CATEGORICAL PERCEPTION OF VOWELS

Ellen Gerrits and Bert Schouten

Utrecht Institute of Linguistics OTS, Utrecht University

ABSTRACT

The present research aimed at investigating the difference in perception between normal, underspecified, and overarticulated, isolated vowels. Unexpectedly, we found observed discrimination to be worse than predicted by the labelling data. There was no difference in the degree of categorical perception between the two vowel conditions.

1. INTRODUCTION

In categorical perception research, the relation between discrimination and classification of a stimulus continuum between two speech sounds is investigated. The first experiment designed to test this relation was performed by Liberman, Harris, Hoffman & Griffith (1957). According to their hypothesis, discrimination of certain speech sounds would be limited by classification; two different stimuli could be discriminated only to the extent that they were classified differently.

One of the often recurring results in categorical perception research is the difference between the perception of consonants, mainly stop consonants, and vowels. Stop consonants are said to be categorically perceived, whereas the perception of vowels is often called continuous. This difference in perception can be explained by the difference in coding between these sounds (Pisoni, 1975; Fujisaki & Kawashima, 1971). The essential acoustic cues for stop consonants are rapidly changing F1 and F2 transitions and a noise burst. In contrast, the major acoustic cues for vowels remain uniform over the entire length of the stimulus. The difference in acoustic cues between vowels and stop consonants has an effect on the availability of auditory memory for these two classes of speech sounds. According to Pisoni (1975), the outcome of a stimulus comparison during discrimination may represent two types of memory components: auditory memory and phonetic memory. This is in line with the dual-process model of Fujisaki & Kawashima (1971), who propose that discrimination can be performed in an auditory or phonetic mode. To explain the difference in perception between stop consonants and vowels they formulated the "cue-duration hypothesis". According to this hypothesis, the major factor responsible for the inferior auditory memory with consonants is the duration of the critical information in the signal. The acoustic cues of the encoded stop consonants (i.e. formant transitions) presumably cannot be stored well in memory. Trace decay of rapidly changing acoustic information is too fast to make an acoustic comparison possible, with the result that discrimination is based on a phonetic mode. This is not the case for the relatively unencoded isolated vowels. An isolated vowel consists of a characteristic steady-state, which is salient in auditory memory, and therefore can be retained long enough for the listener to make an auditory comparison (Pisoni, 1975).

In line with this explanation, Schouten & Van Hessen (1992) propose that the lack of categorical perception of vowels may be due to the nature of the stimulus material that has been used. Up to now, the vowels used as stimulus material have been modelled on productions in isolated words. When produced in isolated words, vowels will be lengthened. The duration of a plosive will be less affected by overarticulation than that of a vowel: the articulation of the burst can hardly be lengthened and the lengthening of the transitions is also restricted because this would lead to a change in phoneme identity. The major cue for vowels, on the other hand, is a relatively constant steady-state, which can be easily changed in duration without changing vowel identity. As a consequence of the long steady-state, vowels become relatively unencoded signals, whereas stop consonants remain coded signals. In running speech, the comparative spectral and temporal reduction of the signal will result in more complex coding. We expect that if coding is more complex, vowel perception will be more categorical (Schouten & Van Hessen, 1992). We hypothesise the spectral and temporal reduction of vowels to force the listeners to make a quick decision about the phoneme category, especially when stimuli are difficult to discriminate. This will strengthen the relationship between discrimination and classification of the same stimuli. To test this hypothesis, we studied the difference in perception between vowels spoken in isolated words and in a fast read text. The question of interest in the current research is whether more normal, underspecified vowels are perceived more categorically than overarticulated vowels.

