ELSEVIER

Lingua 100 (1997) 91–100

Lingua

Some consequences of the representation of words in memory

Morris Halle

Department of Linguistics, MIT, Cambridge, MA 02139, USA

> ... phonetic phenomena are not physical phenomena *per se*, however necessary in the preliminary stages of linguistic research it may be to get at the phonetic facts by way of their physical embodiment. The present discussion is really a special illustration of the necessity of getting behind the sense data of any type of expression in order to grasp the intuitively felt and communicated forms which alone give significance to such expression.
> (Sapir, 1925)

Humans are not born knowing the words of any language. The words we know we have learned from our parents, from other people, from books, or other media. On rare occasions, we invent a word that then becomes part of our language. When we speak we select from our store of words, our lexicon, those that are appropriate for our purpose, and when hearing someone else's speech we again access our lexicon and determine the string of words that was encoded in the acoustic signal that struck our ears. Accessing the lexicon is thus a crucial step in every speech event. It is because of this that speech is usually perceived by both speaker and addressee as a sequence of words, and this illusion – I use this word advisedly – that speech is composed of discrete words is shared even by the most experienced phonetician, one who knows that the acoustic signal is continuous and is not segmented into discrete words. The aim of my essay is to draw some inferences from these mundane facts, especially as they affect the nature of phonetics, which I take to be the study of the articulatory activity involved in speaking and the acoustic signal resulting from this activity. I attempt to show below that the fact that every speech event involves access to items stored in memory shed unexpected light on certain questions of fundamental importance to phonetics.

1

When we learn a word we store in our memory information that enables us to use the word in utterances that we ourselves produce and to recognize it when it appears in an utterance produced by someone else. The information that we need in order to

PII S0024-3841(96)00043-5

do this is of two kinds. On the one hand, we need information about the sound of the word; let us call it the *phonetic information*. In addition, we need information about the meaning of each word, its lexical category (e.g., is it a noun, a verb, adjective, etc.?), its grammatical peculiarities (e.g., is it a "transitive" or "intransitive" verb? does it form its perfect participle with the regular "*ed*-ending", as in *play-ed* or *pass-ed* or with /n/ as in *tak-en, see-n*, etc.?). Let us call the latter the *grammatico-semantic* information. This distinction is important for our purpose since in what follows the focus is on the phonetic information; grammatico-semantic information is touched on only occasionally, if at all.

Although not always made explicit, it has been widely – though not universally – assumed that the phonetic information about each word is stored in memory as a sequence of discrete sounds or *phonemes*.

When we pronounce a sequence of phonemes we execute a specific gymnastics with certain portions of our upper respiratory and alimentary tracts. The movable portions of the tracts that execute the gymnastics are called the *articulators*, and there are six of them: labial (lips), coronal (tongue blade), dorsal (tongue body), velum, tongue root, glottis.[1]

One of the first observations that students in an introductory phonetics course are led to make is that the gymnastics performed in the course of producing different phonemes is readily analyzed into more elementary components or *features* which, in combination with other features, are also found in other speech sounds. For concreteness consider the set of phonemes in the top line of (1), where ŋ represents the nasal consonant at the end of words such as *king* or *long*, *x* stands for the fricative found at the end of a German word such as *Bach*, and *γ* is the voiced counterpart of *x*.

(1)	p b f v m	t d s z n	k g x γ ŋ
articulator	Labial	Coronal	Dorsal
consonantal	+ + + + +	+ + + + +	+ + + + +
nasal	– – – – +	– – – – +	– – – – +
voice	– + – + +	– + – + +	– + – + +
continuant	– – + + –	– – + + –	– – + + –

Every phoneme in (1) is produced with a radical constriction in the mouth which impedes the flow of air through the oral cavity. It is this constriction that differentiates the consonants in (1) from vowels, which are produced without constriction. What differentiates the three sets in (1) one from another is that the constriction in each set is formed with a different articulator. In [pbfvm] this is the *Labial* or lip articulator; in [tdszn] this is the *Coronal* or tongue blade articulator; and in [kgxγŋ] this is the *Dorsal* or tongue body articulator. I have indicated this information in the first line of the table in (1).

