A MODEL FOR THE PHONETIC MENTAL REPRESENTATION OF WORDS

WALTER F. SENDLMEIER

Max-Planck-Institut für Psycholinguistik
Nijmegen, The Netherlands

ABSTRACT

Seven psychological models of word recognition are analysed as to their explicit and implicit assumptions on the phonetic mental representation of words, and are then considered in the light of experimental results concerning the concept of the primary perceptual unit and findings from first language acquisition research. On the basis of these considerations a model for the phonetic mental representation of words is proposed which assumes simultaneous representation of differently sized units in the form of prototypes. The implications of this model for models of word recognition are discussed.

INTRODUCTION

Hardly any of the leading word recognition models contains explicit information on the phonetic mental representation of words. This may be seen as a serious drawback of these models considering that (phonetic) mental representation may not only be regarded as a result of the perception process, but that it functions at the same time as a monitor for perception. Almost all models, however, make more or less clear statements on primary perceptual units to which - at least implicitly - the status of mental representation is ascribed.

- Klatt /1/ assumes in his 'LAFS' (lexical-access-from-spectra) model that the listener is able to distinguish words directly by spectral analysis of the speech signal without having to segment it into smaller units. However, he also assumes that words have an internal structure which can best be described by units of diphone size. An important part of the word recognition process according to Klatt's model is the recognition of the internal diphone structure of a word by a listener. In this model words must thus be mentally represented as diphone sequences in the listener.
- In describing his 'logogen model' Morton /2/ gives the impression that he does not regard any segmentation within word boundaries necessary for the recognition process. Words are held to be represented as holistic entities.
- In the 'cohort model' /3/ it is assumed that words are represented as sequences of discrete units in the listener. The size of these units equals approximately that of single sounds, although statements on the linguistic status of the units and thus on their degree of abstractness (phoneme, allophone or phone) are avoided.
- Forster /4/ was the first to include specifications on the phonetic mental representation of words in his 'search model'. This model is based on the assumption that words in the lexicon are represented as sequences of phonological segements (phonemes).
- Pisoni, Nusbaum, Luce and Slowiaczek /5/ also make explicit statements on the mental representation of words in their 'phonetic refinement theory'. They believe that words are represented in the mental lexicon as sequences of discrete phonetic segments equalling single sounds which are defined in a multi-dimensional space /6/.
- Elman and McClelland /7/ assume that there are processing units of different sizes on different levels. These processing units are acoustic phonetic features, phonemes (allophones) and words. Even though Elman and McClelland assume interactions between these different units during the word recognition process, on closer examination of their 'trace model' these units appear to be hierarchically organized. Thus the question remains, whether the different units are simultaneously present in the sense of a mental representation or whether they have to be deduced one from another in a given sequence.
- Grosjean and Gee /8/ distinguish between units of processing and units of representation, but only make specific statements on the former. In their view, units of processing are the stressed syllable and the phonological word consisting of a stressed syllable and a number of unstressed syllables linked with the stressed syllable. Unfortunaltely, Grosjean and Gee do not specify how these units are related to potential units of mental representation. Considering the importance the authors ascribe to the function of prosodic features in the word recognition process, it seems feasible to deduce that they do not tend to assume that words are phonetically represented in form of sequences of discrete single sounds.

PRIMARY PERCEPTUAL UNITS

As mentioned above, the problem of phonetic mental representation of words is closely linked with the question of the basic (natural) units of speech perception. When, in the early fifties, experimental phoneticians and psychologists started to investigate the relation between the linguistic unit and its processing by the human listener, they were guided by the concept of minimal pairs and the ensuing distinctive feature theory developed by phonologists. Thus they focussed on the smallest isolated and reduced units - presented in form of synthesyzed signals to listeners in the laboratory who were asked to identify and discriminate them. Notwithstanding the valuable results obtained by such studies, one should be aware of the fact that the experiments were based on artificial acoustic phenomena which were as far distant as possible from their natural manifestations.

