Perception of the [m]–[n] distinction in CV syllables

Bruno H. Repp
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

(Received 5 August 1985; accepted for publication 18 February 1986)

The contribution of the nasal murmur and the vocalic formant transitions to perception of the [m]–[n] distinction in utterance-initial position preceding [i,u] was investigated, extending the recent work of Kurowski and Blumstein. A variety of waveform-editing procedures were applied to syllables produced by six different talkers. Listeners' judgments of the edited stimuli confirmed that the nasal murmur makes a significant contribution to place of articulation perception. Murmur and transition information appeared to be integrated at a genuinely perceptual, not an abstract cognitive, level. This was particularly evident in [−l] context, where only the simultaneous presence of murmur and transition components permitted accurate place of articulation identification. The perceptual information seemed to be purely relational in this case. It also seemed to be context specific, since the spectral change from the murmur to the vowel onset did not follow an invariant pattern across front and back vowels.

PACS numbers: 43.71.Es

INTRODUCTION

In a recent study on the perceptual integration of nasal murmur and vocalic formant transition cues to place of articulation of nasal consonants, Kurowski and Blumstein (1984), henceforth K&B, showed that not only did both cues contribute to the perception of the [m]–[n] distinction, but also that their contributions were nearly equal. Their materials were 50 CV syllables uttered by a male speaker of American English, five tokens each of [m,n] followed by [i,e,a,o,u]. Portions of these syllables were presented to listeners as follows: (1) the full murmur (up to the point of consonantal release); (2) the full vowel (i.e., the stimulus portion following the release, which included initial formant transitions); (3) the last six pitch pulses of the murmur; (4) the first six pitch pulses of the vowel; and (5) the last three pulses of the murmur followed by the first three pulses of the vowel (i.e., the six pulses surrounding the release). The principal findings were that (a) the full murmur and the full vowel were about equally informative when presented separately (about 80% correct place of articulation identification); (b) shortening of these stimulus portions to only six pitch pulses led to a nonsignificant decrease in identification scores (about 77% correct); and (c) scores were highest for stimuli that included both the end of the murmur and the beginning of the vowel (89% correct).

Although it was known from earlier studies that the vocalic formant transitions are strong cues to the place of articulation in nasal consonants (e.g., Larkey et al., 1978; Liberman et al., 1954) and also that nasal murmurs in isolation can be identified at levels better than chance (Malecog, 1956; Nakata, 1959), K&B were the first to systematically compare identification of the two stimulus components in isolation and in combination. Their study contrasts with previous work by Malecog (1956), Nord (1976), and Recasens (1983), who used various combinations of conflicting murmurs and transitions to assess their relative contributions. With such stimuli, the transitions almost always emerge as the dominant place of articulation cue. K&B point out that this result could be due to artificial spectral discontinuities occurring at the splicing point, although the mechanism that would lead to perceptual dominance of the transitions over the murmur in such a situation has not been defined. (See Tartter et al., 1983, for a similar argument concerning the perception of stop consonant place of articulation in VCV stimuli.) In any case, K&B avoided this possible problem by combining only murmurs and transitions deriving from the same utterance. This, however, resulted in an ambiguity of their results which they acknowledge: The murmur and the transitions could act as independent cues that are combined at some higher level of processing (cf. Massaro and Oden, 1980; Repp, 1982), or the murmur and the transitions might be integrated at an early perceptual level and thus might constitute a single effective cue. This second possibility was favored by K&B on grounds of parsimony and because it is more compatible with the search for invariant properties that Blumstein and her associates are engaged in (e.g., Blumstein and Stevens, 1979, 1980; Lahiri et al., 1984). These two hypotheses may be called the multiple-cue (or late integration) and single-cue (or early integration) hypotheses, respectively.

The present experiment addressed several issues relevant to these hypotheses, as applied to nasal consonant perception, thereby extending the work of K&B. Although the study was mainly an attempt to replicate the results of K&B using a larger variety of test utterances and conditions, some of the conditions were novel and explored the nature of the perceptual integration process and the role of dynamic stimulus information.

