VOWEL AND NASAL DURATION AS CUES TO VOICING IN WORD-FINAL STOP CONSONANTS: SPECTROGRAPHIC AND PERCEPTUAL STUDIES

LAWRENCE J. RAPHAEL, M. F. DORMAN, *and* FRANCES FREEMAN

Haskins Laboratories, New Haven, Connecticut

CHARLES TOBIN

New York University, New York, New York

To determine durational differences between vowel and nasal segments preceding word-final /t/ and /d/, spectrograms were made of adult speakers' productions of minimal pairs of the type /pɛnt/-/pɛnd/. Vowel, nasal, and vowel plus nasal (vocalic nucleus) durations were greater before /d/ than before /t/. Assuming the voiceless context as a base, the increase in nasal duration in the voiced case was proportionately greater than the increase in vowel duration. This outcome suggests that nasal duration is a more powerful cue to the voicing characteristic of the following consonant than is vowel duration. To test this, adult listeners were presented synthetic CVNC utterances in which the nasal and vowel segments were independently varied in duration over a range of 40 msec to 200 msec and were instructed to label the final stop consonant as either voiced /d/ or voiceless /t/. Although changes in both vowel and nasal duration were sufficient to cue both voiced and voiceless judgments, listeners' categorization of final consonants shifted more rapidly as a result of varying nasal rather than vowel duration. Nasal duration, therefore, appears to be a stronger cue than vowel duration for the word-final voiced-voiceless consonant distinction in CVNC utterances.

The variation in the duration of vowels caused by the voicing characteristic of the consonants that follow them has been reported for many years in the phonetic/linguistic literature (Kenyon, 1951; Thomas, 1958; House, 1961; Heffner, 1964). The phenomenon is usually described as one in which vowels are lengthened in duration when they precede voiced consonants, so that in such minimal pairs as *beat–bead, float–flowed,* and *bus–buzz* the vowel in the first member of each pair is considerably shorter in duration than the vowel in the second member. An approximate estimate of this lengthening effect in English is that the duration of a vowel preceding a voiceless consonant will be from two-thirds to one-half of the duration of that same vowel preceding a voiced consonant (Raphael, 1971). Such an approximation ignores the variance caused by such factors as the intrinsic durations of different vowels (Peterson and Lehiste, 1960) and the manner and place of articulation of the following consonants.

Reprinted from *Journal of Speech and Hearing Research*
September 1975, Vol. 18, No. 3

The rather large differences between the durations of English vowels in voiced versus voiceless environment have interested linguists, phoneticians, and speech scientists for a number of reasons—one of the most important being the potential cue value of vowel duration to the perception of a following consonant as either voiced or voiceless.

The existence of this cue value seems likely when one considers that the vowel duration difference may be the only difference consistently present in minimal pairs such as *bit–bid,* which are distinguished, theoretically, by the voicing or voicelessness of the postvocalic consonant. Those cues to voicing that are operative in other phonetic environments, such as voicing during consonant closure, burst release, and duration of consonant closure, are often either neutralized or subjected to considerable variation in word-final consonants. What this means, of course, if one takes the voiced-voiceless opposition as somehow basic or primary to the perceptual separation of minimal pairs such as *bit–bid,* is that a secondary feature, vowel duration, is a more significant cue than the supposedly primary cue of voicing. Further, the more significant, though secondary, feature does not reside within the articulatory period or the acoustic segments being differentiated perceptually, but rather resides within a preceding, adjacent articulatory/acoustic segment.

The cue value of vowel duration has, in fact, been demonstrated (Denes, 1955; Raphael, 1972) for vowels preceding word-final stops, fricatives, and clusters consisting of stop + stop *(lopped–lobbed)* stop + fricative *(picks–pigs),* and fricative + stop *(bussed–buzzed).*

There are, however, several instances in English where vowels are not the sole component of the vocalic segment preceding consonants. These include instances in which a vowel + resonant consonant sequence precedes a voiced or voiceless consonant, yielding such minimal pairs as *bent–bend, hurt–heard,* and *welt–weld.* What has been little studied at this point (Lehiste, 1972) and what we attempted to investigate in the first of two experiments were the following questions:

1. What are the nature and magnitude of the lengthening effect in vowel + resonant consonant + obstruent consonant syllables?
2. Are the vowel and resonant durations affected differentially by the following consonants?

