Psychological Review
1969, Vol. 76, No. 2, 232–235

AUDITORY OR ARTICULATORY CODING IN VERBAL SHORT-TERM MEMORY [1]

WAYNE A. WICKELGREN [2]

Massachusetts Institute of Technology

Attempts to determine whether the verbal short-term memory (STM) trace is in an auditory system or an articulatory system by comparison of the confusion matrices for auditory recognition and STM are shown to be based on assumptions that are very likely invalid. Attempts to decide this question by means of the effects of noise on STM are also shown to be inconclusive. Finally, the possibility must not be ignored that the trace is in an abstract verbal system that is neither purely auditory nor purely articulatory.

Intrusion errors in short-term recall of verbal items composed of two or more phonemes (e.g., letter names, digits, CV digrams) tend to have a phoneme in common with the correct item (Conrad, 1962, 1964; Wickelgren, 1965a, 1965c). Furthermore, false recognition rates in short-term recognition memory are higher for incorrect test items sharing a phoneme with the correct item than for test items having no phonemes in common with the test item (Wickelgren, 1966b). Confounding perceptual errors have been excluded from the short-term memory (STM) data either by slow visual presentation or by having subjects copy items as they are presented and scoring for recall only correctly copied items. Thus, verbal items appear to be coded in STM, at least in part, as a sequence of phonemes (or allophones).

Intrusion errors in short-term recall of single vowel or consonant phonemes are also systematic, tending to have distinctive features in common with the correct vowel or consonant (Hintzman, 1967; Wickelgren, 1965b, 1966a). Thus, it appears that a phoneme is coded in STM, at least in part, as a set of distinctive features, such as voicing, nasality, place of articulation, and openness of the vocal tract.

Phonemes, allophones, and distinctive features are rather abstract concepts. Even restricting ourselves to events occurring within the nervous system, one cannot say that a phoneme, allophone, or feature is merely a sequence of auditory (sensory) events, because it is also a sequence of speech-motor commands and also the kinesthetic feedback from speech, at other phases of the verbal communication process. Moreover, there may be other functional stages of the verbal communication process which deserve identification as separate levels in the nervous system. Certainly there is a growing school of linguists and psycholinguists that believe it is necessary to postulate abstract levels of processing language material. These abstract verbal systems are very likely not purely sensory or motor modalities, though they undoubtedly receive input from sensory systems and can provide output to motor systems.

So, what does it mean to say that verbal items are coded as phonemes (or allophones) and phonemes (or allophones) are coded as distinctive features in STM? It means that a theoretical analysis found useful in linguistics and speech communication has proved useful in predicting the systematic errors in STM for simple verbal materials. It demonstrates that STM does not take place in some totally different part of the nervous system from those parts that are concerned with other types of linguistic behavior. It indicates that not all of the STM trace for verbal materials is in a visual or writing-motor modality, since the phonemic (or allophonic) and distinctive feature codes have no conceivable definition in visual or writing-motor terms. The results obtained thus far do not logically eliminate the possibility of some replication of verbal STM traces in visual or writing motor systems, but they do suggest that the phonetically coded trace is the principal one.

What existing error data for STM do not answer is the question of whether the verbal STM trace is in (*a*) an auditory (sensory) modality, (*b*) an articulatory (motor or kines-

[1] This work was supported by Grant MH 08890-04, from the National Institute of Mental Health, United States Public Health Service.

[2] The author would like to thank A. M. Liberman and D. L. Hintzman for their comments on earlier versions of this paper. Requests for reprints should be sent to Wayne A. Wickelgren, Department of Psychology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

thetic-sensory) modality, or (*c*) an abstract verbal modality. Existing STM error data can be equally well-interpreted in auditory, articulatory, or abstract verbal terms, and furthermore, it is not clear how the issue can be settled by behavioral experiments.

Two recent experiments by Hintzman (1965, 1967) attempted to decide whether the STM trace was auditory or articulatory. The possibility that the trace might be in an abstract verbal modality was ignored. Also, the articulatory coding hypothesis was phrased in terms of kinesthetic-sensory traces rather than speech-motor traces, without any evidence indicating (*a*) that the speech musculature must be activated in rehearsal in order to obtain the memory-preserving functions of rehearsal or (*b*) that reduction or elimination of the kinesthetic feedback (through anesthetization of the vocal tract) has any effect on verbal STM. However, Hintzman (1967) did recognize both interpretations of the articulatory coding hypothesis. So, let us ignore the distinction between the two kinds of articulatory coding hypotheses and, for the moment, forget about the possibility of abstract verbal coding. Even having done this, Hintzman (1965, 1967) has not established that articulatory coding is more likely than auditory coding for verbal STM.