Although K&B's study was carefully conducted, and incorporated five different vowel contexts, it had two methodological limitations. One is the use of a single talker: The surprisingly high identification scores for isolated murmurs could have reflected a peculiarity of his articulation. The other feature is that the subjects were permitted to respond
A. Talkers and recording procedure

Six talkers participated, three males (AA, TG, SS) and three females (CG, SM, MB). All were native speakers of American English. AA is an experienced graduate student and native speaker of English. The talkers were asked to produce the syllables /ma ma mi mi mI/ in a neutral manner. The recording session was tape-recorded using a Jensen microphone, placed approximately 10 cm from the talker's mouth, and high-quality tape recorder.

B. Stimuli and test sequences

One good token of each syllable was selected from each talker's productions. The basic stimulus set consisted of 36 syllables (6 tokens x 6 utterances) at 10 kHz sampling rate, filtered in 10 kHz bands at 4 kHz. Using a waveform editing program, each file was divided into 16 segments, each file with an initial pitch period. The stimuli were placed at the beginning of the first pitch period, labeled "NR" at the onset of the burst of the first pitch period. This point was defined as the onset of the burst of the first pitch period.

The waveform of each syllable was then analyzed using a waveform editing program. The waveform was divided into 16 segments, each file with an initial pitch period. The stimuli were placed at the beginning of the first pitch period, labeled "NR" at the onset of the burst of the first pitch period. This point was defined as the onset of the burst of the first pitch period.
E. Results and Discussion

The main objective of this study was to investigate the performance of the proposed traffic prediction system under various traffic conditions. The results showed that the system was able to accurately predict traffic flow, even under challenging conditions.

The study was conducted on a network of 10 intersections, with traffic data collected over a period of one week. The data was divided into training and testing sets, with the training set used to train the model and the testing set used to evaluate its performance.

The performance of the system was evaluated using a set of metrics, including accuracy, precision, recall, and F1-score. The results showed that the system achieved high accuracy, with an average accuracy of 95%. The precision and recall values were also high, with values of 97% and 98%, respectively.

The system was able to predict traffic flow with a delay of up to 10 minutes, and the predictions were found to be reliable and accurate. The system was also able to adapt to changes in traffic conditions, such as sudden increases in traffic volume or changes in traffic patterns.

The results of the study indicate that the proposed traffic prediction system has the potential to improve traffic management and reduce congestion in urban areas. The system can be used to optimize traffic signals, plan road closures, and allocate resources more efficiently.

F. Conclusion

In conclusion, the results of this study demonstrate the potential of the proposed traffic prediction system to improve traffic flow and reduce congestion in urban areas. The system is capable of accurately predicting traffic flow, even under challenging conditions, and can be used to optimize traffic signals, plan road closures, and allocate resources more efficiently.

Further research is needed to evaluate the long-term impact of the system on traffic flow and to explore its potential for use in other areas, such as urban planning and emergency response.
The paper discusses the effects of various factors on the performance of speech recognition systems. It notes that the use of a higher percentage of syllables in the output language can lead to a decrease in the overall performance of the system. The paper also mentions the importance of considering the impact of different factors on the performance of speech recognition systems, and recommends that further research be conducted in this area.
C. Excerpts

The data for uncorrected (UC) and percent correct (PC) are shown in Figure 4. The percent correct data were obtained by dividing the total number of correct responses by the total number of responses for each subject. The UC data were then calculated by multiplying the PC data by the number of trials. The results show that the percent correct data are highly correlated with the UC data, with a correlation coefficient of 0.98. This suggests that the percent correct data provide a good estimate of the UC data.

Figure 4. Percent correct and UC scores for each subject. The percent correct scores were obtained by dividing the total number of correct responses by the total number of responses for each subject. The UC scores were then calculated by multiplying the percent correct scores by the number of trials. The results show that the percent correct scores are highly correlated with the UC scores, with a correlation coefficient of 0.98. This suggests that the percent correct scores provide a good estimate of the UC scores.

Perceptual learning proceeds through a process of learning new information, which is then stored in memory. This process is facilitated by the use of appropriate teaching methods, such as structured practice, feedback, and reinforcement. The results of this study suggest that perceptual learning can be effectively taught using a combination of these methods.
en...
null
Table II. Percent correct scores for the stimulus condition. Conditions: M = murmur stimuli (3 m, 3y); Y = vowel stimuli (6 m, 11).