EXPERIMENT I

METHOD

In attempting to determine the magnitude of the lengthening effect for the entire vocalic segment and its component parts, we limited ourselves to spectrographic analyses of monosyllables ending in vowel + nasal + stop sequences. For English, this effectively restricts the utterances studied to those consisting of a vowel + /n/ + /t/ or /d/. Although such sequences as /mp – mb/ *(ample–amble)* and /ŋk – ŋg/ *(tinkle–tingle)* occur in English, they do not provide phonological oppositions in word-final position in monosyllables.

Five native speakers of American English recorded a randomized list of words that included many "dummy" items. The speakers were not told the nature of the experiment. The members of the following minimal pairs were dispersed throughout the list: *can't–canned, daunt–dawned, stunt–stunned, pent–penned, sent–send, paint–painted, pint–pined, mount–mound,* and *burnt–burned.* Eight different vowels were thus included in the data to be analyzed. Spectrograms were made of each of the utterances and inspected to derive the following data (see Figure 1):

1. The duration of the vowel.
2. The duration of the nasal.
3. The duration of the entire vocalic nucleus (vowel + nasal).

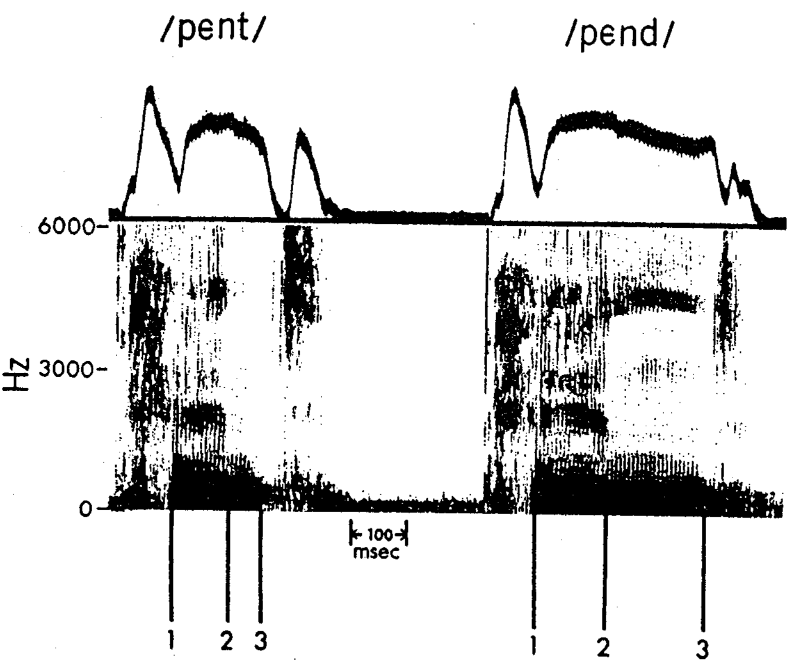

FIGURE 1. Spectrogram marked to delimit vowel duration, 1 to 2; nasal duration, 2 to 3; and vocalic nucleus, 1 to 3.

The onset of the vowel was taken as that point at which the harmonic formant structure of any formant first became visible on the spectrogram. Aspiration was thus excluded from the vowel duration. The onset of the nasal consonant was taken as that point at which the weak low-frequency formant, which characteristizes nasals in general, was present to the exclusion of the formant structure that characterizes the vowel. That is, in those cases where nasalization accompanied vowel articulation, such nasalization was considered as part of the vowel and not as part of the following nasal consonant. Although the presence of a nasalized vowel presents certain difficulties to the segmentation of the acoustic signal, we felt secure in our segmenting decisions, more secure than we felt after reading reports of similar research prior to our analysis (Lehiste, 1972).

