from the relevant cluster comparison. However, if a word was divided ambisyllabically on one of two occasions, it counted 0.50 in the percentage ambisyllabic.

[a] Three words divided ambisyllabically were not included in the table: *pretzel*, /prɛt.tsəl/; *acquainted*, /ək.kweɪnt.əd/; and elephantitis, /ɛl.ə.fənt.taɪ.tɪs/.

[b] The number of examples of each type = n.

sible but the morpheme boundary would be preserved by closing a syllable (e.g., in *final.ize*), and R.B. usually did so (21/25, $p < .001$, sign test). Inspection of individual examples indicated that morphological units were important regardless of the phonological context.

Interview and reliability data. When asked to explain the basis on which she reorders language, R.B. said that she attempts to match the dictionary's syllabic divisions. She also maintained that she visualizes words vividly before reordering them. Both these observations and R.B.'s treatment of geminate consonants indicate a role of orthography. R.B. reported that she sometimes felt inclined to divide words with nongeminate spelling within a consonant, but refrained from such a response. In contrast, at another point in the conversation she asserted that her divisions depended a great deal on the exact *pronunciation* that the tester used. Further, R.B.'s reorderings were unusually true to the phonetic properties of the forward model. For example, [ŋ] was used, x was considered biphonemic, and vowel quality was accurately preserved in most cases. Like the subjects in Study 1, R.B. seems to combine multiple available sources of information such as phonology and orthography to arrive at an unambiguous segmentation on any particular trial. R.B.'s subjective report contains elements from these sources of information.

Finally, when tested twice on a word, R.B. sometimes divided the word differently the second time. Nonetheless, on both trials her three responses usually were fluent and were valid, consistent reorderings. This rules out the possibility that R.B. simply has memorized a large vocabulary of reordered words. Instead, she must rapidly apply segmenting and reordering operations at the time of testing.

General Discussion

It is clear that subjects were able to combine several types of linguistic information to rapidly form an unambiguous representation. If there had been any conflict between types of information, the conflict must have been resolved within milliseconds in almost every case. We begin, therefore, by discussing the separate types of information that seem to be accessible to backward talkers. Second, we examine the process whereby subjects combine information to arrive at the representation used in the backward speech task. Finally, we offer a tentative model of phonological processing in both backward speech and ordinary language.

Types of Information about Speech

Surface phonetics. Subjects displayed a sensitivity to the surface phonetic properties of speech. For example, most subjects preserved at least some characteristics of stress and intonation found in the forward speech stimuli. In both studies, subjects often preserved the sound quality of vowels, so that neutralized vowels in nonstressed syllables remained neutralized in the reversal, and distinct, nonneutralized vowels remained so.

Phonology. Reversals were not dominated by surface phonetics. If they had been, subjects would have identically reversed word pairs such as *medal* versus *metal,* rather than making the /d/–/t/ distinction. Moreover, subjects aspirated or deaspirated voiceless stop consonants depending upon the location of the consonant in the reversed form. These examples suggest that phonological structure was an influential type of knowledge in both studies. In Study 1, this was evident also in subjects' treatment of diphthongs, which were preserved by most native English-speaking subjects, and affricates, which were preserved by all 12 subjects. However, the evidence from native German backward talkers suggests that the status of diphthongs depends on language-specific phonological principles. A further sensitivity to phonological structure was evident in Study 2. For example, R.B.'s syllabic divisions always obeyed consonant sequencing constraints, and were sensitive to stress and to the differential sequencing properties of tense versus lax vowels.

Morphology. There also was a sensitivity to the morphological characteristics of speech. The main evidence of this was that most subjects reversed each word separately, and in Study 2, the fact that morpheme boundaries were a preferred locus of syllable division.

Orthography. Finally, even among these 12 subjects whose backward speech was based primarily on phonemic units, there was a limited sensitivity to orthographic structure. For example, subjects sometimes treated the phoneme pairs /ks/ or /gz/ as single units when they were represented by the letter *x,* and sometimes analyzed the sound pair /ju/ as one versus two units depending on the spelling. In Study 2, R.B. almost always divided words in the middle of consonants having a geminate spelling, but this rarely happened in words without geminate spelling.