<table>
<thead>
<tr>
<th>Condition</th>
<th>[m]</th>
<th>[n]</th>
<th>[m]</th>
<th>[n]</th>
<th>[m]</th>
<th>[n]</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M+V</td>
<td>62</td>
<td>69</td>
<td>70</td>
<td>69</td>
<td>85</td>
<td>86</td>
<td>84</td>
</tr>
<tr>
<td>3M-Y</td>
<td>59</td>
<td>65</td>
<td>52</td>
<td>67</td>
<td>82</td>
<td>88</td>
<td>84</td>
</tr>
<tr>
<td>6M+V</td>
<td>57</td>
<td>67</td>
<td>63</td>
<td>67</td>
<td>77</td>
<td>84</td>
<td>78</td>
</tr>
<tr>
<td>6M-Y</td>
<td>51</td>
<td>65</td>
<td>55</td>
<td>80</td>
<td>84</td>
<td>87</td>
<td>75</td>
</tr>
</tbody>
</table>

The results are presented in Table II. Looking first at the 3M+V results, we see that the average score for these stimuli was higher than for the corresponding [m] stimuli. However, the abrupt decline in the [m] scores for the vowels (3M+V) indicates a specific effect of the vowel that was not present in the [m] stimuli. This result indicates the presence of a vowel effect that is not a result of the specific vowel sounds used in this experiment.

In the 3M-Y condition, the average score was lower than for the 3M+V condition, indicating a vocalic effect. However, the abrupt decline observed in the [m] scores for the vowels (3M-Y) indicates a specific effect of the vowel that was not present in the [m] stimuli. This result suggests the presence of a vowel effect that is not a result of the specific vowel sounds used in this experiment.

The 6M+V condition showed a lower average score compared to the 3M+V condition, indicating a general effect of the vowel. The abrupt decline observed in the [m] scores for the vowels (6M+V) indicates a specific effect of the vowel that was not present in the [m] stimuli. This result suggests the presence of a vowel effect that is not a result of the specific vowel sounds used in this experiment.

In the 6M-Y condition, the average score was lower than for the 3M-Y condition, indicating a general effect of the vowel. The abrupt decline observed in the [m] scores for the vowels (6M-Y) indicates a specific effect of the vowel that was not present in the [m] stimuli. This result suggests the presence of a vowel effect that is not a result of the specific vowel sounds used in this experiment.

In summary, the results indicate that the presence of a vowel effect was present in all conditions, but was more pronounced in the 3M+V and 6M+V conditions. The vowel effect was not present in the [m] stimuli, indicating that the effect was specific to the vowel sounds used in this experiment.
ACKNOWLEDGMENTS

Project supported by NCHRD Grant HD-0019.

Postdoc writer.

Received 3 May 1094. Accepted 1 June 1094.

The manuscript was submitted on 6 June 1996.

The research was supported by NCHRD Grant HD-0019.

Abstract: Here we describe a novel approach to the problem of object recognition in computer vision. We propose a method that integrates edge detection, feature extraction, and classification in a unified framework. The method is based on the assumption that objects in an image can be represented as a set of edge points, and that these points can be classified into different categories based on their spatial and intensity relationships.

The proposed approach combines the advantages of level set methods and neural networks. It uses a level set formulation to model the evolving contour of an object, while the neural network is used to classify the edge points into different categories. The model is trained using a supervised learning approach, where a set of labeled images is used to train the neural network.

The results of our experiments show that the proposed method is effective in recognizing a wide variety of objects, including both simple and complex shapes. The method is also robust to noise and occlusions, which are common in real-world images.

We believe that this work represents a significant step towards the development of a unified framework for object recognition.

Our research was funded by the National Science Foundation under grant No. 123456.

REFERENCES


ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their valuable comments and suggestions. This work was supported by a grant from the National Science Foundation.
K&B used the term long transitions for the initial 60 ms or so is illustrated by A&B's Figure 1, which reports mean second frequency onsets of stop consonants in different vowel environments, but as the center of the vowel and the center of the stop consonant are passed, the transition to the vowel begins, and the degree of overlap between the two is increased. This is also illustrated by the results of the study by K&B, which showed that the degree of overlap between the consonant and vowel was not significantly different, and that the transition to the vowel occurs at the same time as the transition to the consonant. The study did not include comparisons with other studies that have shown different results, such as the study by A&B, which found a significant difference between the two transitions. However, the results of the study by K&B were consistent with the results of previous studies that have shown a similar pattern of results. In conclusion, the results of the study by K&B suggest that the transition to the vowel occurs at the same time as the transition to the consonant, and that the degree of overlap between the two is not significantly different.