A more difficult segmentation task occurred at the boundary between the nasal consonant and the following stop consonant. The problem here centered around the fact that the nasal formants can be extremely variable over a weak range of intensities. They occasionally can seem to disappear for several milliseconds, only to reappear in the middle of what might otherwise have been taken for the closure period of the final stop. Then, too, the presence or absence of a fundamental frequency, indicative of voicing, is not of any consistent use in segmenting, since in many cases where the speakers intended voiceless /t/, vocal fold vibrations occur well into the stop closure period, and since, in the opposing case, intended voiced /d/ revealed little, if any, voicing during the occlusion for the final stop. This apparent inconsistency of glottal pulsing during closure is to be expected if it is in fact the duration of some part or parts of the preceding vocalic element that cues the perception of voicing in the final consonant. That is, the feature that embodies the more important perceptual cue is preserved; the feature bearing a redundant, less important perceptual cue is subject to deletion—or perhaps atrophy would be a more appropriate word. The most satisfactory solution for segmenting the nasal from the final stop seemed to be to mark the termination of the nasal at the last point in time at which one could confidently discern the low-frequency nasal formant, if a fundamental frequency were also present.

RESULTS

Two major findings resulted from the spectrographic analyses. First, both the vowel and the nasal consonant, and thus the entire preceding vocalic segment, were of greater duration when they preceded /d/ than when they preceded /t/. This was true without exception for all speakers and all utterances (see Table 1). Although there were instances where the magnitude of the increase in duration was quite small for either the nasal or the vowel in an utterance (as little as five msec for the vowel and 15 msec for the nasal), the smallest durational difference found between any single vocalic nucleus as it occurred in each voicing environment was 55 msec. It appears that if one of the components in the vocalic nucleus increases only minimally in duration before a voiced consonant, then the other component will undergo a more substantial durational increment.

Second, the vowel and nasal durations were affected differentially by their voicing environments (see Tables 2 and 3). Although both vowel and nasal increased in duration from the voiceless to the voiced environments, the increment of nasal duration was proportionately greater than that of vowel

Table 1. Vowel, nasal, and vowel + nasal durations for five speakers. Values (in msec) are rounded to the nearest 5 msec.

	Speaker 1			*Speaker 2*			*Speaker 3*			*Speaker 4*			*Speaker 5*		
Pairs	*V*	*N*	*V + N*	*V*	*N*	*V + N*	*V*	*N*	*V + N*	*V*	*N*	*V + N*	*V*	*N*	*V + N*
Pent	150	70	220	80	75	155	130	75	205	90	55	145	85	40	125
Pend	175	175	350	155	255	410	140	125	265	140	160	300	110	95	205
Sent	145	65	210	100	65	165	130	70	200	105	70	175	70	65	135
Send	180	130	310	110	195	305	180	110	290	155	110	265	80	115	195
Paint	135	65	200	110	40	150	145	40	185	140	35	175	–	–	–
Pained	255	135	390	160	210	370	175	90	265	205	95	300	–	–	–
Can't	175	60	235	150	75	225	170	45	215	175	30	205	125	45	170
Canned	225	140	365	220	170	390	255	70	325	215	110	325	150	75	225
Daunt	235	60	295	160	90	250	225	75	300	220	30	250	160	30	190
Dawned	260	125	385	215	230	445	255	125	380	260	90	350	200	45	245
Stunt	145	35	180	75	80	155	105	60	165	110	55	165	100	25	125
Stunned	150	160	310	100	175	275	150	100	250	145	145	290	110	80	190
Pint	175	60	235	110	45	155	195	40	235	160	55	215	–	–	–
Pined	245	115	360	205	170	375	245	60	305	300	105	405	–	–	–
Mount	170	80	250	–	–	–	200	40	240	205	50	255	–	–	–
Mound	295	130	425	–	–	–	280	60	340	265	95	360	–	–	–
Burnt	–	–	–	100	45	145	135	50	185	155	75	230	110	25	135
Burned	–	–	–	165	135	300	210	75	285	210	145	355	195	75	270

Table 2. Percentages of vocalic nucleus distributed between vowel and nasal.