2. METHOD

2.1. Materials

The stimulus material consisted of two continua of eight stimuli between the vowels [u] and [i]. The vowels were presented in a /p-V-p/ context. One continuum was generated with the original vowels produced in isolated words (Continuum overarticulated vowels), and the other continuum with the original vowels segmented from a text that was read aloud at a fast speech rate (Continuum underspecified vowels). The stimuli were obtained by spectral interpolation as described by Van Hessen (1988). With this method only the timbre of a spoken (not synthesised) stimulus is manipulated, in such a way that the timbre of the last stimulus of a set is equal of that of a second spoken stimulus. All other characteristics, such as pitch, duration, and voice quality, remain constant. The stimuli sounded completely natural.

2.2. Subjects

The subjects were 19 students of the Faculty of Arts at Utrecht University. They had no known hearing deficits and were all native speakers of Dutch. They were paid a fixed hourly rate.

2.3. Design

The experiment consisted of six tests, three tests for each of two vowel continua, involving the same subjects. The tests were fixed discrimination, roving discrimination, and classification. The subjects took the tests in a fixed order: the classification experiment was always performed after the discrimination tests. Classification involved a forced choice between two alternatives, the vowels [u] and [i]. There was no response-time limit. In the classification test, each stimulus had to be identified 64 times in a random order.

The discrimination task. The prediction test is the most widely used formal criterion of categorical perception (the Haskins model). It requires a close correspondence between the actual discrimination of speech stimuli along a continuum and discrimination performance predicted from the classification results. To avoid labelling strategies, we opted for a task that reduces the load on auditory memory and encourages a direct auditory comparison between the sounds in a trial. Such a task is 4-interval oddity. In this task the A and B stimuli are presented randomly in the two orders AABA or ABAA, with a 50% a priori probability. Stimulus A at the beginning and end of each quadruplet functions as a reference. This way discrimination will be guided by auditory processing of the stimuli and not by phoneme labelling, and will be a diagnostic instrument for determining whether categorical perception occurs.

The subjects' task was to indicate whether stimulus B (the odd ball) was in the second or the third interval. The stimuli in the second and third intervals were always apart by one step along the continuum; the number of comparisons was therefore seven. The intertrial interval was determined by the response time. The interstimulus interval within a trial was 200 ms. In the fixed discrimination experiment all possible combinations of only one stimulus pair (A and B) were presented over and over again during a block of trials. The fixed test consisted of 7 blocks, one for each comparison, which were clearly separated from each other. Each block contained 64 trials, 32 for each of the two possible combinations, AABA and ABAA. In the roving discrimination experiment, the A and B stimuli to be discriminated were drawn randomly from the total range of stimuli and thus varied from trial to trial. In the roving discrimination test, 7x64 trials were presented.

2.4. Procedure

The stimuli were presented to the subjects over headphones in a sound-treated booth. In the discrimination tests, it was stressed that differences between the stimuli would be small, and in most cases could only be detected by listening carefully to all details of a stimulus. The subjects responded by mouse-clicking on one of two response fields (labelled "2" and "3") on a computer screen. After answering, visual feedback of the correct answer was given so that the subject was able to judge and possibly improve his own performance. Discrimination training consisted of 128 trials, and was intended to familiarise subjects with their task. In the fixed discrimination context the first ten trials of every block were considered practice and were not included in the data analysis. In classification, one stimulus was played on each trial, and the subject had to identify it by mouse-clicking on a response field labelled "oe" or "ie" ([u] and [i]). The only training consisted of 16 trials. No feedback about correct responses was given.

3. RESULTS

The discrimination and classification results are displayed in figures 1 and 2. The data in the figures represent the averages of the 19 subjects' individual d' scores. The numbers (n) along the abscissa refer to stimulus pairs, consisting of stimuli (n) and (n+1); n is therefore a number between 1 and 7. Stimuli in pair 1 resemble [u] and stimuli in pair 7 sound like [i]. In order to compare classification and discrimination, the classification scores were transformed into d' values.

We expected the data in figure 1 to be less categorical than in figure 2. The results show that this is clearly not the case: the overarticulated vowels are not perceived less categorically than the underspecified vowels. Neither figure shows any relationship between observed and predicted discrimination, so we can conclude that there is no indication of categorical perception for either of the two vowel conditions.