Nevertheless, the backward speech of these subjects was by no means *dominated* by spelling. This is quite clear in both studies, for example, from subjects' treatment of words with silent letters and homographs, and the use of sounds (e.g., [θ] or [ʃ]) represented by multiple letters.

Toward a Model of Phonological Representational Levels

The fact that subjects were able to combine these various types of information in order to segment and reverse speech is interesting because only some of this information is transparent in the speech signal, and because subjects arrive at these representations quite rapidly. Below, we discuss the source of speech representation, the nature of that representation, and a tentative model of phonological processing.

Source of representation in the backward speech task. In the introduction, we specified two sources of speech representation that could be used in the present task. First, subjects could transfer a lexical representation from long-term memory to working memory. Alternatively, subjects could have a set of analytic procedures that map an auditory image of a word into a string of phonological units put into working memory. Several facts favor the second possibility. Subjects are able to reorder nonsense words and words in foreign languages (Cowan & Leavitt, 1981; Cowan et al., 1982), even though no preestablished lexical representation of these words could exist. Further, the lexical entry is presumably stable from one occasion of use to another, yet in Study 2 the subject sometimes used different syllabic representations for a single word presented on two occasions. Similarly, it is unlikely that either [ks] spelled with an *x* or [ju] spelled with a *u* alternate sporadically between one versus two units in the lexical representation, but there were examples in which both representations were used. Thus, some analytic procedures seem to intervene on-line between the presentation of the stimulus and the presence in working memory of a string of units to be reversed. The subjects may also use the lexical identity, but if so it is probably used only to supplement or clarify the auditory stimulus, which is then analyzed through a set of procedures that are themselves independent of lexical identity.

The nature of representation in backward speech. Subjects rapidly combined various kinds of information to arrive at a single representation of each word to be reversed. The two main types of information available seem to have been (a) phonemic units consistent with the representation that linguists have suggested, and (b) information from the orthography, which sometimes was inconsistent with linguistic phonological analyses. The data suggest that subjects sometimes used orthography as a "notation system" for phonological structure, but that some phonological information was available beyond what was marked by the orthography. The phonology seemed to determine *lower limits* for the size of speech units, but the orthography determined *upper limits* (i.e., the largest units present were phoneme groups represented by a single letter). Demonstrating lower limits, the major English diphthongs and affricates were treated as single units, even when they were represented by multiple letters (as in *choice*). Similarly, monophonemic letter groups like *sh* and *th* were always treated as single units. Exemplifying upper limits, the sequence [ks] was sometimes treated as a single unit, but only when represented by a single letter (e.g., in *tax* but not *tacks*).

These results can be used to assess alternative conceptions of how phonology and orthography are combined. (a) In the first conception, phonology and orthography would be completely separate systems of rules and representations in ordinary language tasks like speaking, listening, and writing. However, they would be weighed or combined during backward speech. (b) Alternatively, the adult might have a single, complex system that includes both phonology and orthography. In the course of learning to read and write, the child's lexical representations and phonological rules would become fundamentally restructured (e.g., the child might form a more abstract phonological system that incorporates regularities of orthography). (c) A third conception, intermediate between the other two, states that there are two levels of phonological representation. There would be a preliterate system of lexical and phonological rules that serve comprehension and production. These would continue without fundamental change as literacy was acquired. However, with literacy would come a secondary level of representation that includes some information from the primary level, but also orthographic regularities and letter-to-sound correspondences. This level would provide access to phonological structure in a way that could be used to aid reading and writing of words whose spelling or pronunciation is unknown, as well as to produce or appreciate relationships between words and factors of style and esthetics in language. The units formed by this secondary level of analysis would not always be the same as the phonemic units of the primary level. In the "backward speech" task this secondary level, rather than the primary phonemic level, would be the direct source of speech segmentation.

