Perceptual Processing Time for Consonants and Vowels

David B. Pisoni
Haskins Laboratories, New Haven

ABSTRACT

Perceptual processing time for brief CV syllables and steady-state vowels was examined in a backward recognition masking paradigm. Subjects were required to identify a 40 msec sound selected from either a consonant set (/ba/, /da/, /ga/) or a vowel set (/i/, /I/, /e/). The target sound was followed by a different sound drawn from the same set after a variable silent interstimulus interval. The second sound interrupted the perceptual processing of the target sound at short interstimulus intervals. Recognition performance improved with increases in the silent interstimulus interval. One experiment examined processing time for consonants and vowels under binaural presentation. Two additional experiments compared consonant and vowel recognition under both binaural and dichotic presentation. The results indicated that: (1) consonants require more processing time for recognition than vowels and (2) binaural and dichotic presentation conditions produce differential effects on consonant and vowel recognition. These findings have several important implications for understanding the recognition process. First, speech perception is not immediate, but is the result of several distinct operations which are distributed over time. Second, speech perception involves various memorial processes and mechanisms which recode and store information at different stages of perceptual analysis.

One of the most basic questions in speech perception concerns the process of recognition. How is a particular speech sound identified as corresponding to a specific phonetic segment? Although many of the current theories of speech perception have focused on the recognition process for some time, they have all been quite vague in their approach to this problem (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Stevens and House, 1972). It is

---

*Paper presented at the 84th meeting of the Acoustical Society of America, Miami Beach, Fla., 1 December 1972. The research was supported by a grant from NICHD to Haskins Laboratories and a PHS grant (S05 RR 7031) to Indiana University.

†Also Indiana University, Bloomington.

[HASIKINS LABORATORIES: Status Report on Speech Research SR-31/32 (1972)]
usually assumed that the recognition process entails a series of stages and operations in which the acoustic stimulus makes contact with either a stored representation in long-term memory or some representation that may be constructed or generated from rules residing in long-term memory (Liberman et al., 1967; Halle and Stevens, 1962). Unfortunately, little empirical work has been directed at examining these hypothetical stages or specifying what types of operations might be involved in the recognition process for speech sounds.

The present study is concerned with mapping out in a quantitative way the earliest stages of perceptual processing for speech sounds. To achieve precise control over early processing, a backward recognition masking procedure was used. With this technique the processing of one stimulus may be interrupted at various times after its presentation by another stimulus and thereby provide information about the temporal course of perceptual processing (Massaro, 1971, 1972).

Figure 1 shows the general features of the backward recognition masking paradigm used in the present series of experiments. On each trial the listener is presented with two successive stimuli but is required to identify only the first stimulus or target sound. The second sound in the sequence serves as the masking stimulus and is presented after some variable silent interstimulus interval. When the mask follows the target at very short intervals it may interrupt or interfere with the processing of the target sound. By varying the duration of the silent interval between the target and mask it is possible to determine the amount of processing time needed for recognition of the target sound. The perceptual processing time for the recognition of brief consonant-vowel (CV) syllables and steady-state vowels was examined in this study because consonants and vowels not only differ in their acoustic properties but have also been shown to have basically different perceptual characteristics (Liberman et al., 1967; Studdert-Kennedy and Shankweiler, 1970).

METHOD

The stimulus conditions employed in these experiments are shown in Figure 2. The consonant stimuli were the CV syllables /ba/, /da/, and /ga/. They were 40 msec in duration and had formant transitions lasting 20 msec. The steady-state vowels used were /ɪ/, /ɪ/, and /ɛ/ and they were also 40 msec in duration. The target sound was selected from either the consonant set or the vowel set. A given target stimulus was then followed by a different stimulus drawn from the same set after a variable silent interstimulus interval. The three sounds within each stimulus condition were arranged in all possible permutations to produce the six stimulus pairs shown in Figure 2. Each pair represented a target and masking sound combination. The intensity relations between target and mask were also manipulated but they will not be discussed here since the effects are not relevant to the major conclusions.

The details of the experimental design are shown in Figure 3. In each of the experiments the intensity and interstimulus interval variables were identical and completely random across trials. Experiment 1 compared consonant and vowel recognition under binaural presentation conditions with the same group of listeners. Experiments 2 and 3 compared binaural and dichotic masking conditions for consonants and vowels with separate groups of listeners. In the
Figure 1: Description of the backward recognition masking paradigm used in the present experiments.
STIMULUS CONDITIONS

I. CONSONANTS VS VOWELS

<table>
<thead>
<tr>
<th>CONSONANT - VOWEL SYLLABLES (40 MSEC)</th>
<th>STEADY - STATE VOWELS (40 MSEC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA - DA - GA</td>
<td>i - I - ε</td>
</tr>
</tbody>
</table>

STIMULUS PAIRS:
1) BA - DA — DA - BA
2) BA - GA — GA - BA
3) DA - GA — GA - DA

STIMULUS PAIRS:
1) i - I — I - i
2) i - ε — ε - i
3) i - ε — ε - I

II. INTENSITY RELATIONS: \( \Delta I = I_{\text{TARGET}} - I_{\text{MASK}} \)

1. \( \Delta I = 0 \text{ dB} \)

2. \( \Delta I = 4.5 \text{ dB} \)

3. \( \Delta I = 9.0 \text{ dB} \)

Figure 2: Description of the stimulus conditions used in the present experiments.
Figure 3: Experimental design of the three experiments.

- **Experiment 1 (N=6)**
  - I. Consonants vs. Vowels
  - II. Intensity
  - III. Interstimulus Interval

- **Experiment 2 (N=9)**
  - I. Consonants
  - II. Intensity
  - III. Interstimulus Interval

- **Experiment 3 (N=15)**
  - I. Binaural vs. Dichotic
  - II. Intensity
  - III. Interstimulus Interval
Figure 4: Average recognition scores for consonants and vowels in Experiment 1 as a function of interstimulus interval.
Figure 5: Average recognition scores for consonants under binaural and dichotic presentation conditions in Experiment 2 as a function of interstimulus interval.
Figure 6: Average recognition scores for vowels under binaural and dichotic presentation conditions in Experiment 3 as a function of interstimulus interval.
binaural conditions target and mask were presented to both ears. In the dichotic condition the target sound was presented to one ear and the mask was presented to the other. Targets and masks were presented equally often to both ears in the dichotic condition.

RESULTS AND DISCUSSION

Figure 4 shows the results of Experiment 1 which compared consonant and vowel stimuli under binaural presentation. Recognition performance is expressed in terms of the percent correct identification of the target sound. Recognition improves with increases in the silent interstimulus interval for both vowels and consonants. However, the masking appears to be much more effective for the consonants than the vowels. This is especially noticeable at short interstimulus intervals and indicates that consonants need more time for recognition than vowels.

The results of the second experiment which examined processing time for the consonant stimuli under binaural and dichotic presentation conditions are shown in Figure 5. The binaural data are almost identical to the findings obtained in the first experiment. However, there is a large and consistent difference between the dichotic and binaural presentation conditions. Performance under dichotic presentation is lower overall than performance under binaural presentation. Moreover, performance under dichotic presentation at short intervals appears to be markedly different than performance under binaural presentation.

The results of the third experiment which studied recognition of vowels under binaural and dichotic presentation conditions are shown in Figure 6. The effect observed in Figure 5 for the consonants is strikingly absent. There is again some masking at short intervals for the vowels but there is no difference between binaural and dichotic presentation conditions.

In summary, when the target stimulus was followed by the masking sound at short interstimulus intervals recognition of the target was interrupted, suggesting that perceptual processing for speech sounds continues even after the stimulus has ended. This finding indicates that speech perception is not a result of immediate stimulation but rather requires a certain amount of processing time for the extraction of relevant features from the acoustic signal.

The present results also reveal that consonants require more processing time for recognition than vowels. Additionally, when binaural and dichotic masking conditions were compared for these classes of speech sounds differences in recognition were obtained for consonants but not for vowels. This last result suggests that there may be an additional stage or stages of perceptual processing needed for consonant recognition that is not needed for vowel recognition.

REFERENCES