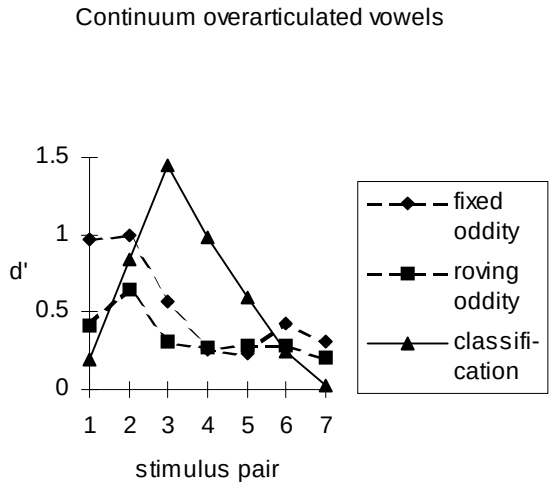

Figure 1. Discrimination and classification results for the overarticulated vowels.

The ordering of results in figures 1 and 2 is not what one would expect. The expected ordering is: fixed discrimination best, followed by roving discrimination, and then classification (Macmillan & Creelman, 1991; Van Hessen & Schouten, 1992). Classification is best here, followed by discrimination. This is rather awkward because it indicates that listeners did hear differences between stimuli while classifying them, but could not hear differences between the same stimuli during discrimination. Apparently, listeners used a phoneme labelling strategy during classification, but could not assign labels to the stimuli during discrimination. Unexpectedly, without labelling listeners were incapable of discriminating the stimuli.

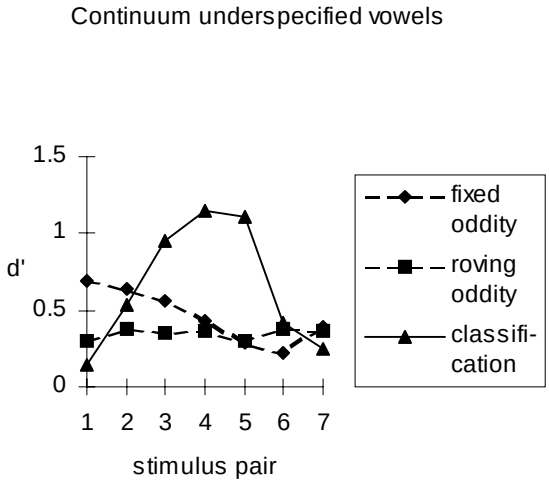

Figure 2. Discrimination and classification results for the underspecified vowels.

Apparently, the acoustic differences between the stimuli were too small for the listeners to profit from psycho-acoustic processing.

The results of a multivariate analysis of variance support these observations. Fixed independent variables were Paradigm (3 levels), Vowel Condition (2 levels), and Stimuli (7 levels: nested under Vowel Condition). Cell variance was over 19 subjects. Any effect reported as significant here had a p-value below .01. The results of the analysis of variance revealed that discrimination and classification performance is not significantly affected by Vowel Condition [F = 0.12]. The differences between the tasks are significant [F(1,2) = 14.62]. There is a significant effect of Stimulus within Vowel Condition and a significant interaction between Paradigm and Stimulus within Vowel Condition [$F_1(1,12) = 5.58$; $F_2(1,24) = 4.56$]. A Newman-Keuls test on the factor Paradigm revealed a significant difference between the means of roving oddity and classification [F(1,5) = 5.20]. A Newman-Keuls test on the factor Stimulus (Continuum overarticulated vowels) showed a significant peak in fixed oddity at stimulus pair 1 and 2, and a significant peak in classification at pair 3 [$F_1(1,6) = 3.81$; $F_2(1,6) = 14.39$]. A similar test on the data from the Continuum underspecified vowels revealed a significant peak in the classification curve at stimulus pairs 3-4-5 [F(1,6) = 4.53].