Pairs	Speaker 1		Speaker 2		Speaker 3		Speaker 4		Speaker 5	
	V	N	V	N	V	N	V	N	V	N
Pent	68.2	31.8	51.6	48.4	63.4	36.6	62.1	37.9	68.0	32.0
Pend	50.0	50.0	37.8	62.2	52.8	47.2	46.7	53.3	53.7	46.3
Sent	69.0	31.0	60.6	39.4	65.0	35.0	60.0	40.0	51.9	48.1
Send	58.1	41.9	36.1	63.9	62.1	37.9	58.5	41.5	41.0	59.0
Paint	67.5	32.5	73.3	26.7	78.4	21.6	80.0	20.0	–	–
Pained	65.4	34.6	43.2	56.8	66.0	34.0	68.3	31.7	–	–
Can't	74.5	25.5	66.7	33.3	79.1	20.9	85.4	14.6	73.5	26.5
Canned	61.6	38.4	56.4	43.6	78.5	21.5	66.2	33.8	66.7	33.3
Daunt	79.7	20.3	64.0	36.0	75.0	25.0	88.0	12.0	84.2	15.8
Dawned	67.5	32.5	48.3	51.7	67.1	32.9	74.3	25.7	81.6	18.4
Stunt	80.6	19.4	48.4	51.6	63.6	36.4	66.7	33.3	80.0	20.0
Stunned	48.4	51.6	36.4	63.6	60.0	40.0	50.0	50.0	57.8	42.2
Pint	74.5	25.5	71.0	29.0	83.0	17.0	74.4	25.6	–	–
Pined	68.1	31.9	54.7	45.3	80.3	19.7	74.1	25.9	–	–
Mount	68.0	32.0	–	–	83.3	16.7	80.4	19.6	–	–
Mound	69.4	30.6	–	–	82.4	17.6	73.6	26.4	–	–
Burnt	–	–	69.0	31.0	73.0	27.0	67.4	32.6	81.5	18.5
Burned	–	–	55.0	45.0	73.7	26.3	59.2	40.8	72.2	27.8

TABLE 3. Percentages of increase for vowel, nasal, and vowel + nasal from voiceless to voiced environments.

Pairs	Speaker 1			Speaker 2			Speaker 3			Speaker 4			Speaker 5		
	V	*N*	*V* + *N*	*V*	*N*	*V* + *N*	*V*	*N*	*V* + *N*	*V*	*N*	*V* + *N*	*V*	*N*	*V* + *N*
Pent Pend	16.7	150.0	59.1	93.7	240.0	64.5	7.7	66.7	29.3	55.6	190.9	106.9	29.4	137.5	64.0
Sent Send	24.1	100.0	47.6	10.0	200.0	84.8	38.5	57.1	45.0	47.6	57.1	51.4	14.3	76.9	44.4
Paint Pained	88.9	107.7	95.0	45.5	425.0	146.7	20.7	125.0	43.2	46.4	171.4	71.4	–	–	–
Can't Canned	28.6	133.3	55.3	46.7	126.7	73.3	50.0	66.7	51.7	22.9	266.7	58.5	20.0	66.7	32.4
Daunt Dawned	10.6	108.3	30.5	34.4	155.6	78.0	13.3	66.7	26.7	18.2	200.0	40.0	25.0	50.0	28.9
Stunt Stunned	3.4	357.1	72.2	33.3	118.8	77.4	42.9	66.7	51.5	31.8	163.6	75.8	10.0	220.0	52.0
Pint Pined	40.0	91.7	53.2	86.4	227.8	141.9	25.6	50.0	29.8	87.5	90.9	88.4	–	–	–
Mount Mound	73.5	62.5	70.0	–	–	–	40.0	50.0	41.7	29.3	90.0	41.2	–	–	–
Burnt Burned	–	–	–	65.0	200.0	106.9	56.6	50.0	54.1	35.5	93.3	54.3	77.3	200.0	100.0

duration. For example, for Speaker 1 (Table 2) it can be seen that in the word *pent* the vowel represents 68.2% of the vocalic nucleus, the nasal 31.8%. But in the word *pend*, the vocalic nucleus is almost evenly divided between vowel and nasal. Table 3 reveals, for the same subject and utterances, that the vowel in *pend* has increased by 16.7% of its duration in *pent*, whereas the nasal has increased by 150.0% of its duration. Nor is this by any means an extreme case, as inspection of the data in the tables reveals.

