Word recognition—uniqueness or deviation? A theoretical note

MARCUS, Stephen M., FRAUENFELDER, Ulrich Hans

Abstract
The structure and distribution of entries in the lexicon impose strong constraints on word recognition. A construct of value in operationalizing such constraints is that of uniqueness point. Computed by reference to a phonetic dictionary, it is the earliest moment at which a word can be uniquely distinguished from all others in terms of a sequential phoneme by phoneme comparison. There is empirical evidence that the uniqueness point bears a close relationship to the recognition point, the moment at which a word can actually be recognized on the basis of incoming acoustic information. In reality, recognition depends upon noisy and unreliable acoustic information, making categorical phonetic decisions difficult, and the value of a uniqueness point, questionable. Minimal deviation is proposed as a new construct to represent the extent of mismatch between a given word and the closest non-identical word in the lexicon. A computer analysis of a large phonetic dictionary shows that on average minimal deviation increases almost constantly with stimulus input subsequent to the uniqueness point. This simulation suggests that the [...]
Manuscript no. 497

Word recognition - uniqueness or deviation?
a theoretical note

Stephen M. Marcus
&
Uli H. Frauenfelder

Submitted to: Language and Cognitive Processes
word recognition - uniqueness or deviation?

Stephen M. Marcus
Institute for Perception Research (IFO)
Eindhoven, The Netherlands

&

Uli H. Frauenfelder
Max Planck Institut für Psycholinguistik
Nijmegen, The Netherlands

Abstract.

The structure and distribution of items in the lexicon imposes strong constraints on the word recognition process. A construct which has proven of value in operationalizing such constraints is that of uniqueness point, defined in the Cohort Model as the earliest point in a word at which it can be uniquely distinguished from all others in a given language (Marsten-Wilson & Welsh, 1978). This model assumes that each segment in the incoming signal can be correctly categorized as it occurs. In practice, the recognition process depends upon acoustic information which is often noisy and unreliable, making such categorical decisions dangerous and the determination of a uniqueness point questionable. The term minimal deviation is defined as the extent of mismatch between a stimulus and the closest non-identical word in the lexicon. A computer analysis of a large phonetic dictionary shows that after the uniqueness point minimal deviation on average increases steadily with subsequent stimulus input. Therefore, it is possible that the effectiveness of uniqueness point as a predictor of word recognition performance is not due to the single mismatch at that moment in the stimulus, but to later deviation. Recognition performance on stimuli in which minimum deviation does not increase past the uniqueness point will allow this to be tested.
An aspect of speech recognition that has been receiving increasing attention is its real-time nature, that is, the way in which stimulus information is used as it arrives in time in mapping onto the appropriate lexical entry. Strong constraints on such a process will be imposed both by the physical structure of the stimulus and also by the structure and distribution of items in the lexicon. This paper will show that current theoretical approaches to the recognition process which employ widely different and apparently conflicting mechanisms can be reconciled by examining structural properties of the lexicon. Some of these properties will be illustrated by the results of a computer analysis.

The Cohort Model and Uniqueness

An attempt to quantify both temporal and informational aspects of word recognition is the Cohort Model (Marslen-Wilson & Welsh, 1978). This model proposes that as stimulus information arrives, a large cohort of those words compatible with the initial two or three phonemes of a word is selected. This cohort is then successively reduced by removing word candidates which do not correspond with later arriving information. This process continues until only one candidate remains, and the stimulus is thus recognized. Seen in this way, the recognition process is both categorical and left-to-right in time; phonemes (or other unspecified perceptual units) are assumed to be categorically recognized as they arrive, and to be used to immediately activate appropriate, and then deactivate inappropriate, word candidates. The value of this approach is that for each word in a given language, a uniqueness point can be defined as that point at which the initial sequence of phonemes is common to that word and no other; recognition is assumed to be crucially
dependent upon the location of this uniqueness point.

This construct has been shown to have considerable empirical validity (for a review, see Marslen-Wilson, 1983). In particular, monitoring reaction times to the phoneme target /t/ at various locations in 360 words were found to be highly correlated to the target location relative to the uniqueness point of that word (Marslen-Wilson, 1983). Ottevanger (1984) found that mispronunciations prior to the uniqueness point were detected significantly slower than those after the uniqueness point for a range of uniqueness point locations in isolated words. Tyler & Wessels (1983) used the gating paradigm to demonstrate that the uniqueness point is also a useful construct in predicting both the amount of acoustic information subjects need to guess a word, and the subjective certainty of this guess.