Because a large proportion of the total variance was explained by cell variance (78,5%) it was decided to take a closer look at the behaviour of the subjects. On the basis of the roving oddity scores the subjects were divided into quartiles. Figures 3 and 4 show the discrimination and classification results of the overarticulated vowels (the results are the same for the two vowel conditions) from the subjects in the upper (four subjects) and lowest quartiles (five subjects). There is a marked difference between the two subject groups. Only the subjects in the upper quartile are able to discriminate stimuli on a level equal to, or higher than, classification performance.

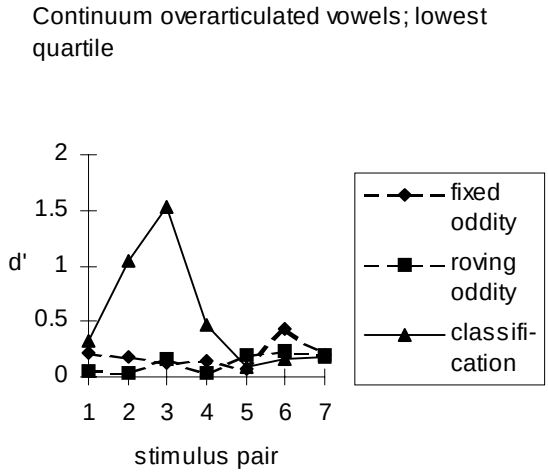

Figure 3. Discrimination and classification results for the vowels produced in isolated words; lowest quartile.

Discrimination performance of the subjects in the remaining three quartiles gradually decreases towards performance at chance level for the lowest quartile. The subjects in the lowest quartile do not hear any difference between the stimuli. Their discrimination performance is at chance level. Overall labelling performance was similar for both quartiles. These results imply that mainly psycho-acoustic processes played a role during discrimination.

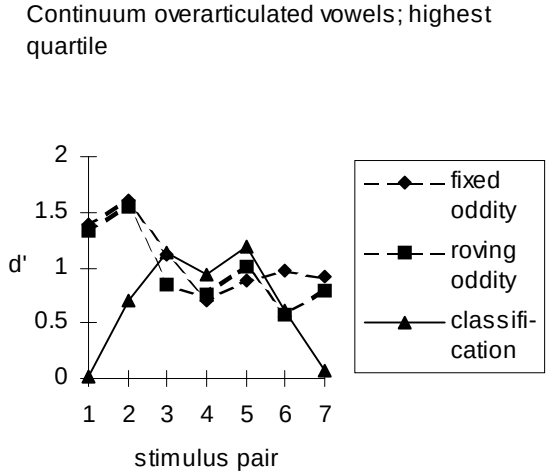

Figure 4. Discrimination and classification results for the vowels produced in isolated words; highest quartile

4. DISCUSSION AND CONCLUSION

In a series of experiments we compared the perception of isolated vowels with more normal, underspecified vowels. We expected that the temporal and spectral reduction of the latter would result in more complex coding and hence more categorical perception. However, our results do not indicate that

normal, underspecified vowels are perceived more categorically than overarticulated vowels, mainly because there is no relationship between observed and predicted discrimination, and thus no indication of categorical perception for either of the two vowel conditions. It could be that the acoustic differences between the two vowel continua were too small to induce differences in perception, in the sense that the vowels from the fast read text were possibly not sufficiently underspecified. However, this is in contrast with the results of numerous studies that have demonstrated that speaking rate affects the acoustic information specifying vowels and that these rate-dependent modifications of the acoustic structure of vowels have perceptual consequences.