A final observation concerns two cases in which a reversal occurs in the data. The pairs *mount–mound* (Speaker 1) and *burnt–burned* (Speaker 3) reveal that the vowel duration increase is proportionately greater than the increase in nasal consonant duration. Inspection of these two cases reveals an extensive nasalization of the vowel in each. It may well be that an alternative strategy to a proportionately greater lengthening of the nasal duration is the addition of nasal resonances to the vowel. Further investigations are needed to determine if there is a regular trading relationship between the duration of vowel nasalization and the duration of the nasal per se.

DISCUSSION

The results of these spectrographic analyses suggest two hypotheses. First, both the individual vowel and nasal durations, as well as their combined duration (that is, that of the vocalic nucleus), are potentially sufficient cues to the voicing characteristic of the following consonant. Second, the proportionately greater change in nasal duration indicates that it should be a more powerful cue than that of vowel duration to the voicing characteristic of the following consonant.

To test these hypotheses, a second experiment was performed in which listeners were presented with synthetic CVNC utterances. The durations of the nasal and vowel segments were independently varied in the utterances and listeners were asked to label the final consonant as voiced /d/ or voiceless /t/.

EXPERIMENT II
METHOD

Subjects

The listeners were 10 college student volunteers who had normal hearing.

Preparation of Stimuli

Two series of stimuli were prepared on the Haskins Laboratories' parallel-resonance synthesizer. The stimuli in each series were variations of *bend* (/bɛnd/), synthesized as follows:

1. Thirty-five msec linear transitions were provided in a three-formant /ɛ/. These transitions were appropriate with respect to both stop manner and

bilabial place of articulation. A 100-Hz fundamental frequency extended throughout the duration of all stimuli.

2. A steady-state vowel, /ɛ/, was provided with $F_1 = 537$ Hz, $F_2 = 1845$ Hz, and $F_3 = 2525$ Hz, followed by linear formant transitions to the nasal consonant /n/.
3. A steady-state nasal consonant, /n/, was provided with low applitude F_1 and F_2 at 260 Hz and 1232 Hz, respectively, and with F_3 at 2358 Hz.
4. A formantless (closure) interval of 100 msec was provided.
5. A three-formant, 10-msec, frictionless burst release, with formant frequencies appropriate to stop manner and alveolar place of articulation, was provided (see Figure 2).

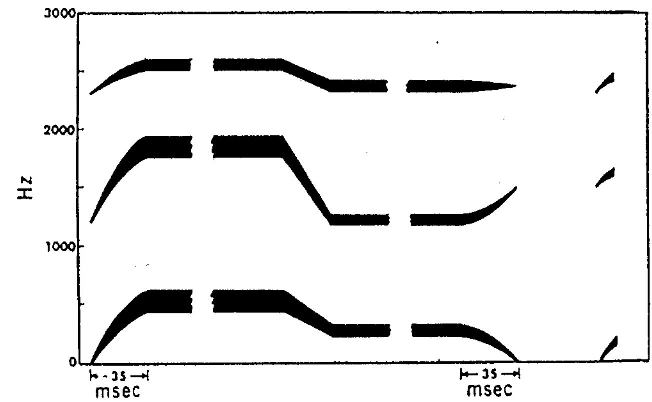

FIGURE 2. Schematic spectrogram of basic CVNC utterance. Both nasal and vowel segments were independently varied in duration from 40 msec to 200 msec in 20 20-msec steps. Formant width indicates relative intensity.