The first two models do not seem to account for our results well. If phonology and orthography were separate sources of information, one would expect complete phonological information to be available to subjects who use a sound-based system of backward speech. The availability additionally of orthographic information should not obscure the fact that a single letter such as *x* or *u* may represent two phonemes. Moreover, the first model would allow that a different, more phonetically accurate form of backward speech should be possible in preliterate speakers, but no such cases were found. Finally, unlike the basically "orthographic" backward talkers, the "phonemic" backward talkers denied that they visualized words during speech reversal (with the exception of R.B.). This suggests that the orthographic influence in basically phonemic backward speech did not result from an independent orthographic system. Similarly, the second model, that literacy fundamentally restructures the phonological system, cannot account for both ordinary language use and backward speech. There is no evidence for any sharp change in vocal speech comprehension or production as a consequence of literacy, or between literate and illiterate adults. In the present data, to assume that [ks] or [ju] could sporadically become one unit with literacy, or that the syllable boundaries of some simple words would change as geminate spellings are acquired, seems odd and contradicts any previous linguistic analysis.

The most satisfactory view seems to be that literate speakers have access to a secondary, "metaphonological" level of speech representation (i.e., the third model). In this model, the subject would operate as follows. The smallest units perceived within speech would be phonemes, but phonemes would not be immediately available for use in working memory (i.e., for backward speech). Instead, the subject would have to apply a secondary analysis to the string of phonemes in order to generate the units available to working memory, and these units would sometimes differ from units in the primary system. In support of this model, it has been shown that illiterate adult speakers do not have access to segmentation sufficient to divide words into phonemes (Morais et al., 1979). As noted above, the relationship between reading and access to phonological units (i.e., metaphonological ability) has often been remarked (e.g., Bradley & Bryant, 1983; Hakes, 1980; Liberman et al., 1977).

It is not necessary to posit that the secondary analysis is uniform across subjects, as one generally expects of the primary linguistic system. Instead, the secondary level merely refers to the units that result when subjects intentionally analyze speech, whether or not subjects are consciously aware of those units.

Role of the primary system. The primary phonological system clearly places constraints upon the secondary system. An indication of these constraints is that subjects preserved as single units the phonemes /θ/, /ʒ/, and /ʃ/, and often /ŋ/, even though English orthography provides no clue to the integrity of these speech units. No subject considered such consonant clusters as /tr/ or /pr/ to represent single units, but all subjects did so with the affricates /dʒ/ and /tʃ/. Thus, there were limits to subjects' analytic abilities that applied systematically.

Consequently, it is possible to draw inferences about the primary system. For instance, if the affricate /dʒ/ were two units in the primary system, we would expect it to be reversed as [ʒd] most of the time (just as /ks/ was usually reversed as [sk], even when represented by the letter *x*). Inasmuch as the segments within /dʒ/ were never reversed, we conclude that it is a single unit in the primary system. The same is true of the affricate /tʃ/. A similar argument applies for English diphthongs, although two subjects did have some awareness of segments within diphthongs. Thus, although subjects' analytic capabilities naturally varied and they did not always adhere to the primary phonological system, this primary system seemed to define the lower limits of segmental information that could be rapidly and consistently accessed. In previous research discussed in the introduction (Halle, 1962; Sapir, 1949; Sherzer, 1970; Treiman, in press) metaphonological types of response similarly have been used to clarify the primary phonological system.

A Model of Processing in Backward Speech versus Ordinary Language

Figure 1 schematizes our conception of the primary phonological system and metaphonological analysis in both backward speech and ordinary language. The perception of spoken language leads to a primary phonological representation. This representation provides the basic information that is input to a metaphonological analysis of speech. However, orthographic information also enters into this metaphonological analysis, perhaps because the speaker assumes that orthography represents speech more closely than it actually does. The metaphonological knowledge permits the subject to place in working memory a sequence

