Syllable Timing and Vowel Perception*

Robert R. Verbrugge+ and David Isenberg

ABSTRACT

Consonantal environment may aid in specifying vowel identity by supplying critical information about timing. Several vowel pairs in American English are distinguished by temporal as well as spectral variables, and these temporal differentiae vary with articulatory rate. Two studies were designed to explore the following paradox: when consonantal formant transitions are introduced into a steady-state vowel, holding syllable duration constant, a response shift is observed toward longer vowel alternatives, even though steady-state duration has been reduced. The first study verified this finding for the vowel pair /e/-/æ/ in comparisons of #V# and bVb continua. Pairs of continua were defined separately by F1 variation and by duration variation, and both continuum types evidenced the paradox to some degree. A second study varied the rate of symmetric consonantal transitions in F1-varying CVC continua (V = /e,æ/, C = /#/, b, w/) in order to test whether transition rate might specify an articulatory rate that effectively scales vowel duration. Vowel responses did not vary monotonically with either transition rate or steady-state duration, but interacted with the perceived identity of the initial consonant. Listeners' judgments may demonstrate a sensitivity to constraints on the relative timing of consonantal and vocalic gestures.

There are several possible explanations for the greater identifiability of vowels in a consonantal environment (see Strange, Verbrugge, Shankweiler and Edman, 1976). One possibility, which is not often considered, is that consonantal environments supply critical information about the timing of the gestures that comprise a syllable. It has been demonstrated that variation in

*This paper was presented at the 95th meeting of the Acoustical Society of America, Providence, R.I., 15-19 May 1978.

+Also University of Connecticut, Storrs.

Acknowledgment: The research was supported by a grant to Haskins Laboratories from the National Institute of Child Health and Human Development (NICHD Grant HD-01994; BRSG Grant RR-05596), by a fellowship to R. R. Verbrugge from the University of Michigan Society of Fellows, and by a fellowship to D. Isenberg from the National Institutes of Health (NS05493-01).

the rate and rhythm of a carrier phrase can alter the identity of a vowel in an embedded syllable [Ainsworth, (1974); Nooteboom, (1974); Verbrugge, Strange, Shankweiler and Edman, (1976); Verbrugge and Shankweiler, (1977)]. It is possible that single syllables, in themselves, contain information about the rate and rhythm of articulation, and that this information (like that carried over longer contexts) can alter the identity of the vowel perceived.

In a study reported by Shankweiler, Verbrugge and Studdert-Kennedy (1978), synthetic steady-state vowels were compared with synthetic b-vowel-b syllables of the same acoustic duration. The major finding of that study was an overall decrease in perceptual errors for the medial vowels in contrast to the steady-state vowels. A second and more subtle finding of that study was a systematic exception to the overall reduction in errors: short vowels were misidentified as long vowels more often in the bVb syllables than in the steady-state vowels. This was a surprising result: the introduction of consonantal transitions shortened the duration of the steady-state, but lengthened the perceived vowel.

Several studies have shown that some fraction of consonantal formant transitions or frication contributes to the perceived duration of a vowel in the same syllable--that is, vowel duration is not to be identified solely with duration of a steady-state [for example, Raphael, Dorman and Liberman, (1975); Mermelstein, Liberman and Fowler, (1977)]. Estimates of the fraction vary widely, showing dependencies on a number of factors, including the manner class, place and syllable position of the consonants employed. However, if we seek a comparable interpretation of the results of Shankweiler, et al. (1978), we are forced to an unusual and troubling conclusion: a fraction greater than 100 percent of the transition is treated as vowel duration. The pattern with transitions is heard as containing a longer vowel than the pattern without transitions, even though the total acoustic durations of the two patterns are the same.

Before lavishing explanations on this paradox, we decided it would be best to verify it in a design that would allow the response shift to be assessed more parametrically. Using an OVE III synthesizer, we designed continua to study the short-long vowel pair /ɛ/ and /æ/. Typical formant patterns are illustrated in Figure 1.