In criticizing the assumption of distinctive features as being psychologically real, in the beginning of the seventies an explicit discussion on the nature of the primary perceptual unit began. It was believed that in reaction time experiments, especially by target monitoring tasks, one could determine linguistic, taxonomically structured units according to their relevance as units in the speech perception process. One of the important results of these experiments is that the reaction times for short sentences, words, syllables and sounds are the same, if the search list consists of units of the same size as the target unit /9, 10, 11/. On the condition that reaction time experiments are an adequate means to reveal information on the primary perceptual unit, it can be deduced that units of different sizes may serve as primary perceptual units. In spite of such results a number of authors still argue for certain units to be the exclusive representatives of primary perception and try to prove their hypotheses by experimental studies /12, 13/.

RESULTS FROM FIRST LANGUAGE ACQUISITION RESEARCH

Another possibility of gaining insight into the phonetic mental representation of words lies in looking at the early stages of the child's language acquisition process. In first language acquisition research it has become quite an unquestioned fact that the child learns a word as bearing meaning corresponding to a certain object or class of objects. It seems plausible to assume that in this learning process the phonetic characteristics are globally perceived; in other words, the child learns the word 'ball', for example, as a phonetic unit and not as a combination of the single sounds /b/+/ɔ:/+/l/ or even as a matrix of 3x9 distinctive features.

Empirical results support this view: For example, Bruce /14/ found in investigations with 5- to 7 1/2-year-old children that during this stage in development holistic processing of words changes to more analytic processing. Liberman, Shankweiler, Fischer and Carter /15/ carried out experiments with 4- and 5-year-olds and found that these children could segment words much more easily into syllables than into single sounds. In using rhyming tests Magnusson, Naucler and Söderpalm /16/ found that preschool children were not able to give metalinguistic judgments on the basis of the phonetic-phonological structure of the words they heard. School children, however, were well able to do this, which may be accounted for by their ability to read and write. These findings, among others, point to the fact that at first the child perceives words phonetically in a global, non-analytic manner.

The prerequisites of a more analytic way of perceiving speech elements, in other words, the insight into the existence of certain recurring features, is only possible on the grounds of a substantial vocabulary. The possibility that an analytic recognition of words may occur in a more advanced stage in the process of cognitive development and that it may be furthered by special training is not questioned. But such perception of speech which analyses different speech signals within word boundaries may only follow global perception in the developmental sequence, and it cannot extinguish the earlier developed global way of perception.

To summarize, in this approach it is assumed that the child begins by recognizing words as global units. More analytic ways of speech perception may be used in later stages of language acquisition with interindividually varying degrees.

A MODEL OF THE MENTAL PHONETIC REPRESENTATION

These considerations lead to the following model of phonetic mental representation. The grown-up speaker/listener has stored a variety of mental representations on the phonetic level, the most important being: words, syllables, single sounds and phonetic features. Figure 1 illustrates the outlines of the model.

It should be noted that the different units are not localised on different levels of representation, but that they are different kinds of representation within one level, i.e. the phonetic level. These different kinds of representation are simultaneously

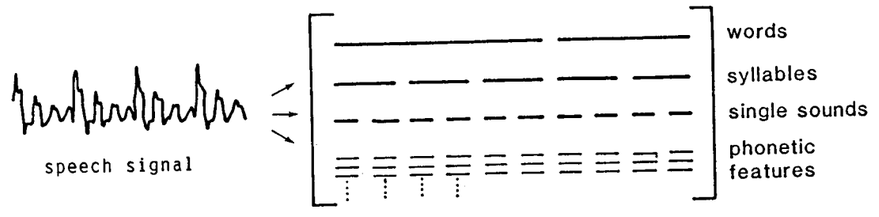

Fig. 1 : Different kinds of mental representation of words on the phonetic level which are simultaneously at the disposal of the listener; the listener focusses that kind of representation first which seems most efficient for word recognition.

at the disposal of the listener/speaker once he has established them. From which kind of representation the listener primarily takes the relevant information for solving a perception task is determined, for example, by the type of task, the context of perception, the speed and/or the complexity of the incoming stimuli etc.. Besides, it seems to make sense to assume that the perceptual activities of a listener vary not only with varying tasks, but that he may also interchangeably focus on different kinds of representation while solving one particular task, for example by recognizing a phrase or a sentence. Thus a listener can switch to single sounds or even phonetic features when discriminating difficult words such as proper names or words of a foreign language, and then he can switch back to words later.

Such a type of model in which a simultaneous representation of stimuli within different systems of similarity and contexts is postulated, is successfully being used in other psychological fields, as for example in the cognitive psychological research on problem solving; it has amply been shown that the flexibility in problem solving is based on the ability to change perspective /17/.