---

[1] For additional discussion of the role of articulators in the specification of speech sounds, see Halle (1995).

The three other features in (1) are so familiar as to require no comment. I shall however have more to say below about the feature *voice*. The conception of the phoneme as a complex of features is found in the work of Jakobson and Trubetzkoy, on the one hand, and in Bloomfield, on the other. Thus, Jakobson 1932 (1962: 231) defines the phoneme as "a set of those concurrent sound properties which are used in a language to distinguish words of unlike meaning". And analogous definitions can be found in Bloomfield's 1933 *Language* (p. 79) and in Trubetzkoy's 1939 *Grundzüge* (p. 35).

In the discussion just above, the features that compose a phoneme have been viewed as instructions for actions by the articulators that result in a specific acoustic event. On this view, the table (1) is a list of instructions for the actions that are required to produce the different phonemes. The instructions are encoded in some neurologically appropriate form. Although little is known at this time about the neurological counterparts of instructions to the different articulators there is no question about their reality. To produce the different sounds in the table (1) it is necessary to execute the particular set of actions listed.

An implication of the view that phonemes are complexes of features is that morphemes are stored in memory as sequences of feature complexes; i.e., as sequences of sets of instructions for actions of the articulators. A modicum of support for this consequence comes from the fact that the articulator actions implied by the features are actually observed when speakers pronounce the morphemes.

A fact of fundamental importance for matters under discussion here is that the representation of a word or affix in memory is not always identical with its phonetic actualization. For example, the English nouns in (2) have one stem in the singular and another in the plural.

(2) life live-s leaf leave-s house hou[z]e-s
moth mo[ð]-s wreath wrea[ð]-s

While the number of English nouns with stem-doublets of this kind is not large, these include some of the most common nouns of the language and the alternations are therefore part of the competence of ordinary speakers. Moreover, the relation between the base and derived forms of the stem is of a kind that is found over and over again among the languages of the world. Finally, the work in phonology of the last sixty years beginning with Sapir's (1933) 'The psychological reality of phonemes' and including such notable studies as Bloomfield's (1939) 'Menomini morphophonemics', Jakobson's (1948) 'Russian conjugation', Chomsky's (1951) 'Morphophonemics of modern Hebrew', Chomsky and Halle's (1968) *The sound pattern of English*, and the many works composed under the direct influence of the latter has documented in great detail that a given morpheme may have a large number of phonetically distinct actualizations and that the relationship between the form of the base morpheme and that of its different actualizations is of the same lawful kind as that exhibited in (2), though of course not nearly as simple.

The existence of morpheme alternants of great variety and complexity is a fundamental and solidly established fact about language. Since the correct choice among morpheme alternants in a given language L is a *sine qua non* for producing correct utterances in L, it follows that knowledge of these alternants and the principles for choosing among them is a crucial part of the knowledge that every fluent speaker of L possesses and constantly uses. Since this knowledge determines what speakers do and what their interlocutors hear the study of the principles that relate the morpheme alternants of a language to each other sheds important light on the correct representation of morphemes in memory. One of the conditions that an adequate theory of phonological representation must therefore satisfy is that it allows for a correct – i.e., for a natural and insightful – characterization of the morpheme alternations in all languages. Meeting this "alternation" condition is no less essential than satisfying the requirement that the theory should allow for the correct characterization of the articulatory activity involved in producing the utterance and of the acoustic effects of this activity.

It is not difficult to show that speakers not only have knowledge of the feature composition of the phonemes but also make constant use of this knowledge. As shown in (3) the plural suffix in English has the two alternants /z/ and /s/.

(3) (a) ski-s ray-s can-s mile-s rib-s drug-s grave-s
(b) stamp-s stick-s print-s cough-s fourth-s

Since speakers readily form correct plurals of nouns that they have never encountered before (for some experimental evidence in support of this proposition, see Berko 1958), we must conclude that speakers have not memorized each plural form separately but make use of a rule or principle. Assuming that the base variant of the plural suffix is /z/, the rule required must have the effect of turning this suffix into /s/ after voiceless consonants. The requisite rule is stated in terms of phonemes in (4a) and in terms of features in (4b).