In one pair of continua, the contrast between /ɛ/ and /æ/ was achieved by varying the steady-state frequency of the first formant (F₁) through nine steps of approximately 15 Hz, over a range from 526 to 744 Hz. This is indicated by the vertical arrows in Figure 1. One continuum consisted of bVb syllables (illustrated on the top) and the second continuum consisted of steady-state vowels (at the bottom). All of the patterns in these continua were 235 msec in total duration.

In a second pair of continua, the contrast between /ɛ/ and /æ/ was achieved by varying the duration of the syllables. For these two series, a fixed, intermediate value of F₁ (702 Hz) was used for all patterns. The duration-varying patterns ranged in 20-msec steps from 175 to 315 msec in total duration, as indicated by the horizontal arrows in Figure 1. The F₁-
Figure 1: Formant control parameters used in synthetic bVb continua (top) and isolated vowel continua (bottom). Vertical arrows indicate range of first-formant variation ($F_1$-varying continua), and horizontal arrows indicate range of variation in total syllable duration (duration-varying continua).
varying and duration-varying continua allowed two independent tests of the paradoxical influence of consonantal environment on vowel perception. For each continuum type, the paradox would be verified if there were a shift toward the longer vowel alternative (/æ/) in the consonantal environment.

Results for the \( F_1 \)-varying continua are shown in Figure 2. The curves present the percentage of /æ/ responses for each step on the \( F_1 \) continuum, averaged over eight listeners. There was a significant difference between the total proportion of /æ/ responses for the bVb continuum and the steady-state continuum (\( t(7) = 2.18, p < .05 \)). Throughout the lower range of \( F_1 \) values, there was a consistent bias toward hearing the longer vowel /æ/ in the bVb syllables.

The results for the duration-varying continua tell a similar though less coherent story. For four of our listeners, the duration variation was not successful in defining a contrast between /e/ and /æ/, that is, no phoneme boundary was discernible on either the bVb or steady-state continuum. It is worth noting, however, that for the four listeners who did show categorization, there was a fairly consistent shift in the boundary similar to that illustrated in Figure 2—that is, the short steady-state vowels became consistently more /æ/-like in consonantal environment. (The average boundary shift was 52 msec for the four listeners.)

These results pose a puzzle for any perceptual theory that treats perceived vowel length as a simple function of the durations of a syllable's acoustic components. How can vowel quality lengthen, when no acoustic measure is lengthened? One possible resolution to the puzzle is to argue that the bVb syllables somehow specify a faster rate of articulation. This, in turn, could scale the interpretation of duration, so that a particular acoustic duration would specify a relatively long event in a fast utterance (in this case, the bVb syllables) and a relatively short event in a slow utterance (the steady-state vowels).

What properties of the bVb syllable structure might carry information about rate? One possibility is the duration of formant transitions—the 35 msec transitions in our bVb patterns may have specified a relatively fast rate of articulation. If so, one would expect that, as transition duration is lengthened, a slower articulatory rate will be specified and the degree of shift toward /æ/ responses will diminish.

To test this possibility, we created a set of seven continua with varying transition durations. The formant patterns are illustrated schematically at the top of Figure 3. The seven continua were defined by introducing progressively longer symmetric transitions, ranging from no transitions to 60-msec transitions, in 10 msec steps. Each of these continua contained nine steps of \( F_1 \) variation, defining a contrast between the vowels /e/ and /æ/. As before, all of the syllables were equalized in total duration at 235 msec.

We also prepared seven continua of steady-state vowels matched in duration to the steady-state regions of the CVC syllables. The lower two patterns in Figure 3 illustrate one of the CVC syllables (in this case with
Figure 2: Average percent /æ/ identification for the F1-varying continua. Each point represents an average for eight listeners, 20 judgments per listener.
Figure 3: Formant control parameters used in synthetic $F_1$-varying continua in transition duration test. Seven such continua were defined by the progressive introduction of symmetric transitions ranging from 0 to 60 msec in duration (indicated by fan-shaped lines at top). Sample CVC and steady-state patterns are illustrated at the center and bottom.
Figure 4: Average percent /æ/ identification for the CVC continua and steady-state continua. Each point represents data pooled over a continuum of nine steps, and pooled over eight listeners, 20 judgments per step per listener.
30-msec transitions) and its corresponding steady-state control.