Hintzman (1967) reasoned essentially as follows: (*a*) Confusion matrices for auditory perception of consonants reflect the code used by the human auditory system and only that code. (*b*) Confusion matrices for verbal STM (eliminating all perceptual errors) reflect the code used by whatever system the verbal STM traces are in and only that code. (*c*) If there are systematic differences between these confusion matrices, then the verbal STM traces must be in some system other than the auditory system. (*d*) Hintzman (1967) and Wickelgren (1966a) have found that both voicing and place of articulation are important features in STM for English consonants (each consonant followed by the vowel /ɑ/). (*e*) Hintzman (1967) then asserted that Miller and Nicely (1955) had found voicing to be a feature in perceptual recognition of consonants in noise, but had found no evidence that place of articulation was an important feature in auditory recognition. Therefore, according to Hintzman, verbal STM is not in the auditory system. Since Hintzman only considers one alternative to auditory coding, namely, articulatory coding, and since "place of articulation" clearly sounds like the name of an articulatory dimension, Hintzman concludes that the coding of verbal STM is articulatory and not auditory.

There are several flaws in Hintzman's (1967) argument, and one potential germ of truth.

In the first place, Hintzman (1967) should not have relied exclusively on the data of Miller and Nicely (1955) to determine the features important in auditory recognition of consonants. Even Miller and Nicely did not rely exclusively on their data to determine their feature system for auditory recognition of English consonants. (They included place of articulation as a feature dimension.) Hintzman (1967) is correct in asserting that there is no compelling evidence in the data of Miller and Nicely for place of articulation as a feature in auditory recognition of English consonants. However, there is some suggestion in Miller and Nicely's data that place of articulation is an important feature dimension in auditory recognition of consonants, and other studies conclusively demonstrate its importance (Harbold, 1960; Liberman, Delattre, & Cooper, 1958; Singh & Black, 1966). It is even known that the second formant transition is an important auditory cue for the dimension of place of articulation (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Delattre, Liberman, & Cooper, 1955; Liberman, Delattre, Cooper, & Gerstman, 1954). Apparently, the white noise used by Miller and Nicely to obtain errors in auditory recognition of consonants had a more devastating effect on the cues for place of articulation than it did upon the cues for other features such as voicing, that is, place cues had a lower signal-to-noise ratio than voicing cues. Thus, the errors tended to preserve voicing much more often than they preserved place.

The more general point is that the nature of the noise one uses to obtain errors in auditory recognition has a systematic effect on the kinds of errors one sees, and no single such study should be used to rule out a potential feature dimension, especially when nothing is offered in its place. Thus, Hintzman's assumption that auditory confusion matrices reflect only auditory coding is false, because the nature of the noise one uses to distort the physical signal has a qualitative, as well as a quantitative, effect on the auditory confusion matrix.

Hintzman's (1967) assumption that auditory confusion matrices reflect only auditory coding may be false for another, more interesting, reason, namely, because of the evidence for the motor theory of speech perception. It is the case that, when the vowel accompanying a consonant is held constant, there is a simple mapping of articulatory dimensions onto the auditory cue dimensions that appear to mediate perceptual recognition. If exactly the same mapping from articulation onto auditory cues were found for consonants in all vowel environments, then it would seem to be impossible ever to use confusion matrices to decide between auditory or articulatory coding in either recognition or STM. However, this is not the case; the auditory cues for the various feature dimensions of consonants are profoundly affected by the vowel environment, while the articulatory attributes are much less affected (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Since the feature dimensions for auditory recognition of consonants seem to be more simply defined in articulatory terms than in auditory terms, it has been argued that the articulatory system plays a large role in speech recognition (Halle & Stevens, 1962; Liberman, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

Since the feature dimensions for verbal STM appear to be highly similar or identical to the feature dimensions for auditory recognition (the exact opposite of what Hintzman claimed), the same argument can be given for the articulatory system playing a large role in verbal STM as was given for the articulatory system playing a large role in auditory recognition. This is what Hintzman (1967) was trying to establish and is the potential germ of truth in his position, though the reasoning is almost exactly the inverse of that which he used.