Probabilistic nature of the recognition process

While such results provide support for the concept of uniqueness point as a useful construct in predicting the time-course of word recognition, there is reason to question some of the simplifying assumptions in the Cohort Model. In particular, the definition of uniqueness point presupposes that phonetic decisions are made in real-time as acoustic information associated with each phoneme arrives. In practice, it seems unlikely that such categorical decisions can reliably be made with the noisy and ambiguous signal which is speech. Miller, Heise & Lichten (1951) have shown that words which are fully intelligible in sentences are often poorly understood when extracted and presented in isolation. Thus, clearly, incoming phonetic information cannot always be categorically recognized solely on the basis of the acoustic signal. In the same way, phonemic information also appears to be evaluated
probabilistically rather than categorically during the process of word recognition (Frenenier & Carter, in prep.; Pisoni & Tash, 1978; Sereeter & Marie, 1979; Massive & Covert, 1983; Whalen, 1983).

A class of models based upon probabilistic decisions is represented by interactive activation models, which have been developed for both visual word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) and auditory word recognition (Elman & McClelland, 1984). In the latter, there are several levels of processing, each made up of a set of interconnected nodes representing distinctive features, phonemes and words. These discrete yet interacting levels of processing are continuously exchanging probabilistic information. Incoming sensory input provides bottom-up excitation of the distinctive feature nodes, which in turn activate phoneme nodes. As the phoneme nodes become excited, they appropriately alter the level of activation of word nodes. Conversely, higher level nodes can provide feedback to lower level nodes; for example, lexical constraints are hypothesised to increase or decrease the level of activation of phoneme nodes. Computer simulations of this model are claimed to mimic many effects found in human performance (Elman & McClelland, 1984).

Uniqueness and Deviation

In a probabilistic framework, what should be decisive for word recognition is the moment at which the cumulative mismatch between the input and all but one lexical candidate reaches some criterion value. In order to allow for possible misperception of the input, this criterion would presumably need to be greater than one phoneme. It would then be possible for errors in perceiving an early phoneme in a word not to be disastrous for word
recognition, since enough evidence may be present later in the word to confirm a match with the correct lexical item and disconfirm all others. In contrast, in the Cohort Model, the only factor of importance is the location of the uniqueness point, the moment at which the stimulus first mismatches all but the appropriate lexical entry by a single phoneme. This location, at first glance, appears to have little relation to the cumulative mismatch required in probabilistic models. How is it then that there is good empirical evidence both for the probabilistic nature of processing and for the efficacy of the uniqueness point as a predictor of word recognition?

By defining the concept of deviation and examining its relation to the location of the uniqueness point in a large lexicon of real words, it will be demonstrated that the efficacy of the uniqueness point is the result not of accident or fortuitous choice of stimuli, but of statistical properties of the lexicon itself.

For practical rather than theoretical reasons (the availability of a computer readable phonemic dictionary), the following section will make use of phonemes in computing uniqueness points and the extent of deviation between lexical items. Though there is no conclusive evidence to favour any specific code in the mapping from sound to word, be it phonemic, syllabic or in terms of some set of distinctive features, let us note that phonemes are defined by their ability to discriminate between different words, or words and non-words (Swadesh, 1934), and may therefore be supposed to correlate with some of the parameters of importance in the recognition process.
A computerised phonetic dictionary was used in order to empirically investigate the relationship between uniqueness point and a measure of deviation between words. Deviation will be defined as the number of phonemes mismatching between a stimulus token and a particular item in the lexicon. This can be determined not only for the entire stimulus, but also at any moment in time by considering only those phonemes from word onset up to that moment. Given a lexicon, deviation can be computed between the stimulus and each lexical entry. Minimum deviation will be used to refer to the deviation between the stimulus and the closest item (other than the stimulus item itself) in the lexicon, also for each moment in time. By definition, for a correct representation of the stimulus, up to the uniqueness point the minimum deviation will be zero, and at the uniqueness point one.

The interesting question is how minimum deviation increases past the uniqueness point. It may be that for most lexical items the minimum deviation continues to increase rapidly, that is, all or most subsequent phonemes differ. It is clear that there are examples for which minimum deviation does not increase, even well past the uniqueness point: though the uniqueness point of the word "mobility" is at the first /i/, it has a minimum deviation of one right up to the end of the word, since the lexicon also includes the word "mobility". Note that in this case, the word which deviates minimally overall from the target word was in fact never in its word initial cohort. In order to determine the way in which minimum deviation increases after the uniqueness point, minimum deviation was determined for each phoneme position past the uniqueness point for each of the words in a 20,000 word phonetic dictionary of American English. Figure 1 shows the results.
As can be seen, averaged over these 20,000 words, minimum deviation increases steadily from uniqueness point onwards. This means that the uniqueness point is, on average, not only the point at which a stimulus ceases to match any lexical entry except the correct one, but also the point after which most subsequent phonemes also fail to find a match. The empirical success of the uniqueness point as an indicator of word recognition performance may in fact not be attributable to the initial deviation, but rather to the correlated measure of cumulative deviation. However, this pattern is not shared by all words; "nobility" has its select companions. Table I lists two sample sets of words of varying length exhibiting large and small minimum deviations at the end of the word, but having roughly the same uniqueness point.