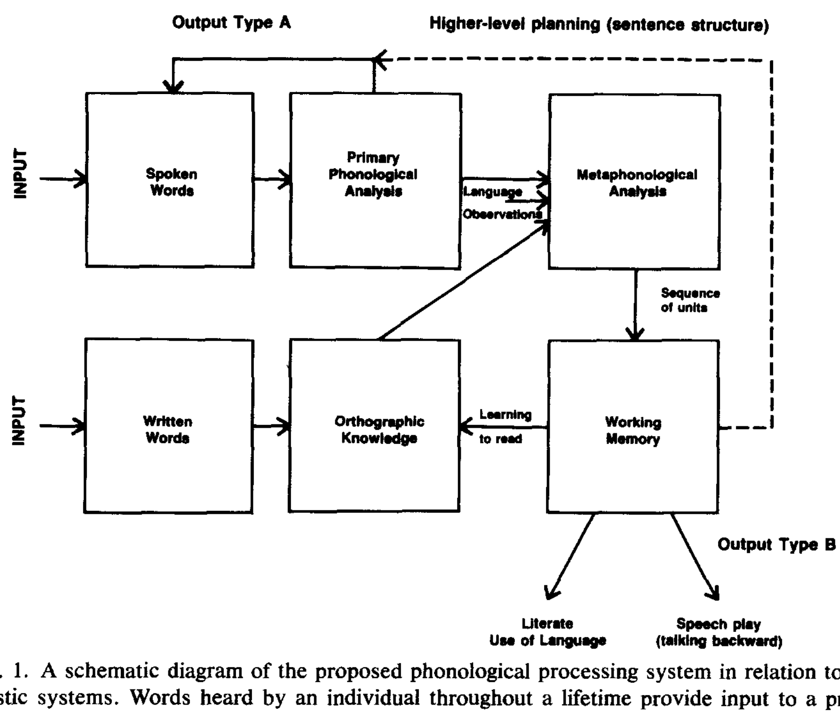

FIG. 1. A schematic diagram of the proposed phonological processing system in relation to other linguistic systems. Words heard by an individual throughout a lifetime provide input to a primary phonological system, which serves as the major basis of ordinary speech (Output Type A). Additionally, the primary system, orthographic information, and any other language observations the individual may make are combined in a secondary, metaphonological system. This system is used to provide speech units in working memory for the deliberate manipulation or interpretation of units, in literacy or in speech games (Output Type B).

of units. This representation in working memory can be used in learning to read (e.g., to become aware of letter–sound correspondences), to produce language in a literate fashion (e.g., to use alliteration and similar devices), and in speech games such as the present backward speech task. However, the selection and ordering of phonemes in ordinary speech presumably does not require this metaphonological level or working memory. Working memory probably is used to plan language at a higher level (e.g., to plan meaning, syntax, and choice of words), but we assume that the selection and ordering of phonemes in speech take place automatically. In Figure 1, ordinary speech is labeled *Output Type A,* whereas the uses of the metaphonological system are collectively labeled *Output Type B.* Although this model undoubtedly is incomplete, it provides a framework consistent with the available evidence.

In conclusion, a general contribution of the present research is to determine which aspects of linguistic structure can be available to deliberate analysis. There is no reason to expect this secondary, deliberate analysis to be coherent and uniform across subjects in the same sense that the primary system should be. Nevertheless, it is appropriate to speak of a metaphonological "level" of representation, because subjects are forced to be explicit about the units of speech. For example, in the word "metal," the subject cannot remain undecided about whether to pronounce the "t" as [t] or as [ɾ], and whether to pronounce "l" as [əl] or [lə]. It is assumed that backward speech is based upon a secondary representation, because the units sometimes disagree with any extant linguistic analysis of spoken language.

These data reveal the segmentation that speakers settle upon after considerable private experimentation, when they presumably wish to use a representation that is as natural and efficient as possible. We suggest that in such a situation, the subject has no choice but to use the primary phonological system that underlies ordinary speech as the elementary input for a secondary, metaphonological process. What is most striking about the data is that subjects can settle upon any representation rapidly enough to permit backward speech and "simultaneous translation" at a normal conversational pace.

REFERENCES

BRADLEY, L., & BRYANT, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature (London)* **301**, 419–421.

CHOMSKY, N., & HALLE, M. (1968). *The sound pattern of English.* New York: Harper & Row.

COWAN, N., BRAINE, M. D. S., & LEAVITT, L. A. (in press). Identifying phonemes and syllables: Evidence from people who rapidly reorder speech. *University of Massachusetts Occasional Papers in Linguistics.*

COWAN, N., & LEAVITT, L. A. (1981). Juggling acts with linguistic units. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior.* Chicago: Chicago Linguistic Society.