However, we do not have any independent knowledge of the coding used at central levels of either the articulatory or auditory system. Thus, it is impossible, at the present time, to make definite decisions regarding whether the feature dimensions underlying either recognition or STM confusion matrices represent auditory or articulatory feature dimensions. If the phoneme is a basic unit of verbal recognition and STM, then auditory coding does seem less likely than articulatory or abstract verbal coding. However, if the basic unit is on the order of an allophone, rather than a phoneme, then auditory coding is in no way made less likely by existing error data. Finally, it is at least as reasonable to assume that the verbal STM trace is in an abstract verbal system as to assume that it is in either an auditory or an articulatory system.

Although error data are not capable, at present, of establishing whether the STM trace is auditory, articulatory, or abstract verbal, perhaps there is some other way. Hintzman (1965) did suggest another way, which, however, will be shown to be inadequate. He reasoned that if the STM trace were auditory, it should be possible to disrupt it by white noise presented at the same time that the subject is being given the items to be remembered via a memory drum. Furthermore, Hintzman reasoned that the tendency to produce phonemically similar intrusions might decrease in the white noise condition as a result of the subject's placing greater reliance on some other storage system than the auditory storage system, provided the storage was auditory rather than articulatory.

The results were clear. White noise had no effect on either the frequency of errors or the tendency to produce phonemically similar errors. Furthermore, subjects appeared to be rehearsing more audibly in the white noise condition.

Does this suggest articulatory coding? I think not. Whether the storage was auditory, articulatory, or verbal, subjects were apparently impelled to rehearse more audibly, perhaps to counteract the noise. By this hypothesis, some combination of more audible rehearsal and greater attenuation of auditory input via attention mechanisms occurred and was successful in combating any interfering effects of the white noise. This is a less parsimonious explanation than Hintzman's (1965) explanation, but it is certainly possible.

However, even if we accept *part* of Hintzman's (1965) conclusion and assume that white noise never enters the area where the STM trace is stored, it is as reasonable to assume that there is an auditory area or abstract verbal area for speech sounds that is protected from white noise input as it is to assume that there is an articulatory area which is protected from white noise input.

All this does not mean that the STM trace is not partly or even primarily articulatory. It may be. Certainly, rehearsal seems to occur in experiments where the rate of presentation is as slow as one or two items per second, so it must be serving a function. But we do not yet know how to determine whether that function is to maintain an active reverberating STM trace, a passive trace in the auditory system, a passive trace in the articulatory system, passive traces in both auditory and articulatory systems, or traces in some abstract verbal system that is neither purely auditory nor purely articulatory.

REFERENCES

CONRAD, R. An association between memory errors and errors due to acoustic masking of speech. *Nature*, 1962, 193, 1314–1315.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology,* 1964, **55**, 75–84.

COOPER, F. S., DELATTRE, P. C., LIBERMAN, A. M., BORST, J., & GERSTMAN, L. J. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America,* 1952, **24**, 597–606.

DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America,* 1955, **27**, 769–773.

HALLE, M. & STEVENS, K. Speech recognition: A model and a program for research. *IRE Transactions of the Professional Group on Information Theory,* 1962, **IT-8**, 155–159.

HARBOLD, G. J. Whispered monosyllabic speech, initial and final consonant confusions. United States Naval School of Aviation Medicine, Pensacola, Florida, Project MR005.13-7003 (Formerly NM 18 02 99) Report #86, 1960.

HINTZMAN, D. L. Classification and aural coding in short-term memory. *Psychonomic Science,* 1965, **3**, 161–162.

HINTZMAN, D. L. Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 312–316.

LIBERMAN, A. M. Some results of research on speech perception. *Journal of the Acoustical Society of America,* 1957, 29, 117–123.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review,* 1967, **74**, 431–461.

LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech,* 1958, **1**, 153–167.

LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S., & GERSTMAN, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs,* 1954, **68**(8, Whole No. 379)

MILLER, G. A., & NICELY, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America,* 1955, **27**, 338–352.

SINGH, S., & BLACK, J. W. Study of 26 intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America,* 1966, 39, 372–387.

WICKELGREN, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology,* 1965, **70**, 102–108. (a)

WICKELGREN, W. A. Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America.* 1965, **38**, 583–588. (b)

WICKELGREN, W. A. Similarity and intrusions in short-term memory for consonant-vowel digrams. *Quarterly Journal of Experimental Psychology,* 1965, 17, 241–246. (c)

WICKELGREN, W. A. Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America,* 1966, **39**, 388–398. (a)

WICKELGREN, W. A. Short-term recognition memory for single letters and phonemic similarity of retroactive interference. *Quarterly Journal of Experimental Psychology,* 1966, **18**, 55–62. (b)

(Received February 27, 1968)