Table I about here

The generality of these results remains to be tested for languages other than American English. Furthermore, the use of other measures of deviation needs to be studied; no account was taken here of the possible similarity of
the mismatching phonemes. This could be done using some perceptual measure (such as those used in the current task) or a separate classification such as into distinctive features (the perceptual status of which is not unequivocal). The current simulation also made use of strict sequential comparison and positional information on each phoneme — no insertion or deletion of phonemes was allowed between the two items. Such a comparison could be replaced by algorithms allowing limited insertion or deletion, or even by approaches in which no explicit sequential order is imposed whatsoever (Marcus, 1981, 1983). Though such extensions would increase the generality of the current results, it seems unlikely that they will dramatically change the fundamental relationship demonstrated here.

Conclusions

To determine whether uniqueness point is an important theoretical construct, as suggested by the Cohort Model, or simply a statistical artefact well correlated with the amount of deviation between a word and other lexical entries, recognition experiments will need to be carried out with words in which uniqueness point and deviation are varied independently. Computer databases will play a crucial role in devising the stimuli for such experiments. These experiments will also allow the perceptual validity of various definitions of "deviation" to be compared.

Whether or not uniqueness point as such turns out to be the crucial variable governing word recognition performance, it may be more than a statistical quirk that the lexicon exhibits the strong relationship shown between uniqueness and deviation. Could it be that a "uniqueness principle" is important in guiding the selection of words for a spoken lexicon — that the
Lexicon evolved such that the left-to-right structure of words leads as rapidly as possible to the definition of a unique candidate, and from then on time to define its pressure and dissimilarity that of any other word. This would have considerable advantages in a real-time system in which word can be recognized well before its offset, allowing a rapid selection to be made and subsequent phonological information to be used to confirm the accuracy of this choice. The Cohort Model would then be not so much a literal description of the recognition process, but rather of the ideal properties of the spoken lexicon.
REFERENCES


Figure 1. Minimum deviation as a function of number of phonemes past the uniqueness point for a 21,000 phoneme section of American English. The standard deviation is plotted around each data point. The diagonal line indicates the relation expected if all subsequent phonemes mismatched after the uniqueness point.
<table>
<thead>
<tr>
<th>Word</th>
<th>Pattern</th>
<th>Uniqueness</th>
<th>Minimum &quot;Competitor&quot; Point</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALTITUDE</td>
<td>@ptxyud</td>
<td>8</td>
<td>1</td>
<td>ALTITUDE</td>
</tr>
<tr>
<td>CONSIDER</td>
<td>k@nsxleSxn</td>
<td>10</td>
<td>1</td>
<td>CONSIDER</td>
</tr>
<tr>
<td>FIDGMENT</td>
<td>f@y@ntt</td>
<td>7</td>
<td>1</td>
<td>FIDGMENT</td>
</tr>
<tr>
<td>BASIS</td>
<td>oesxs</td>
<td>5</td>
<td>1</td>
<td>BASIS</td>
</tr>
<tr>
<td>POSITION</td>
<td>fxzISxn</td>
<td>7</td>
<td>1</td>
<td>POSITION</td>
</tr>
<tr>
<td>ITALIAN</td>
<td>st@lyxn</td>
<td>7</td>
<td>1</td>
<td>ITALIAN</td>
</tr>
<tr>
<td>ULTRA</td>
<td>@brxkkxd@brx</td>
<td>5</td>
<td>6</td>
<td>ULTRA</td>
</tr>
<tr>
<td>BARBITURATE</td>
<td>karb@hYdret</td>
<td>5</td>
<td>6</td>
<td>BARBITURATE</td>
</tr>
<tr>
<td>HEPATITIS</td>
<td>h@pxpatxmxs</td>
<td>5</td>
<td>6</td>
<td>HEPATITIS</td>
</tr>
<tr>
<td>ACCENTUAL</td>
<td>InsEstSxwx</td>
<td>7</td>
<td>5</td>
<td>ACCENTUAL</td>
</tr>
<tr>
<td>NONAGENARIAN</td>
<td>v@ok@byx1Eri</td>
<td>5</td>
<td>6</td>
<td>NONAGENARIAN</td>
</tr>
<tr>
<td>CENTRALIZATION</td>
<td>vEntrIlxkw1zM</td>
<td>7</td>
<td>7</td>
<td>CENTRALIZATION</td>
</tr>
</tbody>
</table>

Example sets of words with small and large minimum deviation at the end of the word yet similar uniqueness points. For the set of words with large word final minimum deviation, the "competitor" is only one of a large number of words exhibiting such a deviation from the specified word.