In the first series of stimuli, the nasal duration was held constant at 130 msec while the vowel duration varied between 40 msec and 200 msec in 20-msec steps. In the second stimulus series, the nasal duration varied in 20-msec steps over a range of 40 msec to 200 msec, while the vowel duration was held constant at 130 msec. Our spectrographic evidence indicated that the 130 msec constant duration for both vowel and nasal was representative of real-speech values.

A separate test tape was prepared for each of the stimulus series. Each tape contained six tokens of each stimulus type, for a total of 54 test items. The stimuli were randomized on the tape. The interstimulus interval was three seconds.

Tasks

Two groups of listeners were tested separately. One group of listeners ($N = 5$) heard the nasal-varying series first, then the vowel-varying series. A second group ($N = 5$) heard the vowel-varying sequence first, then the nasal-varying sequence.

Procedure

The listeners were seated in a semicircle, within a large, sound-attenuated room, facing an AR-4x loudspeaker. The listeners were told they would hear a series of computer-synthesized words, and were instructed to label on a printed response form the final consonant as either /d/ or /t/. After several tokens of the stimuli were presented to familiarize the listeners with the synthetic signals, the first test sequence was begun. At the end of this sequence, the listeners were given a brief rest, then were presented with the second test sequence.

RESULTS

Vowel Duration Series

All of the subjects indicated a change in perception of the final stop consonant with changes in vowel duration. When vowel duration was brief, all listeners reported the stimuli as ending in /t/; when the vowel duration was long, all listeners reported /d/ as the final consonant.

The 50% point in listeners' judgment occurred over a range of vowel durations of from 115 msec to 180 msec (see Figure 3). The mean crossover point for all listeners was 142.5 msec.

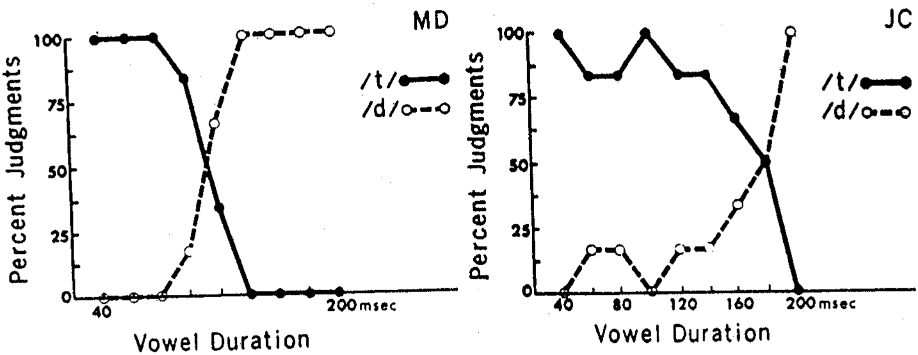

FIGURE 3. Percent /d/ – /t/ identification as a function of vowel duration. The labeling functions of JC and MD represent the subject extremes in cross-over point for /d/ – /t/ judgments.

Nasal Duration Series

All listeners indicated a change in perception of the final consonant with changes in nasal duration. Short duration nasals elicited /t/ identifications, whereas long duration nasals elicited /d/ as a response. The 50% point in listeners' judgments occurred over a range of 90 msec to 115 msec (see Figure 4). The mean crossover point for all listeners was 99.4 msec of nasal duration.

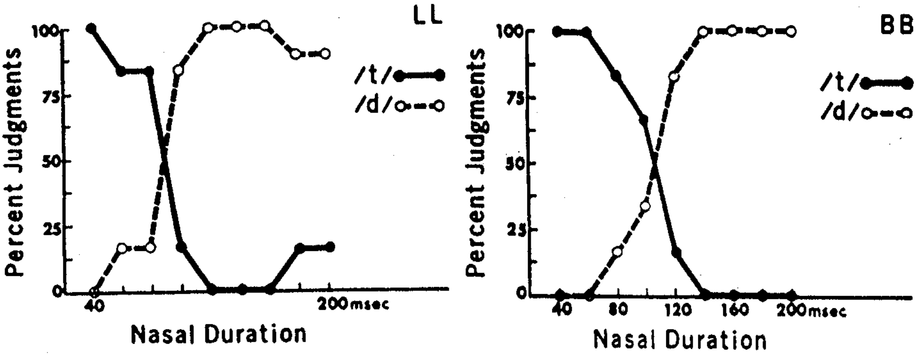

FIGURE 4. Percent /d/ – /t/ identification as a function of nasal duration. The labeling functions of LL and BB represent the subject extremes in cross-over point for /d/ – /t/ judgments.