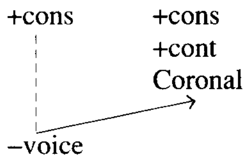
(4) (a) /z/ → /s/ in env. [p, t, k, þ, f] + ____
(b) [+cons, +cont, Coronal] → [−voice]
in env. [−voice] + ____

It is easy to show that the form in (4b) more nearly represents the knowledge that speakers use in producing the plural suffix than does (4a). The evidence, which was brought to my attention by Professor Lise Menn of the University of Colorado, makes use of the fact that English speakers can form plurals of nouns that end with phonemes that are not part of the English phoneme repertory. For example, when asked for the plural of the German name *Bach*, which ends with the phoneme /x/, English speakers normally choose the suffix /s/, rather than /z/. It is obvious that this performance cannot result from the use of rule (4a) because /x/ is not listed in (4a). Since no rule of English could plausibly include reference to a phoneme such as /x/ that is not part of the English phoneme repertory, the performance could not be accounted for by reference to a rule that is formulated in terms of phonemes.

The choice of /s/ to mark the plural of *Bach* is readily explained if the rule is formulated in features terms as in (4b), where the list of phonemes is designated by means of the feature(s) that its member phonemes share. If the latter means is adopted however the set would not be limited to the phonemes actually occurring in a language, but would designate also phonemes that do not occur in the language yet possess the appropriate feature(s). Since, as we have just seen, application of the English plural rule is not limited to words ending with English phonemes, we conclude that English speakers have access to a rule in which sets of phonemes are designated by means of features rather than by listing.[2]

The representation of phonemes as complexes of features allows us to reformulate rule (4b) as in (4c).

(4c)



The formulation (4c) is superior to that in (4b) in that it makes explicit the reason for the change of /z/ to /s/; i.e., the plural suffix assimilates voicelessness from the stem final consonant. Assimilation is not only a very common phenomenon in phonology, it is also an eminently plausible one. It is a manifestation of inertia: an action associated with one segment is carried over to the next.

I note that the preceding discussion of morpheme alternations relied crucially on the standard hypothesis that utterances are composed of phonemes, which themselves are complexes of distinctive features. To replace this hypothesis with one that denies the existence of phonemes and/or features it is necessary at a minimum to show how such an alternative hypothesis accounts for the facts just discussed as well as for other kinds of morpheme alternation that have been documented in the literature. None of the proposals for replacing the standard view of phonemes and features has come close to satisfying this essential criterion of adequacy.

2

As noted above, if phonemes are complexes of features, then words, which are sequences of phonemes, are stored in memory in the form of two-dimensional matrices resembling those in (1). When we wish to say a word we retrieve such a matrix from our memory and use its feature composition as instructions for the different articulatory actions involved in pronouncing the word; these actions produce an

---

[2] The astute reader will have noticed that this alternative is not available for every arbitrary set of phonemes, but only for sets such as /m n ŋ/ or /p t k/ that share one or two features in common. It is a fact of great significance for matters under discussion that the rules that have been discussed by phonologists are overwhelmingly of the latter kind. For some additional discussion of the above example, see Halle (1990).

acoustic signal. The converse happens in perception: hearers perceive an acoustic signal which contains the information enabling them to identify a given word in their memory. Since the word is represented in memory as an array of features and features are instructions for articulatory actions, the hearer must extract these instructions from the acoustic signal.

The need for such a translation from acoustics into articulatory features is a consequence of the decision to treat features as instructions for articulatory actions. This is, of course, not the only possible way to view features. A plausible alternative view may be that features in memory are directly related to properties in the acoustic signal. It is therefore necessary at this point to review the evidence that supports the primacy of articulation over acoustics in phonetics. I note parenthetically that the primacy of articulation over acoustics is a central proposition of the motor theory of speech long advocated by Alvin Liberman of Haskins Laboratories and the University of Connecticut. (See Liberman 1996, and literature cited there.)

Perhaps the most telling evidence in favor of the primacy of articulation over acoustics derives from an understanding of the feature *voice*. Halle and Stevens (1971) argued that there is no feature *voice* and that the phonetic contrast between obstruents produced with and without vocal fold vibration is actually a manifestation of the articulatory action (feature) of vocal fold stiffening.

Voicing or vocal fold vibration is physiologically produced not by individual movements of the folds but rather, like the so-called *Bronx cheer* or *raspberry*, by a form of relaxation oscillation. Pressure is built up below the folds to the point where it separates them and allows air to escape from the lungs. The outflow of air decreases the sub-glottal pressure and causes the elastic folds to come together again. This shuts off the air flow and results in an increase of the sub-glottal pressure to the point where the sequence of events is repeated once again.

Halle and Stevens (1971) started off from the obvious observation that the vocal folds could be made more or less stiff and asked what effect changes in stiffness would have on the vibrations. Unsurprisingly one effect of greater vocal fold stiffness is an increase in the rate of vibration; i.e., sounds produced with stiffer vocal folds have higher pitch than sounds produced with less stiffness in the vocal folds. Somewhat less obvious is the fact that these pitch differences are found only for certain pressure drops across the vocal folds: as the pressure drop across the folds falls below some critical value, the same increase in stiffness that previously had caused a noticeable rise in pitch now results in total cessation of vocal fold vibration; i.e., in voicelessness. I have summarized the basic information in (5).

(5)	large pressure drop	small pressure drop
stiff folds	higher rate of vibration	cessation of vibration
slack folds	lower rate of vibration	presence of vibration

What is responsible for the different pressure drops across the folds? Air flows out of the lungs into the ambient atmosphere because the subglottal pressure exceeds atmospheric pressure. In vowels and other phonemes produced with no blockage of air outflow, the supra-glottal pressure is that of the ambient atmosphere. The situation is different in stops such as [p t k] or fricatives such as [f s x]. Here the outflow of air from the lungs is impeded by a constriction in the mouth. Since the constriction traps air in the mouth, pressure builds up inside the cavity, the pressure drop across the folds decreases, and the vocal fold vibration comes to a halt. To produce vocal fold vibration in obstruents it is necessary to decrease their stiffness.

Changes in vocal fold stiffness thus have two distinct acoustic consequences. In obstruents – i.e., in stops and fricatives – stiffening vs. slacking of the vocal folds results in voicelessness vs. voicing. In all other phonemes – vowels, nasals, liquids – stiffening vs. slacking results in high pitch vs. low pitch sounds. The same articulatory action – stiffening vs. slacking of the vocal folds – results thus in vastly different acoustic consequences: high vs. low pitch in one case, and voicelessness vs. voice in the other. In view of this it is natural to ask whether there is any evidence that might indicate which of these two – the articulatory action or the acoustic consequences – plays the dominant role. The linguistic evidence, some of it reviewed below, indicates that it is the articulatory action that is primary and that the acoustic effects are secondary, thus confirming further the fundamental insight of Liberman's "motor theory of speech".

An example of the linguistic evidence is given by the data from the Songjiang dialect of Chinese in (6) (from Bao, 1990). (The numbers to the right of the different words indicate pitches, with 5 the highest and 1 the lowest.)

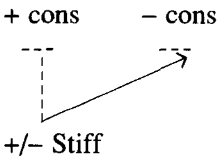
(6) ti 53 'low' di 31 'lift'
ti 44 'bottom', di 22 'brother'
ti 35 'emperor' di 13 'field'

In this dialect the words are differentiated by the three tonal melodies: falling, even, rising.[3] As shown in (6) each of these melodies appears in two versions: a high-pitch version of the left and the low-pitch version on the right. We also see that the high-pitch version occurs after voiceless stops, whereas the low-pitch version occurs after voiced stops. If the acoustic differences were primary, then the fact that the higher pitch is correlated with voicelessness and the lower pitch with voicing, would be a mere accident: the reverse situation would be just as plausible. If, on the other hand, there is only a single feature controlling both the voicing and the pitch distinction, then the facts in (6) have a rather different explanation: they are an instance of assimilation, of spreading the feature [stiff vocal folds] from the word-initial consonant onto the following vowel. The process is represented formally in (7).

---

[3] The fact that the three melodies involve variations in pitch that are independent of those under discussion implies (as noted already by Yip, 1980) that the melody contrasts are controlled by a feature distinct and separate from the one under discussion here. This issue is the subject of current research by Stevens and Halle.