Since different listeners make different experiences in their perceptual surroundings, the degree of their ability to differentiate, i.e. the number of types of representation of a given word they have at their disposal, may differ from one individual to another. This is why the kind of representation on which listeners rely in a successful recognition process may also vary according to properties of the listeners themselves. For example, the knowledge of a phonetically oriented writing system (such as is acquired when learning to read and write an alphabetical writing system) may lead to a more differentiated organization of the mental representation of words. Morais, Cary, Alegria and Bertelson /18/ could in fact show that adult illiterates had much more difficulties in solving certain linguistic tasks involving detailed phonetic analyses than literate adults. What Morais et al. showed for speakers of Portugese, Sendlmeier /19/ could confirm also for native speakers of German. Within the scope of the introduced model these results may be explained in such a way that the adult illiterates have no concept of the single sound the way literates have. This, however, should not lead to the misinterpretation that the one group could listen better than the other. As a matter of fact, illiterates are just as able as literates to distinguish minimal phonetic differences in discrimination tasks, which, however, gives no clue as to the primarily focussed type of representation in the process of word recognition.

Closely related to the question in which size the phonetic perceptual units are represented is the problem of how these representations are present. Here Wertheimer's concept of 'ideal types' /20/ or Rosch's related concept of 'prototypes' /21/ seem to be adequate alternatives to abstract feature matrices.

The representation in form of prototypes is postulated for all kinds of representation of the phonetic level in the model. It seems plausible to assume that a listener generates a prototype from all the ever heard representatives of a category in the sense of a statistical mean during the course of language acquisition. If one supposes that phonetic units of different sizes (up to words) are represented analogously in form of typical prototypes, but not in the sense of a first degree isomorphy, this implies an enormous capacity of the long term memory. Objections by scientists who by referring to - up to now uncertain - principles of economy argue against such a supposition of storage-consuming representation can be rejected in view of an almost unlimited capacity of the human brain /22/. The material basis of an analogous representation in form of prototypes may be seen in neurophysiological correlates of spectral patterns, since it may be taken for certain that the incoming soundwave is subjected to a frequency analysis by the peripheral hearing system.

CONSEQUENCES FOR WORD RECOGNITION MODELS

The presented model of mental representation contains a number of constraints on the process of word recognition. This is due to the fact that structure and process mutually depend on each other. It is up to word recognition models to delineate the rules and mechanisms that characterize the different types of strategies in speech perception. However, in doing so the following facts should not be ignored:

- Word stress patterns are normally used in word retrieval; words seem to be organized in the lexicon according to stress contours /23, 24, 25/.
- Linguistic differences can cause listeners with different languages to develop different perceptual strategies /26/.
- Configurational (prosodic) features of words often hinder the listener from focussing on single sounds in recognizing words /27/.
- Unstressed function words usually are recognized some time after their off-set, in most cases only after taking into account the following stressed syllable /8/.
- The size of the phonetic units used by listeners varies with the complexity of the words in similarity judgments /28/.
- The size of the primary perceptual unit varies with the size of the respective context /29/.

Word recognition models which assume only one kind of primary perceptual unit - phonetic features, single sounds, syllables or words - are confronted with a number of problems when trying to explain findings like the ones listed above. It seems that only such models will be of lasting importance which start from the assumption that the listener has active control over the process of auditory word recognition and that he can focus at will on any kind of representation that seems useful for successful word recognition.

REFERENCES

/1/ D.H. Klatt, "Speech understanding systems and speech perception theory", Journal of Phonetics 7, 1979, 279-312.

/2/ J. Morton, "Word recognition", in J. Morton, J. C.Marshall (eds.), Psycholinguistics 2: Structures and processes.Cambridge, 1979, 107-156.

/3/ W. Marslen-Wilson, A. Welsh, "Processing interactions and lexical access during word recognition in continuous speech", Cognitive Pschology 10, 1978, 29-63.

/4/ K.I. Forster, "Levels of processing and the structure of the language processor", in W.E. Cooper, E.C.T. Walker (eds), Sentence processing: psycholinguistic studies presented to Merrill Garrett. Hillsdale 1979, 27-86.