DISCUSSION

The data support the hypothesis that the word-final voiced-voiceless opposition, at least as represented by /t/ and /d/, can be cued by the duration of either the preceding vowel or the nasal when in a vowel-nasal-stop sequence. By extension, therefore, the duration of the entire vocalic segment of vowel + nasal must be a cue to the perception of voicing in a following consonant.

The data also reveal that the duration of the nasal is a relatively stronger cue than that of vowel duration. This seems clear from a comparison of the amount of variation in nasal versus vowel duration needed to change the listeners' judgments from /t/ to /d/. Clearly, listeners' perceptions of final consonants as voiced or voiceless are more sensitive to variations in nasal duration than to variations in vowel duration. Assuming that the cues to voicing contained in the final consonant segment do have some effect on perception, and recalling that these cues in the synthetic stimuli were appropriate to /d/, it is apparent that the effect of these cues can be either neutralized or enhanced by smaller changes in nasal duration than in vowel duration.

Such an outcome seems reasonable, if for no other reason than the greater proximity of the nasal than of the vowel to the following consonant. If the durations of the vowel and nasal segments are stored sequentially in echoic memory (Crowder and Morton, 1969) during speech perception, then it may simply be the case that the shorter storage time needed for the nasal segment makes it more efficient as a primary cue. The cue value of the more distant (vowel) segment may be reduced by (1) the loss of some of the auditory information from echoic memory during the processing of the closer (nasal) segment, (2) interference from the entry of the nasal information into echoic memory, or (3) a combination of (1) and (2). Any of these mechanisms could also contribute to the fact that although the cue of vowel duration is relatively weaker, under certain conditions it can still be a sufficient cue to the perception of the voicing characteristic of the word-final consonant.

ACKNOWLEDGMENT

This research was supported by NICHD contract HD 01994 to Haskins Laboratories, and a Schuster Fellowship to Lawrence J. Raphael and M. F. Dorman from Herbert H. Lehman College, City University of New York. Requests for reprints may be sent to L. J. Raphael, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06570.

REFERENCES

Crowder, R., and Morton, J., Precategorical acoustic storage (PAS). *Percept. Psychophys.*, 5, 365-373 (1969).

Denes, P., Effect of duration on the perception of voicing. *J. acoust. Soc. Amer.*, 27, 761-764 (1955).

Heffner, R-M. S., *General Phonetics*. Madison, Wis.: Univ. of Wisconsin Press (1964).

House, A. S., On vowel duration in English. *J. acoust. Soc. Amer.*, 33, 1174-1178 (1961).

Kenyon, J. S., *American Pronunciation*. (10th ed.) Ann Arbor, Mich.: George Wahr (1951).

Lehiste, I., Manner of articulation, parallel processing, and the percent of duration. *Working Paper in Linguistics* Computer and Information Science Research Center, Ohio State University, 12, 33-52 (1972).

Peterson, G. E., and Lehiste, I., Duration of syllabic nuclei in English. *J. acoust. Soc. Amer.*, 32, 693-703 (1960).

Raphael, L. J., Vowel duration as a cue to the perceptual separation of cognate sounds in American English. Doctoral dissertation, City Univ. of New York (1971).

Raphael, L. J., Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *J. acoust. Soc. Amer.*, 51, 1296-1303 (1972).

Thomas, C. K., *An Introduction to the Phonetics of American English.* (2nd ed.) New York: Ronald Press (1958).

Received July 5, 1974.
Accepted December 16, 1974.