COWAN, N., & LEAVITT, L. A. (1982). Talking backward: Exceptional speech play in late childhood. *Journal of Child Language,* **9**, 481–495.

COWAN, N., LEAVITT, L. A., MASSARO, D., & KENT, R. (1982). A fluent backward talker. *Journal of Speech and Hearing Research,* **25**, 48–53.

CUTLER, A. (1980). Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in linguistic performance* (pp. 67–80). New York: Academic Press.

ERICSSON, K. A., CHASE, W. G., & FALOON, S. (1980). Acquisition of a memory skill. *Science (Washington, DC)* **208**, 1181–1182.

FALLOWS, D. (1981). Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics,* **17**, 309–317.

FROMKIN, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language,* **47**, 27–52.

GLEITMAN, L. R., & ROZIN, P. (1977). The structure and acquisition of reading I: Relations between orthographies and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading.* Hillsdale, NJ: Erlbaum.

HAKES, D. T. (1980). *The development of metalinguistic abilities in children.* Berlin: Springer-Verlag.

HALLE, M. (1962). Phonology in generative grammar. *Word,* **18**, 54–72.

HOOPER, J. B. (1972). The syllable in phonological theory. *Language,* **48**, 525–540.

HOOPER, J. B. (1976). *An introduction to natural generative phonology.* New York: Academic Press.

HYMAN, L. M. (1975). *Phonology: Theory and analysis*. New York: Holt, Rinehart, & Winston.

JONES, D. (1950). *The phoneme: Its nature and use*. London: Cambridge Univ. Press.

KAHN, D. (1976). *Syllable-based generalizations in English phonology*. Indiana University Linguistics Club.

LADEFOGED, P. (1982). *A course in phonetics* (2nd ed.). New York: Harcourt Brace Jovanovich.

LIBERMAN, A. M. (1982). On finding that speech is special. *American Psychologist*, 37, 148–167.

LIBERMAN, I. Y., SHANKWEILER, D., LIBERMAN, A. M., FOWLER C., & FISCHER, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading*. Hillsdale, NJ: Erlbaum.

LINELL, P. (1979). *Psychological reality in phonology*. London: Cambridge University Press.

MAKKAI, V. B. (1972). *Phonological theory: Evolution and current practice*. Lake Bluff, IL: Jupiter Press.

MARTINET, A. (1964). *Elements of general linguistics*. Chicago: Univ. of Chicago Press.

MCCARTHY, D. (1970). *Manual for the McCarthy scales of children's abilities*. New York: The Psychological Corporation.

MORAIS, J., CAREY, L., ALEGRIA, J., & BERTELSON, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323–331.

PIKE, K. L. (1947). On the phonetic status of English diphthongs. *Language*, 23, 151–159.

PULGRAM, E. (1970). *Syllable, word, nexus, cursus*. The Hague: Mouton.

READ, C. (1978). Children's awareness of language, with emphasis on sound systems. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language*. Berlin: Springer-Verlag.

SAPIR, E. (1949). The psychological reality of phonemes. In D. G. Mandelbaum (Ed.), *Selected writings of Edward Sapir in language, culture, and personality*. Berkeley: Univ. of California Press. (Originally in French in *Journal de Psychologie Normale et Pathologique*, 1933, 30, 247–265.)

SHERZER, J. (1970). Talking backwards in Cuna: The sociological reality of phonological descriptions. *Southwestern Journal of Anthropology*, 26, 343–353.

SLOAT, G., TAYLOR, S. H., & HOARD, J. E. (1978). *Introduction to phonology*. Englewood Cliffs, NJ: Prentice-Hall.

TRAGER, G. L., & SMITH, H. L. (1951). *An outline of English structure*. (*Studies in linguistics: Occasional papers*, No. 3). Norman: Battenburg Press.

TREIMAN, R. (in press). The structure of spoken syllables: Evidence from novel word games. *Cognition*.

WECHSLER, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

(Received July 25, 1984)
(Revision received February 1, 1985)