(7)

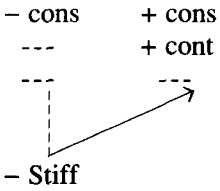


This quite natural account, which is formally similar to the process (4c) responsible for the choice between the plural alternants of English, becomes impossible if [voice] and [stiff vocal folds] are distinct features.

Poser (1981) has shown that in Jabem, a Melanesian language of New Guinea, the converse of the relation in (7) obtains. Whereas in (7) the consonant feature spreads to the vowel and determines its tone, in Jabem it is the vowel tone that determines whether the consonant is voiced or voiceless. Formally, this is described as spreading the feature [+/–stiff vocal folds] from vowel to consonant.

A quite similar process is responsible for the consonant voicing in proto-Germanic accounted for by Verner's Law, as pointed out in an unpublished paper by Noyer (1991). Voicing due to Verner's Law affects continuants that follow a stressless vowel. It is plausible to assume that in proto-Germanic, like in many modern languages, stressed vowels were actualized with high pitch whereas unstressed vowels had low pitch. In terms of the discussion above this means that stressed vowels had the feature [+stiff vocal folds], and unstressed vowels had the feature [–stiff vocal folds]. Voicing by Verner's Law would then be expressed formally by the rule in (8), which differs from (7), mainly in the fact that the consonant assimilates stiffness of the vowel, whereas in (7) it is the vowel that assimilates stiffness from the consonant.

(8)



In sum, there is significant evidence for the proposal that [voice] is not part of the universal set of features and that it is to be replaced everywhere by [stiff vocal folds].[4]

---

[4] In their study of the influence that voicing contrasts in consonants exercise on the pitch contours of adjacent vowels Kingston and Diehl (1994) assume without argument that consonant voicing and vowel pitch are controlled by separate features. The absence of any discussion of this matter is unfortunate since Kingston and Diehl's main point – that in addition to a language-particular phonology, it is necessary to recognize also a language-particular phonetics – is seriously weakened if consonant voicing and vowel pitch are contextual manifestations of the single feature [stiff vocal folds]. Under the latter hypothesis the bulk of the data in the paper are explained as instances of "leaking" of a feature from one phoneme to a neighboring one as, e.g., when nasality spreads from a consonant to an adjacent vowel. On the latter view, the data in the paper provide no evidence for Kingston and Diehl's central claim that there are language-specific influences of one feature on another.

3

Questions have been raised about the motor theory as a result of a series of experiments conducted by Kuhl and associates (see Kuhl and Miller (1975, 1978), Kuhl (1981)), stop + vowel stimuli varying in Voice Onset Time (VOT) were presented to both humans and chinchillas, and it was found that the "form of the labeling function and the 'phonetic boundaries' for chinchillas and English-speaking adults were similar" (Kuhl and Miller, 1975). Since the upper respiratory and alimentary tract in chinchillas is vastly different from that in humans these experiments have been interpreted as constituting a challenge to the claim, central to the motor theory, of the primacy of articulation over acoustics. For example, Miller (1990) writes that these findings are "a clear problem for the motor theory. [They] suggest that at least some aspects of speech perception do not require the operation of a species-specific mechanism" (p. 85). These conclusions seem to me unwarranted for the following reasons.

It was shown by Liberman et al. (1952) that short bursts of noise followed directly by vowel-like stimuli are perceived as stop+vowel sequences. It was shown subsequently by Lisker and Abramson (1964) that the time interval between the burst and the onset of the vowel (commonly referred to as VOT) determines whether the stop was perceived as voiceless or voiced: when the VOT interval is long, the stimuli are identified as beginning with a voiceless stop, i.e., [pa ta ka], but when the VOT interval is short the stimuli are identified as beginning with a voiced stop, i.e., [ba da ga]. Kuhl's experiments showed that chinchillas could be trained to make the same discriminations as humans.