Surprisingly, we found observed discrimination to be worse than predicted by the labelling data. This is in contrast to the traditional finding that the discrimination function is usually higher than predicted by the labelling data (Schouten & Van Hessen, 1992). A first attempt to account for the traditional difference was made in the dual-process model for the discrimination of speech stimuli by Fujisaki & Kawashima (1971). This model explicitly distinguishes between categorical phonemic judgements and judgements based on auditory memory for acoustic stimulus attributes. The authors propose that two perceptual modes are active simultaneously (or in rapid sequence). One of them is strictly categorical and represents phonetic classification and the associated verbal short-term memory. The other mode is completely continuous and represents processes common to all auditory perception. The results of any particular speech discrimination experiment are assumed to reflect a mixture of both components. The part of performance that can be predicted from labelling probabilities is attributed to categorical judgements, whereas the remainder (the deviation from ideal categorical perception) is assigned to memory for acoustic stimulus properties.

In our choice of the discrimination paradigm we were led by the idea that a psycho-acoustic discrimination task was needed to assess categorical perception in an unbiased way. If a discrimination task is used that prevents a direct comparison between successive stimuli, listeners will be encouraged to use a phonetic labelling strategy for discrimination and results are more likely to favour categorical perception. We were even more successful than we expected in preventing such a strategy. We assumed that by using the 4-interval oddity task we would be able to control the weighting of the two processes in the sense that psycho-acoustic processing would be encouraged and thus phonetic labelling would be discouraged. We still expected phonetic labelling to play a dominant role during discrimination, because Fujisaki & Kawashima (1971) claim that labelling of speech stimuli is inevitable even in discrimination experiments in which classification is not part of the listeners' task. However, our results indicate that in our discrimination task listeners were unable to focus on the stimuli as phonetic percepts, and were consequently on average incapable of discriminating them.

With the 4-interval oddity task we succeeded in testing unbiased discrimination. Listeners could not use a labelling strategy and discriminated the stimuli on the basis of purely psycho-acoustic information. If we follow the strict definition of Liberman et al (1957) we have to conclude that the results show no categorical perception: there is no relation between observed discrimination and discrimination predicted by classification. However, the results show that perception is only categorical if listeners are in the phonetic mode, as is the case when classifying speech stimuli and hence, presumably, in normal, spontaneous speech processing. Whenever listeners discriminate between two stimuli in the psycho-acoustic mode, results will be unrelated to phoneme categories represented in long-term memory. The conclusion must be that the results of the present study show "true" categorical perception: most listeners are unable to hear differences between speech stimuli, if they are in a psycho-acoustic mode. In the phonetic mode, however, speech stimuli are perceived categorically.

5. REFERENCES

1. Fujisaki, H. and Kawashima, T. "A model of the mechanisms for speech perception: Quantitative analyses of categorical effects in discrimination", *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 59-68, 1971*.

2. Liberman, A.M., Harris, K., Hoffman, H.S. and Griffith, B. "The discrimination of speech sounds within and across phoneme boundaries", *Journal of Experimental Psychology, 54, 358-368, 1957*.

3. Macmillan, N.A. and Creelman, C.D., *Detection Theory: A User's Guide.* Cambridge University Press, NY, 1991.

4. Massaro, D.W. "Perceptual units in speech recognition", *Journal of Experimental Psychology, 102, 199-208, 1974*.

5. Pisoni, D.B. "Auditory short-term memory and vowel perception", *Memory & Cognition, 3, 7-18, 1975.*

6. Schouten, M.E.H. and Van Hessen, A.J. "Modelling Phoneme Perception, I: Categorical Perception", *Journal of the Acoustical Society of America, 92, 1841-1855, 1992*.

7. Van Hessen, A.J. and Schouten, M.E.H. "Modelling phoneme perception. II: A model of stop consonant discrimination", *Journal of the Acoustical Society of America, 92, 1856-1868, 1992*.

8. Van Hessen, A.J. "A new speech modulation model based on sinusoidal representation", *Progress Report Institute of Phonetics Utrecht, 13(2), 16-28, 1988.*

Correspondence address: Ellen Gerrits, Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands. E-mail: Ellen.Gerrits@let.uu.nl