In any one block of trials, listeners heard a randomized series of either the transition-varying syllables or the steady-state controls. Results for eight listeners are presented in Figure 4. Each point plots the total percentage of /æ/ responses for a particular spectral continuum (for which all patterns had a fixed duration).

The results for the steady-state continua are connected by the dotted line. As steady-state duration decreases from 235 msec at the left to 115 msec at the right, the proportion of /æ/ responses in the steady-state vowels drops monotonically. This is exactly what one would expect: as steady-state duration decreases, there is a biasing toward the shorter vowel alternative.

The results for the transition-varying continua are plotted with a solid line. While steady-state duration again decreases from left to right, there is not a monotonic decrease in /æ/ responses like that observed for the pure steady-state patterns. In fact, /æ/ responses increase over the first four continua, even though the steady-state durations decrease over this range by 60 msec. Thus, the vowel judgments did not change either linearly or monotonically with the durations of syllable components when a consonantal environment was specified.

In examining these functions for vowel identification, it is helpful to know the regions in which particular consonantal neighbors are heard. Across the top of Figure 4, we have indicated the approximate cross-over points (determined in a pilot study) where listeners shift from hearing predominantly isolated vowels at short transition durations, to syllables with initial /b/ at intermediate durations, and syllables with initial /w/ at the long transition durations. The two consonantal environments, /b/ and /w/, both showed substantially higher proportions of /æ/ responses than the patterns heard as isolated vowels, again verifying the paradoxical effect we found before.

Can this effect be attributed to changes in rate specified by transition duration? Our prediction was that /æ/ responses should decrease as transition duration is lengthened, that is, as a slower rate is specified. This prediction can be tested for each CVC environment, initial /b/ and initial /w/. For the syllables with initial /b/, the observed trend was exactly opposite to that predicted: the percentage of /æ/ responses increased as transition duration was lengthened. For the syllables with initial /w/, the trend was not as clear, but there appeared to be a decrease, as predicted, in the /æ/ responses at longer transition durations. Thus there was no consistent evidence that the paradoxical shifts could be attributed to changes in rate, to the extent that rate may be specified by transition durations.

One important unknown at this point is how listeners divide their responses in the boundary regions between isolated vowels, /b/ environments, and /w/ environments. It is possible, for example, that the rising slope in the bV region of Figure 4 is an artifact of the averaging process. The underlying proportions in all three phoneme regions might be stable or falling.
as a function of transition duration. We are currently studying contingencies between consonant and vowel responses to assess these possibilities.¹

Another alternative that we are currently exploring is to view the vowel as an event that extends beyond what is acoustically realized. The duration of a stop-vowel-stop syllable extends throughout the silent closure intervals that precede initial release and follow final closure. Thus, the vowel's duration is not limited to the period of time that the vocal tract has a nonzero output amplitude. As an articulatory event, its duration may encompass part of the silent closure intervals as well. From this perspective, it would not be at all paradoxical for a bVb syllable to specify a longer vowel than is specified by a steady-state pattern of equal acoustic duration. In effect, consonantal transitions might lengthen the perceived vowel, not by specifying a faster rate of articulation, but by specifying a slower one.

Our simple paradox has become absorbingly complex. In our efforts to explain it, we have come to realize how little is known about the timing relationships between consonants and vowels, how these relationships are changed or preserved with changes in rate, and how the acoustics of a syllable specify them to a listener. Answers to these questions may be basic to understanding why consonantal environments facilitate vowel identification. The presence of consonants may make the temporal as well as spectral identity of vowels more determinate.

REFERENCES


¹Our preliminary analyses suggest that the proportions of /æ/ responses decrease for all three phoneme environments (initial /b/, initial /w/ and no initial consonant). These results are consistent with the view that longer transitions specify a slower rate and, in turn, a shorter vowel, but they are also consistent with the simpler view that shorter steady-state duration (which is correlated with the longer transitions) specifies a shorter vowel. However, the latter approach is not able to explain the differences between the three environments—in particular, why syllables with initial /b/ or /w/ are perceived to contain a longer vowel than those heard as steady-state vowels.