A CASE FOR GLOBAL LISTENING STRATEGIES

J.C.T. RINGELING

Dept. of English
Utrecht University
Oudenoord 6
3513 ER Utrecht, The Netherlands

W. EEFTING

Dept. of Phonetics
Utrecht University
Trans 14
3512 JK Utrecht, The Netherlands

ABSTRACT

A case is made for global perceptual strategies. In poor listening conditions subjects appear to be able to perceive and comprehend elliptic speech, albeit with some difficulty. If sufficient semantic information is available, they seem capable of basing themselves on global characteristics in speech sounds, particularly on information related to place of articulation. The study pleads for the formulation of perceptual features to obtain a better insight into the processes operative in speech perception.

I. INTRODUCTION

When Zue (1) showed that a trained spectrogram reader can recognize a substantial number of words from spectral information, the discussion on invariant features in speech perception gained new ground. In the seventies many linguists did not take invariance very seriously, although some invariant features were generally accepted (see e.g. 2 and 3). Naturally, Zue’s success in reading spectrograms was partly caused by extensive use of linguistic expectancy to solve ambiguities, but it made clear that some sort of invariance must be present in speech, although perhaps of a different nature than had traditionally been accepted in terms of linguistic features.

Carlson, Elenius, Granström and Hunnicut (4) and more recently Veenhof and Bloothoof (5) have shown that it may theoretically be possible in many cases to come a long way in arriving at word identification by specifying the acoustic information on the basis of broad phonetic categories. They showed that a classification of phonemes into global categories such as plosives, nasals, fricatives, remaining sonorant or vowel, often provides sufficient information to limit the number of words in a cohort for recognition to take place. This insight that word recognition may be feasible on the basis of a broad phonetic classification has proved helpful in automatic speech recognition (6,7).

However, it is by no means certain if human perception can adequately use a broad phonetic classification in the process of listening to connected speech, and if so, it remains questionable whether listeners base themselves on the same phonetic categories as are frequently adopted in theoretical studies. If we wish to find out what phonetic underlying features can be used in human perception, it is imperative that listening tasks are constructed which vary the amount of acoustic information along global phonetic parameters. An attempt at such a task was an informal study by Ringeling (8), who demonstrated that Dutch listeners could fairly successfully identify sentences in which all consonants had been replaced for consonants that were similar with regard to place of articulation, in such a way that the phonotactic constraints of Dutch were not violated. The use of elliptic speech (see e.g. 9) thus served to manipulate the amount of acoustic information in the speech signal. An English example in ordinary orthography would be the conversion of the saying: ‘no place like home’ to ‘mow crafe wipone’. The resulting sentences sounded Dutch, but could not readily be understood. However, when redundancy of the acoustic signal was reduced by adding noise to the sentences, it turned out that listeners produced much better recognition scores on the same material. One of the most interesting findings of this study was that subjects were rarely aware of the manipulations that had been carried out. This suggests that a global phonetic analysis had taken place on the basis of similarity of place of articulation. In view of the task at hand, which drew heavily on an intensive use of linguistic expectancy, sentences with constraining context were understood much better than those with relatively neutral content.

Van der Woude (10) based a study on this idea. He investigated the theoretical possibility of arriving at unique identifiability of words by grouping consonants together, either on the basis of manner of articulation, or on the basis of place of articulation. On the basis of a random sample of 100 words from 68,000 word tokens (12,000 word types), he found that specification of Dutch words in terms of broad phonetic classes thus defined, did not yield a clear theoretical advantage to either classification. In his definition of patterns, leaving vowel-quality intact, he found...
We therefore expect that in the experiment reported in anneililleggand Isard as early as 1963. showed equally effective, since it is well-known to AS is articulation (see e.g. 11). It was there will be identified correctly more often than on here, sentences with high semantic constraints presented under high levels of noise. If ts be reclassification would seem to qualify as potendent candidates from a 12,000 wordtype lexicog of the basis of what was stated above, we will expect to find a discrepancy in recognition scores based on the amount of phonetic information. If the PLACE-changed and manner-changed consonants lead to unique word patterns, better recognition scores are expected than if those consisted of separate sounds, resulting word patterns leave room for ambiguities.

2. METHOD
2.1 Stimuli
21 Sentences were synthesized using the diphone synthesis system, developed by Eiseleiborn (1). In order to preserve it a natural flow of speech US was possible to arrange the diphone transitions so that no one could tell at a glance where one word ended and the next began. This was turned on 1 second before the signal. The noise was turned on 1 second before the signal started and turned off .5 s after the speech signal had ended.

3. RESULTS
Reactions from the subjects and the amount of missing data (4.5 % of the sentences) indicated that the task was considered quite difficult. In some instances subjects even asked how much time had been used for synthesis. The Feature voiced/voiceless in these elliptic sentences remained unaffected. Some of the sentences were synthesized without manipulation of consonant features.

4. DISCUSSION AND CONCLUSIONS
In this experiment we hoped to learn something about the type of acoustic and non-sensory information that listeners use when faced with circumstances where no acoustic information is available. By using lexical knowledge rather than relying on the phonetic variability of the speech signal, it may be possible to arrive at core-features underlying perception in the future. Some of the features may be rather different from what we have traditionally used in articulatory or phonological terms. It is, for instance, noteworthy that in the synthesis techniques, when applied to perceptual studies, the use of a particular standard or nontypical feature has hardly any effect on the listeners ability to arrive at the correct perceptual features. This is a different from what we have traditionally used in articulatory or linguistic terms. It is, for instance, noteworthy that in the synthesis techniques, when applied to perceptual studies, the use of a particular standard or nontypical feature has hardly any effect on the listeners ability to arrive at the correct perceptual features. This is a different from what we have traditionally used in articulatory or linguistic terms.
variables in use now, are sometimes haphazard and only used ‘because they appear to work’. If we obtain a better understanding of the fundamentally important variables in speech perception, many issues may become more accessible. Notice in this respect that Van der Woude (6) found no theoretical reason to prefer a classification based on PLACE-CHANGED consonants to one that was based on MANNER-CHANGED consonants. But actual perceptual strategies evidently favour a PLACE-CHANGED approach. Nevertheless, we are by no means certain yet, if a PLACE-CHANGED categorization is the best possible approach in global listening strategies. It would be highly counterintuitive if this was not the case, but we will need to lay bare the fundamental features of speech perception first.

REFERENCES:
