

Perception & Psychophysics
1974, Vol. 16, No. 1, 84-86

Discrimination of intensity differences on ^{formant}~~formant~~
transitions in and out of syllable context

M. F. DORMAN*

Haskins Laboratories, New Haven, Connecticut 06510

Discrimination of intensity differences on ^{formant} transitions in and out of syllable context

M. F. DORMAN*

Haskins Laboratories, New Haven, Connecticut 06510

The Ss were presented pairs of stimuli, /bæ/s, /æ/s, or isolated transitions from /bæ/s, which differed in the initial 60 msec of the signals by 0, 7.5, or 9 dB. In the syllable context, the intensity differences were discriminated essentially at chance; in both the vowel and isolated transition conditions, the intensity differences were discriminated essentially perfectly. This outcome suggests that after the acoustic features of a stop-consonant/vowel syllable have been recorded into a phonetic representation, the acoustic information is relatively inaccessible for recall from auditory short-term memory.

According to traditional psychoacoustics, listeners can discriminate among many more acoustic stimuli along nonspeech continua than they can identify absolutely. For example, listeners can discriminate about 1,200 different pitches, yet can identify absolutely only about 7 ± 2 (Miller, 1956; Pollack, 1952).

This type of relationship between identification and discrimination does not hold, however for the stop consonants /ba, da, ga, pa, ta, ka/. In discrimination tests, when listeners are presented with synthetic speech stimuli which differ in extent and direction along the acoustic continuum of the second formant transition, discrimination is essentially perfect between stimuli drawn from different phonetic categories (e.g., /ba, da/), but is near chance for physically different stimuli drawn from the same phonetic category. Thus, the discrimination of differences in extent and direction of the second formant transition (also for the cutback of the first formant, a cue for manner of articulation) is little better than an absolute phonetic categorization (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1971). Liberman and his colleagues have termed this type of relationship between identification and discrimination (which is apparently unique to the perception of speech) "categorical perception" (Mattingly, Liberman, Syrdal, & Halwes, 1971; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

The categorical perception of stop consonants has been interpreted as demonstrating that after the recoding of the acoustic signal into a phonetic representation, auditory information cannot be readily retrieved from short-term auditory memory (Pisoni, 1971). From the studies cited above, however, it is not clear whether all of the acoustic information on the formant transitions suffers a fate in short-term auditory memory similar to the acoustic information which directly cues different phonetic categories. For example, is intensity of the formant transitions (information which does not directly signal phonetic categories) as

poorly discriminated within a phonetic category as changes in the extent and direction of the second formant transition or first formant cutback? The purpose of the present experiment was to determine the discriminability of differences in intensity of formant transitions on a computer-generated, stop-consonant/vowel syllable /bæ/.

METHOD

Subjects

Ten adults, who had previously participated in research with synthetic speech, served as Ss.

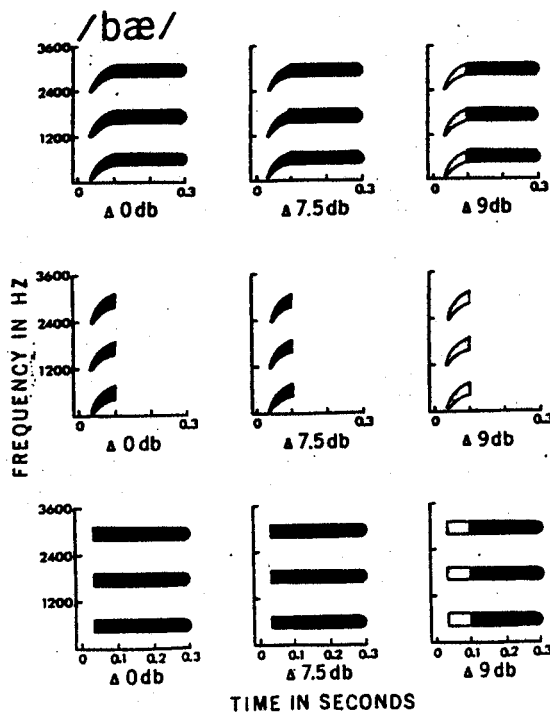


Fig. 1. Schematic spectrographic patterns for /bæ/, the isolated formant transitions from /bæ/ and /æ/. For each stimulus type, the initial 60 msec was either the same intensity as the remaining 240 msec or was 7.5 or 9.0 dB less intense.

*Also Herbert H. Lehman College of the City University of New York. Reprint requests should be addressed to M. F. Dorman, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

/bæ/

20msec $\Delta 0$ db $\Delta 7.5$ db $\Delta 9$ db $\Delta 0$ db $\Delta 7.5$ db $\Delta 9$ db

/æ/

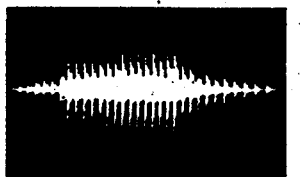
 $\Delta 0$ db $\Delta 7.5$ db $\Delta 9$ db

Fig. 2. Oscillographs of the /bæ/, isolated formant transitions, and /æ/ stimuli.

Preparation of Stimuli

The stimuli were generated on the Haskins Laboratories parallel-resonance synthesizer.

One set of stimuli was created by first synthesizing the syllable /bæ/ (see Fig. 1). For the first stimulus in this set, the overall intensity of the first 60 msec (the duration of the formant transitions) was the same as the remaining 240 msec of the syllable. This will be referred to as a $\Delta 0$ -dB stimulus. For the second and third stimuli in this set, the first 60 msec was 7.5 and 9 dB, respectively, less intense than the initial portion of the $\Delta 0$ -dB stimulus. The following 240 msec of both stimuli was the same intensity as the $\Delta 0$ -dB stimulus.

A second set of stimuli was created by simply eliminating the vowel steady-state formants from the three /bæ/ syllables, leaving the formant transitions isolated (see Fig. 1). The resulting three stimuli no longer sounded like speech, but, rather, like nonspeech "chirps" or pitch glides.

A third set of stimuli was created by synthesizing a 300-msec steady-state vowel /æ/ (see Fig. 1). For the first stimulus in this set, the intensity of the initial 60 msec was as intense as the remaining 240 msec. For the second and third stimuli in this set, the initial 60 msec was 7.5 and 9 dB, respectively, less intense than the initial portion of the $\Delta 0$ -dB stimulus. The following 240 msec of both stimuli was the same intensity as the $\Delta 0$ -dB stimulus. Special care was exercised to equate the amplitude envelopes of the /bæ/ and /æ/ stimuli. Since the amplitude envelope of the $\Delta 0$ -dB /bæ/ reached maximum amplitude approximately 30 msec after stimulus onset, the /æ/ stimuli were synthesized with a 30-msec linear rise time. Figure 2

displays oscillographs of all the stimuli.

Three test sequences were recorded on audio tape. One sequence contained the consonant-vowel stimuli; a second, the isolated formant transition stimuli; and the third, the steady-state vowel stimuli. The stimuli in all of the sequences were arranged in pairs. The first member of each pair was a $\Delta 0$ -dB stimulus, the second member a $\Delta 0$ -, $\Delta 7.5$ -, or $\Delta 9$ -dB stimulus. Each test sequence contained a randomized sequence of 18 pairs of $\Delta 0$ - $\Delta 0$ -dB stimuli, 18 pairs of $\Delta 0$ - $\Delta 7.5$ -dB stimuli, and 18 pairs of $\Delta 0$ - $\Delta 9$ -dB stimuli. The interstimulus interval within stimulus pairs was 500 msec. The interval between successive pairs was 4 sec.

Procedure

Each S was tested individually in a sound-attenuated room. The stimuli were presented via TDH-39 headphones at 72 dB SPL (for the $\Delta 0$ -dB stimuli). Each S heard all three test sequences. The order of the sequences was partially randomized across Ss. The Ss were informed as to the type of sounds they would hear and were instructed to listen for differences in intensity of the initial portions of the stimulus pairs. The Ss wrote either same (S) or different (D) on printed answer sheets. Before each test sequence, the Ss were presented with 18 practice trials of stimulus pairs and were concurrently shown the correct responses on an answer sheet.

RESULTS

The probability of a "different" response when the

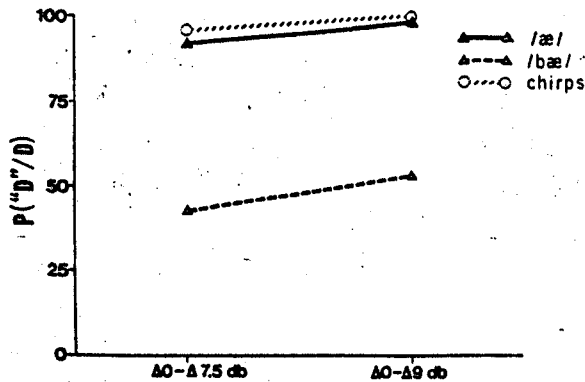


Fig. 3. $P("D"/D)$ for the 7.5- and 9-dB differences in intensity of formant transitions in syllable context, in isolation, and for steady-state vowels.

stimulus pairs were in fact different [i.e., $P("D"/D)$] for the three experimental conditions is shown in Fig. 3. The $\Delta 0$ - $\Delta 7.5$ -dB and $\Delta 0$ - $\Delta 9$ -dB stimulus pairs, in the syllable context /bæ/, were discriminated correctly on 42% and 54% of the trials, respectively. These same intensity differences in the isolated formant context were discriminated correctly on 96% and 100% of the trials, respectively. The intensity differences in the steady-state vowel context were discriminated correctly on 92% and 98% of the trials, respectively.

By scoring the "catch" trials, that is, the $\Delta 0$ - $\Delta 0$ -dB stimulus pairs, in all conditions, the detectability (d') of the intensity differences independent of response bias was computed. In the syllable context, Ss averaged 30% false alarms; in the isolated formant context, 7% false alarms; and in the steady-state vowel context, 3% false alarms. Taking these false-alarm rates into account, the d' scores for the discriminability of the 7.5- and 9.0-dB intensity differences in the syllable context were .32 and .62; in the isolated formant context, 3.72 and 3.79; in the steady-state vowel context, 3.28 and 3.93.

DISCUSSION

The 7.5- and 9-dB differences in intensity of the formant transitions, when in the syllable context /bæ/, were discriminated essentially at chance. However, the 30% false-alarm rate indicates a somewhat conservative criterion for accepting a difference in intensity. (If the Ss were performing at chance with no response bias, a 50% false-alarm rate would be expected.) This suggests that some, although very little, information about the intensity of the formant transitions in the $\Delta 0$ - $\Delta 9$ -dB condition may have been available for recall from short-term auditory memory. This outcome is consistent with the previously cited studies on the discriminability of within-category acoustic information, such as extent

and direction of the second formant transition, which directly cues different phonetic categories (Mattingly et al, 1971; Pisoni, 1971). The poor discrimination was not due to the intensity differences being in themselves not discriminable, since in the isolated formant context, the differences were discriminated essentially perfectly.

It might be argued, however, that the poor discrimination of the intensity differences on the formant transitions was due to backward masking and not due to processes unique to the coding of speech. That is, the steady-state vowel formants, because they follow and are more intense than the formant transitions, mask some of the acoustic information on the formant transitions. This interpretation can be assessed by examining the discriminability of the 7.5- and 9-dB intensity differences in the steady-state vowel context. Backward masking ought to have been maximized in this condition, since the formant frequencies for the initial 60 msec and the following 240 msec were identical.

As shown in Fig. 2, however, Ss discriminated intensity differences in the steady-state vowel context about as well as in the isolated formant context. It does not seem likely, then, that the differences in discriminability of the intensity of the formant transitions, in and out of syllable context, was a function of backward masking.

We may account for these data by assuming that after the acoustic cues for stop consonants have been recoded into a phonetic representation, all of the acoustic information is stored in a relatively inaccessible short-term auditory memory. All auditory information should be then equally inaccessible for recall. This would account for the poor discrimination within a stop consonant category of differences in first and second formant transitions, as well as the poor discrimination of intensity differences on formant transitions.

REFERENCES

- Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, 2, 131-157.
- Miller, G. A. The magical number seven, plus or minus two, or some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-96.
- Pisoni, D. On the nature of categorical perception of speech sounds. PhD thesis, University of Michigan, 1971. (Issued as Supplement to Haskins Laboratories Status Report on Speech Research.)
- Pollack, I. The information in elementary auditory displays. *Journal of the Acoustical Society of America*, 1952, 24, 745-749.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, 77, 234-249.

(Received for publication July 1, 1973;
revision received February 25, 1974.)