/5/ D. Pisoni, H. Nusbaum, P. Luce, L.M. Slowiaczek, "Speech perception,word recognition and the structure of the lexicon", Speech Communication 4, 1985, 75-95.

/6/ M. Treisman, "Space or lexicon? The word frequency effect and the error response frequency effect", Journal of Verbal Learning and Verbal Behavior 17, 1978, 37-59.

/7/ J.L. Elman, J.L. McClelland, "An architecture for parallel processing in speech recognition: the trace model", in M.R. Schroeder (ed.), Speech and Speaker Recognition, Basel, 1985, 6-35.

/8/ F. Grosjean, J.P. Gee, "Another view of spoken word recognition", Cognition 15, 1987, in press.

/9/ H.B. Savin, T.G. Bever, "The nonperceptual reality of the phoneme", Journal of Verbal Learning and Verbal Behavior 9, 1970, 295-302.

/10/ D.J. Foss, P.A. Swinney, "On the psychological reality of the phoneme: perception, identification and consciousness", Journal of Verbal Learning and Verbal Behavior 12, 1973, 246-257.

/11/ D. McNeill, L. Lindig, "The perceptual reality of phonemes, syllables, words and senteces", Journal of Verbal Learning and Verbal Behavior 12, 1973, 419-430.

/12/ D. Norris, A. Cutler, "Juncture detection", Linguistics 23, 1985, 689-705.

/13/ J. Mehler, J.Y. Dommergues, U. Frauenfelder, J. Segui, "The syllables role in speech segmentation", Journal of Verbal Learning and Verbal Behavior 20, 1981, 298-305.

/14/ D. J. Bruce, "The analysis of word sounds by young children", Journal of Educational Psychology 34, 1964, 158-159.

/15/ I.Y. Liberman, D. Shankweiler, F.W. Fischer, B. Carter, "Explicit syllable and phoneme segmentation in the young child", Journal of Experimental Child Psychology 18, 1974, 201-212.

/16/ E. Magnusson, K. Naucler, E.Söderpalm, "Form or substance? The linguistic awareness of preschool children and school children investigated by means of a rhyming test", Abstracts of the 10th international Congress of Phonetic Sciences, Dordrecht, 1983, 633.

/17/ W. Prinz, "Wahrnehmung und Tätigkeitssteuerung", Berlin, 1983

/18/ J. Morais, L. Cary, J. Alegria, P. Bertelson, "Does awareness of speech as a sequence of phones arises spontaneously?" Cognition 7, 1979, 323-331.

/19/ W.F. Sendlmeier, "Psychophonetische Aspekte der Wortwahrnehmung, Hamburg, 1985.

/20/ M. Wertheimer,"Untersuchungen zur Lehre von der Gestalt, II", Psychologische Forschung 4, 1923, 301-350.

/21/ E. Rosch, "Cognitive Reference points", Cognitive Psychology 7, 1975, 532-547.

/22/ W. Penfield, "Consciousness, memory and man's conditioned reflexes", in K.H. Pribram (ed.), On the biology of learning, New York, 1969.

/23/ R. Brown, D.McNeill, "The 'tip of the tongue' phenomenon", Journal of Verbal Learning and Verbal Behavior 5, 1966, 325-337.

/24/ D. Fay, A. Cutler, "Malapropisms and the structure of the mental lexicon", Linguistic Inquiry 8, 1977, 505-520

/25/ E. Engdahl, "Word stress as an organizing principle of the lexicon", in D. Farkas (ed.), Papers from the parasession on the lexicon. Chicago, 1978, 138-147

/26/ A. Cutler, J. Mehler, D. Norris, J. Segui, "A language specific comprehension strategy", Nature 304, 1983, 159-160.

/27/ A.G. Samuel, W.H. Ressler, "Attention within auditory word perception: insights from the phonemic restoration illusion", Journal of Experimental Psychology: Human Perception and Performance 12,1986, 70-79.

/28/ W. F. Sendlmeier, "Auditive judgements of word similarities",Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 40, 1987, in press.

/29/ H. Fujisaki, K. Hirose, H. Udagawa, N. Kanedera, "A new approach to continuous speech recognition based on considerations on human processes of speech perception", ICASSP, IEEE 86, 1986, 1959-1962.

ACKNOWLEDGEMENT

I thank Prof. H. Fujisaki for fruitful discussions on the topic of mental phonetic representation.