The fact that a given form of behavior is observed in several species does not authorize us to conclude that the same neurological mechanisms trigger this behavior in the different species. For example, horses and even elephants can be trained to walk on their hind legs. From this fact, however, no one would conclude that the same neurological mechanisms underlie the bipedal gait of these three different species. By the same token, from the fact that chinchillas exhibit the same VOT discriminations as humans we may not conclude that the behaviors are based on the same neurological mechanisms in the two species. For humans the VOT stimuli sound like <pseudo->words and unlike other ambient noises such as yawns or car rattles, and as a consequence humans analyze VOT stimuli – but not car rattles or yawns – into phonemes and features. No such analyses are available to chinchillas. Since, unlike humans, they have no knowledge of language and have no lists of words in their memories, chinchillas discriminate the VOT stimuli in purely acoustic terms, as they do all other kinds of noises. That they can be trained to discriminate VOT stimuli in the same way as humans is an interesting fact, but since it has not been shown that the discrimination involves the same mechanisms in the two species, it does not allow us to conclude anything about speech perception in humans.

References

Bao, Z., 1990. On the nature of tone. Ph.D. dissertation, Department of Linguistics, MIT.

Bloomfield, Leonard, 1933. Language. New York: Henry Holt.

Bloomfield, Leonard, 1939. Menomini morphophonemics. Travaux du Cercle de Prague 8, 105–115.

Bromberger, Sylvain and Morris Halle, 1997. The contents of phonological signs. In: I.M. Roca (ed.), Derivations and constraints in phonology. Oxford: Oxford University Press.

Chomsky, Noam, 1951. The morphophonemics of Modern Hebrew. New York: Garland Press.

Chomsky, Noam and Morris Halle, 1968. The sound patttern of English. Paperback reprint 1993 Cambridge, MA: MIT Press.

Halle, Morris, 1990. Phonology. In: D. Osherson and H. Lasnik (eds.), Language: An invitation to cognitive science, Vol. 1, 53–68. Cambridge, MA:

Halle, Morris, 1995. Feature geometry and feature spreading. Linguistic Inquiry 26, 1–46.

Halle, Morris and K.N. Stevens, 1971. A note on laryngeal features. MIT RLE Quarterly Progress Report 101, 198–213.

Jakobson, Roman, 1932. Phoneme and phonology. In: Selected writings, Vol. 1, 23–233. The Hague: Mouton, 1962.

Jakobson, Roman, 1947. Russian conjugation. Word 4, 155–167.

Kingston, John and Randy L. Diehl, 1994. Phonetic knowledge. Language 70, 419–454.

Kuhl, Patricia K., 1981. Discrimination of speech by non-human animals. Journal of the Acoustical Society of America 70, 340–349.

Kuhl, Patricia K. and James D. Miller, 1975. Speech perception by the chinchilla. Science 190, 69–72.

Kuhl, Patricia K. and James D. Miller, 1978. Speech perception by the chinchilla: Identification for synthetic VOT stimuli. Journal of the Acoustical Society of America 63, 905–917.

Liberman, Alvin M., 1996. Speech: A special code. Cambridge, MA: MIT Press.

Liberman, Alvin M., Pierre Delattre and Franklin S. Cooper, 1952. The role of selected stimulus-variable in the perception of unvoiced stop consonants. The American Journal of Psychology 65, 497–516.

Lisker, Leigh and Arthur A. Abramson, 1964. A cross-language study of voicing in initial stops. Word 20, 384–422.

Mandelbaum, David G. (ed.), 1949. Selected writings of Edward Sapir. Berkeley and Los Angeles: University of California Press.

Noyer, Rolf, 1991. Verner's law and underspecification theory. Unpublished paper, Department of Linguistics, MIT.

Poser, Willam, 1981. On the directionality of the tone-voice correlation. Linguistic Inquiry 12, 483–488.

Sapir, Edward, 1925. Sound patterns of language. Language 1, 37–51. [Also in D.G. Mandelbaum, 1949, 33–45.]

Sapir, Edward, 1933. The psychological reality of phonemes. In: David G. Mandelbaum (ed.), 1949, 46–72.

Trubetzkoy, Nikolai S., 1939. Grundzüge der Phonologie. Travaux du Cercle Linguistique de Prague 7.

Yip, Moira, 1980. The tonal phonology of Chinese. Ph.D. dissertation, Department of Linguistics, MIT.